Group Decision and Negotiation (2023) 32:889–923
https://doi.org/10.1007/s10726-023-09828-y

Optimal Decision Rules for the Discursive Dilemma

Aureli Alabert[1] · **Mercè Farré**[1] · **Rubén Montes**[1]

Accepted: 17 March 2023 / Published online: 27 April 2023
© The Author(s) 2023

Abstract
We consider the classical problem in truth-tracking judgment aggregation of a conjunctive agenda with two premisses and one conclusion. We study this problem from the point of view of finding the best decision rule according to a quantitative criterion, under very mild restrictions on the set of admissible rules. The members of the deciding committee are assumed to have a certain probability to assess correctly the truth or falsity of the premisses, and the best rule is the one that minimises a combination of the probabilities of false positives and false negatives on the conclusion.

Keywords Judgment aggregation · Truth tracking · Discursive dilemma · Doctrinal paradox

1 Introduction

1.1 Statement of the Problem

We consider the problem in judgment aggregation where there are two clauses P and Q and each member of a committee has to decide between P and its negation $\neg P$ and between Q and its negation $\neg Q$; and that the final goal is to assess the truth value of $C := P \wedge Q$.

In practice, the situation appears when a court is deciding if a defendant is guilty (a set of evidences are all verified), or not-guilty (at least one of the evidences is false); or when a prize or a job position is awarded if and only if several debatable conditions concur; or when several subjective medical indicators determine the presence of an illness or the need of a treatment; and so on.

✉ Aureli Alabert
Aureli.Alabert@uab.cat

Mercè Farré
Merce.Farre@uab.cat

[1] Department of Mathematics, Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalonia, Spain

Table 1 The discursive dilemma: The collective majority voting on the three propositions is inconsistent with the doctrine $C \Leftrightarrow P \wedge Q$

Voter	Proposition P	Proposition Q	Proposition C
1	P	Q	C
2	P	$\neg Q$	$\neg C$
3	$\neg P$	Q	$\neg C$
Majority	P	Q	$\neg C$

The so-called *doctrinal paradox* is the fact that two different reasonable majority-type voting rules may lead to different outcomes. Suppose the voters first decide by simple majority between P and $\neg P$, and separately between Q and $\neg Q$. If both P and Q get the majority, then the conclusion is C, and otherwise it is $\neg C$. This decision rule is called *premiss-based*. Suppose on the other hand that each voter decides directly on C or $\neg C$, and then the collective decision is taken by simple majority on these alternatives. This rule is called *conclusion-based*. There are cases where the premiss-based rule leads to C, while the conclusion-based yields $\neg C$.

A slightly different formulation receives the name of *discursive dilemma*, and is illustrated in Table 1. The committee votes on the three propositions P, Q and C, and each one is decided by majority. The result may be inconsistent with the agreed logical relation between the propositions, despite the individual votes being consistent.

In this paper we study general decision rules for the situation given. We will consider the set of all possible decision rules, subject only to a very mild requirement. They will be called *admissible rules*. We want to study this set as a whole, and find the best rule according to some objective criterion, disregarding whether that rule can or cannot be explained on "logical" or "intuitive" grounds, it is a consequence of some political or sociological idea, or it satisfies some other desirable property.

The mentioned requirement states only that if a member of the committee changes their[1] opinion on a clause in some direction, the conclusion can only eventually change in the same direction.

We take the epistemic point of view that *there is an actual truth* that we want to guess with the highest possible confidence. This is different from the aggregation of preferences as in elections, or in taking decisions on the course of actions, where there is not an absolute truth.

Our objective criterion is related to the minimisation of the combined chances to incur in false positives (deciding C when the reality is $\neg C$) and in false negatives (deciding $\neg C$ when the reality is C). This is explained in detail in Sect. 3 and involves a mathematical (probabilistic) setting where all elements have to be precisely defined. Our point of view is thus "conclusion-centric", in the sense that we do not care about the correct guessing of the premisses.

We emphasise that the adoption of a particular criterion is a modelling choice, and it is what confers the rationale to the best rule under it. The criterion proposed here can be replaced by another one, if deemed better for the situation at hand, and

[1] The singular *they/their* will be used to avoid gender bias.

the philosophy of *finding the best under the chosen criterion* can be applied as well. We consider here, in fact, a family of criteria, parametrised by the relative weight put on false positives and false negatives.

With this optimisation approach, we do not need to talk about *majorities*. The votes of the n members of the committee will be split into four slots: $P \wedge Q$, $P \wedge \neg Q$, $\neg P \wedge Q$, and $\neg P \wedge \neg Q$, and the aggregated number of votes for each possibility will be non-negative integers x, y, z, t, respectively, with $x + y + z + t = n$, being n the number of voters. In the case of three premisses, there will be 8 slots, and in general p premisses would give 2^p different possible votes of each member. A decision rule states, for each possible values of x, y, z, t which decision, C or $\neg C$, is taken.

The number of rules grows exponentially with n. The admissible rules are much less, and they can be implicitly enumerated so that all computations needed to find the optimal rule or a ranking of rules are relatively efficient.

If each committee member could infallibly guess the truth or falsity of each premiss, then the correct truth or falsity of the conclusion will be reached without difficulty. In fact, a single-member committee would suffice. The whole point of having multi-member committees is to alleviate the possibility that the final conclusion be wrong. It is therefore quite natural to use a probabilistic model that starts with the (estimated) probability that the committee members make the correct guessing on each premiss. We call this probability their *competence*, and we assume that it is greater that $\frac{1}{2}$, and that is the same for all members of the committee and for all premisses, although this is easily relaxed, as we will see in the final section.

The collective decision guesses correctly or incorrectly with some probability that depends on the voters' competence and on the real truth value of the premisses. Only the conclusion matters, and only the premisses are voted. One may think, as pointed out by Mongin (2012), that an external judge has to decide on the conclusions after the committee has sent them their individual opinions.

We state here some definitions that are used below: In general, an *agenda* is a logically consistent set of propositions, closed under negation, on which judgments have to be made, and that can be entangled by logical constraints. In our case the agenda is $\{P, \neg P, Q, \neg Q, C, \neg C\}$, with the constraint $C \Leftrightarrow P \wedge Q$. A *judgment* is defined as a mapping from the agenda to the doubleton $\{\text{True}, \text{False}\}$; a *feasible judgment* respects moreover the underlying logical constraints of the propositions,[2]. The *judgment aggregation problem* is then defined as the construction of a feasible reasonable collective judgment from the voters' individual judgments. Formally, an *aggregation rule* F is a mapping that assigns to every *profile* $(J_1, \ldots, J_n)$ of individual judgments J_i of the n voters, a collective judgment $J = F(J_1, \ldots, J_n)$. A *feasible aggregation rule* must assign a feasible judgment to any input of feasible profiles. Feasible rules trivially exist; for instance $J \equiv J_i$ for a given i is such a rule (called a *dictatorship*, for obvious reasons). Banning dictatorships and imposing other mild desirable conditions leads very quickly to non-existence of feasible aggregation rules. The range of possible *voting paradoxes* is the set of non-feasible mappings. In

[2] Lang et al. (2017) call it *consistent judgment* in the sense that it is logically consistent (not a contradiction) when the logical constraints are added.

a *truth-functional* agenda, the propositions are split into a set of premisses, and a set of conclusions. Assigning a Boolean value to all premisses, and applying the logical constraints, the value of all conclusions is determined. When there is an underlying objective truth of each proposition under scrutiny (as in court cases), we are in the realm of *truth-tracking* (or *epistemic*) judgment aggregation, and it is where the present work belongs.

1.2 Related Literature

Doctrinal Paradox/Discursive Dilemma The term *doctrinal paradox* appears first in the works of Kornhauser (1992a, 1992b), and Kornhauser and Sager (1993). They were interested in legal court cases, so that they spoke of *issue-by-issue* and *case-by-case* majority voting.

Pettit (2001) and List and Pettit (2002) formulate the problem in terms of propositional logic, called it the *discursive dilemma*, and coined the terms premiss-based and conclusion-based.

In the Kornhauser–Sager formulation, the committee votes either on the first two propositions (premiss-based/issue-by-issue), or on the third (conclusion-based/case-by-case), and the two results are different. In the List–Pettit formulation, the committee votes on the three propositions, and this leads to a logical inconsistency. The inconsistency comes from the constraint $C \Leftrightarrow P \wedge Q$ (the "doctrine" to which there is a previous agreement). We can see in Table 1 that the individual members of the committee adhere to the doctrine; however the committee as a whole does not. The advantage of the formulation in terms of propositional logic is that it can be generalised to any set of propositions, to the point that the distinction between premisses and conclusions may be unnecessary.

The two classical rules are quite natural, and both can be justified on intuitive or philosophical grounds; see for example, Mongin (2012, section 2). In particular, the conclusion-based rule respects the deliberation of the individual judges; in the premiss-based rule the decision can be fully justified in legal terms. Others rules can be proposed. In Alabert and Farré (2022), a new rule was introduced, which stands in some sense midway between the premiss-based and the conclusion-based rules.

Judgment Aggregation The body of knowledge that has been developed from List–Pettit formulation is known as Judgment Aggregation Theory (or Logical Aggregation Theory, as proposed by Mongin (2012)). In a quite natural way, the backbone of the theory is formed by (im)possibility results on the existence of aggregation rules satisfying certain desirable axioms. List and Pettit (2002) and List and Pettit (2004) already proved results of this kind, extended very soon by Pauly and van Hees (2006), Dietrich (2006) and Nehring and Puppe (2008).

The aggregation problem is described in full generality for example in the preliminaries of Nehring and Pivato (2011) and Lang et al. (2017), and in the complete surveys by Mongin (2012), List and Puppe (20009), List and Polak (2010) and List (2012).

The specialisation to the truth-functional case has been studied mainly in Nehring and Puppe (2008), for independent as well as interdependent premisses, and in Dokow and Holzman (2009) (see also Miller and Osherson (2009)).

Distance-Based Methods and Truth-Tracking From 2006 Pigozzi (2006) and Dietrich and List (2007) another point of view emerged, in which specific judgment rules are proposed, and their properties studied. See Lang et al. (2017) for a partial survey, and the references therein. Most of these rules can be defined as some sort of optimisation with respect to a criterion, i.e. the rule is defined as the one(s) that maximises or minimises a certain quantity, usually a distance or pseudo-distance to the individual profiles, while providing a consistent consensus judgment set.

They can be applied either when the collective judgment set is a decision on the course of actions (as in the adoption of public policies), or in the truth-tracking setting. In the latter, the goal is to get the right values of the pre-existent "state of nature", or at least the right values of the set of conclusions in the truth functional case. In this context, the concept of *competence* of the voters arise naturally: how likely is that a voter guess the correct answer to an issue? And it is also natural to model this likelihood as a probability. Actually, this approach dates back to Condorcet and his celebrated Jury Theorem.

The competence as a parameter has been studied, for example, in Bovens and Rabinowicz (2006) and in Grofman et al. (1983) for the one-issue case. The latter extends the Condorcet theorem in several directions, particularly for the case of unequal competences among voters.

List (2005) computed the probability of appearance of the doctrinal paradox, and the probability of correct truth-tracking as a function of the different states of nature, allowing for different competences in judging both premisses but the same across individuals. Fallis (2005) also observed that the premiss-based rule is better or not than the conclusion-based rule depending on the competence and on the "scenario" (state of nature).

The fact that the probability to guess correctly the truth depend on the unknown true state of nature leads to a modelling choice: Either we specify an a priori probability distribution on the possible states of nature (in our setting, the four states $P \wedge Q$, $P \wedge \neg Q$, $\neg P \wedge Q$ and $\neg P \wedge \neg Q$), or we have to resort to conservative estimations, as in classical (non-Bayesian) statistics. The main approach in this paper is the second, but most of the related literature assumes the first. Notably:

The cited paper by Bovens and Rabinowicz (2006) compares the premiss-based and the conclusion-based rules under the assumption of same competence for both premisses and their negations, and independence of voters, as in the present paper. They impose a Bernoulli prior on each premiss, the same for all.

Hartmann et al. (2010) aims at generalising (List 2005) and Bovens and Rabinowicz (2006) with a conjunctive truth-functional agenda allowing more than two premisses. The authors propose a continuum of distance-based rules, parametrised by the weight of the conclusion relative to the premisses, and containing the premiss- and conclusion-based procedures as extreme cases. The hypotheses are essentially the same as in Bovens and Rabinowicz (2006). Miller and Osherson (2009) also propose a variety of distance metrics, and distinguish between "underlying

metric" and "solution method". Each solution method chooses a loss function to minimise (based on the metric) and a set of eligible rules.

The point of view of Pivato (2013) is that the votes are observations of the 'truth plus noise'. This allows to think of the profile of individual judgments as a statistical sample (at least under the hypothesis of the same noise distribution for all voters), and study the decision rules as statistical estimators.

Truth-Functional Agendas and Truth-Tracking The abovementioned papers List (2005), Fallis (2005), Bovens and Rabinowicz (2006), Hartmann et al. (2010) and, partially, Miller and Osherson (2009), deal with truth-functional agendas. Nehring and Puppe (2008) focusses on truth-functional agendas with logically interdependent premisses.

Combined with a truth-functional agenda, the truth-tracking setting can still be concerned either with guessing the truth of all propositions ('getting the right answer for the right reasons') or only on the conclusion ('getting the right answer for whatever reasons'). Bovens and Rabinowicz (2006) discuss the merits of premiss- and conclusion-based procedures for both goals. Bozbay et al. (2014) and Bozbay (2019) also study both aims, for independent and for interrelated issues, respectively. The cited work by Hartmann et al. (2010) is conclusion-centric ("whatever reasons"), while Pigozzi et al. (2009), being conclusion-centric, applies later a procedure based on Bayesian networks to get the premisses that "interpret" the previously decided conclusions.

Distance-based methods are nothing else that the minimisation of a *loss function* that measures the dissatisfaction with every possible consistent outcome. Equivalently, one may maximise *utility functions*. Both are capable to account for the consequences of the decisions, and thus allow to set up more complete models, in line with Statistical Decision Theory (Berger 1985). Different loss functions or utilities give rise to possibly different optimal rules, and it is a modelling task to choose the right loss function for the problem at hand.

In this sense, Fallis (2005) writes about the 'epistemic value', highlighting that guessing correctly a proposition may have a different value that guessing correctly its contrary; Bozbay (2019) uses a simple 0–1 utility function to indicate incorrect-correct guessing (of all propositions or of the conclusions alone); Hartmann et al. (2010) tries giving different utilities to false positives and false negatives on the conclusion to assess the performance of their continuum of metrics; finally, Bovens and Rabinowicz (2006), in the discussion section, suggest introducing different utilities to each correct guessing to compare the premiss-based and the conclusion-based voting rules in each practical case.

Our proposal in this paper is an optimisation criterion in the truth-tracking, conclusion-centric case of a conjunctive agenda. The best decision rule minimises a combination of false positives and false negatives, and any two rules can be easily compared according to this criterion. No prior on the states of nature needs to be established, although it can be accommodated without difficulty. In the theoretical results we assume the same competence level of all committee members and all

premisses, and independence among premisses. In practice, the specific computation of the score of each rule can be done under much less assumptions. In any case, the loss function fully determines the optimal rule.

We do not consider strategic voting; we assume that everyone votes honestly each of the premisses. Strategic voting is conceivable even in our simple setting: someone who honestly would vote for P and $\neg Q$ could change to $\neg P$ and $\neg Q$ just to push more for the $\neg C$ conclusion). Strategic voting is considered in Bozbay et al. (2014), de Clippel and Eliaz (2015), Terzopoulou and Endriss (2019) and Bozbay (2019).

1.3 Organisation of the Paper

The remainder of the paper is organised as follows: In Sect. 2 we study the structure of the sets of voting tables and of decision rules, which are both partially ordered sets with an order induced by the admissibility condition.

Section 3 introduces in the first part our probabilistic model, based on the probability that each committee member guesses correctly the truth value of each premiss, and the concepts of false positive and false negative when the true state of nature is unknown. In the second part, we introduce the family of optimisation criteria, parametrised by a relative weight assigned to the two errors.

Section 4 contains our main theoretical results. It turns out that it is relatively easy to determine whether a given voting table leads to the conclusion C or $\neg C$ in the optimal rule. Moreover, this depends only on two numbers: the difference between votes to $P \wedge Q$ and to $\neg P \wedge \neg Q$, and the difference between votes to $P \wedge \neg Q$ and to $\neg P \wedge Q$. This simplifies the structure of the set of voting tables, and shortens the evaluation of the rules. In this section we also characterise completely the set of values of competence and weight for which the premiss-based rule is optimal.

Section 5 explains details on the actual computations, and describes the accompanying software, downloadable from https://discursive-dilemma.sourceforge.io. Finally, in Sect. 6 we discuss the results and propose possible extensions. Some marginal computations and checks have been left to an "Appendix".

2 The Set of Decision Rules

A possible voting result of a committee with n members assessing on issues P and Q will be a *table*, denoted $\begin{bmatrix} x & y \\ z & t \end{bmatrix}$, or (x, y, z, t) to save space, with non-negative integer entries, adding up to n, representing the quantity of votes received by the options

$P \wedge Q$, $P \wedge \neg Q$, $\neg P \wedge Q$ and $\neg P \wedge \neg Q$, respectively. The set of all such tables will be denoted by $\mathbb{T}$.

A *decision rule* can be thought of as a mapping $\mathbb{T} \longrightarrow \{0, 1\}$. Tables mapped to 1 are those that entail the decision $C = P \wedge Q$; those mapped to 0 represent the opposite, $\neg C$. Sometimes we will call them *positive tables* and *null tables*, and denote by $\mathbb{T}^+$ and $\mathbb{T}^0$ the respective sets. The decision rule can also be seen as the subset of positive tables, and we will make use of both interpretations.

There are $N = \frac{n!}{x!y!z!t!}$ ways to fill a voting table, and 2^N decision rules, as many as subsets of the set of tables. This is a huge number, already for $n = 3$, but the set of "reasonable rules" will be much more modest.

Two tables (x, y, z, t) and (x, z, y, t) are *transposed* of each other. Since P and Q will play symmetric roles, it makes sense to admit only rules that assign the same decision to both tables.

Besides this symmetry, we impose another condition for the admissibility of a decision rule, which is a monotonicity-type condition. Suppose that, given a positive voting table (x, y, z, t), one of the voters of $\neg P$ changes their choice to P, or one of the voters of $\neg Q$ changes to Q. The new table would support better the conclusion C than the older; hence it makes sense to impose that the new table be also a positive table. To formulate the condition in a mathematically practical way, let us introduce the partial order on $\mathbb{T}$ generated by the four relations

$$\begin{aligned}
(x, y, z, t) &\leq (x, y, z + 1, t - 1) \\
(x, y, z, t) &\leq (x, y + 1, z, t - 1) \\
(x, y, z, t) &\leq (x + 1, y - 1, z, t) \\
(x, y, z, t) &\leq (x + 1, y, z - 1, t) \,,
\end{aligned} \tag{1}$$

that means, the smallest partial order $\leq$ that satisfies relations (1) above, for all x, y, z, t for which they make sense. A partially ordered set is also called a *poset*, for short. The relations (1) are called the *transitive reduction* of the poset $(\mathbb{T}, \leq)$. Posets can be represented by *Hasse diagrams*, which are directed graphs with the transitive reduction represented by arrows pointing in the increasing direction. The case of committee size $n = 3$ is depicted in Fig. 1, where transposed tables have been identified; they occupy the same spot and are not comparable. When two tables $T, S \in \mathbb{T}$, satisfy $T \leq S$ and $T \neq S$, we shall obviously write $T < S$, or $S > T$.

We thus arrive to the following reasonable definition of admissibility:

Definition 2.1 A decision rule $r : \mathbb{T} \to \{0, 1\}$ is *admissible* if:

1. It takes the same value on transposed tables:

$$r(x, y, z, t) = r(x, z, y, t) \,.$$

2. It is order-preserving on the partially ordered set $(\mathbb{T}, \leq)$:

$$(x, y, z, t) \leq (x', y', z', t') \Rightarrow r(x, y, z, t) \leq r(x', y', z', t') \,.$$

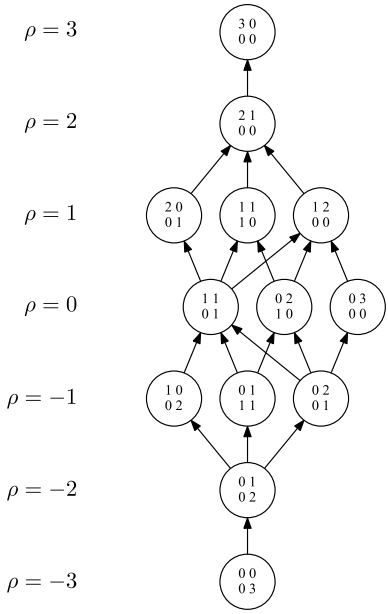

Fig. 1 Hasse diagram of the poset $(\mathbb{T}, \leq)$ for committee size $n = 3$, and transposed tables identified. The arrows correspond to the transitive reduction. All other relations are deduced by transitivity. The rank function $\rho(x, y, z, t) = x - t$ is also represented

Example 2.2 The classical premiss-based rule r_{pb} is defined by

$$r_{pb}(x, y, z, t) = 1 \text{ if and only if } x + y > z + t \text{ and } x + z > y + t,$$

whereas the conclusion-based rule r_{cb} is given by

$$r_{cb}(x, y, z, t) = 1 \text{ if and only if } x > y + z + t.$$

It is readily checked that both rules are admissible in the sense of Definition 2.1. In Alabert and Farré (2022), another admissible rule was introduced, called *path-based*, and defined by

$$r_{hb}(x, y, z, t) = 1 \textit{ if and only if } x > z + t \textit{ and } x > y + t.$$

As an example of a non-admissible rule, consider declaring C true if and only if the votes for $P \wedge Q$ are more than any other combination of premisses (i.e. $x > y$ and $x > z$ and $x > t$). □

We will see later that the second admissibility condition is not restrictive with respect to our optimisation criterion: given a non-order-preserving rule, there exists an order-preserving one that performs better.

The poset $(\mathbb{T}, \leq)$ is *ranked* (also called *graded*), i.e. there exists a *rank function* ρ compatible with the order relation: It satisfies $T < S \Rightarrow \rho(T) < \rho(S)$, and if S is an *immediate successor* of T (there are no elements in between), then

$\rho(S) = \rho(T) + 1$. In the Hasse diagram, each rank can be pictured as a "level" in the graph (see again Fig. 1).

To prove that $(\mathbb{T}, \leq)$ is a ranked poset, we use the known result that a finite poset admits a rank function if and only if all maximal chains have the same length.[3] Recall that a *chain* is a totally ordered subset of the poset. A *maximal chain* is a chain with maximal cardinality.

A poset is *connected* if for every two elements T and S there is a finite sequence $T = U_1, \ldots, U_n = S$ of elements such that U_i and U_{i+1} are comparable, i.e. either $U_i \leq U_{i+1}$ or $U_{i+1} \leq U_i$. The poset $(\mathbb{T}, \leq)$ is connected, because we can transform a table into any other one by moving votes one at a time through the transitive reduction.

Proposition 2.3 $(\mathbb{T}, \leq)$ *is a ranked poset, with* $\rho(x, y, z, t) = x - t$ *as a rank function.*

Proof There is a unique minimal element, namely $m = (0, 0, 0, n)$, and a unique maximal element $M = (n, 0, 0, 0)$. Since $(\mathbb{T}, \leq)$ is connected, all maximal chains start and finish in these elements. To go from m to M, each vote must make two steps, one of them up or to the left of the table (that means, from t to y or z), and the other one to the left or up respectively (from y or z to x). These individual movements are the transitive reduction of the partial order $\leq$, and therefore there are no other tables in between. Since there are n votes, we need $2n$ steps to move all votes from the minimal to the maximal element, and in consequence any maximal chain has exactly $2n + 1$ elements.

Notice that a movement towards an immediate successor imply subtracting one unit to t or adding one unit to x, but not both. Therefore $\rho(T) = x - t$ is a rank function for $(\mathbb{T}, \leq)$. □

The rank function of a ranked poset is not unique, but it is completely determined by setting the rank of any element of the poset.

The set $\mathcal{A}$ of admissible decision rules possesses also a natural partial order: $r \leq s$ if $r(T) \leq s(T)$ for all tables $T \in \mathbb{T}$. This is the usual partial order in a set of real functions on any domain. Since the range of decision rules mappings is $\{0, 1\}$, the relation $r \leq s$ means that the set of positive tables relative to r is included in the set of positive tables relative to s. It can be said that r is less liberal (or more conservative) than s in the sense explained in the Introduction. In terms of the risk of opting for C when it is wrong, $\leq$ is the relation "to be less risky or equal to".

Example 2.4 Refer to the rules of Example 2.2. The premiss-based rule is more liberal than the path-based rule, and this one is in turn more liberal than the conclusion-based rule. In other words, we are saying that, in $(\mathcal{A}, \leq)$,

$$r_{cb} \leq r_{hb} \leq r_{pb}.$$

[3] Called the *Jordan–Dedekind property*.

This can be seen using the characterisations given in Example 2.2 (see Alabert and Farré 2022, Proposition A.2 for the proof). To precise a little more, one can check that rules r_{hb} and r_{cb} coincide for committee sizes $n = 3, 5$, and they are different for $n \geq 7$. Rule r_{pb} is always strictly greater than r_{hb}. □

An *upper set* is a subset U of a poset such that $x \in U, x < y \Rightarrow y \in U$. It is immediate to see, from the second condition of admissibility, the following equivalence.

Proposition 2.5 *A decision rule* $r : \mathbb{T} \to \{0, 1\}$ *is order-preserving if and only if* $\{T \in \mathbb{T} : r(T) = 1\}$ *is an upper set of* $(\mathbb{T}, \leq)$.

Given a poset, the family of its upper sets, with the inclusion order relation, is a *complete lattice*: A partially ordered set in which all subsets have a *supremum* (a least upper bound) and an *infimum* (a greatest lower bound). Applied to our case, we are saying that the union and the intersection of upper sets are upper sets or, in terms of rule mappings, that the maximum (= sum) and the minimum (= product) of admissible rules are admissible rules.

An *antichain* is a subset $\mathbb{S}$ of a poset such that any two elements of $\mathbb{S}$ are not comparable. Antichains and upper sets are related in the following way: The minimal elements of any upper set form and antichain; conversely, any antichain A determines the upper set

$$\{x : (\exists y)(y \in A \wedge x \geq y)\} .$$

The empty antichain is also considered, and corresponds in our case to the rule $r \equiv 0$.

For finite posets, the correspondence between antichains and upper sets is bijective. Enumerating upper sets is therefore equivalent to enumerating antichains. Even computing the number of upper sets is not easy in general. For example, in the well-known poset of the subsets of a given set of k elements, with the inclusion relation, the number of upper sets (called the *Dedekind numbers*, see OEIS Foundation https://oeis.org/), is not known for $k > 8$.

3 Probabilistic Model and Optimisation Criterion

We want to find the best of all admissible rules, according to some quantitative criterion, formulated in terms of a probabilistic model. In this section the criterion will be introduced, and the next one will be devoted to the characterisation of the optimal rule.

3.1 Probabilistic Model

Suppose $C = P \wedge Q$ is the true state of nature. If for some rule r and a table of votes $T = (x, y, z, t)$, we have $r(T) = 1$, we say that this is a *true positive* (TP). Otherwise, if $r(T) = 0$, it is a *false negative* (FN). Similarly, if $\neg C$ is the true state,

$r(T) = 0$ will be a *true negative* (TN) and $r(T) = 1$ will be a *false positive* (FP). Ideally, a good decision rule should minimise somehow the occurrence of false positives and false negatives. To assess the likelihood of these occurrences we need a probabilistic model in which to evaluate the probability of appearance of FP and FN. To that end, we need an estimate of the probability that the members of the committee guess correctly the true value of the premisses P and Q.

The probability that a committee member vote the correct value of the premisses will be called its *competence*. We will assume that all committee members have the same competence, a number strictly between $\frac{1}{2}$ and 1. Notice that a competence less than $\frac{1}{2}$ does not make sense, because in that case we can reverse all opinions of the committee, and we get another committee with competence greater than $\frac{1}{2}$. If it were exactly $\frac{1}{2}$, there is a trivial solution that will be pointed out later; if it is 1, then a one-member committee is enough and they are always right. We will also assume that the committee size n is odd and two additional independence conditions. Specifically, we assume in the sequel the following hypotheses:

(H1) *Odd committee size*: The number of voters is an odd number, $n = 2m + 1$, with $m \geq 1$.
(H2) *Equal competence*: The competence θ satisfies $\frac{1}{2} < \theta < 1$ and it is the same for all voters and for both premisses P and Q.
(H3) *Mutual independence among voters*: The decision of each voter does not depend on the decisions of the other voters.
(H4) *Independence between P and Q*: For each voter, the decision on one premiss does not influence the decision on the other.

Formally, hypotheses (H2)–(H4) can be rephrased by saying that for each voter in the committee and each premiss, there is a random variable that takes the value 1 if the voter believes the clause is true, and zero otherwise, and all these random variables are stochastically independent and identically distributed. Their specific distribution depends on the true state of nature. See Sect. 6 for possible relaxations of these hypotheses.

Proposition 3.1 *Assume hypotheses* (*H*1)–(*H*4), *with committee size n*, *and competence* θ. *Then the probability that the votes result in a particular table* $T = (x, y, z, t)$ *is*, *under the different states of nature*,

$$\text{if } P \wedge Q, \text{ then} \quad \frac{n!}{x!y!z!t!}\theta^{2x+y+z}(1-\theta)^{y+z+2t}\,, \tag{2}$$

$$\text{if } P \wedge \neg Q, \text{ then} \quad \frac{n!}{x!y!z!t!}\theta^{x+2y+t}(1-\theta)^{x+2z+t}\,, \tag{3}$$

$$\textit{if } \neg P \wedge Q, \textit{ then } \quad \frac{n!}{x!y!z!t!}\theta^{x+2z+t}(1-\theta)^{x+2y+t}\,, \tag{4}$$

$$\textit{if } \neg P \wedge \neg Q, \textit{ then } \quad \frac{n!}{x!y!z!t!}\theta^{y+z+2t}(1-\theta)^{2x+y+z}\,, \tag{5}$$

where ! *means the factorial of a number.*

Proof See Proposition A.4 of Alabert and Farré (2022). □

Let us denote $\mathbb{P}_{P\wedge Q}$ the probabilities computed under the state of nature $P \wedge Q$. According to the proposition above, the probability of obtaining a true positive when rule r is employed is the sum of the probabilities (2) for all tables T such that $r(T) = 1$:

$$\mathbb{P}_{P\wedge Q}(\mathrm{TP}) = \sum_{\{r(x,y,z,t)=1\}} \frac{n!}{x!y!z!t!}\theta^{2x+y+z}(1-\theta)^{y+z+2t}\,.$$

Therefore, the probability of incurring a false negative is

$$\mathbb{P}_{P\wedge Q}(\mathrm{FN}) = 1 - \mathbb{P}_{P\wedge Q}(\mathrm{TP}) = \sum_{\{r(x,y,z,t)=0\}} \frac{n!}{x!y!z!t!}\theta^{2x+y+z}(1-\theta)^{y+z+2t}\,.$$

We cannot proceed in a completely analogous way to define true negatives and false positives, because $\neg(P \wedge Q)$ is not a state of nature, but an ensemble of three states, each of which may yield different probabilities. At this point, there are two possible modelling paths, according to the information available: Either there is no further information about the true state of nature (or we do not want to use it); or, there is enough information to postulate an "a priori" probability π on the states of nature, and we can follow a Bayesian approach.

The main line in this paper is the first path, always applicable. Let us deviate for a moment and sketch the second one, which corresponds to a situation considered, among others, in Terzopoulou and Endriss (2019), Bovens and Rabinowicz (2006) and Bozbay (2019). In the Bayesian approach, $\mathbb{P}_{P\wedge Q}$ is interpreted as a conditional probability given $P \wedge Q$, and analogously for the other ones, that we denote $\mathbb{P}_{P\wedge\neg Q}$, $\mathbb{P}_{\neg P\wedge Q}$, and $\mathbb{P}_{\neg P\wedge\neg Q}$. Hence, the probability of a true negative in this setting will be

$$\begin{aligned}\mathbb{P}_{\neg(P\wedge Q)}(\mathrm{TN}) = {} & \mathbb{P}_{P\wedge\neg Q}(\mathrm{TN}) \cdot \pi(P \wedge \neg Q) \\ & + \mathbb{P}_{\neg P\wedge Q}(\mathrm{TN}) \cdot \pi(\neg P \wedge Q) \\ & + \mathbb{P}_{\neg P\wedge\neg Q}(\mathrm{TN}) \cdot \pi(\neg P \wedge \neg Q)\,,\end{aligned} \tag{6}$$

and then the probability of a false positive is given by

$$\mathbb{P}_{\neg(P\wedge Q)}(\mathrm{FP}) = 1 - \mathbb{P}_{\neg(P\wedge Q)}(\mathrm{TN}) \tag{7}$$

For example, if π is assumed to give the same probability to all three negative states, then $\mathbb{P}_{\neg(P\wedge Q)}(\mathrm{TN})$ will be the arithmetic mean of the probabilities of TN under each state. This is the chosen prior in Terzopoulou and Endriss (2019); that of Bovens and Rabinowicz (2006) is different, and Bozbay (2019) completely forbids the result $\neg P \wedge \neg Q$. In general, if the committee knows the prior, the independence in the judgments of the premisses (H4) cannot be assumed.

Now, using expressions (3), (4) and (5), the probabilities of a true negative under the three negative states are

$$\mathbb{P}_{P\wedge\neg Q}(\mathrm{TN}) = \sum_{\{r(x,y,z,t)=0\}} \frac{n!}{x!y!z!t!}\theta^{x+2y+t}(1-\theta)^{x+2z+t} \tag{8}$$

$$\mathbb{P}_{\neg P\wedge Q}(\mathrm{TN}) = \sum_{\{r(x,y,z,t)=0\}} \frac{n!}{x!y!z!t!}\theta^{x+2z+t}(1-\theta)^{x+2y+t} \tag{9}$$

$$\mathbb{P}_{\neg P\wedge\neg Q}(\mathrm{TN}) = \sum_{\{r(x,y,z,t)=0\}} \frac{n!}{x!y!z!t!}\theta^{y+z+2t}(1-\theta)^{2x+y+z} \tag{10}$$

and the probability of a false positive is then computed from (6) and (7).

After this digression, let us turn to our main setting. For the non-Bayesian situation, we can resort to the following analogy with the classical theory of Statistical Hypothesis Testing: Suppose one has to decide if there is enough evidence that a certain population parameter is equal to a value C, as provided by a sample drawn from the population. To this end, one computes how likely the observed sample could have been produced by the value of the parameter in the complement set $\neg C$ which is "the closest" to C. If that likelihood is acceptable (by some numerical threshold), the decision is to stick to the "null" (status quo) conclusion $\neg C$. If it is not acceptable, C is proclaimed as the new estimated conclusion.

Translating the analogy to our case, we must ask ourselves which of the states of nature belonging to the complement set $\neg(P \wedge Q)$ is closest to $P \wedge Q$. Intuitively, $P \wedge \neg Q$ and $\neg P \wedge Q$ are equally close, and are closer than $\neg P \wedge \neg Q$. This is rigorously stated in the next proposition. Although intuitive, the rigorous proof is a little bit technical. We use a probabilistic procedure called *coupling*, that transforms inequalities about probabilities into inequalities about random variables.

These arguments support the *definition* of the "probability" of a false positive as the probability that a rule decides C when $P \wedge \neg Q$ is the true state of nature: $\mathbb{P}_{P\wedge\neg Q}(\mathrm{FP}) := 1 - \mathbb{P}_{P\wedge\neg Q}(\mathrm{TN})$. It can be also thought as the maximum of the probabilities of a false positive for all possible choices of a prior π on the states of nature. It is therefore a conservative estimate of the possible error, in response to the lack of information about the underlying truth.

Proposition 3.2 *Under hypothesis* $(H1)$–$(H4)$, *for any admissible decision rule* r,

$$\mathbb{P}_{P\wedge\neg Q}\{r=1\} = \mathbb{P}_{\neg P\wedge Q}\{r=1\} \geq \mathbb{P}_{\neg P\wedge\neg Q}\{r=1\}\,.$$

Proof The first equality is clear from Definition 2.1, item 1. We only need to prove the inequality on the right. Let $\mathbb{M}_{\neg P\wedge\neg Q}$ and $\mathbb{M}_{P\wedge\neg Q}$ be two probability measures defined on the subsets of a set Ω, and let T be a random variable $T : \Omega \to \mathbb{T}$ such that the law of T under $\mathbb{M}_{\neg P\wedge\neg Q}$ is $\mathbb{P}_{\neg P\wedge\neg Q}$ and the law of T under $\mathbb{M}_{P\wedge\neg Q}$ is $\mathbb{P}_{P\wedge\neg Q}$. That is, using (5) and (3),

$$\mathbb{M}_{\neg P\wedge\neg Q}\{\omega : T(\omega) = (x, y, z, t)\} = \mathbb{P}_{\neg P\wedge\neg Q}\{(x, y, z, t)\} = \frac{n!}{x!y!z!t!}\theta^{y+z+2t}(1-\theta)^{2x+y+z} ,$$

$$\mathbb{M}_{P\wedge\neg Q}\{\omega : T(\omega) = (x, y, z, t)\} = \mathbb{P}_{P\wedge\neg Q}\{(x, y, z, t)\} = \frac{n!}{x!y!z!t!}\theta^{x+2y+t}(1-\theta)^{x+2z+t} .$$

Suppose we could define another random variable $S : \Omega \to \mathbb{T}$ such that

(a) $T(\omega) \leq S(\omega)$, for all $\omega \in \Omega$, and
(b) The law of S under $\mathbb{M}_{\neg P\wedge\neg Q}$ coincides with the law of T under $\mathbb{M}_{P\wedge\neg Q}$.

Then, since r is order-preserving, we will have $\{\omega : r(S(\omega)) = 1\} \supseteq \{\omega : r(T(\omega)) = 1\}$, and the conclusion

$$\mathbb{P}_{P\wedge\neg Q}\{r(T) = 1\} = \mathbb{P}_{\neg P\wedge\neg Q}\{r(S) = 1\} \geq \mathbb{P}_{\neg P\wedge\neg Q}\{r(T) = 1\} .$$

Let us prove the existence of $S : \Omega \to \mathbb{T}$ with the properties a) and b) above, and we are done: Let $T_1, \ldots, T_n$ be independent identically distributed random variables $T_i : \Omega \to \mathbb{T}$ with the same law as T but for the vote of one individual. We will switch to table notation again, for clarity, in the rest of the proof.

Let $S_1, \ldots, S_n$ be another collection of independent identically distributed random variables $S_i : \Omega \to \mathbb{T}$, defined as follows:

$$\text{If } T_i = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \text{ then } S_i = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\text{If } T_i = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \text{ then } S_i = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$\text{If } T_i = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \text{ then } S_i = \begin{cases} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} & \text{with probability } \frac{1-\theta}{\theta} \\ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} & \text{with probability } \frac{2\theta-1}{\theta} \end{cases}$$

$$\text{If } T_i = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \text{ then } S_i = \begin{cases} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} & \text{with probability } \frac{1-\theta}{\theta} \\ \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} & \text{with probability } \frac{2\theta-1}{\theta} \end{cases}$$

We have clearly that $T_i(\omega) \leq S_i(\omega)$, for all $\omega \in \Omega$.

Let us compute the law of S_i under $\mathbb{M}_{\neg P\wedge\neg Q}$, using conditional probabilities to the value of T_i:

$$
\begin{aligned}
\mathbb{M}_{\neg P\wedge\neg Q}\Big\{S_i=\begin{bmatrix}1&0\\0&0\end{bmatrix}\Big\} &= \mathbb{M}_{\neg P\wedge\neg Q}\Big\{S_i=\begin{bmatrix}1&0\\0&0\end{bmatrix}\Big/T_i=\begin{bmatrix}1&0\\0&0\end{bmatrix}\Big\}\cdot\mathbb{M}_{\neg P\wedge\neg Q}\Big\{T_i=\begin{bmatrix}1&0\\0&0\end{bmatrix}\Big\}\\
&\quad+\mathbb{M}_{\neg P\wedge\neg Q}\Big\{S_i=\begin{bmatrix}1&0\\0&0\end{bmatrix}\Big/T_i=\begin{bmatrix}0&0\\1&0\end{bmatrix}\Big\}\cdot\mathbb{M}_{\neg P\wedge\neg Q}\Big\{T_i=\begin{bmatrix}0&0\\1&0\end{bmatrix}\Big\}\\
&= (1-\theta)^2+\frac{2\theta-1}{\theta}\cdot\theta(1-\theta)=\theta(1-\theta)\ .\\
\mathbb{M}_{\neg P\wedge\neg Q}\Big\{S_i=\begin{bmatrix}0&1\\0&0\end{bmatrix}\Big\} &= \mathbb{M}_{\neg P\wedge\neg Q}\Big\{S_i=\begin{bmatrix}0&1\\0&0\end{bmatrix}\Big/T_i=\begin{bmatrix}0&1\\0&0\end{bmatrix}\Big\}\cdot\mathbb{M}_{\neg P\wedge\neg Q}\Big\{T_i=\begin{bmatrix}0&1\\0&0\end{bmatrix}\Big\}\\
&\quad+\mathbb{M}_{\neg P\wedge\neg Q}\Big\{S_i=\begin{bmatrix}0&1\\0&0\end{bmatrix}\Big/T_i=\begin{bmatrix}0&0\\0&1\end{bmatrix}\Big\}\cdot\mathbb{M}_{\neg P\wedge\neg Q}\Big\{T_i=\begin{bmatrix}0&0\\0&1\end{bmatrix}\Big\}\\
&= \theta(1-\theta)+\frac{2\theta-1}{\theta}\cdot\theta^2=\theta^2\ .\\
\mathbb{M}_{\neg P\wedge\neg Q}\Big\{S_i=\begin{bmatrix}0&0\\1&0\end{bmatrix}\Big\} &= \mathbb{M}_{\neg P\wedge\neg Q}\Big\{S_i=\begin{bmatrix}0&0\\1&0\end{bmatrix}\Big/T_i=\begin{bmatrix}0&0\\1&0\end{bmatrix}\Big\}\cdot\mathbb{M}_{\neg P\wedge\neg Q}\Big\{T_i=\begin{bmatrix}0&0\\1&0\end{bmatrix}\Big\}\\
&= \frac{1-\theta}{\theta}\cdot\theta(1-\theta)=(1-\theta)^2\ .\\
\mathbb{M}_{\neg P\wedge\neg Q}\Big\{S_i=\begin{bmatrix}0&0\\0&1\end{bmatrix}\Big\} &= \mathbb{M}_{\neg P\wedge\neg Q}\Big\{S_i=\begin{bmatrix}0&0\\0&1\end{bmatrix}\Big/T_i=\begin{bmatrix}0&0\\0&1\end{bmatrix}\Big\}\cdot\mathbb{M}_{\neg P\wedge\neg Q}\Big\{T_i=\begin{bmatrix}0&0\\0&1\end{bmatrix}\Big\}\\
&= \frac{1-\theta}{\theta}\cdot\theta^2=\theta(1-\theta)\ .
\end{aligned}
$$

We see that the law of each S_i under $\mathbb{M}_{\neg P\wedge\neg Q}$ coincides with that of T_i under $\mathbb{M}_{P\wedge\neg Q}$. Now, we have that

$$
T(\omega) := \sum_{i=1}^{n} T_i(\omega) \le \sum_{i=1}^{n} S_i(\omega) =: S(\omega)\ , \quad \omega\in\Omega\ ,
$$

and that T has the law given by (5) under $\mathbb{M}_{\neg P\wedge\neg Q}$, and by (3) under $\mathbb{M}_{P\wedge\neg Q}$, and S has the law given by (3) under $\mathbb{M}_{\neg P\wedge\neg Q}$.

Hence, T and S are the random variables we were looking for, and the proof is complete. □

As a corollary, since the maximum is always greater than any weighted mean in (6), we get that $\mathbb{P}_{P\wedge\neg Q}\{r=1\}$ is greater or equal than the probability of a false positive computed with any prior distribution on the set $\neg(P\wedge Q)$.

In a completely analogous way, one can also prove that for admissible rules the probability to conclude C is greater when the true state of nature is $P\wedge Q$ than with any other state.

We do not need any more the subindexes in the probabilities of false positives and false negatives, since $\mathbb{P}(\mathrm{FN})$ always refers to the state $P\wedge Q$, and $\mathbb{P}(\mathrm{FP})$ refers to the state $P\wedge\neg Q$ (or to the given prior on $\neg(P\wedge Q)$ in the Bayesian case). Instead, we will subindex $\mathbb{P}$ by the rule employed. For reference in the sequel, we repeat here the formulae for FP and FN: For any rule $r:\mathbb{T}\to\{0,1\}$,

$$
\mathbb{P}_r(\mathrm{FP}) = \sum_{\{r(x,y,z,t)=1\}} \frac{n!}{x!y!z!t!}\theta^{x+2y+t}(1-\theta)^{x+2z+t}\ , \tag{11}
$$

$$\mathbb{P}_r(\mathrm{FN}) = \sum_{\{r(x,y,z,t)=0\}} \frac{n!}{x!y!z!t!}\theta^{2x+y+z}(1-\theta)^{y+z+2t} \ . \tag{12}$$

3.2 Optimisation Criterion

We want to obtain the best decision rule, under the probabilistic model stated above and the optimisation criterion what we develop in this subsection. This criterion was introduced in Alabert and Farré (2022) and is based on minimising a weighted sum of the probability to commit a false positive and the probability to commit a false negative. It can be thought as a multi-objective optimisation problem, but that point of view does not contribute any practical insight.

Any rule $r : \mathbb{T} \to \{0, 1\}$ (admissible or not) has associated probabilities of producing a False Positive $\mathbb{P}_r(\mathrm{FP})$ and a False Negative $\mathbb{P}_r(\mathrm{FN})$ according to formulae (11)–(12). Despite the simplified notation, recall that these two probabilities stem from different states of nature. If both failures are considered equally harmful, it is natural to look for the admissible rule $r \in \mathcal{A}$ that minimises the sum

$$\mathbb{P}_r(\mathrm{FP}) + \mathbb{P}_r(\mathrm{FN}) \ . \tag{13}$$

If one of them is considered worse that the other, one can take a weighted sum

$$L_w(r) := w\mathbb{P}_r(\mathrm{FP}) + (1-w)\mathbb{P}_r(\mathrm{FN}) \ , \tag{14}$$

where w is a real number, $0 < w < 1$, as the loss function that to minimise. For example, if a false positive is deemed twice as harmful as a false negative, $w = \frac{2}{3}$ is the suitable value. Note that the weight w is a modelling choice relative to each particular application, and it is supposed to be fixed in advance of the voting stage.

In Statistics, the combination (13) of probabilities of the two types of error is called the *area of the triangle*, a term that comes from its origin in signal processing and the graphical methodology called *Receiving Operating Characteristics* (ROC). We refer the reader to Fawcett (2006) for a simple introduction to ROC.

If r and s are two admissible rules, and $r \leq s$ (equivalently, $U_r \subseteq U_s$, where U_r and U_s are the upper sets defining r and s respectively), then obviously, from (11) and (12),

$$\mathbb{P}_r(\mathrm{FP}) \leq \mathbb{P}_s(\mathrm{FP})$$
$$\mathbb{P}_r(\mathrm{FN}) \geq \mathbb{P}_s(\mathrm{FN}) \ .$$

This means that $r \mapsto \mathbb{P}_r(\mathrm{FP})$ and $r \mapsto \mathbb{P}_r(\mathrm{FN})$ are respectively an increasing function and a decreasing function defined on the poset of admissible rules $(\mathcal{A}, \leq)$. Moreover the rule $r \equiv 0$ (always conclude $\neg C$) satisfies $\mathbb{P}_r(\mathrm{FP}) = 0$ and $\mathbb{P}_r(\mathrm{FN}) = 1$, and the rule $r \equiv 1$ (always conclude C) satisfies $\mathbb{P}_r(\mathrm{FP}) = 1$ and $\mathbb{P}_r(\mathrm{FN}) = 0$.

As we said before, the value $\theta = \frac{1}{2}$ can be excluded because it is trivial: If $w < \frac{1}{2}$, the decision must be always C; if $w > 1/2$, the decision must be $\neg C$; and if $w = 1/2$,

then the problem is completely equivalent to a single coin toss. See the "Appendix" for the details.

4 Main Results

Suppose we have an admissible rule r, with sets $\mathbb{T}^+$ of positive tables and $\mathbb{T}^0$ of null tables (recall the definitions in Sect. 2). If we choose a table $T \in \mathbb{T}^0$ and move it to $\mathbb{T}^+$, we are increasing the probability of a false positive and at the same time decreasing the probability of a false negative. In doing that, we also have to move its transposed table, to maintain the first condition of admissibility. This movement may result in a decrease or an increase of the loss function L_w.

Definition 4.1 Let $r : \mathbb{T} \to \{0, 1\}$ be an admissible decision rule, with positive set $\mathbb{T}^+$ and null set $\mathbb{T}^0$. If moving a table T and its transposed table (and no other) from $\mathbb{T}^0$ to $\mathbb{T}^+$ results in a decrease of the loss function, we will say that we have a *good table*.

Here "good" only means that the voting table T "should be supporting decision C". Thus, it seems that the set $\mathbb{T}^+$ of the optimal rule must consist of the good tables and no others. However, we still have to see that the rule defined in this way is indeed admissible.

Theorem 4.2 *The rule whose positive set $\mathbb{T}^+$ consists exactly of the good tables is admissible and optimal.*

This is a consequence of the following two lemmas, which are interesting in their own. Lemma 4.3 characterises the good tables in terms of w and θ, and confirms that a table and its transposed are both good or both bad. Lemma 4.4 proves that the second condition of admissibility is also satisfied, in view of Proposition 2.5.

Lemma 4.3 *Given weight $0 < w < 1$ and competence $\frac{1}{2} < \theta < 1$, a table $T = (x, y, z, t) \in \mathbb{T}$ is a good table if and only if*

$$\left(\frac{\theta}{1-\theta}\right)^{(y-z)-(x-t)} + \left(\frac{\theta}{1-\theta}\right)^{(z-y)-(x-t)} < \frac{2(1-w)}{w}\,. \tag{15}$$

Proof Using formulae (11) and (12), the change in the quantity $w\mathbb{P}(\mathrm{FP}) + (1-w)\mathbb{P}(\mathrm{FN})$ when T and its transposed table are moved from the null to the positive set is given by

$$\begin{aligned}\frac{n!}{x!y!z!t!}\Big[&w\theta^{x+2y+t}(1-\theta)^{x+2z+t} + w\theta^{x+2z+t}(1-\theta)^{x+2y+t}\\ &- 2(1-w)\theta^{2x+y+z}(1-\theta)^{y+z+2t}\Big]\end{aligned} \tag{16}$$

Dropping the factorials and dividing by

$$\theta^{2x+y+z}(1-\theta)^{y+z+2t} ,$$

we see that (16) is negative when

$$\theta^{y+t-x-z}(1-\theta)^{x+z-y-t} + \theta^{z+t-x-y}(1-\theta)^{x+y-z-t} < \frac{2(1-w)}{w} ,$$

and this is immediately equivalent to (15). In the case $y = z$, the quantity (16) should be divided by two, but the conclusion is the same. □

Lemma 4.4 *The set of good tables is an upper set of* $(\mathbb{T}, \leq)$.

Proof It is easy to see that, for each fixed $\frac{1}{2} < \theta < 1$, the function

$$H(x, y, t, z) = \left(\frac{\theta}{1-\theta}\right)^{(y-z)-(x-t)} + \left(\frac{\theta}{1-\theta}\right)^{(z-y)-(x-t)} \tag{17}$$

is decreasing in $(\mathbb{T}, \leq)$. It is enough to check it for the pairs of the transitive reduction (1). Moving from the smallest to the greatest table of the pair, one of the terms in (17) remains unchanged whereas the other one is multiplied by $\left(\frac{\theta}{1-\theta}\right)^{-2} < 1$. Therefore, if a table is good, according to Lemma 4.3 a greater table in the poset $(\mathbb{T}, \leq)$ is also good, and the good tables indeed form an upper set. □

Notice that the optimal rule according to L_w among those satisfying the first admissibility condition only, automatically satisfies the second.

The next theorem determines when, under the optimal decision rule, a voting table leads to a verdict of $C = P \wedge Q$ or the opposite, for the symmetric case $w = 1/2$. This is a further characterisation of the condition (15). We prove this special case because the statement and the proof are neater, and the extension to general $0 < w < 1$ is straightforward, as will be seen after the theorem.

In words, the theorem says that: if votes in favour of $P \wedge Q$ are less than those in favour of $\neg P \wedge \neg Q$ or there is a tie, then the decision must be $\neg C$; otherwise, if the difference is greater than the difference in absolute value between votes for $P \wedge \neg Q$ and votes for $\neg P \wedge Q$, then the decision must be C; otherwise, the decision must be C if the competence θ of the committee is below a certain threshold (which can be computed to any desired accuracy), and $\neg C$ if it is above.

Theorem 4.5 *Assume* $w = 1/2$, *and let* θ *be any competence level,* $\frac{1}{2} < \theta < 1$. *Assume r is an optimal rule and denote*

$$\rho := x - t \text{ and } \alpha := |y - z| .$$

Then, given a voting table $T = (x, y, z, t)$,

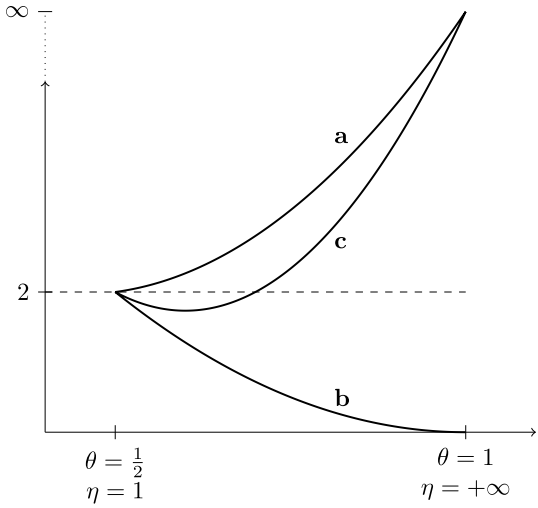

Fig. 2 Different possible shapes for function G of (18) according to the three cases in the proof of Theorem 4.5

(a) *if* $\rho \leq 0$, *then* $r(T) = 0$.
(b) *if* $\rho > \alpha$, *then* $r(T) = 1$.
(c) *if* $0 < \rho < \alpha$, *then there exists* $\theta_0 \in (\frac{1}{2}, 1)$ *such that* $r(T) = 1$ *for* $\theta < \theta_0$, *and* $r(T) = 0$ *for* $\theta > \theta_0$.

And these are all possible cases.

Proof Starting with the last claim, these are all possible cases because $n = x + y + z + t$ odd implies that α and ρ have different parity. In particular, $\rho \neq \alpha$ and $\rho \neq -\alpha$.

Consider the bijective increasing transformation $\eta = \frac{\theta}{1-\theta}$, which maps $(\frac{1}{2}, 1)$ onto $(1, \infty)$. The left-hand side of (15) can then be written as a function

$$G(\eta) := \eta^{-\rho-\alpha} + \eta^{-\rho+\alpha}, \quad \eta \in (1, \infty), \tag{18}$$

with ρ and α integers, $\alpha \geq 0$. The restriction $x + y + z + t = n$ implies that

$$0 \leq \alpha \leq n$$
$$-(n - \alpha) \leq \rho \leq n - \alpha$$

with ρ and α of different parity, as already noted. According to Lemma 4.3, the table T is good for values of η such that $G(\eta) < 2$, and the optimal rule r should assign it the value 1; for values of η such that $G(\eta) > 2$, the table is bad and we must have $r(T) = 0$.

Clearly, G is differentiable on $(1, \infty)$ and $\lim_{\eta \searrow 1} G(\eta) = 2$, for all ρ and α. Also, the derivative of G can always be written as

$$G'(\eta) = \eta^{-\rho-1}\left[(\alpha-\rho)\eta^{\alpha} - (\alpha+\rho)\eta^{-\alpha}\right], \tag{19}$$

and we have $\lim_{\eta\downarrow 1} G'(\eta) = -2\rho$. Thus, G takes the value 2 at the left boundary of the domain of interest, and starts from there decreasing or increasing according to the sign of ρ.

Let us now proceed with the three cases of the statement. Please refer to Fig. 2.

Tables of type **a**: $\rho \leq 0$.
If $\rho = 0$, then α cannot be zero, by parity. We have $\alpha \geq 1$ and G is clearly increasing, with $\lim_{\eta\to\infty} G(\eta) = +\infty$. If $\rho < 0$, G is also increasing for all $\alpha \geq 0$ and $\lim_{\eta\to\infty} G(\eta) = +\infty$ again. The table is bad in both situations.

Tables of type **b**: $\rho > \alpha$.
Both exponents in (18) are negative, and G is therefore decreasing. Moreover $\lim_{\eta\to\infty} G(\eta) = 0$. The table is good.

Tables of type **c**: $0 < \rho < \alpha$.
Now G' is negative near $\eta = 1$; therefore G is decreasing at least on some interval to the right of 1. Solving for η in $G'(\eta) = 0$, and taking into account that $\alpha + \rho > \alpha - \rho > 0$, we find that

$$\eta^* = \left(\frac{\alpha+\rho}{\alpha-\rho}\right)^{1/2\alpha} \in (1, \infty) \tag{20}$$

is the only critical point of G, which is a minimum since $\lim_{\eta\to\infty} G(\eta) = +\infty$. Thus, there exists a unique point $\eta_0 \in (\eta^*, \infty)$ where $G(\eta_0) = 2$. The table is good for $\eta \in (1, \eta_0)$, and bad for $\eta \in (\eta_0, \infty)$.
In other words, for a competence value θ less than $\theta_0 := \frac{\eta_0}{1+\eta_0}$, the table is good. For $\theta > \theta_0$, the table is bad. This finishes the proof.

□

The case for general $0 < w < 1$ is very easy to explain with the help of Fig. 2. For $w < \frac{1}{2}$, the dashed horizontal line is above level 2. All tables are good for small enough competence levels θ (the decision must always be C). The tables of type **a** and **c** are bad for θ greater than some θ_0. The tables of type **b** are all good for all competence levels.

For $w > \frac{1}{2}$, the dashed line is below level 2, and all tables are bad for low enough competence levels (the decision must always be $\neg C$). Tables **a** stay bad for all θ, and tables **b** turn good after some point. For tables of type **c**, two things may happen: Either they are always bad or, as θ increases, they have an interval of "goodness" before turning bad again.

All intersection points are easily computed to any desired precision by solving numerically for η the equation $G(\eta) = \frac{2(1-w)}{w}$ on $(1, \infty)$. See Sect. 5 for more details.

Lemma 4.3 allows a notable conceptual, notational and computational simplification: Since function G in (18) only depends on $\rho = x - t$ and $\alpha = |y - z|$, the tables in $\mathbb{T}$ with the same ρ and α will all be good or bad, once θ and w are fixed. If, on the contrary, two given tables do not share these values, they produce two different functions G.

This allows to consider an equivalence relation in $(\mathbb{T}, \leq)$ that gives rise to a quotient ranked poset, reducing in this way the complexity of the Hasse diagram and the computations. Define

$$(x, y, z, t) \sim (x', y', z', t') \Longleftrightarrow x - t = x' - t' \text{ and } |y - z| = |y' - z'| \,.$$

In particular, this equivalence relation identify transposed tables. The elements of the quotient set $\mathbb{T}/\sim$ are classes of voting tables and can be represented by the pair (ρ, α). We can write $T \in (\rho, \alpha)$ if T is in the class represented by (ρ, α).

Now define the preorder relation in $\mathbb{T}/\sim$ given by

$$(\rho, \alpha) \leq (\rho', \alpha') \Longleftrightarrow \text{there exist } T \in (\rho, \alpha) \text{ and } T' \in (\rho', \alpha') \text{ such that } T \leq T' \,.$$

We use the same symbol '$\leq$' for both relations in $\mathbb{T}$ and $\mathbb{T}/\sim$, since there is no possible confusion. It can be proved in general that a relation defined in this way in the quotient set is reflexive and transitive, therefore a *preorder*. In general it is not antisymmetric.

We shall prove that in our case the antisymmetry holds, so that we have again a partial order. To this end, we make use of the following lemma (see Hallam 2015). A proof is included in the "Appendix", for the reader convenience.

Lemma 4.6 *Let $(X, \leq)$ be a finite poset, $\sim$ an equivalence relation on X, and the preorder $\preceq$ on $X/\sim$ defined as: $\bar{x} \preceq \bar{y} \Leftrightarrow \exists x \in \bar{x}, \exists y \in \bar{y} : \; x \leq y$.*

Assume that if $\bar{x} \preceq \bar{y}$ in $X/\sim$, then for all $x \in \bar{x}$, there exists $y \in \bar{y}$ such that $x \leq y$ in X. Then, $(X/\sim, \preceq)$ is a poset.

It is not difficult to show that the hypothesis of the lemma holds in our case; see the "Appendix".

Since we are identifying tables in the same rank level, the resulting quotient poset is also ranked, with the same rank function ρ.

Example 4.7 Figure 3 illustrates the resulting poset of (ρ, α)-tables, for $n = 3, 5, 7$, with different intensities according to types **a**, **b**, **c** of Theorem 4.5. □

The relations (1) defining the transitive reduction in $(\mathbb{T}, \leq)$ translate to

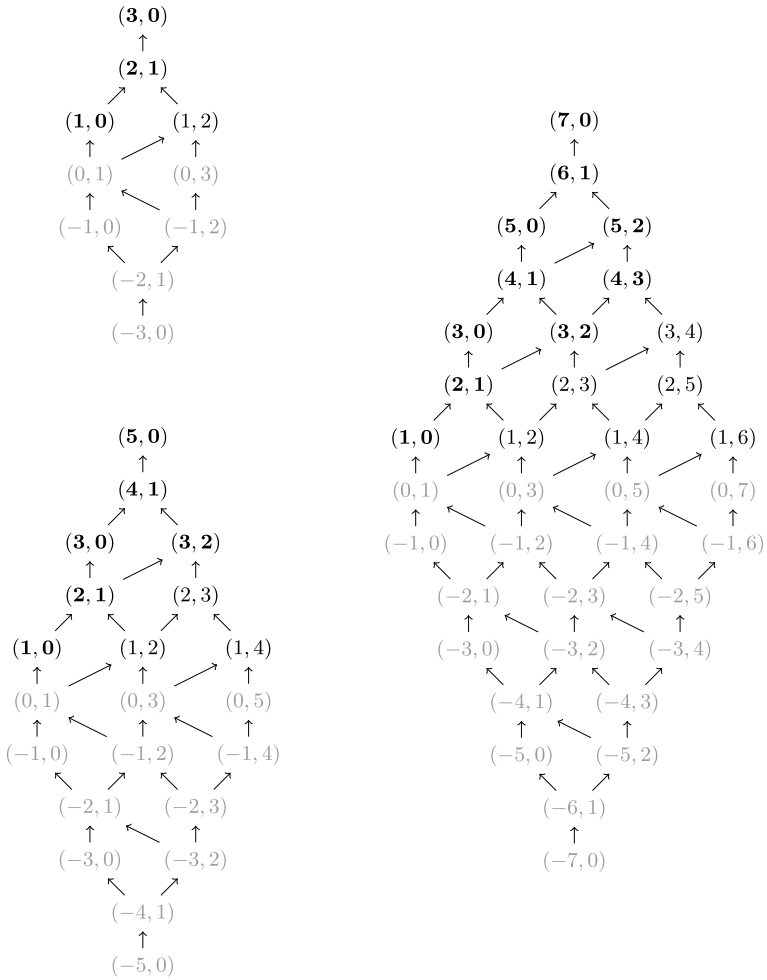

Fig. 3 Hasse diagrams for $n = 3, 5, 7$, in the quotient poset of (ρ, α)-tables. In boldface, tables of type **b**, in normal type those of type **c**, and greyed out those of type **a**

$$\begin{aligned}(\rho, \alpha) &\leq (\rho + 1, \alpha - 1)\\ (\rho, \alpha) &\leq (\rho + 1, \alpha + 1)\end{aligned} \tag{21}$$

in the quotient poset, and it is easy to see, while proving that Lemma 4.6 is applicable to $(\mathbb{T}, \leq)$ (see the "Appendix"), that (21) is precisely the transitive reduction in the quotient poset. This makes the Hasse diagrams like those of Fig. 3 very easy to generate for any n.

The classical premiss-based rule coincides with the one formed exactly by the tables of type **b** (see Example 2.2). This suggests that it is possible to characterise

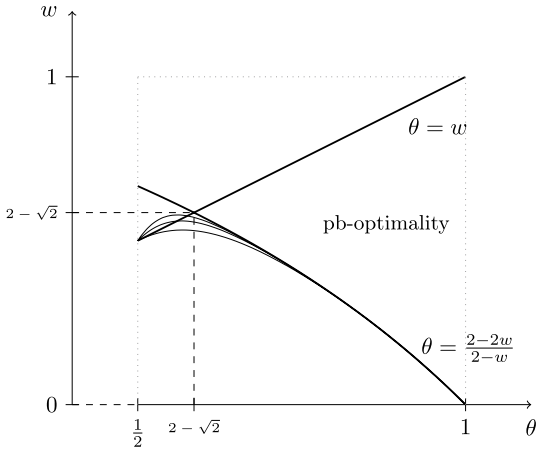

Fig. 4 Region of optimality of the premiss-based (pb) rule, in the natural coordinates (θ, w). The thinner curves correspond to $n = 3, 5, 7$ approaching monotonically the curve $\theta = \frac{2-2w}{2-w}$ as $n \to \infty$. If $\theta < w$, some tables of type **b** must leave the set $\mathbb{T}^+$; if (θ, w) falls below the lower curve, other tables must join those of type **b** in the set $\mathbb{T}^+$; and both things may happen at the same time, for $\theta < 2 - \sqrt{2}$

exactly under what conditions on w and θ the premiss-based rule is the optimal one. The result is given in Theorem 4.9. It will be a consequence of the following lemma. In the sequel we will denote $G_{\rho,\alpha}$ the function defined in (18).

Lemma 4.8 *In the poset* $(\mathbb{T}/\sim, \leq)$ *of* (ρ, α)*-tables*,

1. *The subset of tables of type* **b** *has a unique minimal element*: *Table* $(1, 0)$.
2. *The union of tables of type* **a** *and* **c** *has a unique maximal element*: *Table* $(\frac{n-1}{2}, \frac{n+1}{2})$.

Proof The statements are equivalent to say that $G_{\rho,\alpha} \leq G_{1,0}$ for $\rho > \alpha \geq 0$, and that $G_{\rho,\alpha} \geq G_{\frac{n-1}{2},\frac{n+1}{2}}$ for $\rho < \alpha$ with $\alpha \geq 0$. For the first inequality, both exponents $-\rho - \alpha$ and $-\rho + \alpha$ are negative, hence $G_{\rho,\alpha}(\eta) \leq \eta^{-1} + \eta^{-1} = G_{1,0}(\eta)$; for the second, the first exponent is positive, and the second is greater or equal to $-n$, hence $G_{\rho,\alpha}(\eta) \geq \eta + \eta^{-n} = G_{\frac{n-1}{2},\frac{n+1}{2}}(\eta)$. □

Theorem 4.9 *Let* $0 < w < 1$ *and* $\frac{1}{2} < \theta < 1$ *be the given weight and competence level*, *and n the committee size*. *The premiss-based rule is optimal if and only if*

$$\theta \geq w \quad \text{and} \quad \frac{\theta}{1-\theta} + \left(\frac{\theta}{1-\theta}\right)^{-n} \geq \frac{2(1-w)}{w} \tag{22}$$

Proof Denote, as before, $\eta := \frac{\theta}{1-\theta}$, and set also $\xi := \frac{2(1-w)}{w}$. In view of Lemma 4.8, the necessary and sufficient condition for the premiss-based rule to be optimal is that the point (η, ξ) lie above the curve $G_{1,0}$ and below the curve $G_{\frac{n-1}{2},\frac{n+1}{2}}$. That means $2\eta^{-1} \leq \xi \leq \eta + \eta^{-n}$, or equivalently,

$$2\left(\frac{\theta}{1-\theta}\right)^{-1} \leq \frac{2(1-w)}{w} \leq \frac{\theta}{1-\theta} + \left(\frac{\theta}{1-\theta}\right)^{-n}.$$

The first inequality is equivalent to $\theta \geq w$, and we are done. □

Table 2 Positive tables of type **c** when $w = \frac{1}{2}$ and $\theta \in (\frac{1}{2}, \theta_0)$

(ρ, α)	(3, 4)	(2, 3)	(1, 2)	(2, 5)	(1, 4)	(1, 6)
θ_0	0.6658	0.6628	0.6478	0.5449	0.5326	0.5141

A simpler sufficient condition, independent of the committee size, is given in the next corollary. Figure 4 illustrates both theorem and corollary.

Corollary 4.10 *Let* $0 < w < 1$ *and* $\frac{1}{2} < \theta < 1$ *be as in Theorem* 4.9. *If*

$$\theta \geq w \quad \text{and} \quad \theta \geq \frac{2(1-w)}{2-w},$$

then the premiss-based rule is optimal, for all committee sizes.

Proof The second inequality results from ignoring the negative exponential term in (22). □

Example 4.11 From the corollary, in the balanced case $w = \frac{1}{2}$, one can be sure that the premiss-based rule is optimal as soon as the competence level is above $\frac{2}{3}$. If this is not the case, other voting tables (those of type **c**) have to be successively added to the set of positive tables, as the competence level decreases. Table 2 shows the critical θ_0, to four decimal places, for the tables of type **c** with $n \leq 7$. Notice that the value of θ_0 is independent of n.

One might conjecture that the tables of type **c** enter the optimal rule following the increasing value of $-\rho + \alpha$, an among those with the same value, following the decreasing value of $\rho + \alpha$. This is true up to $n = 11$. For $n = 13$ this regularity breaks down and table (3, 10) enters before (1, 6), at $\theta_0 = 0.5160$. Hence, we do not find here any computational shortcut to determine the optimal rule for general n, even in the case $w = \frac{1}{2}$. □

By contrast, the classical conclusion-based rule is *never* optimal: For any n, a pair of tables (x, y, z, t) and (x', y', z', t') can be found that lead to different results according to conclusion-based, and yet they belong to the same (ρ, α) class. For instance, for $n = 3$, we have (2, 0, 0, 1) leading to C and (1, 1, 1, 0) leading to $\neg C$; but both belong to the class (1, 0) and should have the same status in the optimal rule. However, it is worth noting that the conclusion-based rule can be better than the premiss-based for certain values of w and θ.

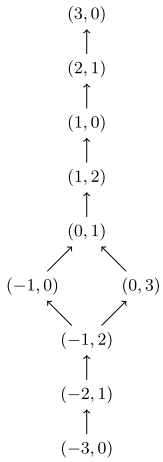

Fig. 5 Partial order in $\mathbb{T}$ induced by the optimality condition (see Proposition 5.1)

5 Computations and Software

The optimality condition adds some more relations to the partial order defined by the admissibility requirement. This simplifies further the computation of the optimal rule.

Proposition 5.1 *If* (ρ, α) *is a positive table in the optimal rule, then* $(\rho, \alpha - 2)$ *is also a positive table in the optimal rule, whenever these values make sense.*

Adding the relations $(\rho, \alpha) \leq (\rho, \alpha - 2)$ *to* $(\mathbb{T}, \leq)$, *the new partial order has the transitive reduction defined by*

$$\begin{aligned} (\rho, \alpha) &\leq (\rho, \alpha - 2) \\ (\rho, \alpha) &\leq (\rho + 1, \alpha + 1) \\ (n - 1, 1) &\leq (n, 0) \end{aligned}$$

Proof We need to see that for the functions $G_{\rho,\alpha}$ defined in (18), we have $G_{\rho,\alpha-2} \leq G_{\rho,\alpha}$ in their whole domain $(1, \infty)$. The inequality can be expressed as

$$0 \leq \eta^{2\alpha} - \eta^{2\alpha-2} - \eta^2 + 1 = (\eta^{2\alpha-2} - 1)(\eta^2 - 1)\,,$$

which is obviously true for all $\eta > 1$ and $\alpha \geq 2$.

The first relation in the transitive reduction (21) is no longer present in the new transitive reduction, since now there is an element in between:

$$(\rho, \alpha) \leq (\rho, \alpha - 2) \leq (\rho + 1, \alpha - 1)$$

except in $(n - 1, 1) \leq (n, 0)$. □

The resulting Hasse diagram is "thinner", and the total number of upper sets is reduced. As an example, the case $n = 3$ is depicted in Fig. 5. There are only

Position	Rule Antichain	Name	Loss Function
1	[1 1 1 0, 2 0 0 1]	premiss-based	0.2731
2	[1 2 0 0, 1 1 1 0, 2 0 0 1]		0.2881
3	[1 1 1 0]		0.2999
4	[1 1 0 1]		0.3013
5	[2 0 0 1]	conclusion-based, path-based	0.3058

Position	Rule Antichain	Name	Loss Function
1	[1 0]	premiss-based	0.2731
2	[1 0, 1 2]		0.2881
3	[0 1]		0.3229
4	[2 1]		0.3327
5	[1 0, 0 3]		0.3368

Fig. 6 The two rankings produced by the Python code: For rules in the form (x, y, z, t) and for rules in the form (ρ, α). In this example, the input was $n = 3$, $w = 0.5$, $\theta = (0.6, 0.7, 0.8)$

twelve upper sets left after this simplification. The new poset is still ranked, but $\rho = x - t$ is no longer a rank function.

This reduction is relevant if we are only interested in the optimal rule. If we want to build instead a ranking of rules, then Proposition 5.1 is not useful.

We have built a program in Python, with a graphical interface, that allows the user to input the values of n, w and θ, and produces a ranking of decision rules. It can be currently found as a public Mercurial repository in https://discursive-dilemma.sourceforge.io/, or requested directly to the authors. The program allows to specify different competence levels for the different committee members (an extension discussed in the next section), so that in fact formulae (11) and (12) are not used, but instead the probability to get a voting table (x, y, z, t) is computed taking into account all possible permutations of voters.

As an example of output, see Fig. 6. Two rankings are produced, the first corresponding to voting tables and rules in extended form (x, y, z, t), and the second in compact form (ρ, α). They are not a direct translation of each other, since a rule that cannot be expressed in compact form (because members of the same (ρ, α) class are assigned different conclusions) may actually be better than the next rule respecting the equivalence relation.

The rules are expressed by means of the antichain that determines the upper set of positive tables. A name is printed if the rule is one of premiss-based, conclusion-based or path-based, and the value of the loss function (14) of each rule is also given. Notice that in this example the rules in positions 3 to 5 in the extended version are not expressible in compact form, but are better than the third rule in the second ranking. Of course, the optimal rule will always coincide in both rankings.

At the moment, the program only outputs up to the five best rules, but this is an arbitrary parameter that can be easily changed in the source code. Also, we have not made any special effort for efficiency. It has been conceived only as a playground and checking tool.

Concerning the computational complexity of producing the admissible decision rules, let us note first that the total number of voting tables is equal to $\frac{1}{24}(2n^3 + 15n^2 + 34n + 21)$ in the original poset $(\mathbb{T}, \leq)$, after identifying transposed tables. In the quotient poset $(\mathbb{T}/\sim, \leq)$, this number is reduced to $\frac{1}{2}(n+1)(n+2)$.

Since admissible rules are in bijection with upper sets, and these in turn are determined by their antichains, we first identify the latter. For this task we make use of the Python package `networkx`, which contains a function `antichains()`. The generation of the corresponding upper set and the evaluation of each table contained in it is very easy. The contribution to the loss function of tables previously computed is stored to speed up the computations. The maximum cardinality of an antichain (after identifying transposed tables) is $\frac{1}{8}(n+3)(n+1)$ in $(\mathbb{T}, \leq)$ and $\frac{1}{2}(n+1)$ in $(\mathbb{T}/\sim, \leq)$. We sketch in the "Appendix" the computation of these numbers.

Apart from having the optimal rule or a ranking of the best rules, one might be just interested in knowing which conclusion has to be assigned to a given voting table T under the optimal rule. This is very easy by asserting inequality (15), in the case of equal competences. Otherwise, the probabilities of a False Positive and a False Negative have to be computed for that table, taking into account the different competences, and then determine their contribution to the loss function.

Finally, in the equal competences case, one may like to determine, given a fixed weight w, the intervals of competence θ where $r(T) = 0$ or 1 for the optimal rule r. We need to find the root or roots of the equation

$$G_{\rho,\alpha}(\eta) = \eta^{-\rho-\alpha} + \eta^{-\rho+\alpha} = \frac{2(1-w)}{w} \,. \tag{23}$$

This is very easy numerically. The functions involved are simple to evaluate so that pure bisection, for instance, is very fast. We only need to have a bracket where the roots are guaranteed to lie. Indeed, they are readily found:

Tables of type **a**: $(\rho \leq 0)$.

If $w \geq \frac{1}{2}$ the table is bad.

If $w < \frac{1}{2}$, the unique root of (23) is less than the root of $\eta^{-\rho+\alpha} = \frac{2(1-w)}{w}$. Therefore, the solution to (23) will be found between 1 and

$$\left(\frac{2(1-w)}{w}\right)^{\frac{1}{-\rho+\alpha}} \,. \tag{24}$$

Tables of type **b**: $(\rho \geq \alpha)$.

If $w \leq \frac{1}{2}$ the table is good.

If $w > \frac{1}{2}$, the unique root of (23) is less than the root of $2\eta^{-\rho+\alpha} = \frac{2(1-w)}{w}$, and it will be found between 1 and

$$\left(\frac{w}{1-w}\right)^{\frac{1}{\rho-\alpha}}.$$

Tables of type **c**: $(0 < \rho < \alpha)$.

If $w \leq \frac{1}{2}$ there is a unique root, and the same bound (24) is valid.

If $\frac{1}{2} < w \leq \frac{2}{3}$, we first check if the minimum η^* of $G_{\rho,\alpha}$ (see (20)) satisfies $G_{\rho,\alpha}(\eta^*) > \frac{2(1-w)}{w}$, in which case there are no roots of (23) and the table is bad; or $G_{\rho,\alpha}(\eta^*) < \frac{2(1-w)}{w}$ and there are two roots: The first between 1 and η^*, and the second between η^* and (24).

If $w > \frac{2}{3}$, the table is bad, because then $\frac{2(1-w)}{w} < 1$, whereas $G_{\rho,\alpha}(\eta^*) > 1$ always.

6 Conclusions and Discussion

We have studied the simplest conjunctive truth-functional agenda, origin of the discursive dilemma between the two classical decision rules. We proposed a method to obtain the best rule (or a ranking of rules) by minimising a loss function that combines false positives and false negatives. Actually, we have introduced a family of loss functions, parametrised by the number $0 < w < 1$.

We have seen that, given a voting table, it is very easy to determine the decision that the optimal decision rule dictates, by applying Theorem 4.5, in the symmetric case, and finding the root(s) of equation (23) if necessary, in general.

The decision rules considered satisfy very mild and reasonable conditions of symmetry and monotonicity (Definition 2.1). In fact, the second condition is not necessary a priori if one is only interested in the best rule and not in ranking rules. In that case, monotonicity appears a posteriori as a property of the optimal rule.

Generically, the optimal rule will be unique, but specific values of weight w and competence θ may lead to ties in the evaluation of the loss function, in particular in its minimum value. To make the exposition simpler, we have avoided mentioning this possibility throughout the paper.

The loss function is a modelling choice; an object to be decided upon a priori and that makes clear what is the criterion under which a rule will be declared "best". It contains all the information that the mathematical model needs to know. With this approach, we draw attention away from the properties of specific decision rules, or from a set of desirable axioms, to focus on the consequences of the decisions.

In any real instance the loss function must be chosen to reflect what the best rule is intended to achieve. Both the choice of the loss function and the optimisation point of view as a whole can be a matter of extensive discussion. But in any case we strongly believe that our approach is worth considering for problems of judgment aggregation in general.

The classical premiss-based and conclusion-based decision rules are linked to the concept of majority: The value of each clause is the one most voted. But we do not use majorities in determining the best rule. We rather support a qualified majority principle: the loss function quantifies the bad consequences of a wrong decision; in turn, this determines the "threshold" for this qualified majority. Not a threshold premiss-by-premiss, but a threshold (or a *cut*, in terms of graph theory) in the partially ordered set of voting tables. In an extreme case, when the competence of the committee is too low, an unanimous vote in one direction may still lead to the contrary decision, if the consequence of deciding the first is too bad. On the other hand, an ample range of values of competence and weight do support the application of the classical premiss-based rule (Corollary 4.10 and Fig. 4).

In Alabert and Farré (2022), where the present point of view was first introduced to compare three specific decision rules, some possible extensions and open problems were discussed substantially. We summarise them here:

- *Different competence for each voter* This is the simplest extension. If J_k is the voting table consisting only on the vote of voter k, with competence level θ_k, the resulting table is $J_1 + \cdots + J_n$, whose probability law can also be computed, and the probabilities of false positive and false negative will be

$$\mathbb{P}_r(\mathrm{FP}) = \sum_{\{r(x,y,z,t)=1\}} \mathbb{P}_{P\wedge\neg Q}\{J_1 + \cdots + J_n = (x,y,z,t)\}\ ,$$
$$\mathbb{P}_r(\mathrm{FN}) = \sum_{\{r(x,y,z,t)=0\}} \mathbb{P}_{P\wedge Q}\{J_1 + \cdots + J_n = (x,y,z,t)\}\ .$$

  Our software already computes the ranking of rules in this more general situation.
- *Different competence for each premiss or state of nature.* The competence of a voter may be in fact a vector $\theta = (\theta_P, \theta_{\neg P}, \theta_Q, \theta_{\neg Q})$ of competences depending on the premiss and/or the true state of nature. In List (2005), the probability of appearance of the doctrinal paradox is studied also when the competence is different on P and Q. The computation of $\mathbb{P}_r(\mathrm{FP})$ and $\mathbb{P}_r(\mathrm{FN})$ is more involved in this case, but still feasible.
- *Non-independence between voters* If the committee members do not vote independently, perhaps through a deliberation process with influential individuals, then the full joint law of the vector $(J_1, \ldots, J_n)$ of individual voting tables is needed to compute the law of the sum $J_1 + \cdots + J_n$ under the different states of nature. Boland (1989) studied this situation for the voting of a single question assuming the presence of a "leader" in the committee. Other works that studied epistemic social choice with correlated voters in the last decade include Peleg and Zamir (2012), Dietrich and Spiekermann (2013a), Dietrich and Spiekermann (2013b) and Pivato (2017).
- *Non-independence between premisses* In practical examples, the premisses P and Q can very well be interconnected, in the sense that believing that P is true or false can change the perception on the truth or falsity of Q. This may lead to a different competence in asserting Q depending on the decision on P. Then the

joint law of the competences under the four states of nature are needed to complete the computations.

The extreme case where one combination of premisses is impossible it is treated in Bozbay (2019), where in addition abstentions are allowed.

- *More than two premisses* Conceptually, there is no difficulty in extending the setting to a conjunctive agenda with any number of premisses $P_1, \dots, P_s$. A voting table will be an element of $\mathbb{T} = \{(x_1, \dots, x_{2^s}) \in \mathbb{N}^{2^s} : \sum_{i=1}^{2^s} x_i = n\}$. The concepts of admissible rule and of false positive are easily extended, and the loss function is the same. However, the analogue of the left-hand side expression of equation (15), and consequently the function G used in the proof of Theorem 4.5, are not so easy to obtain explicitly. There are many more symmetries in the voting arrays to take into account, and this requires further non-trivial work.

Note that *disjunctive agendas*, in which the conclusion is true if and only if at least one premiss is true, are dual to the conjunctive case, by negation of the doctrine (see List 2005, Bovens and Rabinowicz 2006, or Miyashita 2021). They can be considered easily within our framework. We do not know how sensitive the results will be to other type of truth-functional agendas.

An extension with an obvious practical interest is allowing abstentions, or committees with an even number of members. It is clear that enforcing an opinion on all clauses of the agenda may be inconvenient or simply impossible. These so-called *incomplete judgments* have been considered in Gärdenfors (2006), Dietrich and List (2008), Terzopoulou and Endriss (2019) and Bozbay (2019).

It is natural to ask which desirable properties satisfies the optimal rule of a given criterion. We leave this as an open question. In relation to the classical axioms of judgment aggregation and their (im)possibility theorems (see e.g. List 2012), and since here we are centred in reaching a right conclusion for whatever reasons, collective rationality can only be achieved by assigning a value to the premisses *after* deciding on the conclusion (see Pigozzi et al. 2009); but then the properties of monotonicity (in the classical sense), unanimity and systematicity need not be satisfied on the whole agenda. On the other hand, the anonymity requirement is trivially met in our setting. In any case, the advantage of the optimisation model is the immediate existence of decision rules; each of the rules is evaluated through a real-valued loss function, hence at least one rule with a minimal value must exist. Distance-based methods to reach consensus share this feature.

Appendix

- *Proof of the equivalence between admissibility and upper sets* (*Proposition* 2.5)

Given $r \in \mathcal{A}$, the set $\{u \in \mathbb{T} : r(u) = 1\}$ is an upper set of $\mathbb{T}$. We want to see that for all $x \in S = \{u \in \mathbb{T} : r(u) = 1\}$ and for all $y \in \mathbb{T}$, such that $x < y$, we must have $y \in S$. Take $x \in S$ and $y \in \mathbb{T}$ such that $x < y$. Since $r \in \mathcal{A}$ we have $1 = r(x) < r(y)$, hence $r(y) = 1$, and therefore $y \in S$.

- *Proof of the final claim in Sect. 3.2*

In case $\theta = \frac{1}{2}$, the problem is trivial because denoting $\Sigma = \sum_{r(x,y,z,t)=0} \frac{n!}{x!y!z!t!}(\frac{1}{2})^{2(x+y+z+t)}$, we have $L_w(r) = w(1-\Sigma) + (1-w)\Sigma = w + (1-2w)\Sigma$.

If $w < \frac{1}{2}$, the minimum of L_w is equal to w, achieved when $\Sigma = 0$, that is, when $r \equiv 1$; if $w > \frac{1}{2}$, the minimum of L_w is equal to $1-w$, achieved when $\Sigma = 1$, that is, when $r \equiv 0$; if $w = \frac{1}{2}$, the function L_w is constant and equal to $\frac{1}{2}$, that is, all rules are equally good and we might as well toss a coin.

- *Proof that* $(\rho, \alpha) \neq (\rho', \alpha')$ *implies* $G_{\rho,\alpha} \neq G_{\rho',\alpha'}$ (*Sect.* 4)

Since $G'_{\rho,\alpha}(1) = -2\rho$, two different ρ yield for sure two different functions. Suppose $\rho = \rho'$ and $\alpha' > \alpha \geq 0$. Then, $\lim_{\eta\to\infty} \eta^{\rho-\alpha} G_{\rho,\alpha}(\eta) < \infty$ and $\lim_{\eta\to\infty} \eta^{\rho-\alpha} G_{\rho,\alpha'}(\eta) = \infty$, hence $G_{\rho,\alpha}$ and $G_{\rho,\alpha'}$ must be different.

- *Proof of Hallam's lemma* (*Lemma* 4.6)

His proof can be found in Hallam's PhD thesis (Hallam 2015); we include it here for the reader's convenience, and because his statement does not correspond completely to the proof.

Suppose $\bar{x} \preceq \bar{y}$ and $\bar{y} \preceq \bar{x}$. Then $x_1 \leq y_1$ for some $x_1 \in \bar{x}$ and $y_1 \in \bar{y}$. But $\bar{y} \preceq \bar{x}$ implies there must be some $x_2 \in \bar{x}$ such that $y_1 \leq x_2$. But then, there must be some $y_2 \in \bar{y}$ such that $x_2 \leq y_2$. And so on. At some point we must have an equality of elements, since the set is finite. Then the classes $\bar{x}$ and $\bar{y}$ must coincide.

- *Proof that the hypothesis of Hallam's lemma holds in our case*

We can assume everywhere that $y \geq z$. Otherwise, at any time we can interchange the roles of y and z. Assume that $(\rho_0, \alpha_0) \leq (\rho_1, \alpha_1)$, with elements $(x_0, y_0, z_0, t_0) \in (\rho_0, \alpha_0)$ and $(x_1, y_1, z_1, t_1) \in (\rho_1, \alpha_1)$, such that $(x_0, y_0, z_0, t_0) \leq (x_1, y_1, z_1, t_1)$, and assume in addition that they belong to the transitive reduction in $(\mathbb{T}, \leq)$. It is easy to find that there are two cases: Either (I) $\rho_1 = \rho_0 + 1$ and $\alpha_1 = \alpha_0 - 1$, or (II) $\rho_1 = \rho_0 + 1$ and $\alpha_1 = \alpha_0 + 1$.

Take a generic element $(x, y, z, t) \in (\rho_0, \alpha_0)$. In case (I), we have in particular $\alpha_0 > 0$, which implies $y > 0$. Then $(x, y, z, t) \leq (x+1, y-1, z, t) \in (\rho_1, \alpha_1)$. In case (II), if $t > 0$, take $(x, y+1, z, t-1) \in (\rho_1, \alpha_1)$, and if $t = 0$, take $(x+1, y, z-1, t) \in (\rho_1, \alpha_1)$. Note that $z = t = 0$ cannot happen.

If the elements of (ρ_0, α_0) and (ρ_1, α_1) are not related by the transitive reduction, the argument can be iterated through a chain of elements related by the transitive reduction. Note that we have also deduced en passant the transitive reduction of the quotient poset.

- *Computation of the number of voting tables* (*Sect.* 5)

The *Whitney numbers* W_ρ of a finite ranked poset are defined as the number of elements in rank level ρ. We compute the total number of tables in $\mathbb{T}$ by computing first the Whitney numbers. We assume transposed tables are identified.

For odd positive ranks ρ, the possible values of the pair (x, t) are $(\rho + r, r)$, for $r = 0, 1, \ldots, \frac{n-\rho}{2}$.

For each fixed r, the possible values of the pair (y, z), with $y \geq z$, are $(n - \rho - 2r - s, s)$, for $s = 0, 1, \frac{n-\rho-2r}{2}$. Thus, there are $\frac{n-\rho-2r}{2} + 1$ such pairs. adding up these quantities from $r = 0$ to $\frac{n-\rho}{2}$, yields $\frac{1}{8}(n - \rho + 4)(n - \rho + 2)$.

A similar counting gives $\frac{1}{8}(n - \rho + 3)(n - \rho + 1)$ for even non-negative ranks ρ, which is the same number as the odd rank immediately above. The case of negative ranks is deduced by symmetry. Therefore we can write

$$W_\rho = \begin{cases} \frac{1}{8}(n - \rho + 4)(n - \rho + 2), & \text{if } \rho = 1, 3, \ldots, n \\ W_{\rho+1}, & \text{if } \rho = 0, 2, \ldots, n-1 \\ W_{-\rho}, & \text{if } -n \leq \rho \leq -1 \end{cases}$$

A simple but tedious computation, adding up all the Whitney numbers for $-n \leq \rho \leq n$, yields the total number of tables $\sum_{\rho=-n}^{n} W_\rho = \frac{1}{24}(2n^3 + 15n^2 + 34n + 21)$.

The number $\max_\rho W_\rho$ is the maximal cardinality of an antichain: Indeed, since there is a unique minimal and a unique maximal element in $(\mathbb{T}, \leq)$, any element of the poset is comparable to the minimal and maximal element of the poset, and therefore to some element of the most populated rank level. This implies that there cannot be more that $\max_\rho W_\rho$ elements in any antichain. Ranks -1, 0, and 1 are the most populated, and its Whitney number is $\frac{1}{8}(n + 3)(n + 1)$.

For completeness, let us just mention that the number of tables in the original poset $(\mathbb{T}, \leq)$, without identifying transposed tables, is $\frac{1}{24}(4n^3 + 24n^2 + 44n + 24)$, and the most populated rank has $\frac{1}{4}(n + 3)(n + 1)$ elements. In the quotient poset of (ρ, α)-tables, there are $\frac{1}{2}(n + 2)(n + 1)$ tables and a maximum of $\frac{1}{2}(n + 1)$ members in any antichain. All computations are similar to those shown here.

Funding Open Access Funding provided by Universitat Autonoma de Barcelona. No funding was received for this particular work, apart from the regular salaries, administrative support and premises of our employer, Universitat Autònoma de Barcelona.

Code Availability This paper has associated code in a public repository (Sourceforge), with https://discursive-dilemma.sourceforge.io.

Declarations

Conflict of interest The authors have no conflict of interest related to the contents of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Alabert A, Farré M (2022) The doctrinal paradox: comparison of decision rules in a probabilistic framework. Soc Choice Welf 58(4):863–895

Berger JO (1985) Statistical decision theory and Bayesian analysis. Springer series in statistics, 2nd edn. Springer, New York

Boland PJ (1989) Majority systems and the Condorcet Jury Theorem. J R Stat Soc Ser D (Stat) 38(3):181–189

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

Bozbay İ (2019) Truth-tracking judgment aggregation over interconnected issues. Soc Choice Welf 53(2):337–370

Bozbay İ, Dietrich F, Peters H (2014) Judgment aggregation in search for the truth. Games Econ Behav 87:571–590

de Clippel G, Eliaz K (2015) Premise-based versus outcome-based information aggregation. Games Econ Behav 89:34–42

Dietrich F (2006) Judgment aggregation: (im)possibility theorems. J Econ Theory 126(1):286–298

Dietrich F, List C (2007) Judgment aggregation by quota rules: majority voting generalized. J Theor Politics 19

Dietrich F, List C (2008) Judgment aggregation without full rationality. Soc Choice Welf 31(1):15–39

Dietrich F, Spiekermann K (2013) Epistemic democracy with defensible premises. Econ Philos 29(1):87–120

Dietrich F, Spiekermann K (2013) Independent opinions? On the causal foundations of belief formation and jury theorems. Mind 122(487):655–685

Dokow E, Holzman R (2009) Aggregation of binary evaluations for truth-functional agendas. Soc Choice Welf 32(2):221–241

Fallis D (2005) Epistemic value theory and judgment aggregation. Episteme 2(1):39–55

Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27(8):861–874

Gärdenfors P (2006) A representation theorem for voting with logical consequences. Econ Philos 22(2):181–190

Grofman B, Owen G, Feld SL (1983) Thirteen theorems in search of the truth. Theory Decis 15(3):261–278

Hallam J (2015) Quotient Posets and the characteristic polynomial. PhD thesis, Michigan State University

Hartmann S, Pigozzi G, Sprenger J (2010) Reliable methods of judgement aggregation. J Logic Comput 20(2):603–617

Kornhauser LA (1992) Modeling collegial courts I: path-dependence. Int Rev Law Econ 12(2):169–185

Kornhauser LA (1992) Modeling collegial courts, II. Legal doctrine. J Law Econ Org 8(3):441–470

Kornhauser LA, Sager LG (1993) The one and the many: adjudication in collegial courts. Calif Law Rev 81(1):1–51

Lang J, Pigozzi G, Slavkovik M, van der Torre L, Vesic S (2017) A partial taxonomy of judgment aggregation rules and their properties. Soc Choice Welf 48(2):327–356

List C (2005) The probability of inconsistencies in complex collective decisions. Soc Choice Welf 24(1):3–32

List C (2012) The theory of judgment aggregation: an introductory review. Synthese 187(1):179–207

List C, Pettit P (2002) Aggregating sets of judgments: an impossibility result. Econ Philos 18(1):89–110

List C, Pettit P (2004) Aggregating sets of judgments: two impossibility results compared. Synthese 140(1):207–235

List C, Polak B (2010) Introduction to judgment aggregation. J Econ Theory 145(2):441–466

List C, Puppe C (2009) Judgment aggregation. In: Anand P, Pattanaik P, Puppe C (eds) Handbook of rational and social choice. Oxford University Press

Miller MK, Osherson D (2009) Methods for distance-based judgment aggregation. Soc Choice Welf 32(4):575–601

Miyashita M (2021) Premise-based vs conclusion-based collective choice. Soc Choice Welf 57(2):361–385

Mongin P (2012) The doctrinal paradox, the discursive dilemma, and logical aggregation theory. Theory Decis 73(3):315–355

Nehring K, Pivato M (2011) Incoherent majorities: the McGarvey problem in judgement aggregation. Discrete Appl Math 159(15):1488–1507
Nehring K, Puppe C (2008) Consistent judgement aggregation: the truth-functional case. Soc Choice Welf 31(1):41–57
OEIS Foundation. The on-line encyclopedia of integer sequences, sequence number A000372. https://oeis.org/
Pauly M, van Hees M (2006) Logical constraints on judgement aggregation. J Philos Logic 35(6):569–585
Peleg B, Zamir S (2012) Extending the Condorcet jury theorem to a general dependent jury. Soc Choice Welf 39(1):91–125
Pettit P (2001) Deliberative democracy and the discursive dilemma. Philos Issues 11(1):268–299
Pigozzi G (2006) Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. Synthese 152(2):285–298
Pigozzi G, Slavkovik M, van der Torre L (2009) A complete conclusion-based procedure for judgment aggregation. In: Rossi F, Tsoukias A (eds) Algorithmic decision theory. Springer, Berlin, pp 1–13
Pivato M (2013) Voting rules as statistical estimators. Soc Choice Welf 40(2):581–630
Pivato M (2017) Epistemic democracy with correlated voters. J Math Econ 72:51–69
Terzopoulou Z, Endriss U (2019) Optimal truth-tracking rules for the aggregation of incomplete judgments. In: Proceedings of the 12th international symposium on algorithmic game theory (SAGT-2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.