

# Scoring rules for judgment aggregation

Franz Dietrich

CNRS, Paris, UEA, Norwich

26 December 2011

# Scoring rules for judgment aggregation

Franz Dietrich[1]

December 2011

**Abstract**

This paper studies a class of judgment aggregation rules, to be called 'scoring rules' after their famous counterpart in preference aggregation theory. A scoring rule delivers the collective judgments which reach the highest total 'score' across the individuals, subject to the judgments having to be rational. Depending on how we define 'scores', we obtain several (old and new) solutions to the judgment aggregation problem, such as distance-based aggregation, premise- and conclusion-based aggregation, truth-tracking rules, and a Borda-type rule. Scoring rules are shown to generalize the classical scoring rules of preference aggregation theory.

*JEL Classification*: D70, D71

*Keywords*: judgment aggregation, social choice, scoring rules, Hamming rule, Borda rule, premise- and conclusion-based rules

## 1 Introduction

The judgment aggregation problem consists in merging many individuals' yes/no judgments on some interconnected propositions into collective yes/no judgments on these propositions. The classical example, born in legal theory, is that three jurors in a court trial disagree on which of the following three propositions are true: the defendant has broken the contract ($p$); the contract is legally valid ($q$); the defendant is liable ($r$). According to a universally accepted legal doctrine, $r$ (the 'conclusion') is true if and only if $p$ and $r$ (the two 'premises') are both true. So, $r$ is logically equivalent to $p \wedge q$. The simplest rule to aggregate the jurors' judgments – namely propositionwise majority voting – may generate logically inconsistent collective judgments, as Table 1 illustrates. There are of course nu-

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

Table 1: The classical example of logically inconsistent majority judgments

merous other possible 'agendas', i.e., kinds of interconnected propositions a group might face. Preference aggregation is a special case with propositions of the form '$x$ is better than

[1]CNRS, Cerses, Paris, France & UEA, Norwich, U.K. Mail: post@franzdietrich.net. Web: www.franzdietrich.net.

$y$' (for many alternatives $x$ and $y$), where these propositions are interconnected through standard conditions such as transitivity. In this context, Condorcet's classical *voting paradox* about cyclical majority preferences is nothing but another example of inconsistent majority judgments. Starting with List and Pettit's (2002) seminal paper, a whole series of contributions have explored which judgment aggregation rules can be used, depending on, firstly, the agenda in question, and, secondly, the requirements placed on aggregation, such as anonymity, and of course the consistency of collective judgments. Some theorems generalize Arrow's Theorem from preference to judgment aggregation (Dietrich and List 2007, Dokow and Holzman 2010; both build on Nehring and Puppe 2010a and strengthen Wilson 1975). Other theorems have no immediate counterparts in classical social choice theory (e.g., List 2004, Dietrich 2006a, 2010, Nehring and Puppe 2010b, Dietrich and Mongin 2010).

It is fair to say that judgment aggregation theory has until recently been dominated by 'impossibility' findings, as is evident from the *Symposium on Judgment Aggregation* in Journal of Economic Theory (C. List and B. Polak eds., 2010, vol. 145(2)). The recent conference 'Judgment aggregation and voting' (Freudenstadt, 2011) however marks a visible shift of attention towards constructing concrete aggregation rules and finding 'second best' solutions in the face of impossibility results. The new proposals range from a first Borda-type aggregation rule (Zwicker 2011) to, among others, new distance-based rules (Duddy and Piggins 2011) and rules which approximate the majority judgments when these are inconsistent (Nehring, Pivato and Puppe 2011).

The present paper contributes to the theory's current 'constructive' effort by investigating a class of aggregation rules to be called *scoring rules*. The inspiration comes from classical scoring rules in preference aggregation theory; these rules generate collective preferences which rank an alternative according to the sum-total 'score' it receives from the group members (where the 'score' could be defined in different ways, leading to different classical scoring rules). In a general judgment aggregation problem, however, there are no 'alternatives'; so our scoring rules are based on assigning scores to *propositions*, not alternatives. Nonetheless, our scoring rules are related to classical scoring rules, and generalize them, as will be shown.

The paradigm underlying our scoring rules – i.e., the maximization of total score of collective judgments – differs from standard paradigms in judgment aggregation, such as the premise-, conclusion- or distance-based paradigms. Nonetheless, it will turn out that several existing rules can be re-modelled as scoring rules, and can thus be 'rationalized' in terms of the maximization of total scores. Of course, the way scores are being assigned to propositions – the '*scoring*' – differs strongly across rules; for instance, the Hamming rule and the premise-based rule can each be viewed as a scoring rule, but with respect to two very different scorings. This paper explores various plausible scorings. It uncovers the scorings which implicitly underlie several well-known aggregation rules, and suggests other scorings which generate novel aggregation rules. For instance, a particularly natural scoring, to be called *reversal scoring*, will lead to a new generalization of Borda rule from preference aggregation to judgment aggregation. The problem of how to generalize Borda rule has been a long-lasting open question in judgment aggregation theory. Recently, an interesting, though so far incomplete, proposal was made by Zwicker (2011) (who told me that also Conal Duddy and Ashley Piggins have independent work in progress about this). Surprisingly, Zwicker's and the present Borda generalizations are distinctively different.

Though large, the class of scoring rules is far from universal: some important aggrega-

tion rules fall outside this class (notably the mentioned rule approximating the majority judgments, by Nehring, Pivato and Puppe 2011). I will also investigate a natural generalization of scoring rules, to be called *set scoring rules*, which are based on assigning scores to entire judgment sets rather than single propositions (judgments). Set scoring rules are for instance interesting in the context of *epistemic* ('truth-tracking') aggregation models, where they have recently been studied by Pivato (2011).

I could have written this paper by focusing exclusively on one specific application of scoring rules (for instance, the problem of extending Borda rule). However, I chose to give the paper a broader scope, not only to do justice to the diverse applications of scoring rules, but also to be useful at the theory's current stage of searching for concrete mechanisms. I hope that the ideas and perspectives offered below will be stimulating and inspiring.

After this introduction, Section 2 defines the general framework, Section 3 analyses various scoring rules, Section 4 goes on to analyse several set scoring rules, and Section 5 draws some conclusions about where we stand in terms of concrete aggregation procedures.

## 2 Agenda, aggregation rules, and examples

I now introduce the framework, following List and Pettit (2002) and Dietrich (2007).[2] We consider a set of $n$ ($\geq 2$) individuals, denoted $N = \{1, ..., n\}$. They need to form collective judgments on an agenda of propositions. In short, an *agenda* is a set of interconnected propositions which is closed under negation. Formally, it is an arbitrary set $X$ (whose elements we call 'propositions') such that

- $X$ is closed under negation: for every proposition $p$ in $X$ there is a specified proposition denoted $\neg p$ ('not $p$') in $X \backslash \{p\}$, where $\neg\neg p = p$;
- $X$ is endowed with logical interconnections: there is a specification of which subsets of $X$ are 'consistent', i.e., formally, there is a system $\mathcal{C}$ of subsets called 'consistent'.

A set $A \subseteq X$ (a 'judgment set') is *complete* if it contains a member of each pair $p, \neg p \in X$, and *(fully) rational* if it is complete and consistent. The set of all rational judgment sets is denote by $\mathcal{D}$.[3]

As usual, we assume that the consistency notion satisfies standard regularity conditions: no set $\{p, \neg p\}$ is consistent (C1, 'self-entailment'); subsets of consistent sets are consistent (C2, 'monotonicity'); $\varnothing$ is consistent and each consistent set can be extended to a complete and consistent set (C3, 'completability'). It follows that the consistent sets are precisely the subsets of rational sets: $\mathcal{C} = \{C \subseteq A : A \in \mathcal{D}\}$. In fact, the systems $\mathcal{D}$ and $\mathcal{C}$ are interdefinable, so that, given that we assume C1-C3, we could start from $\mathcal{D}$ instead of $\mathcal{C}$ as the primitive.[4]

[2] To be precise, I use a slimmer variant of their models: I do not explicitly introduce the logic $\mathbf{L}$ in which propositions are formed.

[3] Our notion of an 'agenda' is very general. The propositions in $X$ might be syntactic propositions (logical sentences), or semantic propositions (modelled for instance as sets of worlds), or arbitrary attributes that an agent may or may not possess. It is often natural to regard the agenda $X$ as a subset of a logic $\mathbf{L}$ from which it inherits the negation operator and the logical interconnections. This logic is general: it could for instance be standard propositional logic, standard predicate logic, or various modal or conditional logics (see Dietrich 2007).

[4] Instead of starting from the system of consistent sets $\mathcal{C}$ satisfying C1-C3 and deriving the system $\mathcal{D}$ of rational judgment sets, we could equivalently have started from $\mathcal{D}$ (any non-empty system of sets containing exactly one member from each pair $p, \neg p \in X$) and derived the system $\mathcal{C} := \cup_{A \in \mathcal{D}} \{C : C \subseteq A\}$ (which then automatically satisfies C1-C3). So, in algebraic terms, the agenda is definable either as the structure

Further, let $X$ be finite. Notationally, a judgment set $A \subseteq X$ is often abbreviated by concatenating its members in any order (so, $p\neg q\neg r$ is short for $\{p, \neg q, \neg r\}$); and the negation-closure of a set $Y \subseteq X$ is denoted

$$Y^{\pm} \equiv \{p, \neg p : p \in Y\}.$$

I now give two standard examples, to which I shall repeatedly refer.

**Example 1: the standard 'doctrinal paradox agenda'.** The agenda is

$$X = \{p, q, r\}^{\pm}.$$

Logical interconnections are defined relative to the external constraint $r \leftrightarrow (p \wedge q)$. So,

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

**Example 2: the preference agenda**. For an arbitrary, finite set of alternatives $K$, the *preference agenda* is defined as

$$X = X_K = \{xPy : x, y \in K, x \neq y\},$$

where the negation of a proposition $xPy$ is of course $\neg xPy = yPx$, and where logical 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 $K$ uniquely corresponds a judgment set, denoted $A_{\succ} = \{xPy \in X : x \succ y\}$, and the set of all rational judgment sets is

$$\mathcal{D} = \{A_{\succ} : \ \succ \text{ is a strict linear order over } K\}.$$

A *(multi-valued) aggregation rule* is a correspondence $F$ which to every profile of 'individual' judgment sets $(A_1, ..., A_n)$ (from some domain, usually $\mathcal{D}^n$) assigns a set $F(A_1, ..., A_n)$ of 'collective' judgment sets. Typically, the output $F(A_1, ..., A_n)$ is a singleton set $\{C\}$, in which case we identify this set with $C$ and write $F(A_1, ..., A_n) = C$. If $F(A_1, ..., A_n)$ contains more than one judgment set, there is a 'tie' between these judgment sets. An aggregation rule is called *single-valued* or *tie-free* if it always generates a single judgment set. A standard (single-valued) aggregation rule is *majority rule*; it is given by

$$F(A_1, ..., A_n) = \{p \in X : p \in A_i \text{ for more than half of the individuals } i\}$$

and generates inconsistent collective judgment sets for many agendas and profiles. If both individual and collective judgment sets are rational (i.e., in $\mathcal{D}$), the aggregation rule defines a correspondences $\mathcal{D}^n \rightrightarrows \mathcal{D}$, and in the case of single-valuedness a function $\mathcal{D}^n \to \mathcal{D}$.[5]

$(X, \neg, \mathcal{C})$ or, equivalently, as the structure $(X, \neg, \mathcal{D})$. A future challenge is to relax the conditions C1-C3 by studying, e.g., judgment aggregation in non-monotonic logics (in which case $\mathcal{C}$ must be the primitive).

[5] More generally, dropping the requirement of collective rationality, we have a correspondence $\mathcal{D}^n \rightrightarrows 2^X$, where $2^X$ is the set of *all* judgment sets, rational or not. As usual, I write '$\rightrightarrows$' instead of '$\to$' to indicate a *multi*-function.

# 3 Scoring rules

Scoring rules are particular judgment aggregation rules, defined on the basis of a so-called scoring function. A *scoring function* – or simply a *scoring* – is a function $s : X \times \mathcal{D} \to \mathbb{R}$ which to each proposition $p$ and rational judgment set $A$ assigns a number $s_A(p)$, called the *score* of $p$ given $A$ and measuring how $p$ performs ('scores') from the perspective of holding judgment set $A$. As an elementary example, so-called *simple scoring* is given by:

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

so that all accepted propositions score 1, whereas all rejected propositions score 0. This and many other scorings will be analysed. Let us think of the score of a *set* of propositions as the sum of the scores of its members. So, the scoring $s$ is extended to a function which (given the agent's judgment set $A \in \mathcal{D}$) assigns to each set $C \subseteq X$ the score

$$s_A(C) \equiv \sum_{p \in C} s_A(p).$$

A scoring $s$ gives rise to an aggregation rule, called the *scoring rule w.r.t.* $s$ and denoted $F_s$. Given a profile $(A_1, ..., A_n) \in \mathcal{D}^n$, this rule determines the collective judgments by selecting the rational judgment set(s) with the highest sum-total score across all judgments and all individuals:

$$\begin{aligned} F_s(A_1, ..., A_n) &= \text{judgment set(s) in } \mathcal{D} \text{ with highest total score} \\ &= \text{argmax}_{C \in \mathcal{D}} \sum_{p \in C, i \in N} s_{A_i}(p) = \text{argmax}_{C \in \mathcal{D}} \sum_{i \in N} s_{A_i}(C). \end{aligned}$$

By a *scoring rule* simpliciter we of course mean an aggregation rule which is a scoring rule w.r.t. some scoring. Different scorings $s$ and $s'$ can generate the same scoring rule $F_s = F_{s'}$, in which case they are called *equivalent*. For instance, $s$ is equivalent to $s' = 2s$.[6]

## 3.1 Simple scoring and the Hamming rule

We first consider the most elementary definition of scoring, namely *simple* scoring (1). Table 2 illustrates the corresponding scoring rule $F_s$ for the case of the agenda and profile of our doctrinal paradox example. The entries in Table 2 are derived as follows. First, enter

| | Score of... | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $p$ | $\neg p$ | $q$ | $\neg q$ | $r$ | $\neg r$ | $pqr$ | $p\neg q\neg r$ | $\neg pq\neg r$ | $\neg p\neg q\neg r$ |
| Indiv. 1 ($pqr$) | 1 | 0 | 1 | 0 | 1 | 0 | 3 | 1 | 1 | 0 |
| Indiv. 2 ($p\neg q\neg r$) | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 3 | 1 | 2 |
| Indiv. 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 |

Table 2: Simple scoring (1) for the doctrinal paradox agenda and profile

the score of each proposition $(p, \neg p, q, ...)$ from each individual (1, 2 and 3). Second, enter

[6] More generally, certain increasing transformations have no effect. As one may show, scorings $s$ and $s'$ are equivalent (i.e., $F_s = F_{s'}$) whenever there are coefficients $a > 0$ and $b_p \in \mathbb{R}$ ($p \in X$) with $b_p = b_{\neg p}$ for all $p \in X$ such that $s'$ is given by $s'_A(p) = as_A(p) + b_p$.

each individual's score of each judgment set by taking the row-wise sum. For instance, individual 1's score of $pqr$ is $1+1+1=3$, and his score of $p\neg q\neg r$ is $1+0+0=1$. Third, enter the group's score of each proposition by taking the column-wise sum. For instance, the group's score of $p$ is $1+1+0=2$. Finally, enter the group's score of each judgment set, by taking either a vertical or a horizontal sum (the two give the same result), and add a star '*' in the field(s) with maximal score to indicate the winning judgment set(s). For instance, the group's score of $pqr$ using a vertical sum is $3+1+1=5$, and using a horizontal sum it is $2+2+1=5$. Since the judgment sets $pqr$, $p\neg q\neg r$ and $\neg pq\neg r$ all have maximal group score, the scoring rule delivers a tie:

$$F(A_1, A_2, A_3) = \{pqr, p\neg q\neg r, \neg pq\neg r\}.$$

This is a tie between the premise-based outcome $pqr$ and the conclusion-based outcomes $p\neg q\neg r$ and $\neg pq\neg r$. Were we to add more individuals, the tie would presumably be broken in one way or the other. In large groups, ties are a rare coincidence.

To link simple scoring to distance-based aggregation, suppose we measure the distance between two rational judgment sets by using some *distance function* ('metric') $d$ over $\mathcal{D}$.[7] The most common example is *Hamming distance* $d = d_{\text{Ham}}$, defined as follows (where by a 'judgment reversal' I mean the replacement of an accepted proposition by its negation):

$$\begin{aligned} d_{\text{Ham}}(A,B) &= \text{number of judgment reversals needed to transform } A \text{ into } B \\ &= |A \backslash B| = |B \backslash A| = \frac{1}{2}\,|A \triangle B|\,. \end{aligned}$$

For instance, the Hamming-distance between $pqr$ and $p\neg q\neg r$ (for our doctrinal paradox agenda) is 2.

Now the *distance-based rule* w.r.t. distance $d$ is the aggregation rule $F_d$ which for any profile $(A_1, ..., A_n) \in \mathcal{D}^n$ determines the collective judgment set(s) by minimizing the sum-total distance to the individual judgment sets:

$$\begin{aligned} F_d(A_1, ..., A_n) &= \text{judgment set(s) in } \mathcal{D} \text{ with minimal sum-distance to the profile} \\ &= \text{argmin}_{C \in \mathcal{D}} \sum_{i \in N} d(C, A_i). \end{aligned}$$

The most popular example, *Hamming rule* $F_{d_{\text{Ham}}}$, can be characterized as a scoring rule:

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

## 3.2 Classical scoring rules for preference aggregation

I now show that our scoring rules generalize the classical scoring rules of preference aggregation theory. Consider the preference agenda $X$ for a given set of alternatives $K$ of finite size $k$. Classical scoring rules (such as Borda rule) are defined by assigning scores to alternatives in $K$, not to propositions $xPy$ in $X$. Given a strict linear order $\succ$ over $K$, each alternative $x \in K$ is assigned a score $SCO_{\succ}(x) \in \mathbb{R}$. The most popular example is of course *Borda scoring*, for which the highest ranked alternative in $K$ scores $k$, the second-highest $k-1$, the third-highest $k-2$, ..., and the lowest 1. Given a profile $(\succ_1, ..., \succ_n)$

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

of individual preferences (strict linear orders), the collective ranks the alternatives $x \in X$ according to their sum-total score $\sum_{i \in N} SCO_{\succ_i}(x)$. To translate this into the judgment aggregation formalism, recall that each strict linear order $\succ$ over $K$ uniquely corresponds to a rational judgment set $A \in \mathcal{D}$ (given by $xPy \in A \Leftrightarrow x \succ y$); we may therefore write $SCO_A(x)$ instead of $SCO_{\succ}(x)$, and view the classical scoring $SCO$ as a function of $(x, A)$ in $K \times \mathcal{D}$. Formally, I define a *classical scoring* as an arbitrary function $SCO : K \times \mathcal{D} \to \mathbb{R}$, and the *classical scoring rule* w.r.t. it as the judgment aggregation rule $F \equiv F_{SCO}$ for the preference agenda which for every profile $(A_1, ..., A_n) \in \mathcal{D}^n$ returns the rational judgment set(s) that rank an alternative $x$ over another $y$ whenever $x$ has a higher sum-total score than $y$:[8]

$$F(A_1, ..., A_n) = \{C \in \mathcal{D} : C \text{ contains all } xPy \in X \text{ s.t. } \sum_{i \in N} SCO_{A_i}(x) > \sum_{i \in N} SCO_{A_i}(y)\}.$$

Now, any given classical scoring $SCO$ induces a scoring $s$ in our (proposition-based) sense. In fact, there are two canonical (and, as we will see, equivalent) ways to define $s$: one might define $s$ either by

$$s_A(xPy) = SCO_A(x) - SCO_A(y), \tag{2}$$

or, if one would like the lowest achievable score to be zero, by

$$s_A(xPy) = \max\{SCO_A(x) - SCO_A(y), 0\} = \begin{cases} SCO_A(x) - SCO_A(y) & \text{if } xPy \in A \\ 0 & \text{if } xPy \notin A \end{cases} \tag{3}$$

(where the last equality assumes that $SCO_A(x) > SCO_A(y) \Leftrightarrow xPy \in A$ for all $x$, $y$ and $A$, a property that is so natural that we might have built it into the definition of a 'classical scoring' $SCO$). This allows us to characterize classical scoring rules in terms of proposition-based rather than alternative-based scoring:

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

### 3.3 Reversal scoring and a Borda rule for judgment aggregation

Given the agent's judgment set $A$, let us think of the score of a proposition $p \in X$ as a measure of how 'distant' the negation $\neg p$ is from $A$; so, $p$ scores high if $\neg p$ is far from $A$, and low if $\neg p$ is contained in $A$. More precisely, let the score of a proposition $p$ given $A \in \mathcal{D}$ be the number of judgment reversals needed to reject $p$, i.e., the number of propositions in $A$ that must (minimally) be negated in order to obtain a consistent judgment set containing $\neg p$. So, denoting the judgment set arising from $A$ by negating the propositions in a subset

[8] A technical difference between the standard notion of a scoring rule in preference aggregation theory and our judgment-theoretic rendition of it arises when there happen to exist distinct alternatives with identical sum-total score. In such cases, the standard scoring rule returns collective *indifferences*, whereas our $F_{SCO}$ returns a *tie* between *strict* preferences. From a formal perspective, however, the two definitions are equivalent, since to any weak order corresponds the set (tie) of all strict linear orders which linearize the weak order by breaking its indifferences (in any cycle-free way). The structural asymmetry between input and output preferences of scoring rules as defined standardly (i.e., the possibility of indifferences at the collective level) may have been one of the obstacles – albeit only a small, mainly psychological one – for importing scoring rules and Borda aggregation into judgment aggregation theory.

$R \subseteq A$ by $A_{\neg R} = (A \backslash R) \cup \{\neg r : r \in R\}$, so-called *reversal scoring* is defined by

$$\begin{aligned} s_A(p) &= \text{number of judgment reversals needed to reject } p \\ &= \min_{R \subseteq A: A_{\neg R} \in \mathcal{D} \& p \notin A_{\neg R}} |R| = \min_{A' \in \mathcal{D}: p \notin A'} |A \backslash A'| = \min_{A' \in \mathcal{D}: p \notin A'} d_{\mathrm{Ham}}(A, A'). \end{aligned} \tag{4}$$

For instance, a rejected proposition $p \notin A$ scores zero, since $A$ itself contains $\neg p$ so that it suffices to negate *zero* propositions ($R = \varnothing$). An accepted proposition $p \in A$ scores 1 if $A$ remains consistent by negating $p$ ($R = \{p\}$), and scores more than 1 otherwise ($R \supsetneq \{p\}$). Table 3 illustrates reversal scoring for our doctrinal paradox example. For instance, individual 1's judgment set $pqr$ leads to a score of 2 for proposition $p$, since in order for him to reject $p$ he needs to negate not just $p$ (as $\neg pqr$ is inconsistent), but also $r$ (where $\neg pq\neg r$ is consistent). The scoring rule delivers a tie between the judgment sets

| | Score of... | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $p$ | $\neg p$ | $q$ | $\neg q$ | $r$ | $\neg r$ | $pqr$ | $p\neg q\neg r$ | $\neg pq\neg r$ | $\neg p\neg q\neg r$ |
| Indiv. 1 ($pqr$) | 2 | 0 | 2 | 0 | 2 | 0 | 6 | 2 | 2 | 0 |
| Indiv. 2 ($p\neg q\neg r$) | 1 | 0 | 0 | 2 | 0 | 2 | 1 | 5 | 2 | 4 |
| Indiv. 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 |

Table 3: Reversal scoring (4) for the doctrinal paradox agenda and profile

$p\neg q\neg r$ and $\neg pq\neg r$. This is a tie between two conclusion-based outcomes; the premise-based outcome $pqr$ is rejected (unlike for simple scoring in Section 3.1).

The remarkable feature of reversal scoring rule is that it generalizes Borda rule from preference to judgment aggregation. Borda rule is initially only defined for the preference agenda $X$ (for a given finite set of alternatives), namely as the classical scoring rule w.r.t. Borda scoring; see the last subsection. The key observation is that reversal scoring is intimately linked to Borda scoring:

**Remark 1** *In the case of the preference agenda (for any finite set of alternatives), reversal scoring $s$ is given by (3) with $SCO$ defined as classical Borda scoring.*

Let me sketch the simple argument – it should sound familiar to social choice theorists. Let $s$ be reversal scoring, $X$ the preference agenda for a set of alternatives $K$ of size $k < \infty$, and $SCO$ classical Borda scoring. Consider any $xPy \in X$ and $A \in \mathcal{D}$. If $xPy \in X \backslash A$, then $\neg xPy = yPx \in A$, which implies $s_A(xPy) = 0$, as required by (3). Now suppose $xPy \in A$. Clearly, $SCO_A(x) > SCO_A(y)$. Consider the alternatives in the order $\succ$ established by $A$:

$$x_k \succ x_{k-1} \succ \cdots \succ x \succ \cdots \succ y \succ \cdots \succ x_1,$$

where $x_j$ is the alternative with $SCO_A(x_j) = j$. Step by step, we now move $y$ up in the ranking, where each step consists in raising the position (score) of $y$ by one. Each step corresponds to negating one proposition in $A$, namely the proposition $zPy$ where $z$ is the alternative that is currently being 'overtaken' by $y$. After exactly $SCO_A(x) - SCO_A(y)$ steps, $y$ has 'overtaken' $x$, i.e., $xPy$ has been negated. So, $s_A(xPy)$ is *at most* $SCO_A(x) - SCO_A(y)$. It is *exactly* $SCO_A(x) - SCO_A(y)$, since, as the reader may check, no smaller number of judgment reversals allows $y$ to 'overtake' $x$ in the ranking.

Remark 1 and Proposition 2 imply that reversal scoring allows us to extend Borda rule to arbitrary judgment aggregation problems:

**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).*

I note that one could use a perfectly equivalent variant of reversal scoring $s$ which, in the case of the preference agenda, is related to classical Borda scoring $SCO$ via (2) instead of (3):

**Remark 2** *Reversal scoring $s$ is equivalent (in terms of the resulting scoring rule) to the scoring $s'$ given by*

$$s'_A(p) = s_A(p) - s_A(\neg p) = \begin{cases} s_A(p) & \text{if } p \in A \\ -s_A(\neg p) & \text{if } p \notin A, \end{cases}$$

*and in the case of the preference agenda (for any finite set of alternatives) this scoring is given by*

$$s'_A(xPy) = SCO_A(x) - SCO_A(y)$$

*with $SCO$ defined as classical Borda scoring.*

For comparison, I now sketch Zwicker's (2011) interesting approach to extending Borda rule to judgment aggregation – let me call such an extension a 'Borda-Zwicker' rule. The motivation derives from a geometric characterization of Borda preference aggregation obtained by Zwicker (1991). Let me write the agenda as $X = \{p_1, \neg p_1, p_2, \neg p_2, ..., p_m, \neg p_m\}$, where $m$ is the number of 'issues'. Each profile gives rise to a vector $\mathbf{v} \equiv (v_1, ..., v_m)$ in $\mathbb{R}^m$ whose $j^{\text{th}}$ entry $v_j$ is the *net support for* $p_j$, i.e., the number of individuals accepting $p_j$ minus the same number for $\neg p_j$. Now if $X$ is the preference agenda for any finite set of alternatives $K$, then each $p_j$ takes the form $xPy$ for certain alternatives $x, y \in K$. Each preference cycle can be mapped to a vector in $\mathbb{R}^m$; for instance, if $p_1 = xPy$, $p_2 = yPz$ and $p_3 = xPz$, then the cycle $x \succ y \succ z \succ x$ becomes the vector $(1, 1, -1, 0, ..., 0) \in \mathbb{R}^m$. The linear span of all vectors corresponding to preference cycles defines the so-called 'cycle space' $\mathbf{V}_{cycle} \subseteq \mathbb{R}^m$, and its orthogonal complement defines the 'cocycle space' $\mathbf{V}_{cocycle} \subseteq \mathbb{R}^m$. Let $\mathbf{v}_{cocycle}$ be the orthogonal projection of $\mathbf{v}$ on the cocycle space $\mathbf{V}_{cocycle}$. Intuitively, $\mathbf{v}_{cocycle}$ contains the 'consistent' or 'acyclic' part of $\mathbf{v}$. The upshot is that the Borda outcome can be read off from $\mathbf{v}_{cocyle}$: for each $p_j = xPy$, the Borda group preference ranks $x$ above (below) $y$ if the $j^{\text{th}}$ entry of $\mathbf{v}_{cocyle}$ is positive (negative). Zwicker's strategy for extending Borda rule to judgment aggregation is to define a subspace $\mathbf{V}_{cycle}$ analogously for agendas other than the preference agenda; one can then again project $\mathbf{v}$ on the orthogonal complement of $\mathbf{V}_{cycle}$ and determine collective 'Borda' judgments according to the signs of the entries of this projection. This approach has proved successful for simple agendas, in which there is a natural way to define $\mathbf{V}_{cycle}$. Whether the approach is viable for general agendas (i.e., whether $\mathbf{V}_{cycle}$ has a useful general definition) seems to be open so far.[9]

A Borda-Zwicker rule is not just constructed differently from a scoring rule in our sense, but, as I conjecture, it also cannot generally be remodelled as a scoring rule, since most interesting scoring rules use information that goes beyond the information contained in the profile's 'net support vector' $\mathbf{v} \in \mathbb{R}^m$. (Even more does the required information go beyond the projection of $\mathbf{v}$ on the orthogonal complement of $\mathbf{V}_{cycle}$.)

[9] One might at first be tempted to generally define $\mathbf{V}_{cycle}$ as the linear span of those vectors which correspond to the agenda's *minimal inconsistent subsets*. Unfortunately, this span is often the entire space $\mathbb{R}^m$, an example for this being our doctrinal paradox agenda.

In summary, there seem to exist two quite different approaches to generalizing Borda aggregation. One approach, taken by Zwicker, seeks to filter out the profile's 'inconsistent component' along the lines of the just-described geometric technique. The other approach, taken here, seeks to retain the principle of score-maximization inherent in Borda aggregation (with scoring now defined at the level of propositions, not alternatives, as these do not exist outside the world of preferences). The normative core of the scoring approach is to use information about someone's *strength* of accepting a proposition (as measured by the score), just as Borda preference aggregation uses information about someone's *strength* of preferring one alternative $x$ over another $y$ (as measured by the score of $xPy$, i.e., the difference between $x$'s and $y$'s score). Whether strength or intensity of preference is a permissible or even meaningful concept is a notoriously controversial question; the purely ordinalist approach takes a sceptical stance here. This is where Borda preference aggregation differs from Condorcet's rule of pairwise majority voting, which uses only the (ordinal) information of *whether* someone prefers an alternative over another, without attempting to extract strength-of-preference information from that person's full preference relation.

### 3.4 A generalization of reversal scoring

Recall that the reversal score of a proposition $p$ can be characterized as the distance by which one must deviate from the current judgment set in order to reject $p$ – where 'distance' is understood as Hamming-distance. It is natural to also consider other kinds of a distance. Relative to any given distance function $d$ over $\mathcal{D}$, one may define a corresponding scoring by

$$\begin{aligned} s_A(p) &= \text{distance by which one must depart from } A \text{ to reject } p \\ &= \min_{A' \in \mathcal{D}: p \notin A'} d(A, A'). \end{aligned} \tag{5}$$

This provides us with 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.

Interestingly, if we adopt Duddy and Piggins' (2011) distance function, i.e., if $d(A, A')$ is the number of *minimal consistent modifications* needed to transform $A$ into $A'$,[10] then scoring (5) reduces to simple scoring (1), and so the scoring rule reduces to the Hamming rule by Proposition 1. So, ironically, while Duddy and Piggins had introduced their distance in the different context of distance-based aggregation to develop an alternative to Hamming rule, when we use their distance (instead of Hamming's) in our context of scoring rules we are led back to Hamming rule.

### 3.5 Scoring based on logical entrenchment

We now consider scoring rules which explicitly exploit the logical structure of the agenda. Let us think of the score of a proposition $p$ ($\in X$) given the judgment set $A$ ($\in \mathcal{D}$) as the degree to which $p$ is logically entrenched in the belief system $A$, i.e., as the 'strength' with

[10] Judgment sets $A, B \in \mathcal{D}$ are *minimal consistent modifications* of each other if the set $S = A \backslash B$ of propositions in $A$ which need to be negated to transform $A$ into $B$ is non-empty and minimal (i.e., $A$ couldn't have been transformed into a consistent set by negating only a strict non-empty subset of $S$). For our doctrinal paradox agenda, the judgment sets $pqr$ and $p\neg q\neg r$ are minimal consistent modifications of each other, and hence have Duddy-Piggins-distance of 1.

which $A$ entails $p$. We measure this strength by the number of ways in which $p$ is entailed by $A$, where each 'way' is given by a particular judgment *subset* $S \subseteq A$ which entails $p$, i.e., for which $S \cup \{\neg p\}$ is inconsistent. If $A$ does not contain $p$, then no judgment subset – not even the full set $A$ – can entail $p$; so the strength of entailment (score) of $p$ is zero. If $A$ contains $p$, then $p$ is entailed by the judgment subset $\{p\}$, and perhaps also by very different judgment subsets; so the strength of entailment (score) of $p$ is positive and more or less high.

There are different ways to formalise this idea, depending on precisely which of the judgment subsets that entail $p$ are deemed relevant. I now propose four formalizations. Two of them will once again allow us to generalize Borda rule from preference to judgment aggregation. These generalizations differ from that based on reversal scoring in Section 3.3.

Our first, naive approach is to 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:

$$\begin{aligned} s_A(p) &= \text{number of judgment subsets which entail } p \\ &= |\{S \subseteq A : S \text{ entails } p\}|. \end{aligned} \tag{6}$$

If $p \notin A$ then $s_A(p) = 0$, while if $p \in A$ then $s_A(p) \geq 2^{|X|/2-1}$ since $p$ is entailed by at least all sets $S \subseteq A$ which contain $p$, i.e., by at least $2^{|A|-1} = 2^{|X|/2-1}$ sets. One might object that this definition of scoring involves redundancies, i.e., 'multiple counting'. Suppose for instance $p$ belongs to $A$ and is logically independent of all other propositions in $A$. Then $p$ is entailed by several subsets $S$ of $A$ – all $S \subseteq A$ which contain $p$ – and yet these entailments are essentially identical since all premises in $S$ other than $p$ are irrelevant.

I now present three refinements of scoring (6), each of which responds differently to the mentioned redundancy objection. In the first refinement, we count two entailments of $p$ as different only if they have no premise in common. This leads to what I call *disjoint-entailment scoring*, formally defined by:

$$\begin{aligned} s_A(p) &= \text{number of } \textit{mutually disjoint} \text{ judgment subsets entailing } p \\ &= \max\{m : A \text{ has } m \text{ mutually disjoint subsets each entailing } p\}. \end{aligned} \tag{7}$$

In the mentioned case where $p$ ($\in A$) is logically independent of all other propositions in $A$, we now avoid 'multiple counting': $s_A(p)$ is only 1, as one cannot find different mutually disjoint judgment subsets entailing $p$. For our doctrinal paradox agenda and profile, the scoring rule delivers a tie between the two conclusion-based outcomes $p\neg q\neg r$ and $\neg pq\neg r$,

| | Score of... | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $p$ | $\neg p$ | $q$ | $\neg q$ | $r$ | $\neg r$ | $pqr$ | $p\neg q\neg r$ | $\neg pq\neg r$ | $\neg p\neg q\neg r$ |
| Indiv. 1 ($pqr$) | 2 | 0 | 2 | 0 | 2 | 0 | 6 | 2 | 2 | 0 |
| Indiv. 2 ($p\neg q\neg r$) | 1 | 0 | 0 | 2 | 0 | 2 | 1 | 5 | 2 | 4 |
| Indiv. 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 |

Table 4: Disjoint-entailment scoring (7) for the doctrinal paradox agenda and profile

as illustrated in Table 4. For instance, individual 2 has judgment set $p\neg q\neg r$, so that $p$ sores 1 (it is entailed by $\{p\}$ but by no other disjoint judgment subset), $\neg q$ scores 2 (it

is disjointly entailed by $\{\neg q\}$ and $\{p, \neg r\}$), $\neg r$ scores 2 (it is disjointly entailed by $\{\neg r\}$ and $\{\neg q\}$), and all rejected propositions score zero (they are not entailed by any judgment subsets).

Disjoint-entailment scoring turns out to match reversal scoring for our doctrinal paradox agenda (check that Tables 3 and 4 coincide), as well as for the preference agenda (as shown later). Is this pure coincidence? The general relationship is that the disjoint-entailment score of a proposition $p$ is always *at most* the reversal score, as one may show.[11]

While this refinement of naive entailment scoring (6) avoids 'multiple counting' by only counting entailments with mutually disjoint sets of premises, the next two refinements use a different strategy to avoid 'multiple counting'. The new strategy is to count only those entailments whose sets of premises are *minimal* – with minimality understood either in the sense that no premises can be removed, or in the sense that no premises can be logically weakened. To begin with the first sense of minimality, I say that a set *minimally entails* $p$ ($\in X$) if it entails $p$ but no strict subset of it entails $p$, and I define *minimal-entailment scoring* by

$$\begin{aligned} s_A(p) &= \text{number of judgment subsets which } \textit{minimally} \text{ entail } p \\ &= \left|\{S \subseteq A : S \text{ minimally entails } p\}\right|. \end{aligned} \tag{8}$$

If for instance $p$ is contained in $A$, then $\{p\}$ minimally entails $p$,[12] but strict supersets of $\{p\}$ do not and are therefore not counted. For our doctrinal paradox agenda, this scoring happens to coincide with reversal scoring and disjoint-entailment scoring. Indeed, Table 3 resp. 4 still applies; e.g., for individual 2 with judgment set $p\neg q\neg r$, $p$ still scores 1 (it is minimally entailed only by $\{p\}$), $\neg q$ still scores 2 (it is minimally entailed by $\{\neg q\}$ and by $\{p, \neg r\}$), $\neg r$ still scores 2 (it is minimally entailed by $\{\neg r\}$ and by $\{\neg q\}$), and all rejected propositions still score zero (they are not minimally entailed by any judgment subsets).

Scoring (8) is certainly appealing. Nonetheless, one might complain that it still allows for certain redundancies, albeit of a different kind. Consider the preference agenda with set of alternatives $K = \{x, y, z, w\}$, and the judgment set $A = \{xPy,\ yPz,\ zPw,\ xPz,\ yPw,\ xPw\}$ ($\in \mathcal{D}$). The proposition $xPw$ is minimally entailed by the subset $S = \{xPy, yPz, zPw\}$. While this entailment is minimal in the (set-theoretic) sense that we cannot *remove* premises, it is non-minimal in the (logical) sense that we can *weaken* some of its premises: if we replace $xPy$ and $yPz$ in $S$ by their logical implication $xPz$, then we obtain a weaker set of premises $S' = \{xPz, zPw\}$ which still entails $xPw$. We shall say that $S$ fails to 'irreducibly' entail $xPw$, in spite of minimally entailing it. In general, a set of propositions is called *weaker* than another one (which is called *stronger*) if the second set entails each member of the first set, but not vice versa. A set $S$ ($\subseteq X$) is defined to *irreducibly* (or *logically minimally*) entail $p$ if $S$ entails $p$, and moreover there is no subset $Y \subsetneq S$ which can be weakened (i.e., for which there is a weaker set $Y' \subseteq X$ such that $(S \backslash Y) \cup Y'$ still entails $p$). Each irreducible entailment is a minimal entailment, as is seen by taking $Y' = \varnothing$.[13] In the previous example, the set $\{xPy, yPz, zPw\}$ minimally, but not irreducibly entails $xPw$, and the set $\{xPz, zPw\}$ irreducibly entails $xPw$.

[11] The reason is that, given $m$ mutually disjoint judgment subsets which each entail $p$, the reversal score of $p$ is at least $m$ since one must negate at least one proposition from each of these $m$ sets in order to consistently reject $p$.

[12] Assuming that $p$ is not a tautology, i.e., that $\{\neg p\}$ is consistent. (Otherwise, $\varnothing$ minimally entails $p$.)

[13] Assuming $X$ contains no tautology, i.e., no $p$ such that $\{\neg p\}$ is inconsistent.

*Irreducible-entailment scoring* is naturally defined by

$$\begin{aligned} s_A(p) &= \text{number of judgment subsets which } \textit{irreducibly} \text{ entail } p \\ &= |\{S \subseteq A : S \text{ irreducibly entails } p\}|. \end{aligned} \tag{9}$$

This scoring matches reversal scoring and both previous scorings in the case of our doctrinal paradox example: Table 3 resp. 4 still applies. But for many other agendas these scorings all deviate from one another, resulting in different collective judgments. As for the preference agenda, we have already announced the following result:

**Proposition 4** *Disjoint-entailment scoring (7) and irreducible-entailment scoring (9) match reversal scoring (4) in the case of the preference agenda (for any finite set of alternatives).*

Propositions 3 and 4 jointly have an immediate corollary.

**Corollary 1** *The scoring rules w.r.t. scorings (7) and (9) both generalize Borda rule, i.e., match it in the case of the preference agenda (for any finite set of alternatives).*

### 3.6 Propositionwise scoring and a way to repair quota rules with non-rational outputs

We now consider a special class of scorings: *propositionwise* scorings. This will allow us to relate scoring rules to the well-known judgment aggregation rules called *quota rules* – in fact, to 'repair' these rules by rendering their outcomes rational across all profiles.

I call scoring $s$ *propositionwise* if the score of a proposition $p \in X$ only depends on whether $p$ is accepted, i.e., if $s_A(p) = s_B(p)$ whenever $A$ and $B$ (in $\mathcal{D}$) both contain $p$ or both do not contain $p$. Equivalently, scoring is propositionwise just in case for each $p \in X$ there is a pair of real numbers $s_+(p), s_-(p)$ such that

$$s_A(p) = \begin{cases} s_+(p) & \text{for all } A \in \mathcal{D} \text{ containing } p \\ s_-(p) & \text{for all } A \in \mathcal{D} \text{ not containing } p. \end{cases} \tag{10}$$

Intuitively, $s_+(p)$ is the score of an accepted proposition $p$, and $s_-(p)$ is the score of a rejected proposition $p$. Typically, of course, $s_+(p) > s_-(p)$. An example is simple scoring: there, $s_+(p) = 1$ and $s_-(p) = 0$.

How do propositionwise scoring rules behave? They derive a proposition $p$'s sum-total score 'locally', i.e., based only on people's judgments about $p$. This property stands in obvious analogy to a well-studied axiom on aggregation rules, namely the axiom of *propositionwise* or *independent* aggregation, which prescribes that the collective judgment about any given proposition $p$ is derived 'locally', i.e., again based only on people's judgments about $p$. Can we therefore relate propositionwise scoring to independent aggregation? The paradigmatic independent aggregation rules are the *quota rules*.[14] A quota rule is a (single-valued) aggregation rule which is given by an acceptance threshold $m_p \in \{1, ..., n\}$ for each proposition $p \in X$. The quota rule corresponding to the so-called *threshold family* $(m_p)_{p \in X}$ is denoted $F_{(m_p)_{p \in X}}$ and accepts those propositions $p$ which are supported by at least $m_p$ individuals: for each profile $(A_1, ..., A_n) \in \mathcal{D}^n$,

$$F_{(m_p)_{p \in X}}(A_1, ..., A_n) = \{p \in X : |\{i : p \in A_i\}| \geq m_p\}.$$

[14] They are the only independent rules which are anonymous, monotonic and unanimity-preserving.

Special cases are unanimity rule (given by $m_p = n$ for all $p$), majority rule (given by the majority threshold $m_p = \lceil (n+1)/2 \rceil$ for all $p$), and more generally, *uniform* quota rules (given by a uniform threshold $m_p \equiv m$ for all $p$). A uniform quota rules is also referred to as a supermajority rule if $m$ exceeds the majority threshold, and a submajority rule if $m$ is below the majority threshold. Note that supermajority rules may generate incomplete collective judgment sets, while submajority rule may accept both members of a pair $p, \neg p \in X$, a drastic form of inconsistency. If one wishes that exactly one member of each pair $p, \neg p \in X$ is accepted, the thresholds of $p$ and $\neg p$ should be 'complements' of each other: $m_p = n + 1 - m_{\neg p}$.

A non-trivial question is how the acceptance thresholds would have to be set to ensure that the collective judgment set satisfies some given degree of rationality, such as to be (i) consistent, or (ii) deductively closed, or (iii) consistent and deductively closed, or even (iv) fully rational, i.e., in $\mathcal{D}$. These questions have been settled (see Nehring and Puppe 2010a for (iv), and, subsequently, Dietrich and List 2007b for (i)-(iv)). Unfortunately, for many agendas the thresholds would have to be set at 'extreme' and normatively unattractive levels. Worse, often *no* thresholds achieve (iv) (see Nehring and Puppe 2010a). For our doctrinal paradox agenda $X = \{p, q, r\}^{\pm}$ only the extreme thresholds $m_p = m_q = m_r = n$ and $m_{\neg p} = m_{\neg q} = m_{\neg r} = 1$ achieve (iv), and for the preference agenda (with more than two alternatives) *no* thresholds achieve (iv).

Given that quota rules with 'reasonable' thresholds typically violate many of the conditions (i)-(iv), one may want to depart from ordinary quota rules by modifying ('repairing') them so that they always generate rational outputs. This can be done by using propositionwise scoring rules. Given an arbitrary quota rule with threshold family $(m_p)_{p \in X}$, one can specify a propositionwise scoring such that the scoring rule replicates the quota rule whenever the quota rule generates a rational output, while 'repairing' the output otherwise. How must we calibrate $s_+(p)$ and $s_-(p)$ in order to achieve this? The idea is that individuals who accept $p$ should contribute a positive score $s_+(p) > 0$, while those who reject $p$ should contribute a negative score $s_-(p) < 0$. The absolute sizes of $s_+(p)$ and $s_-(p)$ should be calibrated such that the sum-total score of $p$ becomes positive (helping the scoring rule to accept $p$) exactly when the quota rule accepts $p$, i.e., when at least $m_p$ individuals accept $p$. Specifically, we set:

$$s_A(p) = \begin{cases} s_+(p) = n + 1 - m_p & \text{for all } A \in \mathcal{D} \text{ containing } p \\ s_-(p) = -m_p & \text{for all } A \in \mathcal{D} \text{ not containing } p. \end{cases} \tag{11}$$

Intuitively, the higher the acceptance threshold $m_p$ is, the smaller the positive contribution $s_+(p)$ is and the larger the negative contribution $s_-(p)$ is (in absolute value); hence, the more individuals accepting $p$ are needed for $p$'s sum-total score to get positive, and the harder it becomes for the scoring rule to accept $p$. This scoring does the intended job:

**Proposition 5** *For every threshold family $(m_p)_{p \in X}$, the scoring rule w.r.t. scoring (11) matches the quota rule $F_{(m_p)_{p \in X}}$ at all profiles where the quota rule generates rational outputs (and still generates rational outputs at all other profiles).*

As an example, consider our doctrinal paradox agenda $X = \{p, q, r\}^{\pm}$ with $n = 3$ individuals, and suppose the quota rule departs only slightly from propositionwise majority voting: all propositions $t$ in $X \backslash \{\neg r\}$ keep a majority threshold of $m_t = 2$, but $\neg r$ receives a unanimity threshold $m_{\neg r} = 3$. This quota rule manages to never generate logically

inconsistent collective judgment sets,[15] but does so at the expense of allowing collective incompleteness. Indeed, for our example profile, the quota rule returns the collective judgment set $pq$, which is silent on the choice between $r$ nor $\neg r$. As illustrated in Table 5, the scoring rule w.r.t. (11) restores collective rationality by leading to the premise-based

| | Score of... | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $p$ | $\neg p$ | $q$ | $\neg q$ | $r$ | $\neg r$ | $pqr$ | $p\neg q\neg r$ | $\neg pq\neg r$ | $\neg p\neg q\neg r$ |
| Indiv. 1 ($pqr$) | 2 | -2 | 2 | -2 | 2 | -3 | 6 | -3 | -3 | -7 |
| Indiv. 2 ($p\neg q\neg r$) | 2 | -2 | -2 | 2 | -2 | 1 | -2 | 5 | -3 | 1 |
| Indiv. 3 ($\neg pq\neg r$) | -2 | 2 | 2 | -2 | -2 | 1 | -2 | -3 | 5 | 1 |
| Group | 2 | -2 | 2 | -2 | -2 | -1 | 2* | -1 | -1 | -5 |

Table 5: Scoring (11) for the doctrinal paradox agenda and profile

outcome $pqr$. To read the table, note that scoring (11) is given by $s_+(t) = 2$ and $s_-(t) = -2$ for all $t$ in $X\backslash\{\neg r\}$, $s_+(\neg r) = 1$ and $s_-(\neg r) = -3$.

How does our scoring rule 'repair' those special quota rules which use a uniform threshold $m \equiv m_p$ ($p \in X$), such as majority rule?

**Remark 3** *For a uniform threshold $m \equiv m_p$, the scoring rule w.r.t. scoring (11) is the Hamming rule, or equivalently, the simple scoring rule.*

This remark follows from Proposition 1 and the fact that, for a uniform threshold $m \equiv m_p$, scoring (11) is equivalent to simple scoring by footnote 6.

Finally, I note that the scoring rules w.r.t. (11) is not the only scoring rule which can 'repair' the quota rule $F_{(m_p)_{p\in X}}$ – though it might be the most plausible one, as long as we do not wish to introduce additional parameters. If, however, we are prepared to introduce additional parameters, scoring (11) can be generalized: for each $p \in X$ let $\alpha_p > 0$ be a coefficient measuring how important it is that the scoring rule is faithful to the quota rule's collective judgment on $p$; and let scoring be defined by

$$s_A(p) = \begin{cases} s_+(p) = \alpha_p(n + 1 - m_p) & \text{if } p \in A \\ s_-(p) = -\alpha_p m_p & \text{if } p \notin A. \end{cases} \tag{12}$$

The earlier scoring (11) is obviously a special case in which all $\alpha_p$ are 1. Proposition 5 still holds for this generalized kind of propositionwise scoring. The scoring rule will tend to match the quota rule on propositions $p$ with high importance coefficient $\alpha_p$, while modifying ('repairing') the quota rule at propositions $p$ with low $\alpha_p$.

## 3.7 Premise- and conclusion-based aggregation

I have just mentioned the possibility of a differential treatment of propositions when 'repairing' a quota rule. This possibility is particularly salient in the popular context of premise- or conclusion-based aggregation.[16] One may indeed view the classical premise- and conclusion-based rules as two (rival) ways of repairing the simplest of all quota rules –

[15] This follows from Nehring and Puppe's (2010) *intersection property*, generalized to possibly incomplete collective judgment sets (Dietrich and List 2007b).

[16] See for instance List (2004), Dietrich and Mongin (2010) and Nehring and Puppe (2010b).

majority rule – by privileging certain propositions over others, namely premise propositions or conclusion propositions, respectively.

Let me put this precisely. Consider majority voting, i.e., the quota rule with a uniform majority threshold $m \equiv m_p$ (the smallest integer above $n/2$). To restore collective rationality, we again endow each proposition $p \in X$ with a 'coefficient of importance', but now let this coefficient be determined by whether $p$ has a 'premise' or 'conclusion' status. Formally, suppose the agenda is partitioned into two negation-closed sets, the set $P$ of 'premise propositions' and the set $X \backslash P$ of 'conclusion propositions'. In the case of our doctrinal paradox agenda $X = \{p, q, r\}^{\pm}$, we have $P = \{p, q\}^{\pm}$. Each premise proposition $p \in P$ has the importance coefficient $\alpha_p \equiv \alpha_{\text{premise}}$, and each conclusion proposition $p \in X \backslash P$ has the importance coefficient $\alpha_p \equiv \alpha_{\text{conclusion}}$, for fixed parameters $\alpha_{\text{premise}}, \alpha_{\text{conclusion}} \geq 0$. In this scenario, the scoring (12) becomes equivalent (by footnote 6) to the scoring given by

$$s_A(p) = \begin{cases} \alpha_{\text{premise}} & \text{for accepted premise propositions } p \in A \cap P \\ \alpha_{\text{conclusion}} & \text{for accepted conclusion propositions } p \in A \backslash P \\ 0 & \text{for rejected propositions } p \notin A. \end{cases} \tag{13}$$

By calibrating the two importance coefficients, we can influence the relative weights of premises and conclusions. If we give *far more* importance to premise propositions ($\alpha_{\text{premise}} \gg \alpha_{\text{coclusion}}$) or to conclusion propositions ($\alpha_{\text{coclusion}} \gg \alpha_{\text{premise}}$), the scoring rule reduces to the premise- or conclusion-based rule, respectively. To substantiate this claim, one needs to define both rules. For simplicity, I restrict attention to our doctrinal paradox agenda $X = \{p, q, r\}^{\pm}$ with $P = \{p, q\}^{\pm}$ (though more general $X$ and $P$ could be considered[17]). In this case, assuming for simplicity that the group size $n$ is odd,

- the *premise-based rule* is the aggregation rule which for each profile in $\mathcal{D}^n$ delivers the (unique) judgment set in $\mathcal{D}$ containing each premise proposition accepted by a majority;
- the *conclusion-based rule* is the aggregation rule which for each profile in $\mathcal{D}^n$ delivers the judgment set (or sets) in $\mathcal{D}$ containing the conclusion proposition accepted by a majority.[18]

These two rules have the following characterizations as scoring rules:

**Remark 4** *For our doctrinal paradox agenda $X = \{p, q, r\}^{\pm}$ with set of premise propositions $P = \{p, q\}^{\pm}$, and for an odd group size, the scoring rule w.r.t. scoring (13) is*

- *the premise-based rule if and only if $\alpha_{\text{premise}} > (n-2)\alpha_{\text{conclusion}}$,*
- *the conclusion-based rule if and only if $\alpha_{\text{conclusion}} > \alpha_{\text{premise}} = 0$.*

This result lets premise- and conclusion-based aggregation appear in a rather extreme light: each rule is based on somewhat unequal importance coefficients $\alpha_{\text{premise}}$ and $\alpha_{\text{conclusion}}$, deeming one type of proposition to be overwhelmingly more important than the other. It might therefore be interesting to consider more equilibrated values of the importance coefficients, so as to achieve a compromise between democracy at the premise level and democracy at the conclusion level.

[17] Our analysis generalizes easily to any $X$ and $P$ such that (i) the premise propositions in $P$ are logically independent, and (ii) complete judgments across the premise propositions in $P$ uniquely determine the judgments on the conclusion propositions in $X \backslash P$.

[18] In the literature, the conclusion-based procedure is usually taken to be *silent* on the premises, i.e., to return an incomplete judgment set not in $\mathcal{D}$. I have replaced this silence by a tie between all compatible judgments on the premise propositions.

# 4 Set scoring rules: assigning scores to entire judgment sets

An interesting generalization of scoring rules is obtained by assigning scores directly to entire judgment sets rather than single propositions. A *set scoring function* – or simply *set scoring* – is a function $\sigma$ which to every pair of rational judgment sets $C$ and $A$ assigns a real number $\sigma_A(C)$, the *score* of $C$ given $A$, which measures how well $C$ performs ('scores') from the perspective of holding the judgment set $A$. Formally, $\sigma : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$. The most elementary example, to be called *naive* set scoring, is given by

$$\sigma_A(C) = \begin{cases} 1 & \text{if } C = A \\ 0 & \text{if } C \neq A. \end{cases} \tag{14}$$

Any set scoring $\sigma$ gives rise to an aggregation rule $F_\sigma$, the *set scoring rule* (or *generalized scoring rule*) *w.r.t.* $\sigma$, which for each profile $(A_1, ..., A_n) \in \mathcal{D}^n$ selects the collective judgment set(s) $C$ in $\mathcal{D}$ having maximal sum-total score across individuals:

$$F_\sigma(A_1, ..., A_n) = \text{argmax}_{C \in \mathcal{D}} \sum_{i \in N} \sigma_{A_i}(C).$$

An aggregation rule is a *set scoring rule* simpliciter if it is the set scoring rule w.r.t. to some set scoring $\sigma$. Set scoring rules generalize ordinary scoring rules, since to any ordinary scoring $s$ corresponds a set scoring $\sigma$, given by

$$\sigma_A(C) \equiv \sum_{p \in C} s_A(p),$$

and the ordinary scoring rule w.r.t. $s$ coincides with the set scoring rule w.r.t. $\sigma$.

## 4.1 Naive set scoring and plurality voting

*Plurality rule* is the aggregation rule $F$ which for every profile $(A_1, ..., A_n) \in \mathcal{D}^n$ declares the most often submitted judgment set(s) as the collective judgment set(s):

$$\begin{aligned} F(A_1, ..., A_n) &= \text{most frequently submitted judgment set(s)} \\ &= \text{argmax}_{C \in \mathcal{D}} \left| \{i : A_i = C\} \right|. \end{aligned}$$

This rule is of course normatively questionable;[19] but it deserves our attention, if only because of its simplicity and the recognized importance of plurality voting in social choice theory more broadly. Plurality rule can be construed as a set scoring rule:

**Remark 5** *The naive set scoring rule is plurality rule.*

## 4.2 Distance-based set scoring

Set scoring rules generalize distance-based aggregation. Given an arbitrary distance function $d$ over $\mathcal{D}$ (not necessarily the Hamming-distance), all that is needed is to consider what I call *distance-based* set scoring, defined by

$$\sigma_A(C) = -d(C, A). \tag{15}$$

[19] It ignores the internal structure of judgment sets, hence 'throws away' much information.

So, $C$ scores high if it is close to the judgment set held, $A$. This renders sum-score-maximization equivalent to sum-distance-minimization:

**Remark 6** *For every given distance function over $\mathcal{D}$, the distance-based set scoring rule is the distance-based rule.*

So, all distance-based rules can be modelled as set scoring rules (but not vice versa[20]). As an example, consider the so-called *discrete* distance,[21] defined by

$$d(A,B) = \begin{cases} 0 & \text{if } A = B \\ 1 & \text{if } A \neq B. \end{cases}$$

Here, distance-based set scoring (15) is equivalent to naive set scoring (14), since the two differ only by a constant (of one). So, joining Remarks 5 and 6, we may view plurality rule either as the naive set scoring rule or as the discrete-distance-based rule.

## 4.3 Approximating the individual scores

Given an ordinary scoring $s$, we have so far aimed for collective judgments with high total score. But this is not the only plausible aim or approach. We now turn to an altogether different approach. Rather than using $s$ to assign scores only from each individual's perspective, we now assign scores also from the collective perspective, i.e., from the perspective of the aggregate judgment set. Instead of wanting the collective judgments to achieve highest total individual score, we now want them to approximate the individuals' judgments in the sense that collective scores resemble individual scores. In short, propositions which score high for individuals should score high for the collective, and propositions which score low for individuals should score low for the collective. This new approach has its own intuitive appeal. But is it really totally different? As will turn out, aggregation rules which follow this approach – I call them 'score-approximation rules' as opposed to 'scoring rules' – can be viewed as a particular kind of set scoring rules.

Given an ordinary scoring $s$, we can represent judgment sets in $\mathcal{D}$ as vectors in $\mathbb{R}^X$, by identifying each judgment set $A$ in $\mathcal{D}$ with its *score vector*, i.e., the vector in $\mathbb{R}^X$ whose $p^{\text{th}}$ component is the score of $p$, $s_A(p)$.[22] The score vector corresponding to $A \in \mathcal{D}$ is denoted $A^s \equiv (s_A(p))_{p \in X} \in \mathbb{R}^X$. Having represented judgment sets as vectors of numbers, we can apply standard algebraic and geometric operations, such as adding judgment sets, taking their average, or measuring their distance – where, of course, sums or averages of (score vectors of) judgment sets in $\mathcal{D}$ may be 'infeasible', i.e., not correspond to any judgment set in $\mathcal{D}$.

The *score-approximation rule* w.r.t. scoring $s$ is defined as the aggregation rule $F$ which for every profile $(A_1, ..., A_n) \in \mathcal{D}^n$ chooses the collective judgment set(s) whose

[20] In trying to re-model an arbitrary set scoring rule $F_\sigma$ as a distance-based rule, one might be tempted to define the 'distance' between $A$ and $B$ as $d_\sigma(A,B) := \sigma_A(A) - \sigma_A(B)$. If $d_\sigma$ turns out to define a proper distance function (see fn. 7), then we obtain a distance-based rule $F_{d_\sigma}$, which coincides with the set scoring rule $F_\sigma$. But for many plausible set scorings $\sigma$, $d_\sigma$ has little in common with a distance function, violating up to all three axioms, notably symmetry and the triangle inequality.

[21] This metric derives its name from the fact that it induces the discrete topology on whatever set it is defined on (such as $\mathbb{R}$ instead of $\mathcal{D}$).

[22] This identification is one-to-one as long as the scoring has the (very plausible) property that $s_A(p) > s_A(\neg p)$ whenever $p \in A$.

score vector comes closest to the group's average score vector $\frac{1}{n}\sum_{i\in N} A_i^s$ in the sense of Euclidean distance in $\mathbb{R}^X$:

$$\begin{aligned} F(A_1,...,A_n) &= \text{j.s. closest to the average individual j.s. in score vector terms} \\ &= \text{argmin}_{C\in\mathcal{D}} \left\| C^s - \frac{1}{n}\sum_{i\in N} A_i^s \right\|. \end{aligned}$$

Viewed geometrically as an operation in $\mathbb{R}^X$, the collective score vector is the orthogonal projection of the average score vector $\frac{1}{n}\sum_i A_i^s$ on the set $\mathcal{D}^s \equiv \{A^s : A \in \mathcal{D}\} \subseteq \mathbb{R}^X$ of feasible score vectors.[23]

As an illustration, consider once again reversal scoring for our doctrinal paradox agenda. Table 6 reports the score vector of each judgment set (including the one not submitted by

| | $p$ | $\neg p$ | $q$ | $\neg q$ | $r$ | $\neg r$ | distance to group's average |
|---|---|---|---|---|---|---|---|
| $pqr$ (indiv. 1) | 2 | 0 | 2 | 0 | 2 | 0 | $\sqrt{58}/3 \approx 2.54$ |
| $p\neg q\neg r$ (indiv. 2) | 1 | 0 | 0 | 2 | 0 | 2 | $\sqrt{37}/3 \approx 2.03$ |
| $\neg pq\neg r$ (indiv. 3) | 0 | 2 | 1 | 0 | 0 | 2 | $\sqrt{37}/3 \approx 2.03$ |
| $\neg p\neg q\neg r$ (no indiv.) | 0 | 1 | 0 | 1 | 0 | 3 | $7/3 \approx 2.33$ |
| group's average | 1 | $\frac{2}{3}$ | 1 | $\frac{2}{3}$ | $\frac{2}{3}$ | $\frac{4}{3}$ | |

Table 6: The score-approximation rule (w.r.t. reversal scoring) for the doctrinal paradox agenda and profile

any individual), and its distance to the group's average score vector. By minimizing this distance, the rule delivers a tie between the two conclusion-based outcomes $p\neg q\neg r$ and $\neg pq\neg r$. The premise-based outcome $pqr$ looks worse than ever: it is even farther from the average than the never-submitted outcome $\neg p\neg q\neg r$.

Now that we have two rival ways of aggregating based on a scoring $s$ – namely, the scoring rule and the score-approximation rule – the question is whether any connection can be established. The score-approximation rule can be construed as a *set* scoring rule, namely in virtue of the set scoring given by

$$\sigma_A(C) = -\left\|C^s - A^s\right\|^2. \tag{16}$$

Here, $C$ is taken to score high if it is close to $A$ in terms of the *squared* Euclidean distance of score vectors.

**Proposition 6** *For any scoring $s$, the score-approximation rule w.r.t. $s$ is the set scoring rule w.r.t. set scoring (16).*

As an application, let $s$ be simple scoring (1). Here, the set scoring (16) is expressible as an increasing affine transformation of the set scoring corresponding to simple scoring, i.e., of the set scoring $\sigma'$ given by[24]

$$\sigma'_A(C) = \sum_{p\in C} s_A(p) = |C\cap A|.$$

[23] Formally, $F(A_1,...,A_n)^s = \text{PROJ}_{\mathcal{D}^s}(\frac{1}{n}\sum_i A_i^s)$, where the orthogonal projection of $x\in\mathbb{R}^X$ on $Y\subseteq\mathbb{R}^X$ is defined as $\text{PROJ}_Y(x) := \arg\min_{y\in Y}\|y-x\|$.

[24] Since $\sigma_A(C) = -\left(\sqrt{|C\triangle A|}\right)^2 = -|C\triangle A| = -2|C\backslash A| = -2(|C| - |C\cap A|) = -|X| + 2|C\cap A|$.

So, the set scoring rule $F_\sigma$ coincides with the simple scoring rule $F_s$, and hence with the Hamming rule $F_{d_{\text{Ham}}}$ by Proposition 1. Thus, as a corollary of Propositions 1 and 6, the Hamming rule can be characterized not just as a scoring rule but also as a score-approximation rule, both times using the same scoring:

**Corollary 2** *The Hamming rule is the scoring rule and the score-approximation rule, both times w.r.t. simple scoring.*

## 4.4 Probability-based set scoring

I close the analysis by taking a brief (skippable) excursion into an important, but different approach to judgment aggregation: the *epistemic* or *truth-tracking* approach. In this approach, each proposition $p \in X$ is taken to have an objective, but unknown truth value ('true' or 'false'), and the goal of aggregation is to track the truth, i.e., to generate true collective judgments.[25] The truth-tracking perspective has a long history elsewhere in social choice theory (e.g., Condorcet 1785, Grofman *et al.* 1983, Austen-Smith and Banks 1996, Dietrich 2006b, Pivato 2011); but within judgment aggregation theory specifically, rather little work has been done on the epistemic side (e.g., Bovens and Rabinowicz 2006b, List 2005, Bozbay *et al.* 2011).

The epistemic approach warrants the use of particular set scoring rules. To show this, I import standard statistical estimation techniques (such as maximum-likelihood estimation), following the path taken by other authors in the context of preference aggregation (e.g., Young 1995) and other aggregation problems (e.g., Dietrich 2006b, Pivato 2011). My goal is to give no more than a brief introduction to what could be done. The results given below are essentially variants of existing results; see in particular Pivato (2011).[26]

For each combination $(A_1, ..., A_n, T) \in \mathcal{D}^n \times \mathcal{D}$ of $n+1$ judgment sets, let $\Pr(A_1, ..., A_n, T) > 0$ measure the probability that people submit the profile $(A_1, ..., A_n)$ *and* the set of true propositions is $T$, where of course $\sum_{(A_1,...,A_n,T)\in\mathcal{D}^n\times\mathcal{D}} \Pr(A_1, ..., A_n, T) = 1$. From this joint probability function we can, as usual, derive various marginal and conditional probabilities, such as the probability that the truth is $T \in \mathcal{D}$, $\Pr(T) = \sum_{(A_1,...,A_n)\in\mathcal{D}^n} \Pr(A_1, ..., A_n, T)$, the probability that the profile is $(A_1, ..., A_n)$, $\Pr(A_1, ..., A_n) = \sum_{T\in\mathcal{D}} \Pr(A_1, ..., A_n, T)$, the conditional probability $\Pr(T|A_1, ..., A_n) = \frac{\Pr(A_1,...,A_n,T)}{\Pr(A_1,...,A_n)}$ (called the *posterior* probability of $T$ given the 'data' $A_1, ..., A_n$), and the conditional probability $\Pr(A_1, ..., A_n|T) = \frac{\Pr(A_1,...,A_n,T)}{\Pr(T)}$ (called the *likelihood* of the 'data' $A_1, ..., A_n$ given $T$).

The *maximum-likelihood rule* is the aggregation rule $F : \mathcal{D}^n \rightrightarrows \mathcal{D}$ which for each profile $(A_1, ..., A_n) \in \mathcal{D}^n$ defines the collective judgments such that their truth would make the observed profile ('data') maximally likely:

$$F(A_1, ..., A_n) = \text{argmax}_{T\in\mathcal{D}} \Pr(A_1, ..., A_n|T).$$

The *maximum-posterior rule* is the aggregation rule $F : \mathcal{D}^n \rightrightarrows \mathcal{D}$ which for each profile $(A_1, ..., A_n) \in \mathcal{D}^n$ defines the collective judgments such that they have maximal posterior

[25] The epistemic perspective is usually contrasted with the *procedural* perspective, which takes the goal of aggregation to be to generate collective judgments which reflect the individuals' judgments in a procedurally fair way. To illustrate the contrast between the two perspectives, suppose that all individuals hold the same judgment set $A$. Then $A$ is clearly the right collective judgment set from the perspective of procedural fairness. But from an epistemic perspective, all depends on whether people's unanimous endorsement of $A$ is sufficient evidence for $A$ being true.

[26] Proposition 7 follows from proofs in Pivato (2011), and is also related to Dietrich (2006).

probability of truth conditional on the observed profile ('data'):

$$F(A_1, ..., A_n) = \text{argmax}_{T \in \mathcal{D}} \Pr(T|A_1, ..., A_n).$$

Both of these rules correspond to well-established statistical estimation procedures.

Let us now make two standard, but restrictive assumptions on probabilities. We assume that voters are 'independent' and 'equally competent' (in analogy to the assumptions of Condorcet's classical jury theorem[27]). Formally, for every $T \in \mathcal{D}$,

(IND) the individual judgment sets are independent conditional on $T$ being the true judgment set, i.e., $\Pr(A_1, ..., A_n|T) = \Pr(A_1|T) \cdots \Pr(A_n|T)$ for all $A_1, ..., A_n \in \mathcal{D}$ ('*independence*')

(COM) for each $A \in \mathcal{D}$, each individual has the same probability, denoted $\Pr(A|T)$, of submitting the judgment set $A$ conditional on $T$ being the true judgment set ('*equal competence*').

Condition (COM) in particular implies that individuals have the same (conditional) probability of holding the *true* judgment set; but nothing is assumed about the size of this probability of 'getting it right'. The just-defined aggregation rules turn out to be set scoring rules in virtue of defining the score of $T \in \mathcal{D}$ given $A \in \mathcal{D}$ by, respectively,

$$\sigma_A(T) = \log \Pr(A|T) \tag{17}$$

$$\sigma_A(T) = \log \Pr(A|T) + \frac{1}{n} \log \Pr(T). \tag{18}$$

**Proposition 7** *If voters are independent (IND) and equally competent (COM), then*

- *the maximum-likelihood rule is the set scoring rule w.r.t. set scoring (17),*
- *the maximum-posterior rule is the set scoring w.r.t. set scoring (18).*

# 5 Concluding remarks

I hope to have convinced the reader that scoring rules, and more generally set scoring rules, form interesting positive solutions to the judgment aggregation problem. They for instance allow us to generalize Borda aggregation to judgment aggregation (the simplest method being to use reversal scoring). Figure 1 summarizes where we stand by depicting different classes of rules (scoring rules, set scoring rules, and distance-based rules) and positioning several concrete rules (such as Hamming rule). While the positions of most rules in Figure 1 have been established above or follow easily, a few positions are of the order of conjectures. This is so for the placement of our Borda generalization *outside* the class of distance-based rules.[28]

Though several old and new aggregation rules are scoring rules (or at least set scoring rules), there are important counterexamples. One counterexample is the mentioned rule

[27] The classical Condorcet jury theorem is essentially concerned with a simple judgment aggregation problem with a binary agenda $X = \{p, \neg p\}$.

[28] For technical correctness, I also note two details about how to read Figure 1. First, for trivial agendas, such as a single-issue agenda $X = \{p, \neg p\}$, several rules of course become equivalent, and distinctions drawn in Figure 1 disappear. More precisely, by positioning a rule outside a class of rules (e.g., by positioning plurality rule outside the class of scoring rules), I am of course not implying that for *all* agendas the rule does not belong to the class, but that for some (in fact, *most*) agendas this is so. Second, in placing propositionwise scoring rules among the distance-based rules, I made a very plausible restriction: $s_+(p) > s_-(p)$ for each $p \in X$.

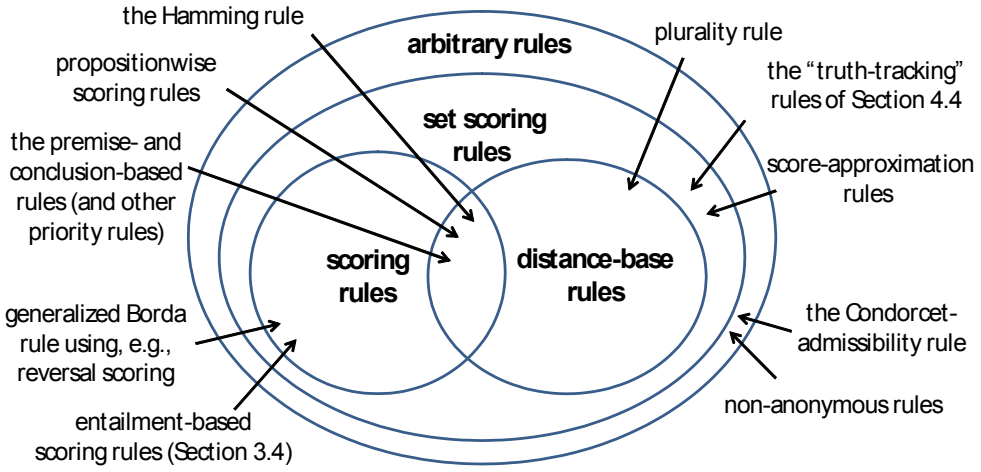


Figure 1: A map of judgment aggregation possibilities

introduced by Nehring *et al.* (2011) (the so-called Condorcet-admissibility rule, which generates rational judgment set(s) that 'approximate' the majority judgment set). Other counterexamples are non-anonymous rules (such as rules prioritizing experts), and rules that return boundedly rational collective judgments (such as rules returning incomplete but still consistent and deductively closed judgments). The last two kinds of counterexamples suggest two generalizations of the notion of a scoring rule. Firstly, scoring might be allowed to depend on the individual; this leads to 'non-anonymous scoring rules'. Secondly, the search for a collective judgment set with maximal total score might be done within a larger set than the set $\mathcal{D}$ of fully rational judgment sets (such as the set of consistent but possibly incomplete judgment sets); this leads to 'boundedly rational scoring rules'. The same generalizations could of course be made for set scoring rules. Much work is ahead of us.

# 6 References


Austen-Smith, D., Banks, J. (1996) Information Aggregation, Rationality, and the Condorcet Jury Theorem, *American Political Science Review* 90: 34–45

Bovens, L., Rabinowicz, W. (2006) Democratic answers to complex questions: an epistemic perspective, *Synthese* 150(1): 131-153

Bozbay, I., Dietrich, F., Peters, H. (2011) Judgment aggregation in search for the truth, working paper, London School of Economics

Condorcet, Marquis, D. (1785) *Essai sur l'application de l'analyse á la probabilité des décisions rendues á la pluralité des voix*

Dietrich, F. (2006a) Judgment aggregation: (im)possibility theorems, *Journal of Economic Theory* 126(1): 286-298

Dietrich, F. (2006b) General representation of epistemically optimal procedures, *Social Choice and Welfare* 26(2): 263-283

Dietrich, F. (2010) The possibility of judgment aggregation on agendas with subjunctive implications, *Journal of Economic Theory* 145(2): 603-638

Dietrich, F., List, C. (2007a) Arrow's theorem in judgment aggregation, *Social Choice*

*and Welfare* 29(1): 19-33

Dietrich, F., List, C. (2007b) Judgment aggregation by quota rules: majority voting generalized, *Journal of Theoretical Politics* 19(4): 391-424

Dietrich, F., List, C. (2010) Majority voting on restricted domains, *Journal of Economic Theory* 145(2): 512-543

Dietrich, F., Mongin, P. (2010) The premise-based approach to judgment aggregation, *Journal of Economic Theory* 145(2): 562-582

Duddy, C., Piggins, A. (2011) A measure of distance between judgment sets, working paper, National University of Ireland Galway

Grofman, B., Owen, G., Feld, S. L. (1983) Thirteen Theorems in Search of the Truth, *Theory and Decision* 15: 261–278

List, C. (2004) A Model of Path Dependence in Decisions over Multiple Propositions, *American Political Science Review* 98(3): 495-513

List, C. (2005) The probability of inconsistencies in complex collective decisions, *Social Choice and Welfare* 24(1): 3-32

List, C., Pettit, P. (2002) Aggregating sets of judgments: an impossibility result, *Economics and Philosophy* 18(1): 89-110

List, C., Polak, B. eds. (2010) *Symposium on Judgment Aggregation*, Journal of Economic Theory 145(2)

Myerson, R. B. (1995) Axiomatic derivation of scoring rules without the ordering assumption, *Social Choice and Welfare* 12 (1): 59–74

Nehring, K., Puppe, C. (2010a) Abstract Arrowian Aggregation, *Journal of Economic Theory* 145: 467-494

Nehring, K., Puppe, C. (2010b) Justifiable Group Choice, *Journal of Economic Theory* 145: 583-602

Nehring, K., Pivato, M., Puppe, C. (2011) Condorcet admissibility: indeterminacy and path-dependence under majority voting on interconnected decisions, working paper, University of California at Davis

Pivato, M. (2011) Voting rules as statistical estimators, Working Paper, Trent University, Canada

Wilson R (1975) On the Theory of Aggregation, *Journal of Economic Theory* 10: 89-99

Young, H. P. (1995) Optimal voting rules, *Journal of Economic Perspectives* 9 (1): 51–64

Zwicker, W. (1991) The voters' paradox, spin, and the Borda count, *Mathematical Social Sciences* 22, 187-227

Zwicker, W. (2011) Towards a "Borda count" for judgment aggregation, extended abstract, presented at the conference 'Judgment Aggregation and Voting', Karlsruhe Institute of Technology

# 7 Appendix: proofs

*Proof of Proposition* 1. The Hamming-distance between $A, C \in \mathcal{D}$ can be written as

$$d_{\text{Ham}}(A,C) = \frac{1}{2}\left|A \triangle B\right| = \frac{1}{2}(|X| - (|A \cap B| + \left|\overline{A} \cap \overline{B}\right|)).$$

Now, since $A$ and $C$ each contains exactly one member of each pair $\{p, \neg p\} \subseteq X$, we have $p \in A \cap C \Leftrightarrow \neg p \in \overline{A} \cap \overline{C}$, and so, $|A \cap C| = \left|\overline{A} \cap \overline{C}\right|$. Hence, $d_{\text{Ham}}(A,C) = \frac{1}{2}|X| - |A \cap C|$. So, for each profile $(A_1, ..., A_n) \in \mathcal{D}^n$, minimizing $\sum_{i \in N} d_{\text{Ham}}(A_i, C)$ is

equivalent to maximizing $\sum_{i\in N}|A_i\cap C|$. Hence, rewriting each $|A_i\cap C|$ as $\sum_{p\in C}s_{A_i}(p)$ where $s$ is simple scoring (1), it follows that $F_{d_{\text{Ham}}}(A_1,...,A_n)=F_s(A_1,...,A_n)$. ■

Before proving Proposition 2, I start with a lemma.

**Lemma 1** *Consider the preference agenda (for any finite set of alternatives $K$), any classical scoring $SCO$, and the scoring $s$ given by (3). For all distinct $x,y\in K$ and all $A\in\mathcal{D}$,*

$$SCO_A(x)-SCO_A(y)=s_A(xPy)-s_A(yPx). \tag{19}$$

*Proof.* This follows easily from (3). ■

Two elements of a set of alternatives $K$ are called *neighbours* w.r.t. a strict linear order $\succ$ over $K$ if they differ and no alternative in $K$ is ranked strictly between them. In the case of the preference agenda (for a set of alternatives $K$), the strict linear order over $K$ corresponding to any $A\in\mathcal{D}$ is denoted $\succ_A$.

*Proof of Proposition* 2. Consider the preference agenda $X$ for a set of alternatives $K$ of finite size $k$, and let $SCO$ be any classical scoring. I show that $F_{SCO}=F_s$ for each scoring $s$ satisfying (19), and hence for the scoring (3) (since it satisfies (19) by Lemma 1) and the scoring (2) (since a half times it satisfies (19)).

Consider any scoring $s$ satisfying (19). Fix a profile $(A_1,...,A_n)\in\mathcal{D}^n$; I show $F_s(A_1,...,A_n)=F_{SCO}(A_1,...,A_n)$. The proof is in three claims.

*Claim 1.* For all $a,b\in K$ and $C,C'\in\mathcal{D}$, if $C\backslash C'=\{aPb\}$, then

$$\sum_{i\in N}SCO_{A_i}(a)-\sum_{i\in N}SCO_{A_i}(b)=\sum_{i\in N,p\in C}s_{A_i}(p)-\sum_{i\in N,p\in C'}s_{A_i}(p).$$

Consider $a,b\in K$ and $C,C'\in\mathcal{D}$ such that $C\backslash C'=\{aPb\}$. For each individual $i\in N$, we by (19) have

$$SCO_{A_i}(a)-SCO_{A_i}(b)=s_{A_i}(aPb)-s_{A_i}(bPa),$$

which, noting that $C'=(C\backslash\{aPb\})\cup\{bPa\}$, implies that

$$SCO_{A_i}(a)-SCO_{A_i}(b)=\sum_{p\in C}s_{A_i}(p)-\sum_{p\in C'}s_{A_i}(p).$$

Summing over all individuals, the claim follows, q.e.d.

*Claim 2.* $F_s(A_1,...,A_n)\subseteq F_{SCO}(A_1,...,A_n)$.

Consider any $C\in F_s(A_1,...,A_n)$. We have to show that $C\in F_{SCO}(A_1,...,A_n)$, i.e., that for all distinct $x,y\in K$,

$$\sum_{i\in N}SCO_{A_i}(x)>\sum_{i\in N}SCO_{A_i}(y)\Rightarrow xPy\in C,$$

or equivalently,

$$yPx\in C\Rightarrow\sum_{i\in N}SCO_{A_i}(y)\geq\sum_{i\in N}SCO_{A_i}(x).$$

Said in yet another way, we have to show that

$$\sum_{i\in N}SCO_{A_i}(x_k)\geq\sum_{i\in N}SCO_{A_i}(x_{k-1})\geq\cdots\geq\sum_{i\in N}SCO_{A_i}(x_1),$$

where I have labelled the alternatives $x_1, x_2, ..., x_k$ such that $x_k \succ_C x_{k-1} \succ_C \cdots \succ_C x_1$. Consider any $t \in \{1, ..., k-1\}$, and write $a$ for $x_{t+1}$ and $b$ for $x_t$. Let $C'$ be the judgment set arising from $C$ by replacing $aPb$ with its negation $bPa$. Now $C' \in \mathcal{D}$; this is because $a$ and $b$ are neighbours w.r.t. $\succ_C$, which guarantees that $C'$ corresponds to a strict linear order (namely to the same one as for $C$ except that $b$ now ranks above $a$). Since $C \in F_s(A_1, ..., A_n)$, $C$ has maximal sum-total score within $\mathcal{D}$; in particular,

$$\sum_{i \in N, p \in C} s_{A_i}(p) \geq \sum_{i \in N, p \in C'} s_{A_i}(p),$$

which by Claim 1 implies the desired inequality,

$$\sum_{i \in N} SCO_{A_i}(a) \geq \sum_{i \in N} SCO_{A_i}(b), \text{ q.e.d.}$$

*Claim 3.* $F_{SCO}(A_1, ..., A_n) \subseteq F_s(A_1, ..., A_n)$.

Consider any $C \in F_{SCO}(A_1, ..., A_n)$. To show that $C \in F_s(A_1, ..., A_n)$, we consider an arbitrary $C' \in \mathcal{D} \backslash \{C\}$ and have to show that $C$ has an at least as high sum-total score as $C'$:

$$\sum_{i \in N, p \in C} s_{A_i}(p) \geq \sum_{i \in N, p \in C'} s_{A_i}(p). \tag{20}$$

To prove this, we first transform $C$ gradually into $C'$ in $m \equiv |C' \backslash C|$ steps, where each step consists in a single judgment reversal, i.e., in the replacement of a single proposition $xPy$ ($\in C \backslash C'$) by its negation $yPx$ ($\in C' \backslash C$). This defines a sequence of judgment sets $C_0, ..., C_m$, where $C_0 = C$ and $C_m = C'$, and where for each step $t \in \{1, ..., m\}$ there is a proposition $x_t P y_t$ such that $C_t = (C_{t-1} \backslash \{x_t P y_t\}) \cup \{y_t P x_t\}$. Note that $\{x_t P y_t : t = 1, ..., m\} = C \backslash C'$. By a standard relation-theoretic argument, we may assume that in each step $t$ the judgment reversal consists in switching the relative order of two *neighbouring* alternatives; i.e., $x_t, y_t$ are neighbours w.r.t. the old and new relations $\succ_{C_{t-1}}$ and $\succ_{C_t}$. This guarantees that each step $t$ generates a set $C_t$ such that $\succ_{C_t}$ is still a strict linear order, i.e., such that $C_t \in \mathcal{D}$.

Now for each step $t$, by Claim 1 we have

$$\sum_{i \in N} SCO_{A_i}(x_t) - \sum_{i \in N} SCO_{A_i}(y_t) = \sum_{i \in N, p \in C_{t-1}} s_{A_i}(p) - \sum_{i \in N, p \in C_t} s_{A_i}(p),$$

and also, since $y_t P x_t \notin C$ and $C \in F_{SCO}(A_1, ..., A_n)$, we have

$$\sum_{i \in N} SCO_{A_i}(y_t) \leq \sum_{i \in N} SCO_{A_i}(x_t);$$

it follows that

$$\sum_{i \in N, p \in C_{t-1}} s_{A_i}(p) - \sum_{i \in N, p \in C_t} s_{A_i}(p) \geq 0.$$

Summing this inequality over all steps $t \in \{1, ..., m\}$, we obtain

$$\sum_{i \in N, p \in C_0} s_{A_i}(p) - \sum_{i \in N, p \in C_m} s_{A_i}(p) \geq 0,$$

which is equivalent to the desired inequality (20) since $C_0 = C$ and $C_m = C'$. ■

*Proof of Remark 2.* Let $s'$ be defined from reversal scoring $s$ in the specified way.

*Claim 1.* $s'$ and $s$ are equivalent.

Consider any profile $(A_1, ..., A_n) \in \mathcal{D}^n$. I show for all $C, D \in \mathcal{D}$ that

$$\sum_{i \in N, p \in C} s_{A_i}(p) \geq \sum_{i \in N, p \in D} s_{A_i}(p) \Leftrightarrow \sum_{i \in N, p \in C} s'_{A_i}(p) \geq \sum_{i \in N, p \in D} s'_{A_i}(p).$$

Consider any $C, D \in \mathcal{D}$. I prove that $\Delta \geq 0 \Leftrightarrow \Delta' \geq 0$, where

$$\begin{aligned} \Delta &\equiv \sum_{i \in N, p \in C} s_{A_i}(p) - \sum_{i \in N, p \in D} s_{A_i}(p) \geq 0, \\ \Delta' &\equiv \sum_{i \in N, p \in C} s'_{A_i}(p) - \sum_{i \in N, p \in D} s'_{A_i}(p) \geq 0. \end{aligned}$$

We have

$$\Delta = \sum_{i \in N} \left\{ \sum_{p \in C} s_{A_i}(p) - \sum_{p \in D} s_{A_i}(p) \right\} = \sum_{i \in N} \left\{ \sum_{p \in C \setminus D} s_{A_i}(p) - \sum_{p \in D \setminus C} s_{A_i}(p) \right\}.$$

So, noting that $D \setminus C = \{\neg p : p \in C \setminus D\}$, we obtain

$$\Delta = \sum_{i \in N} \sum_{p \in C \setminus D} (s_{A_i}(p) - s_{A_i}(\neg p)).$$

By an analogous reasoning,

$$\Delta' = \sum_{i \in N} \sum_{p \in C \setminus D} (s'_{A_i}(p) - s'_{A_i}(\neg p)).$$

Hence, using the definition of $s'$,

$$\begin{aligned} \Delta' &= \sum_{i \in N} \sum_{p \in C \setminus D} ([s_{A_i}(p) - s_{A_i}(\neg p)] - [s_{A_i}(\neg p) - s_{A_i}(p)]) \\ &= 2 \sum_{i \in N} \sum_{p \in C \setminus D} (s_{A_i}(p) - s_{A_i}(\neg p)) \\ &= 2\Delta. \end{aligned}$$

So, $\Delta \geq 0 \Leftrightarrow \Delta' \geq 0$, q.e.d.

*Claim 2.* If $X$ is the preference agenda, $SCO$ is classical Borda scoring, $A \in \mathcal{D}$, and $xPy \in X$, then $s'_A(xPy) = SCO_A(x) - SCO_A(y)$.

Let $X$, $SCO$, $A$ and $xPy$ be as specified. If $xPy \in A$, then

$$\begin{aligned} s'(xPy) &= s(xPy) \text{ by definition of } s' \\ &= SCO_A(x) - SCO_A(y) \text{ by Remark 1, as } xPy \in A. \end{aligned}$$

If $xPy \notin A$, i.e., $yPx \in A$, then

$$\begin{aligned} s'(xPy) &= -s(yPx) \text{ by definition of } s' \\ &= -(SCO_A(y) - SCO_A(x)) \text{ by Remark 1, as } yPx \in A \\ &= SCO_A(x) - SCO_A(y). \blacksquare \end{aligned}$$

*Proof of Proposition* 4. Let $X$ be the preference agenda for some set of alternatives $K$ of size $k < \infty$. Let $s^{\text{rev}}$, $s^{\text{dis}}$ and $s^{\text{irr}}$ be reversal, disjoint-entailment, and irreducible-entailment scoring, respectively. Consider any $A \in \mathcal{D}$, denote the corresponding strict linear

order by $\succ$, let $x_1, ..., x_k$ be the alternatives in the order given by $x_k \succ x_{k-1} \succ \cdots \succ x_1$, and consider any $p \in X$, say $p = x_i P x_{i'} \in X$.

*Claim 1.* $s_A^{\text{rev}}(p) = s_A^{\text{irr}}(p)$.

By the argument given in footnote 11, $s_A^{\text{rev}}(p) \geq s_A^{\text{dis}}(p)$. I now show that $s_A^{\text{dis}}(p) \geq s_A^{\text{rev}}(p)$. This inequality is trivial if $p \notin A$, since then $s_A^{\text{rev}}(p) = 0$ (as $\neg p \in A$). Now suppose $p \in A$. By Remark 1, $s_A^{\text{rev}}(p) = i - i'$. So we need to show that $s_A^{\text{dis}}(p) \geq i - i'$. Consider the $i - i'$ judgment subsets $S_1, ..., S_{i-i'} \subseteq A$ defined as follows: for each $j \in \{1, ..., i - i'\}$,

$$S_j \equiv \{x_i P x_{i-j}, x_{i-j} P x_{i'}\} \subseteq A,$$

where $S_{i-i'}$ is interpreted as the set $\{x_i P x_{i'}\}$ (rather than the set $\{x_i P x_{i'}, x_{i'} P x_{i'}\}$, which is not well-defined since $x_{i'} P x_{i'}$ is not a proposition in $X$). Since these judgment subsets are pairwise disjoint and each of them entails $p$ $(= x_i P x_{i'})$, we have $s_A^{\text{dis}}(p) \geq i - i'$, q.e.d.

*Claim 2.* $s_A^{\text{rev}}(p) = s_A^{\text{irr}}(p)$.

If $p \notin A$, then $s_A^{\text{rev}}(p) = s_A^{\text{irr}}(p)$ since $s_A^{\text{rev}}(p) = 0$ (as $\neg p \in A$) and $s_A^{\text{irr}}(p) = 0$ (as $A$ does not entail $p$). Now suppose $p \in A$. Then, as already mentioned, $s_A^{\text{rev}}(p) = i - i'$ by Remark 1. So we need to show that $s_A^{\text{irr}}(p) = i - i'$. As one may show, each of the just-defined sets $S_1, ..., S_{i-i'}$ irreducibly entails $p$ $(= x_i P x_{i'})$. So it remains to show that no other judgment subset irreducibly entails $p$. Suppose $S \subseteq A$ irreducibly entails $p$. I have to show that $S \in \{S_1, ..., S_{i-i'}\}$. As is easily checked, the set $S \cup \{\neg p\}$ $(= S \cup \{x_{i'} P x_i\})$ is minimal inconsistent. Hence, this set is *cyclic*, i.e., of the form $S \cup \{\neg p\} = \{y_1 P y_2, y_2 P y_3, ..., y_{m-1} P y_m, y_m P y_1\}$ for some $m \geq 2$ and some distinct alternatives $y_1, ..., y_m \in K$ (see Dietrich and List 2010). Without loss of generality, assume $y_1 = x_i$ and $y_m = x_{i'}$, so that $y_m P y_1 = x_{i'} P x_i$ and

$$S = \{y_1 P y_2, y_2 P y_3, ..., y_{m-1} P y_m\}.$$

If $m = 2$, then $S = \{y_1 P y_2\} = \{x_i P x_{i'}\}$, which equals $S_{i-i'}$, and we are done. If $m = 3$, then $S = \{y_1 P y_2, y_2 P y_3\} = \{x_i P y_2, y_2 P x_{i'}\}$. Since $S$ is by assumption included in $A$, it follows that $A$ ranks $y_2$ between $x_i$ and $x_{i'}$. So there is a $j \in \{1, ..., i - i' - 1\}$ such that $y_2 = x_{i-j}$. Hence, $S$ is the set $\{x_i P x_{i-j}, x_{i-j} P x_{i'}\} = S_j$, and we are done again. Finally, $m$ cannot exceed 3, since otherwise the set $S$ $(= \{x_i P y_2, y_2 P y_3, ..., y_{m-1} P x_{i'}\})$ would entail $p$ $(= x_i P x_{i'})$ *non-irreducibly*, since the set arising from $S$ by replacing $x_i P y_2$ and $y_2 P y_3$ with their implication $x_i P y_3$ still entails $p$. ■

*Proof of Proposition 5.* Consider any threshold family $(m_p)_{p \in X}$ $(\in \{1, ..., n\}^X)$, and define scoring $s$ by (11). Consider a profile $(A_1, ..., A_n) \in \mathcal{D}^n$ for which $C^* \equiv F_{(m_p)_{p \in X}}(A_1, ..., A_n)$ belongs to $\mathcal{D}$. We have to show that $F_s(A_1, ..., A_n) = C^*$. For each proposition $p \in X$, writing the number of individuals accepting $p$ as $n_p \equiv |\{i : p \in A_i\}|$, the sum-total score of $p$ is given by

$$\begin{aligned} \sum_{i \in N} s_{A_i}(p) &= \sum_{i \in N : p \in A_i} (n + 1 - m_p) + \sum_{i \in N : p \notin A_i} (-m_p) \\ &= n_p(n + 1 - m_p) + (n - n_p)(-m_p) \\ &= nn_p + n_p - nm_p. \\ &= n(n_p - m_p) + n_p; \end{aligned}$$

and so,

$$\sum_{i \in N} s_{A_i}(p) \begin{cases} > 0 & \text{if } n_p \geq m_p, \text{ i.e., if } p \in C^* \\ < 0 & \text{if } n_p < m_p, \text{ i.e., if } p \notin C^*. \end{cases} \tag{21}$$

Now we have $\{C^*\} = \text{argmax}_{C\in\mathcal{D}} \sum_{p\in C, i\in N} s_{A_i}(p)$, because for each $C \in \mathcal{D}\backslash\{C^*\}$,

$$\sum_{p\in C^*, i\in N} s_{A_i}(p) - \sum_{p\in C, i\in N} s_{A_i}(p) = \sum_{p\in C^*\backslash C} \underbrace{\sum_{i\in N} s_{A_i}(p)}_{>0 \text{ by (21)}} - \sum_{p\in C\backslash C^*} \underbrace{\sum_{i\in N} s_{A_i}(p)}_{<0 \text{ by (21)}} > 0.$$

So, $F_s(A_1, ..., A_n) = \{C^*\} \equiv C^*$. ■

*Proof of Remark 4.* Consider this $X$ and $P$, let $n$ be odd, and let $s$ be scoring (13). I write $\alpha_{\text{pr}}$ for $\alpha_{\text{premise}}$ and $\alpha_{\text{co}}$ for $\alpha_{\text{conclusion}}$. Whenever I consider a profile $(A_1, ..., A_n) \in \mathcal{D}^n$, I write $N_t := \{i : t \in A_i\}$ for all $t \in X$, and I write $\mathcal{MAJ}$, $\mathcal{PRE}$, $\mathcal{CON}$ and $\mathcal{SCO}$ for the outcome of majority rule, premise-based rule, conclusion-based rule, and the scoring rule w.r.t. (13), respectively. Note that for all $(A_1, ..., A_n) \in \mathcal{D}^n$ the sum-total score of a $C = \{p', q', r'\} \in \mathcal{D}$ (where $p' \in \{p, \neg p\}$, $q' \in \{q, \neg q\}$ and $r' \in \{r, \neg r\}$) is given by

$$\sum_{i\in N, t\in C} s_{A_i}(t) = (|N_{p'}| + |N_{q'}|)\alpha_{\text{pr}} + |N_r|\,\alpha_{\text{co}}. \tag{22}$$

*Claim 1.* [$\mathcal{PRE} = \mathcal{SCO}$ for all profiles in $\mathcal{D}^n$] if and only if $\alpha_{\text{pr}} > (n-2)\alpha_{\text{co}}$.

First, assume $\mathcal{PRE} = \mathcal{SCO}$ for all profile in $\mathcal{D}^n$. As one may check, there is a profile such that $|N_p| = |N_q| = \frac{n+1}{2}$ and $|N_r| = 1$. For this profile, $\mathcal{PRE} = \{p, q, r\}$. So, $\mathcal{SCO} = \{p, q, r\}$. Hence, the sum-total score of $\{p, q, r\}$ exceeds that of $\{\neg p, q, \neg r\}$. By (22), these two sum-total scores can be written, respectively, as

$$\begin{aligned}
\sum_{i\in N, t\in\{p,q,r\}} s_{A_i}(t) &= \frac{n+1}{2}\alpha_{\text{pr}} + \frac{n+1}{2}\alpha_{\text{pr}} + \alpha_{\text{co}} = (n+1)\alpha_{\text{pr}} + \alpha_{\text{co}} \\
\sum_{i\in N, t\in\{\neg p,q,\neg r\}} s_{A_i}(t) &= \frac{n-1}{2}\alpha_{\text{pr}} + \frac{n+1}{2}\alpha_{\text{pr}} + (n-1)\alpha_{\text{co}} = n\alpha_{\text{pr}} + (n-1)\alpha_{\text{co}}.
\end{aligned}$$

Hence,

$$(n+1)\alpha_{\text{pr}} + \alpha_{\text{co}} > n\alpha_{\text{pr}} + (n-1)\alpha_{\text{co}},$$

or equivalently, $\alpha_{\text{pr}} > (n-2)\alpha_{\text{co}}$.

Conversely, assume $\alpha_{\text{pr}} > (n-2)\alpha_{\text{co}}$. Consider any profile. We have to show that $\mathcal{PRE} = \mathcal{SCO}$.

*Case 1*: $\mathcal{MAJ} \in \mathcal{D}$. Check that it follows that $\mathcal{PRE} = \mathcal{MAJ}$, and also that $\mathcal{SCO} = \mathcal{MAJ}$. So, $\mathcal{PRE} = \mathcal{SCO}$.

*Case 2*: $\mathcal{MAJ} \notin \mathcal{D}$. Check that it follows that $\mathcal{MAJ} = \{p, q, \neg r\}$. Hence $\mathcal{PRE} = \{p, q, r\}$. We thus have to show that $\mathcal{SCO} = \{p, q, r\}$, i.e., that

$$\begin{aligned}
\Delta_1 &\equiv \sum_{i\in N, t\in\{p,q,r\}} s_{A_i}(t) - \sum_{i\in N, t\in\{\neg p,q,\neg r\}} s_{A_i}(t) > 0 \\
\Delta_2 &\equiv \sum_{i\in N, t\in\{p,q,r\}} s_{A_i}(t) - \sum_{i\in N, t\in\{p,\neg q,\neg r\}} s_{A_i}(t) > 0 \\
\Delta_3 &\equiv \sum_{i\in N, t\in\{p,q,r\}} s_{A_i}(t) - \sum_{i\in N, t\in\{\neg p,\neg q,\neg r\}} s_{A_i}(t) > 0.
\end{aligned}$$

By (22),

$$\Delta_1 = (|N_p| - |N_{\neg p}|)\alpha_{\text{pr}} + (|N_r| - |N_{\neg r}|)\alpha_{\text{co}} = (2\,|N_p| - n)\alpha_{\text{pr}} + (2\,|N_r| - n)\alpha_{\text{co}}. \tag{23}$$

In this, as $p \in \mathcal{MAJ}$ we have $|N_p| \geq (n+1)/2$; and further, as $p, q \in \mathcal{MAJ}$ the sets $N_p$ and $N_q$ each contain a majority, so that $N_p \cap N_q \neq \varnothing$, which (since $N_p \cap N_q \subseteq N_r$) implies $|N_r| \geq 1$. Using these lower bounds for $|N_p|$ and $|N_r|$, we obtain

$$\Delta_1 \geq ((n+1)-n)\alpha_{\text{pr}} + (2-n)\alpha_{\text{co}} = \alpha_{\text{pr}} + (2-n)\alpha_{\text{co}} > 0.$$

The proof that $\Delta_2 > 0$ is analogous. Finally, by (22),

$$\Delta_3 = (|N_p| - |N_{\neg p}|)\alpha_{\text{pr}} + (|N_q| - |N_{\neg q}|)\alpha_{\text{pr}} + (|N_r| - |N_{\neg r}|)\alpha_{\text{co}}.$$

Since $|N_q| > |N_{\neg q}|$ (since $q \in \mathcal{MAJ}$), it follows using (23) that $\Delta_3 > \Delta_2$, and hence, that $\Delta_3 > 0$, q.e.d.

*Claim 2.* [$\mathcal{CON} = \mathcal{SCO}$ for all profiles in $\mathcal{D}^n$] if and only if $\alpha_{\text{co}} > \alpha_{\text{pr}} = 0$.

Unlike in the proof of Claim, there may be ties, and so we treat $\mathcal{CON}$ and $\mathcal{SCO}$ as subsets of $\mathcal{D}$, not elements. First, if $\alpha_{\text{co}} > \alpha_{\text{pr}} = 0$, then it is easy to show that $\mathcal{CON} = \mathcal{SCO}$ for each profile. Conversely, suppose it is not the case that $\alpha_{\text{co}} > \alpha_{\text{pr}} = 0$. Then either $\alpha_{\text{co}} = \alpha_{\text{pr}} = 0$ or $\alpha_{\text{pr}} > 0$. In the first case, clearly $\mathcal{CON} \neq \mathcal{SCO}$ for some profiles, since $\mathcal{SCO}$ is always $\mathcal{D}$. In the second case, again $\mathcal{CON} \neq \mathcal{SCO}$ for some profiles: for instance, if each individual submits $\neg pq\neg r$ then $\mathcal{SCO} = \{\neg pq\neg r\}$ while $\mathcal{CON} = \{\neg pq\neg r, p\neg q\neg r, \neg p\neg q\neg r\}$. ■

*Proof of Proposition 6.* It will sometimes be convenient to write a vector $D = (D_1, ..., D_n) \in \mathbb{R}^n$ as $\langle D_i \rangle$. The mean and variance of this vector $D$ are denoted and defined by, respectively,

$$\overline{D} \equiv \frac{1}{n}\sum_{i \in N} D_i \text{ and } Var(D) \equiv \frac{1}{n}\sum_{i \in N}(D_i - \overline{D})^2.$$

In this notation, the average square deviation of a constant $c \in \mathbb{R}$ from the components in $D$ is $\overline{\langle (c - D_i)^2 \rangle}$ and satisfies

$$\overline{\langle (c - D_i)^2 \rangle} = (c - \overline{D})^2 + Var(D), \tag{24}$$

by the following argument borrowed from statistics:

$$\begin{aligned}
\overline{\langle (c - D_i)^2 \rangle} &= \overline{\langle (c - \overline{D} + \overline{D} - D_i)^2 \rangle} \\
&= \overline{\langle (c - \overline{D})^2 + 2(c - \overline{D})(\overline{D} - D_i) + (\overline{D} - D_i)^2 \rangle} \\
&= (c - \overline{D})^2 + 2(c - \overline{D})\overline{\langle \overline{D} - D_i \rangle} + \overline{\langle (\overline{D} - D_i)^2 \rangle} \\
&= (c - \overline{D})^2 + 0 + Var(D).
\end{aligned}$$

Now consider any scoring $s$ and let the set scoring $\sigma$ be defined by (16). Consider any profile $(A_1, ..., A_n) \in \mathcal{D}^n$ and any $C \in \mathcal{D}$. Under $\sigma$, the sum-total score of $C$ can be written as

$$\begin{aligned}
\sum_{i \in N} \sigma_{A_i}(C) &= -\sum_{i \in N} \|C^s - A_i^s\|^2 \\
&= -\sum_{i \in N}\sum_{p \in X} (C_p^s - A_{ip}^s)^2 \\
&= -n \sum_{p \in X} \frac{1}{n}\sum_{i \in N} (C_p^s - A_{ip}^s)^2.
\end{aligned}$$

Here, the inner expression can be re-expressed as

$$\frac{1}{n}\sum_{i\in N}(C_p^s - A_{ip}^s)^2 = \overline{\langle (C_p^s - A_{ip}^s)^2 \rangle} = (C_p^s - \overline{\langle A_{ip}^s \rangle})^2 + Var(\langle A_{ip}^s \rangle),$$

where the last equality applies (24) with $c = C_p^s$ and $D = \langle A_{ip}^s \rangle$. It follows that

$$\begin{aligned}
\sum_{i\in N}\sigma_{A_i}(C) &= -n\sum_{p\in X}\left\{(C_p^s - \overline{\langle A_{ip}^s \rangle})^2 + Var(\langle A_{ip}^s \rangle)\right\} \\
&= -n\sum_{p\in X}(C_p^s - \overline{\langle A_{ip}^s \rangle})^2 + d \text{ (for some } d \text{ independent of } C) \\
&= -n\left\|C - \overline{\langle A_i^s \rangle}\right\|^2 + d.
\end{aligned}$$

Maximizing this expression w.r.t. $C \in \mathcal{D}$ is equivalent to minimizing its strictly decreasing transformation $\left\|C - \overline{\langle A_i^s \rangle}\right\|$ w.r.t. $C \in \mathcal{D}$. So, the set scoring rule w.r.t. $\sigma$ delivers the same collective judgment set(s) $C$ as the score-approximation rule w.r.t. $s$. ■

*Proof of Proposition 7.* Assume (IND) and (COM) and consider a profile $(A_1, ..., A_n) \in \mathcal{D}^n$.

Firstly, using (IND), the likelihood of the profile given $C \in \mathcal{D}$ can be written as

$$\Pr(A_1, ..., A_n|T) = \prod_{i\in N}\Pr(A_i|T).$$

Maximizing this expression (w.r.t. $T \in \mathcal{D}$) is equivalent to maximizing its logarithm,

$$\sum_{i\in N}\log\Pr(A_i|T),$$

which is precisely the sum-total score of $T$ under set scoring (17).

Secondly, writing $\pi$ for the profile's probability $\Pr(A_1, ..., A_n)$, the posterior probability of $T \in \mathcal{D}$ given the profile can be written as

$$\Pr(T|A_1, ..., A_n) = \frac{1}{\pi}\Pr(T)\Pr(A_1, ..., A_n|T) = \frac{1}{\pi}\Pr(T)\prod_{i\in N}\Pr(A_i|T).$$

Maximizing this expression (w.r.t. $T \in \mathcal{D}$) is equivalent to maximizing its logarithm, and hence, to maximizing

$$\log\Pr(T) + \sum_{i\in N}\log\Pr(A_i|T) = \sum_{i\in N}(\log\Pr(A_i|T) + \frac{1}{n}\log\Pr(T)),$$

which is the sum-total score of $T$ under set scoring (18). ■