



# The doctrinal paradox: comparison of decision rules in a probabilistic framework

Aureli Alabert<sup>1</sup> · Mercè Farré<sup>1</sup>

Received: 17 February 2020 / Accepted: 21 October 2021 / Published online: 14 November 2021  
© The Author(s) 2021, corrected publication 2021

## Abstract

The *doctrinal paradox* is analysed from a probabilistic point of view assuming a simple parametric model for the committee's behaviour. The well known *premise-based* and *conclusion-based* majority rules are compared in this model, by means of the concepts of *false positive rate* (FPR), *false negative rate* (FNR) and *Receiver Operating Characteristics* (ROC) space. We introduce also a new rule that we call *path-based*, which is somehow halfway between the other two. Under our model assumptions, the premise-based rule is shown to be the best of the three according to an optimality criterion based in ROC maps, for all values of the model parameters (committee size and competence of its members), when equal weight is given to FPR and FNR. We extend this result to prove that, for unequal weights of FNR and FPR, the relative goodness of the rules depends on the values of the competence and the weights, in a way which is precisely described. The results are illustrated with some numerical examples.

## 1 Introduction

### 1.1 The doctrinal paradox

The *Condorcet Jury Theorem* (attributed to Condorcet (1785)) states that “if  $n$  jurists act independently, each with probability  $\theta > \frac{1}{2}$  of making the correct decision, then the probability that the jury (deciding by majority rule) makes the correct decision increases monotonically to 1 as  $n$  tends to infinity”. See for instance Boland (1989) and Karotkin and Paroush (2003), and the references therein, for precise statements, proofs, and extensions of this principle.

---

✉ Aureli Alabert  
Aureli.Alabert@uab.cat

Mercè Farré  
farre@mat.uab.cat

<sup>1</sup> Department of Mathematics, Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalonia

The *doctrinal paradox* (a name introduced by Kornhauser (1992) arises in some situations when a committee or jury has to answer a compound question divided in two subquestions or premises,  $P$  and  $Q$ . The point of interest is in deciding between the acceptance of both premises  $P \wedge Q$  ( $P$  and  $Q$ ) and the acceptance of the opposite  $\neg(P \wedge Q) = \neg P \vee \neg Q$  (*not P or not Q*). In view of the Condorcet Jury Theorem, some kind of majority rule seems appropriate for this two-premises problem. However, in some cases, the same set of individual decisions leads to different collective decisions depending on the manner in which the individual opinions are aggregated.

Classically, two standard decision procedures are considered in the literature: the *conclusion-based* and the *premise-based* procedures (*Conc* and *Prem*, respectively, for short). In *Conc*, each committee member or judge decides on both questions and votes  $P \wedge Q$  or  $\neg(P \wedge Q)$ . Then, simple majority wins. In the *Prem* procedure, each committee member decides  $P$  or  $\neg P$  first, and then a joint decision about this premise is taken by simple majority. Similarly, each member chooses between  $Q$  or  $\neg Q$ , and a joint decision is taken again by simple majority. If  $P$  and  $Q$  are separately chosen by a (perhaps differently formed) majority, then  $P \wedge Q$  is proclaimed. Otherwise,  $\neg(P \wedge Q)$  is the conclusion.

Procedure *Conc* is sometimes referred in the literature as the *case-by-case* rule (Kornhauser 1992; Kornhauser and Sager 1993). In fact, it is a reduction to the one-premise Condorcet case. Procedure *Prem* is then referred as the *issue-by-issue* rule.

Both procedures look reasonable, but they may give rise to different results, hence the “paradox”. The simplest example is the case of a 3-member committee, when there is one vote for  $P \wedge Q$ , one for  $P \wedge \neg Q$  and one for  $\neg P \wedge Q$ . The *Prem* rule leads to decide in favour of  $P \wedge Q$ , whereas the *Conc* rule leads to the contrary.

In general, if we have a committee with  $n$  members, we can summarise their votes as in Table 1, where  $x, y, z$  and  $t$  are the number of votes received by each of the options  $P \wedge Q, P \wedge \neg Q, \neg P \wedge Q$ , and  $\neg P \wedge \neg Q$  respectively. We will assume throughout the paper that  $n$  is an odd number:  $n = 2m + 1, m \geq 1$ . The doctrinal paradox appears when the following conditions are simultaneously satisfied:

$$x + y > m, \quad x + z > m, \quad x \leq m.$$

In the sequel Table 1 will be called a *voting table*, but we will write it simply as a matrix

**Table 1** Distribution of  $n$  votes on two premises

	$Q$	$\neg Q$	Totals
$P$	$x$	$y$	$x + y$
$\neg P$	$z$	$t$	$z + t$
Totals	$x + z$	$y + t$	$n$

$$\begin{bmatrix} x & y \\ z & t \end{bmatrix},$$

and also as a vector  $(x, y, z, t)$ , to save space.

We now introduce a new decision rule which is also reasonable and, as we will see later, lies in some sense in between of the classical *Prem* and *Conc* rules. We call it path-based (*Path*, for short), and it can be defined as follows:  $P \wedge Q$  is proclaimed if the number of voters that individually conclude  $P \wedge Q$  is greater than those who conclude  $\neg P$ , and greater than those who conclude  $\neg Q$ , separately. That is, in the notation of Table 1, if  $x > z + t$  and  $x > y + t$ . Comparing with *Prem*, only the number of supporters of  $P \wedge Q$  can be used to beat  $\neg P$ , without using the votes for  $P \wedge \neg Q$ , and similarly to beat  $\neg Q$  without using the votes for  $\neg P \wedge Q$ ; it is therefore a stronger requirement to conclude  $P \wedge Q$ . Comparing with *Conc*, in order to conclude  $P \wedge Q$ , in *Path* the votes for  $P \wedge Q$  do not need to beat the sum of all other options, but only those who deny  $P$  and those who deny  $Q$ , separately, which is a weaker statement.

We can justify morally this new rule by saying that supporters of the conclusion  $P \wedge Q$  must form a majority against the detractors of  $P$ , no matter its position about  $Q$ , and symmetrically for the other premise. It is not our intention to stand for a different “reasonable” rule, but to make it visible that, apart from the two classic rules, some others can be considered. For these specific three rules, we will show that any one of them is the best, depending on the adopted criterion of optimality, drawn from a family of perfectly reasonable criteria.

Our goal is to compare the performance of the three decision rules, for different committee sizes and different individual competence of its members. To this end, we define a theoretical framework consisting of a probabilistic model where the *competence* of a judge is defined as the probability that he/she takes the correct decision about each single premise. It is assumed that a “true state of nature” or “absolute truth” exists, which is one of the four possibilities that combine  $P$ ,  $Q$  and their negations.

Our performance criterion is based on the concepts of true and false positive and negative rates and the *Receiver Operating Characteristics* (ROC) space. They have their origin in the field of electrical engineering and are commonly used in medicine, machine learning and other scientific disciplines (see e.g. Fawcett 2006; Hand and Till 2001). We believe that its application to the doctrinal paradox is completely new, and that it provides an acceptable framework to decide which one of a given set of rules is the best to get the right conclusion. As will be apparent later, our analysis can be applied to any given set of rules, beyond those considered here.

We want to stress the fact that we treat conclusion and premises at a different level. We concentrate in assessing different rules in their ability to get the conclusion right; not the premises. This is reflected in the consideration of false positives and false negatives of the conclusion only. We also note, however, that the computation of these false positives and negatives depends directly on the ability of the committee to get the premises right.

## 1.2 Related literature

The problem of a committee assessing the truth or falsity of the three logical clauses  $P$ ,  $Q$  and  $R$ , with the constraint  $R \Leftrightarrow P \wedge Q$  is only an instance of the broader situation in which a collective decision is to be built from individual decisions in a community. The theory of *judgement aggregation* aims at studying and shedding light into these kind of problems. We refer the reader to the surveys (List 2012) and (List and Puppe 2009) for an overview of the field and its recent developments.

The doctrinal paradox is correspondingly a particular case of a general impossibility theorem inside that theory (see, e.g. List and Pettit 2002): Under reasonable assumptions, there exist individual logically consistent decisions on  $P$ ,  $Q$  and  $R$  that lead to collective inconsistent decisions. Dietrich and List (2007) prove Arrow's impossibility result on preference aggregation as a corollary of this impossibility in judgement aggregation. See also Camps et al. (2012) for a new approach to the problem of constrained judgement aggregation in a general setting.

The concept of decision rule that we introduce in the next section is somewhat narrower than that of aggregation rule in judgement aggregation theory, but sufficient and adapted to our purposes. We do not go further explaining judgement aggregation theory concepts since we focus specifically in the doctrinal paradox with a simple model of behaviour of the committee members. For instance, we disregard strategic behaviour, considered in Dietrich and List (2007), de Clippel and Eliaz (2015), Ahn and Oliveros (2014) and Terzopoulou and Endriss (2019), or the epistemic or behavioural perspective, studied in Bovens and Rabinowicz (2004), Bovens and Rabinowicz (2006), and Bonnefon (2010).

We consider that a true state of nature exists (not known, but certain) and that the committee members are seeking this absolute truth (the so called *truth-tracking* preference, see Bozbay (2019) and the references therein).

Sometimes, as in the recent papers (Bozbay 2019; Terzopoulou and Endriss 2019), the state of nature is thought of as a random experiment, with an assumed prior probability distribution on the set of possible states, which in our case is the set  $\{P \wedge Q, \neg P \wedge Q, P \wedge \neg Q, \neg P \wedge \neg Q\}$ . This Bayesian approach is justified in applications where there is indeed a previous experience, independent of the decision to be currently made. We do not assume any prior probability. As a consequence, the negative conclusion  $\neg(P \wedge Q)$  is in fact composed by three different states of nature, and we use the concepts of classical statistics to state the notion of risk when concluding that  $P \wedge Q$  is false when in fact it is true. Some more notes on the Bayesian approach are pointed out in the discussion section.

Judgement Aggregation Theory frequently takes as starting point the concept of *agenda*, a consistent set of propositions, closed under negation, on which judgements have to be made (List and Puppe 2009; Dietrich 2007; List and Pettit 2002). Moreover the propositions may be linked by logical restrictions. In our case, the agenda is  $\{P, \neg P, Q, \neg Q, P \wedge Q, \neg(P \wedge Q)\}$ . In this language, the doctrinal paradox can be stated by saying that the majority rule can be inconsistent, in the sense that if all pairs of formulae in the agenda are decided by a majority rule, then the accepted formulae could be logically inconsistent.

The aggregation problem is described in full generality in Nehring and Pivato (2011), starting with the concept of *judgement*, defined as a mapping from the set of propositions to the doubleton {True, False}, and that of *feasible judgement*, a mapping that respects the underlying logical constraints of the propositions. The *judgement aggregation problem* is then defined as to find a “best” feasible mapping from the voters individual judgements. If the mapping is built by propositionwise majority, a non-feasible mapping may arise. The range of possible *voting paradoxes* is the set of possible non-feasible mappings. In our case, the only such mapping, assuming the voters respect the underlying logic, is  $P \mapsto \text{True}$ ,  $Q \mapsto \text{True}$ ,  $(P \wedge Q) \mapsto \text{False}$ .

The truth-functional judgement aggregation problem is the special case when there is one or more propositions called *conclusions*, that are functionally determined by the values of other propositions, called *premises*. This functional dependence is not necessarily of conjunctive type; more complicated relations between them can be in force. In this paper we address the simplest non-trivial problem, in which the conjunction of two premises are equivalent to the conclusion.

We review some more literature in the discussion section when presenting possible extensions of the present work.

### 1.3 Organisation of the paper

The paper is organised as follows: In Sect. 2, we define and characterise with precision the *Prem*, *Conc* and *Path* decision rules, and explain what we consider to be an *admissible rule* in the application context we are dealing with. We show that the three rules considered are admissible, and that there exist non-admissible (though not completely irrational) decision rules.

The specific model assumptions are given in Sect. 3. Although the doctrinal paradox cannot be avoided, one can speak of the “best rule”, once some theoretical model is defined and some reasonable performance criterion is chosen. Of course, different criteria gives rise to different “best rules”, and this is again unavoidable.

The concept of true and false positives and negatives and that of ROC space are introduced in Sect. 4. Translated to our setting, the false positive rate FPR will be the probability of accepting  $P \wedge Q$  when it is false, and the false negative rate FNR the probability of rejecting  $P \wedge Q$  when it is true.

Section 5 contains the main results of the paper and their proofs: Rule *Prem* is the best in the ROC setting under a symmetric criterion which gives the same weight to FPR and FNR; in case of unequal weights, any one of the three rules can be the best, depending on the relation between the competence parameter and the specific weights.

Section 6 contains some numerical computations and figures, showing that all values of interest resulting from the probabilistic model can be explicitly obtained. More than that, the simple hypotheses on the model that we impose in Sect. 3 can be relaxed to a great extent and the explicit computations can still be carried out without difficulty with adequate computing resources. This is explained in more detail in the final discussion in Sect. 7, together with other considerations and open problems.

To make the exposition smooth, we postpone most of the technical statements and their proofs to an appendix.

## 2 Decision rules

In this section, we give a detailed characterization of the *Prem*, *Path* and *Conc* rules outlined in the introduction, and formalise the concept of admissible decision rule. We assume throughout the paper that the committee size is an odd number  $n = 2m + 1$ , with  $m \geq 1$ . The simple majority for a single binary question is therefore achieved by any number of committee members greater than  $m$ .

**Definition 2.1** Assume that the opinions of the committee are summarised as in Table 1. Then, we define the following decision rules:

$R_1$  : The *premise-based* rule (Prem),

$$\text{Decide } P \wedge Q \text{ if and only if } x + y > z + t \text{ and } x + z > y + t. \quad (1)$$

$R_2$  : The *path-based* rule (Path),

$$\text{Decide } P \wedge Q \text{ if and only if } x > z + t \text{ and } x > y + t. \quad (2)$$

$R_3$  : The *conclusion-based* rule (Conc),

$$\text{Decide } P \wedge Q \text{ if and only if } x > y + z + t. \quad (3)$$

In the sequel, we shall use the following equivalent expressions, whose proof is straightforward and detailed in the Appendix (Proposition A.1):

$$R_1 : \text{Decide } P \wedge Q \text{ if and only if } x > m - y \wedge z$$

$$R_2 : \text{Decide } P \wedge Q \text{ if and only if } x > m - \left\lfloor \frac{y \wedge z}{2} \right\rfloor$$

$$R_3 : \text{Decide } P \wedge Q \text{ if and only if } x > m$$

where  $\lfloor x \rfloor$  denotes the *integer part* of  $x$ , i.e. the largest integer not greater than  $x$ , and  $x \wedge y$  stands for the minimum of  $x$  and  $y$ . (The context will distinguish the uses of  $\wedge$  as the minimum of two values or the logical operator ‘and’.)

From the characterisation of Proposition A.1, it is clear that the condition of rule  $R_3$  to choose  $P \wedge Q$  is more restrictive than that of  $R_2$ , and the latter in turn is more restrictive than the condition of  $R_1$ . Furthermore, rules  $R_2$  and  $R_3$  are equivalent when  $n = 3$  or  $5$ , and they are different for  $n \geq 7$ . Rules  $R_1$  and  $R_2$  are not equivalent for any  $n \geq 3$ . These facts will be stated as a proposition after a formal definition of decision rule:

**Definition 2.2** A *decision rule* is a mapping from the set  $\mathbb{T}$  of all voting tables into  $\{0, 1\}$ , where 1 means deciding  $P \wedge Q$ , and 0 means the opposite.

If the committee has  $n$  members, there are  $N = \binom{n+3}{3}$  ways to fill the voting table, and  $2^N$  possible decision rules. The number  $N$  can be deduced by a combinatorial argument considering the number of ways to express  $n$  as the sum of four integers, including zero (the so-called *weak compositions* of a number).

Since we assume that  $P$  and  $Q$  must have the same relevance in the final decision, it is natural to impose that a decision rule must yield the same result if we interchange the number of votes for  $P \wedge \neg Q$  and  $Q \wedge \neg P$ , i.e.  $y$  and  $z$ .

Furthermore, we would like to consider only decision rules satisfying the following rationality property: If a table leads to decision 1 and a committee member that has voted for  $P \wedge \neg Q$  or  $Q \wedge \neg P$  changes the vote to  $P \wedge Q$ , the decision for the new table should also be 1; analogously, if the decision was 0 and the same vote changes to  $\neg P \wedge \neg Q$ , then the decision for the new table should also be 0. This condition is easily implemented by considering only rules that preserve the partial order  $\leq$  on  $\mathbb{T}$  generated by the four relations (using the matrix notation of Sect. 1.1)

$$\begin{aligned} \begin{bmatrix} x & y \\ z & t \end{bmatrix} \leq \begin{bmatrix} x+1 & y-1 \\ z & t \end{bmatrix} & \quad \begin{bmatrix} x & y \\ z & t \end{bmatrix} \leq \begin{bmatrix} x+1 & y \\ z-1 & t \end{bmatrix} \\ \begin{bmatrix} x & y \\ z & t \end{bmatrix} \leq \begin{bmatrix} x & y+1 \\ z & t-1 \end{bmatrix} & \quad \begin{bmatrix} x & y \\ z & t \end{bmatrix} \leq \begin{bmatrix} x & y \\ z+1 & t-1 \end{bmatrix} \end{aligned}$$

These considerations lead us to define the following concept of admissible rule. More often than not, we will use the vector notation  $(x, y, z, t)$  instead of the tabular form, to save space.

**Definition 2.3** A decision rule  $R : \mathbb{T} \rightarrow \{0, 1\}$  will be called an *admissible rule* if:

1. It does not distinguish between transposed tables:

$$R(x, y, z, t) = R(x, z, y, t).$$

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

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

The resulting partial order for  $n = 3$  is represented in Fig. 1, where we have already identified tables that merely interchange the values of  $y$  and  $z$ .

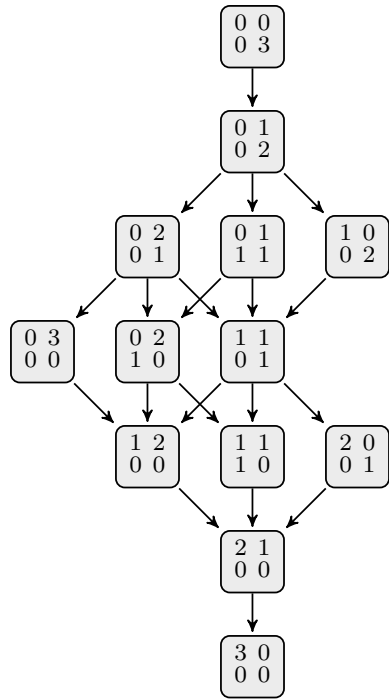
We will write  $R \leq R'$  whenever  $R(T) \leq R'(T)$  for all tables  $T \in \mathbb{T}$ , and  $R < R'$  whenever  $R \leq R'$  and  $R \neq R'$ . Rules  $R_1, R_2, R_3$  of Definition 2.1 are admissible and satisfy  $R_3 \leq R_2 \leq R_1$  (see Appendix, Proposition A.2).

As an example of a non-admissible rule, consider requiring that  $P \wedge Q$  gets more votes than each one of the other options:

$$R_0(x, y, z, t) = 1 \iff x > y \text{ and } x > z \text{ and } x > t.$$

Indeed, with  $n \geq 5$ , one has  $(2, 1, 1, 1) < (2, 2, 1, 0)$ , but applying  $R_0$  to both sides reverses the inequality. This contradicts the second condition of Definition 2.3.

**Fig. 1** The partially ordered set  $(\mathbb{T}, \leq)$  for  $n = 3$ . We have identified tables which are transposed of each other



The relation of  $R_0$  with the other three rules can be summarised as follows: In general,  $R_3 \leq R_2 \leq R_0$ . For  $n = 3$ , the relations  $R_3 = R_2 = R_0 < R_1$  hold true, with  $R_0$  and  $R_1$  differing on  $(1, 1, 1, 0)$ . For  $n = 5$ , we have  $R_3 = R_2 < R_0 < R_1$  with  $R_2$  and  $R_0$  differing on  $(2, 1, 1, 1)$ , and  $R_0$  and  $R_1$  differing on  $(1, 2, 2, 0)$  and on  $(2, 2, 1, 0)$ . Starting with  $n \geq 7$ ,  $R_0$  and  $R_1$  are no more comparable (neither  $R_0 \leq R_1$  nor  $R_1 \leq R_0$ ), and all four rules are different. All these relations are easily checked.

### 3 Probabilistic model of committee voting

Our probabilistic model is the simplest possible, and it is described by the four conditions below. Under these conditions, we can develop the theory of ROC optimality without unnecessary complications, and produce comprehensible examples. As we will discuss in Sect. 7, these conditions can be very much relaxed and the computations of the ROC analysis can be carried out automatically without problems.

A framework similar to ours can be found in List (2005), where the main goal is to compute the probability of appearance of the paradox, and to investigate the behaviour of this probability when the committee size grows to infinity, in the spirit of the classical Condorcet theorem.

We assume that a true “state of nature” exists, in which one of the four exclusive events  $P \wedge Q$ ,  $P \wedge \neg Q$ ,  $\neg P \wedge Q$  and  $\neg P \wedge \neg Q$  is in force.

We assume the following conditions:



- (C1) *Odd committee size*: The number of voters is an odd number,  $n = 2m + 1$ , with  $m \geq 1$ .
- (C2) *Equal competence*: The probability  $\theta$  of choosing the correct alternative when deciding between  $P$  and  $\neg P$  is the same for all voters and satisfies  $\frac{1}{2} < \theta < 1$ . The same competence  $\theta$  is assumed when deciding between  $Q$  and  $\neg Q$ .
- (C3) *Mutual independence among voters*: The decision of each voter does not depend on the decisions of the other voters.
- (C4) *Independence between  $P$  and  $Q$* : For each voter, the decision on one premise does not influence the decision on the other.

Formally, conditions (C2)–(C4) can be rephrased by saying that for each voter  $k$  in the committee and each clause  $c \in \{P, Q\}$ , there is a random variable that takes the value 1 if the voter believes the clause is true, and zero otherwise, and all these variables are stochastically independent and identically distributed. Their specific distribution depends on the true state of nature.

Under these hypotheses, we can obtain the probability of all possible distribution of votes in a table, for each given state of nature. As it is customary in probability and statistics, we distinguish between random variables represented by capital letters  $X, Y$ , etc, and their observed values, represented by small letters  $x, y$ , etc. If  $(X, Y, Z, T)$  is the random vector representing the counts in Table 1 in the probabilistic framework just defined, its probability law is multinomial (see Appendix, Proposition A.4). From the law of  $(X, Y, Z, T)$ , it is easy to compute the law of any given decision rule  $R : \mathbb{T} \rightarrow \{0, 1\}$ .

Notice that the multinomial law holds irrespective of the existence of a background absolute truth or of the competence concept. It only needs independence between voters, and the existence of a vector of probabilities  $(p_x, p_y, p_z, p_t)$  adding up to 1, the same for all voters, representing the probability of opting for each of the four options. List (2005) studies the probability of appearance of the doctrinal paradox in this more general situation and shows that slightly different values of the vector of probabilities may lead to very different values of the probability of appearance of the paradox when  $n \rightarrow \infty$ . Applied to our case, his results imply that, if  $P \wedge Q$  is true, the probability of appearance of the paradox (disagreement between premise-based and conclusion-based rules) tends to 0 when the competence  $\theta$  is greater than  $\sqrt{0.5}$ , and tends to 1 when it is lower; if  $P \wedge Q$  is not true, then it always tends to 0. Interestingly, he also computes the expectation of appearance of the paradox when the vector of probabilities is assumed to follow a non-informative uniform prior on the simplex.

#### 4 True and false rates and ROC analysis

*Receiver operating characteristics* (ROC) plots were introduced to visualize and compare binary classifiers in signal detection (see, e.g. Egan 1975) and its use extends to medical tests, machine learning and other disciplines where binary decisions have to be taken under uncertainty (see Fawcett 2006 for an introductory

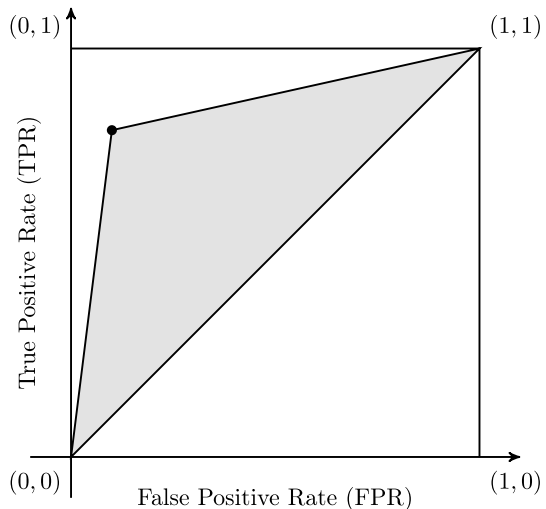
presentation of ROC plots). The term *classifier* is also used as a synonym of *decision rule*.

In signal detection theory, propositions are related to the emission/reception of a binary digit. Denote by  $\hat{\mathbf{0}}$  and  $\hat{\mathbf{1}}$  the bit received and by  $\mathbf{0}$  and  $\mathbf{1}$  the bit actually sent. The *true positive rate* (TPR) is defined as the probability of receiving  $\hat{\mathbf{1}}$  when  $\mathbf{1}$  is the true bit emitted, and the *true negative rate* (TNR) as the probability of receiving  $\hat{\mathbf{0}}$  when  $\mathbf{0}$  is the bit sent. Analogously, the *false positive rate* (FPR) and the *false negative rate* (FNR) are, respectively, the probabilities of receiving  $\hat{\mathbf{1}}$  when  $\mathbf{0}$  is the true digit, and of receiving  $\hat{\mathbf{0}}$  when  $\mathbf{1}$  is the true digit. From these definitions, it is clear that a decision rule such that  $\text{TPR} \approx 1$  and  $\text{FPR} \approx 0$  has a “good performance”.

In classical statistics, decision rules appear in the context of *hypothesis testing*, where the TPR is the *power* of the test, the greater the better, under the restriction that the FPR (called the *type-I error*) does not exceed a fixed small value (the *significance level*). The *type-II error* corresponds to the FNR. The two types of errors are thus treated in a non-symmetric way. In medicine, the TPR and the TNR are respectively called *sensitivity* and *specificity*.

In the ROC graph, several classifiers can be compared on the basis of the pair (FPR, TPR) represented in the unit square  $[0, 1] \times [0, 1]$ , the so-called *ROC space* (see Fig. 2). Usually, the rates are estimated from sample data. “Good” decision rules are expected to correspond to points close to the upper left corner (0, 1) of the unit square. Different measures of the proximity to that corner can be considered. The most widely used is the area of the shaded triangle in Fig. 2, defined by the points (0, 0), (1, 1) and (FPR, TPR). The closer the area to 0.5, the better the classifier is considered. Points on the diagonal of the square correspond to completely random classifiers, for which the probability of true and false positives are equal. Points below the diagonal line represent classifiers that perform worse than random. Following Fawcett (2006), the classifiers plotted near the (0, 0) corner can be said to be “conservative”, because they make few positive classifications (true or false).

**Fig. 2** The black dot represents a classifier in the ROC space, represented by its coordinates (FPR, TPR), and the shaded area is AOT



For the same reason, classifiers plotted near to the (1, 1) corner are sometimes called “liberal” because they tend to have a higher number of false positives.

It is immediate from Fig. 2 that the *area of the triangle*, that will be denoted by AOT, can be expressed in terms of the rates as follows:

$$\text{AOT} := \frac{1}{2}(\text{TPR} - \text{FPR}) \tag{4}$$

$$= \frac{1}{2} - \left( \frac{\text{FPR}}{2} + \frac{\text{FNR}}{2} \right). \tag{5}$$

In the definition of AOT, the roles of the rates FPR and FNR are symmetric. In some situations, it may be desirable to assign different weights to these errors. This leads to the concept of *weighted area* of the triangle, WAOT. Indeed, fixing a weight value  $w \in (0, 1)$ , one can define, by analogy with formula (5),

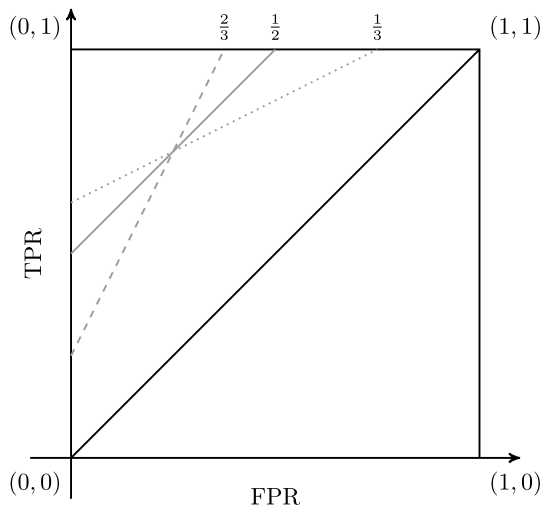
$$\text{WAOT}_w := \frac{1}{2} - (w \cdot \text{FPR} + (1 - w) \cdot \text{FNR}). \tag{6}$$

For any  $w \in (0, 1)$ ,  $\text{WAOT}_w$  takes values in  $[-\frac{1}{2}, \frac{1}{2}]$ , negative “areas” corresponding to points below the diagonal. If  $w > \frac{1}{2}$ , the weighted area  $\text{WAOT}_w$  penalizes false positives more than false negatives; and if  $w < \frac{1}{2}$ , it is the other way round. The points of the ROC space yielding the same value of WAOT are straight lines, with slope equal to  $w/(1 - w)$ , see Figure 3.

Unequal weights are useful in some practical situations: For instance, in court of justice cases, it is common that false positives (declaring guilty an innocent defendant) are considered worst than false negatives; in medical tests, the two errors often play an obvious asymmetric role too.

In some applications, the rates TPR, FPR of a given classifier can be estimated on the basis of a “test sample” in which the actual states of nature are known (0

**Fig. 3** Level lines for the value  $\text{WAOT}_w = 0.25$  with weights  $w = \frac{2}{3}, \frac{1}{2}, \frac{1}{3}$



or  $\mathbf{1}$  in each observation) and the outputs of the classifier ( $\hat{\mathbf{0}}$  or  $\hat{\mathbf{1}}$ ) are compared against the actual states. The results are often summarised in a table known as *confusion matrix*. In social applications, as is the case of court cases, the actual states are supposed to be unknown and there might not be test samples available. However, the rates FPR and TPR can be defined and computed exactly under our model assumptions as it is proved in the Appendix (Propositions A.5, A.6 and A.7). First, we translate the ROC analysis vocabulary to our probabilistic framework.

**Definition 4.1** Assume the model conditions (C1)–(C4) of Sect. 3. The true positive rate associated to a decision rule  $R$  is defined as the probability to “decide  $P \wedge Q$ ” under the state of the nature “ $P \wedge Q$  true”, and it depends on  $n$ ,  $\theta$  and the decision rule:

$$\text{TPR}(n, \theta, R) := \mathbb{P}_{P \wedge Q}\{R(X, Y, Z, T) = 1\}, \quad (7)$$

where  $\mathbb{P}_{P \wedge Q}$  denotes the multinomial law of Proposition A.4, with parameters corresponding to the state of nature “ $P \wedge Q$  true”.

Under the model conditions, the true positive rates (7) for each rule  $R_1$ ,  $R_2$  and  $R_3$  can be expressed in terms of the multinomial probabilities (14). The explicit expressions are stated in the Appendix (Proposition A.5).

Notice that the inequalities  $R_3 \leq R_2 \leq R_1$  induce the corresponding inequalities among the true positive rates:

$$\text{TPR}(n, \theta, R_3) \leq \text{TPR}(n, \theta, R_2) \leq \text{TPR}(n, \theta, R_1).$$

False positives can arise under the three different states of nature contained in the negation  $\neg(P \wedge Q)$ . To define the false positive rate FPR, we adopt the conservative approach, taking the maximum of the probabilities of accepting  $P \wedge Q$  under each of the states. As shown in the Appendix, Proposition A.6, this maximum always corresponds to the case when one of the clauses  $P$  or  $Q$  is true and the other one is false. This is intuitive noticing that the state  $\neg P \wedge \neg Q$  is “the less likely one” to choose  $P \wedge Q$ .

**Definition 4.2** Let  $R$  be any one of the rules  $R_1$ ,  $R_2$  or  $R_3$ . We define the false positive rate as:

$$\text{FPR}(n, \theta, R) := \mathbb{P}_{P \wedge \neg Q}\{R(X, Y, Z, T) = 1\}. \quad (8)$$

By Proposition A.6 again, one can write  $\neg P \wedge Q$  instead of  $P \wedge \neg Q$  in this definition. Furthermore, defining FPR as the largest of the different probabilities of accepting  $P \wedge Q$  when it is false, we are placing ourselves in the most unfavourable position and thus FPR will control the maximum risk. This is in accordance with classical statistics practice, and the sensible choice in the absence of any a priori knowledge on the state of nature. In the discussion section we comment on the relation between this setting and the alternative Bayesian approach.

The computation of (8) for rules  $R_1, R_2, R_3$  are done in the Appendix, Proposition A.7, and we have the ordering

$$FPR(n, \theta, R_3) \leq FPR(n, \theta, R_2) \leq FPR(n, \theta, R_1) ,$$

as with the positive rates.

We now define formally the criteria under which the decision rules will be compared.

**Definition 4.3** Let  $R$  be any one of the rules  $R_1, R_2$  or  $R_3$ . We define the area of the triangle as:

$$AOT(n, \theta, R) := \frac{1}{2} ( TPR(n, \theta, R) - FPR(n, \theta, R) ) . \tag{9}$$

Fix  $w \in (0, 1)$ . We define the weighted area of the triangle as:

$$WAOT_w(n, \theta, R) := \frac{1}{2} - ( w FPR(n, \theta, R) + (1 - w)FNR(n, \theta, R) ) .$$

### 5 Main results

In this section we will use the concepts from ROC analysis introduced in Sect. 4 as a numeric criterion to compare the relative goodness of decision rules. Theorem 5.2 establishes the preference order of the three rules considered, under the criterion of greater area of the triangle, where it is seen that  $R_1$  is uniformly the best. This is still true when the false negatives are more penalised than the false positives (Corollary 5.3). If false positives are deemed worse, the situation is more complex and interesting; it will be covered by Theorem 5.4. All proofs are in the Appendix.

**Definition 5.1** A rule  $R$  is *AOT-better* than a rule  $R'$  if and only if, for all  $n$  odd and  $\theta > \frac{1}{2}$ ,

$$AOT(n, \theta, R) \geq AOT(n, \theta, R') , \tag{10}$$

and the inequality is strict for some value of  $n$  or  $\theta$ .

Under our model assumptions, it is now shown that rule  $R_1$  is AOT-better than  $R_2$  and that  $R_2$  is AOT-better than  $R_3$ :

**Theorem 5.2** *Under the model conditions (C1)–(C4), for all  $n \geq 3$  odd and for all  $\theta > \frac{1}{2}$ ,*

$$AOT(n, \theta, R_3) \leq AOT(n, \theta, R_2) < AOT(n, \theta, R_1)$$

*and the first inequality is strict for  $n \geq 7$ .*

For the weighted area of the triangle defined by formula (6), and weights  $w < \frac{1}{2}$  (that means, when false negatives are considered more harmful than false positives), the relations between  $R_1$ ,  $R_2$  and  $R_3$  are the same as with AOT (case  $w = \frac{1}{2}$ ), as stated in the next Corollary 5.3. However, for  $w > \frac{1}{2}$ , none of the rules gives a greater WAOT than another, uniformly in  $n \geq 3$  and  $\frac{1}{2} < \theta < 1$ ; this will be precisely stated in Theorem 5.4, Lemma A.8, and the numerical examples of Section 6.

**Corollary 5.3** *Under the model conditions (C1)–(C4), for all  $n \geq 3$  odd, and for all  $\theta > \frac{1}{2}$  and  $w < \frac{1}{2}$ ,*

$$\text{WAOT}_w(n, \theta, R_3) \leq \text{WAOT}_w(n, \theta, R_2) < \text{WAOT}_w(n, \theta, R_1),$$

and the first inequality is strict for  $n \geq 7$ .

The case  $w > \frac{1}{2}$  is different. The relation between the WAOT of  $R_1$  and  $R_2$  is still the same of the AOT if the competence  $\theta$  stands above a certain threshold  $C(w)$ , with  $\frac{1}{2} < C(w) < w$ , and similarly with  $R_2$  and  $R_3$ . But not necessarily for  $\theta$  below that threshold. This is made more precise in the next theorem.

**Theorem 5.4** *Fix  $n \geq 3$ . For every weight  $\frac{1}{2} < w < 1$ , there exists  $C_1(w)$ , smaller than  $w$  (except that  $C_1(w) = w$  if  $n = 3$ ), such that*

$$\theta > C_1(w) \Rightarrow \text{WAOT}_w(n, \theta, R_1) > \text{WAOT}_w(n, \theta, R_2).$$

*Fix  $n \geq 7$ . For every weight  $\frac{1}{2} < w < 1$ , there exists  $C_2(w)$ , smaller than  $w$ , such that*

$$\theta > C_2(w) \Rightarrow \text{WAOT}_w(n, \theta, R_2) > \text{WAOT}_w(n, \theta, R_3).$$

## 6 Examples

In this section we illustrate the above theory with some numeric computations and figures.

It is clear from the previous sections that none of the rules considered is best for all pairs  $(\theta, w)$  of competence and weight. In fact, no two rules  $R_i$  and  $R_j$  are comparable, uniformly in  $\theta$  and  $w$  (except for  $R_2$  and  $R_3$  when  $n \leq 5$ , because they coincide). Indeed, Table 2 shows the different possible orders under the WAOT criterion for three fixed competence values and for varying  $w \in (0, 1)$ , and committee size  $n = 11$ .

Notice in Table 2 that in cases where  $w$  should be close to 1 (for instance in criminal cases) rule  $R_3$  might be better than the others, specially for low competence levels. This can be also seen in Fig. 4b.

In Table 3, again with committee size  $n = 11$ , the values of TPR, FPR, and AOT are computed to four decimal places for a large range of competence values, using (15–17) and (19–21). The last column is the value of WAOT for a fixed weight  $w = 0.75$ ; in other words, false positives penalises the performance measure three

**Table 2** For each  $\theta$ , intervals of weights  $w$  where each order of rules holds true. The committee size is  $n = 11$ . The symbol  $<$  in the first column means here the relation “worse than” with respect to the WAOT criterion.

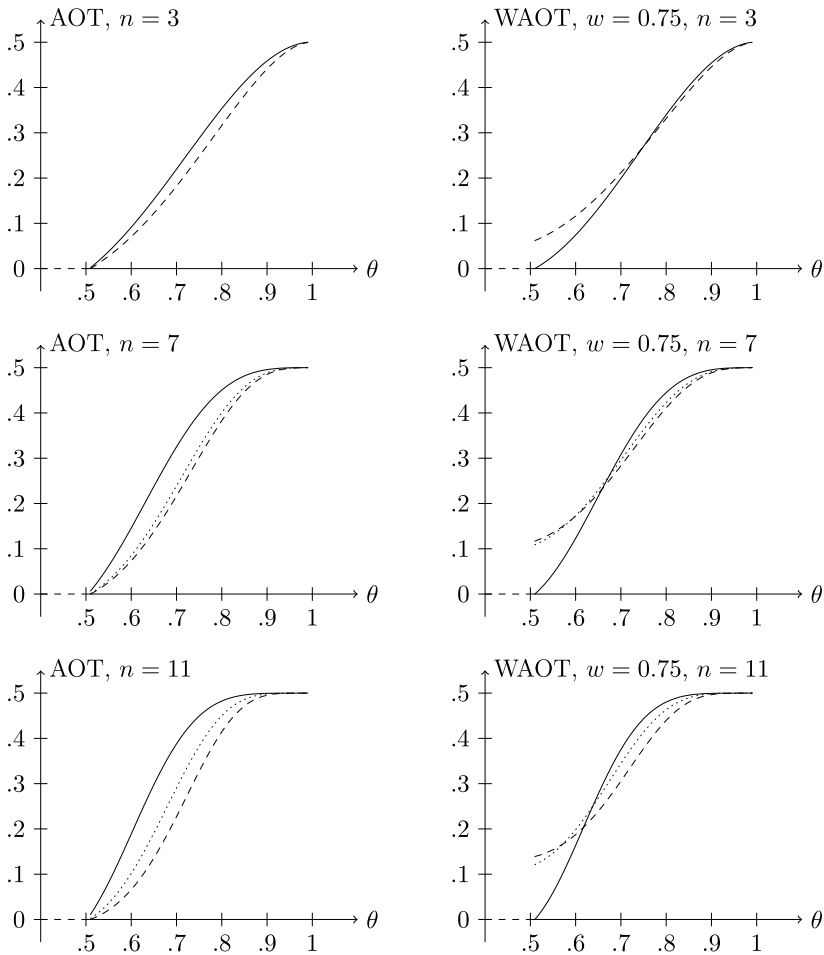
	$\theta = 0.6$	$\theta = 0.75$	$\theta = 0.90$
$R_3 < R_2 < R_1$	$0 < w < 0.6930$	$0 < w < 0.8722$	$0 < w < 0.9576$
$R_3 < R_1 < R_2$	$0.6930 < w < 0.7184$	$0.8722 < w < 0.9154$	$0.9576 < w < 0.9847$
$R_1 < R_3 < R_2$	$0.7184 < w < 0.8215$	$0.9154 < w < 0.9820$	$0.9847 < w < 0.9995$
$R_1 < R_2 < R_3$	$0.8215 < w < 1$	$0.9820 < w < 1$	$0.9995 < w < 1$

**Table 3** Comparison of decision rules  $R_1, R_2$  and  $R_3$  in the ROC map for a fixed number of voters  $n = 11, w = 0.75$  in the column WAOT, and different competence levels  $\theta$ . Notice that  $R_1$  has the largest values of AOT for any fixed  $\theta$ , that is, the sequence of ordered rules from best to worst is always 1, 2, 3

$n$	$\theta$	rule	TPR	FPR	AOT	WAOT
11	0.55	1	0.4008	0.2323	0.0843	0.1760
11	0.55	2	0.1372	0.0589	0.0391	0.2401
11	0.55	3	0.0811	0.0327	0.0242	0.2457
11	0.60	1	0.5678	0.1857	0.1910	0.2526
11	0.60	2	0.2569	0.0480	0.1044	0.2782
11	0.60	3	0.1661	0.0283	0.0689	0.2703
11	0.65	1	0.7247	0.1266	0.2991	0.3363
11	0.65	2	0.4197	0.0340	0.1928	0.3294
11	0.65	3	0.2984	0.0219	0.1382	0.3081
11	0.70	1	0.8497	0.0721	0.3888	0.4083
11	0.70	2	0.6050	0.0207	0.2922	0.3857
11	0.70	3	0.4729	0.0148	0.2291	0.3571
11	0.75	1	0.9325	0.0331	0.4497	0.4583
11	0.75	2	0.7779	0.0105	0.3837	0.4366
11	0.75	3	0.6649	0.0084	0.3282	0.4099
11	0.80	1	0.9768	0.0115	0.4827	0.4856
11	0.80	2	0.9051	0.0042	0.4505	0.4731
11	0.80	3	0.8339	0.0037	0.4151	0.4557
11	0.85	1	0.9947	0.0026	0.4960	0.4967
11	0.85	2	0.9735	0.0012	0.4862	0.4925
11	0.85	3	0.9446	0.0011	0.4717	0.4853
11	0.90	1	0.9994	0.0003	0.4996	0.4996
11	0.90	2	0.9965	0.0002	0.4982	0.4990
11	0.90	3	0.9910	0.0002	0.4954	0.4976
11	0.95	1	1.0000	0.0000	0.5000	0.5000
11	0.95	2	0.9999	0.0000	0.5000	0.5000
11	0.95	3	0.9997	0.0000	0.4999	0.4999

times more than false negatives. In column AOT, both errors penalise in the same proportion.

Under the AOT criterion,  $R_1$  (*Prem*) is always better than  $R_2$  (*Path*), and  $R_2$  is better than  $R_3$  (*Conc*), as Theorem 5.2 claims. The numbers in the table give an



**(a)** AOT. Rule  $R_1$  is uniformly best in  $\theta$ . **(b)** WAOT. For any  $w > 1/2$ , there is no a uniformly best or worst rule.

**Fig. 4** AOT and WAOT for each of the rules  $R_1$  (solid),  $R_2$  (dotted),  $R_3$  (dashed) as a function of  $\theta$ , for several committee sizes  $n$

idea of the extent of the difference, suggesting that  $R_2$  and  $R_3$  are closer together than  $R_1$  and  $R_2$ . We also see that for very high values of  $\theta$  all rules get closer and approach fast to the perfect value 0.5.

For the WAOT criterion, with  $w = 0.75$ , and low competence values of the jury, we see that it is better to use rules  $R_2$  or  $R_3$ . At some point, between  $\theta = 0.60$  and  $\theta = 0.65$ , the order of AOT is re-established and preserved till the end of the table. Of course the exact value can be computed, and turns out to be 0.6374 (to four decimal places).



A simple illustration of the evolution of the AOT for the three rules we are considering, for several committee sizes, can be seen in Fig. 4a. For  $n = 3, 7, 11$  the AOT value for the three rules is drawn against  $\theta$ . Notice that the largest absolute differences in AOT take place around the middle values of the competence range. That means, for  $0.6 \lesssim \theta \lesssim 0.8$ , say, it is when the selection of the decision rule is most critical.

The analogous Fig. 4b shows the same curves, in the case  $w = 0.75$ . Rule  $R_1$  is the worst in the lower end of  $\theta$  values, and the best in the upper end. Rule  $R_3$  does the opposite.

Combining the committee sizes  $n = 3, 7, 11$  and the competence values  $\theta = 0.60, 0.75, 0.90$ , in Fig. 5 we draw the triangles in ROC space of the three decision rules. In this picture, it can be observed that the area of the triangle determined

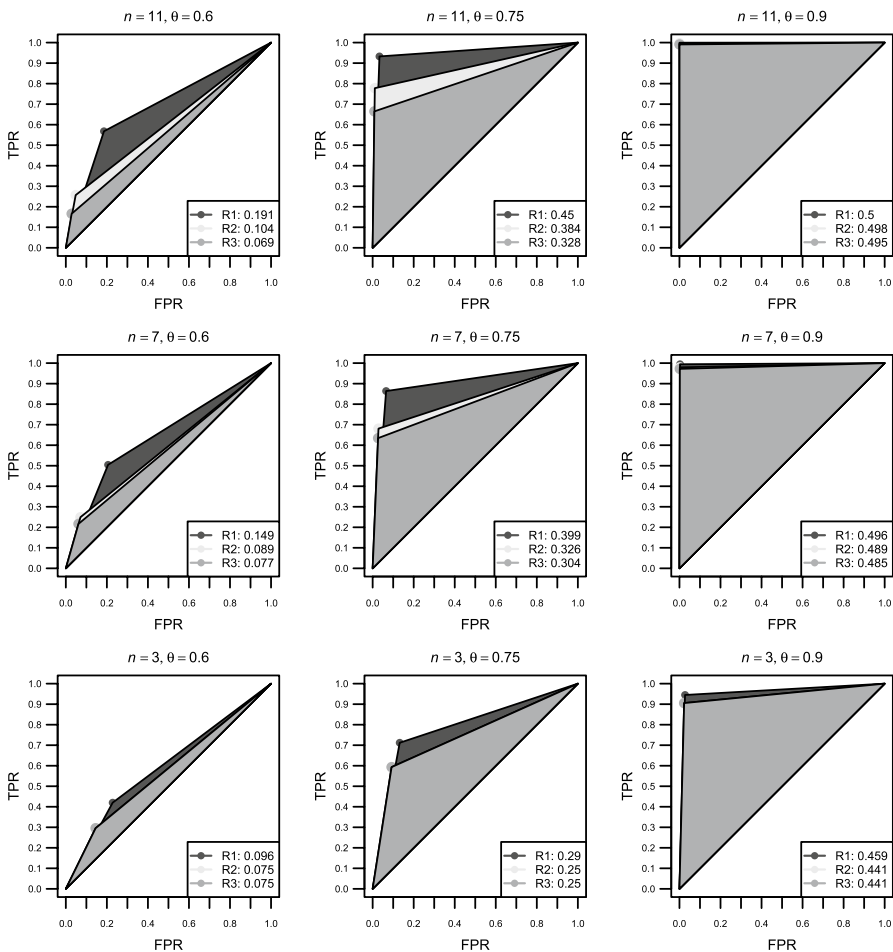
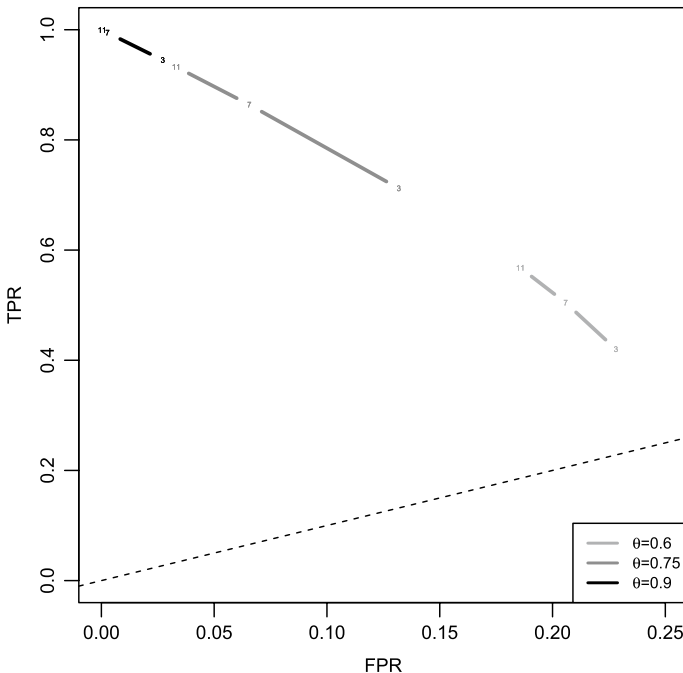


Fig. 5 ROC representation of  $R_1$ ,  $R_2$  and  $R_3$ , for several committee sizes  $n$  and different competences  $\theta$ . Values in the legend are the AOT

by rule  $R_1$  is larger than the area determined by  $R_2$ , which in its turn is larger than the area determined by rule  $R_3$  for  $n > 5$ , and that the triangles of  $R_2$  and  $R_3$  coincide for  $n = 3$  and  $n = 5$ .

The ROC analysis helps in comparing several decision rules, but it is also useful to visualize the performance of a given rule depending on the parameters. For example, taking three values for the competence,  $\theta = 0.60, 0.75, 0.90$ , and three values for the committee size  $n = 3, 7, 11$ , the ROC representation in Fig. 6 displays a curve going from near the diagonal when  $\theta = 0.60$  to near the corner  $(0, 1)$ , when  $\theta = 0.90$ . From the figure it is apparent that the quality of the voters in terms of the competence is definitely much more important than its quantity. (Karatkin and Paroush 2003 arrive to the same conclusion in the single-premise case.)

It is also natural to ask, for a given rule and competence value  $\theta$ , what is the minimum number of voters  $n$  that ensures that the TPR, TNR and AOT ( $w = 0.5$ ) reach a certain given threshold  $k$ . This minimum is easy to find out. For example, Table 4 give the numbers, for decision rule  $R_1$  (*Prem*), a range of  $\theta$  values, and quite demanding thresholds. Notice that to ensure a certain threshold of TPR the size of the committee must be greater than to ensure the same threshold for TNR. This is



**Fig. 6** ROC representation of  $R_1$  for different numbers of voters  $n = 3, 7, 11$  (the number is used as the location in ROC space) and several competences  $\theta = 0.60, 0.75, 0.90$  in different shades of grey. Notice that the range in the horizontal axis has been rescaled and the dashed line represents the diagonal of the unit square. The closer to the corner  $(0, 1)$ , the lower the risk of erroneous classification (high TPR and low FPR). Therefore, we see a good behaviour with small but highly competent committees; but if the committee quality is low, its size has to be drastically increased to reduce the risk

**Table 4** For each indicator, TPR, TNR and AOT, minimum committee size needed to reach the threshold  $k$ , for rule  $R_1$  and different competence values  $\theta$

	$\theta$							
	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
For TPR, $k=0.95$	95	41	23	13	9	7	5	3
For TNR, $k=0.95$	65	29	15	9	7	5	3	3
For AOT, $k=0.45$	81	35	19	13	7	5	3	3

because for any fixed size  $n$ , the probability that rule  $R_1$  produce a true negative is greater than the probability of producing a true positive.

### 7 Conclusions and discussion

In this paper, we have defined a theoretical framework, based on a probabilistic model, that describes the behaviour of a committee confronted with a compound yes-no question. The application of the Receiver Operating Characteristics (ROC) space, a concept originating in signal processing, and adopted in several other fields, seems to be new in judgement aggregation research. It allows an objective assessment of the quality of a group decision on a complex issue, based on the (possibly subjective) competence of the members of the group. It also allows to compare different decision rules, both for symmetric or asymmetric penalising weights on the false positives and the false negatives.

The main results deal with to the comparison of the particular rules  $R_1, R_2$  and  $R_3$  defined in Sect. 2 in terms of the quantities AOT and  $WAOT_w$  in the ROC space introduced in Sect. 4. AOT is a particular case of  $WAOT_w$  when false positives and false negatives are equally weighted ( $w = \frac{1}{2}$ ). Putting together Theorem 5.2 and Corollary 5.3, we have shown that rule  $R_1$  is better than rule  $R_2$ , and rule  $R_2$  is in its turn strictly better than rule  $R_3$ , for all competence values  $\theta > \frac{1}{2}$ , if the weight  $w$  on false positives is less or equal than  $\frac{1}{2}$ . Rule  $R_1$  (premise-based) has been already considered superior than  $R_3$  (conclusion-based) according to other criteria (for example, by the deliberative democracy doctrine, see e.g. Dietrich and List 2007; List 2006).

Furthermore, Lemma A.8 establishes that  $R_1$  is still better than  $R_2$  and  $R_2$  better than  $R_3$  for some values of the weight  $w$  greater than the competence, but less than another quantity  $D(\theta)$  which depends only on the competence  $\theta$ . On the other hand, for  $w$  beyond  $D(\theta)$ , any one of the three rules can be the best. Notice that Theorem 5.4 states these facts in a more natural way: once fixed the relative importance of the two errors FPR and FNR, the competence  $\theta \in (\frac{1}{2}, 1)$  of the committee determines the relative goodness of the three rules. The numerical experiments of Sect. 6 show that there are several possible goodness orders of the rules and, in particular, both premise-based and conclusion-based can be the best and the worst of the three.

The simplicity of the model has allowed us to focus on the methodology of ROC space, but the assumptions can be easily weakened in several ways and the computations can be adapted without much difficulty. For example:

- *Different voters' competence*  $\theta$ . If competences are different, the law of  $(X, Y, Z, T)$  described in Proposition A.4 is no longer multinomial. The vote of each committee member  $k = 1, \dots, n$  is a random vector  $J_k$  equal to one of the four possible vote schemes with certain probabilities  $(p_x^k, p_y^k, p_z^k, p_t^k)$ , independently. If  $\theta_k$  is the competence of member  $k$ , then, under the true state  $P \wedge Q$ , this vector of probabilities is

$$(\theta_k^2, \theta_k(1 - \theta_k), \theta_k(1 - \theta_k), 1 - \theta_k^2).$$

The law of  $(X, Y, Z, T)$  will be the law of  $J_1 + \dots + J_n$ , which can be computed for any given set of parameters  $\theta_k$ . The same can be done for the states of nature  $P \wedge \neg Q, \neg P \wedge Q$ , and  $\neg(P \wedge Q)$ .

- The true and false positive rates under the vector of competences  $\theta = (\theta_1, \dots, \theta_n)$  and rule  $R$  will be given, following Definitions 4.1 and 4.2, by

$$\text{TPR}(n, \theta, R) = \sum_{R(i,j,k,\ell)=1} \mathbb{P}_{P \wedge Q} \{J_1 + \dots + J_n = (i, j, k, \ell)\},$$

and

$$\text{FPR}(n, \theta, R) = \sum_{R(i,j,k,\ell)=1} \mathbb{P}_{P \wedge \neg Q} \{J_1 + \dots + J_n = (i, j, k, \ell)\}.$$

A study of dichotomous decision making under different individual competences, within a probabilistic framework, can be found in Sapir (1998).

- *Non-independence between voters*. If the committee members do not vote independently, then, in order to make exact computations, one must have the joint probability law of the vector  $(J_1, \dots, J_n)$ , which take values in the  $n$ -fold Cartesian product of  $\{P \wedge Q, P \wedge \neg Q, \neg P \wedge Q, \neg(P \wedge Q)\}$ . From the joint law, the distribution of the sum  $J_1 + \dots + J_n$  can always be made explicit, for each state of nature and taking into account the given vector of competences  $(\theta_1, \dots, \theta_n)$ . And from there, the values of FPR, FNR and AOT can be also obtained for any rule. Boland (1989) studied this situation of non-independence and diverse competence values for the voting of a single question, and assuming the existence of a “leader” in the committee. He generalises Condorcet theorem when the correlation coefficient between voters does not exceed a certain threshold. That situation is completely different from ours. Non-independence of voters may also arise when some voters have information on other voters' preferences and vote strategically (see for instance Terzopoulou and Endriss 2019). Other works that have studied epistemic social choice models with correlated voters are Ladha (1992), Ladha (1993), Ladha (1995), Dietrich and List (2004), Peleg and Zamir (2012), Dietrich and Spiekermann (2013), Dietrich and Spiekermann (2013), Pivato (2017).
- *Non-independence between the premises*. The premises may depend on each other in the sense that believing that  $P$  is true or false changes the perception on the veracity or falsity of  $Q$ . An extreme example of dependence is the classical

$P$  = “existence of a contract” and  $Q$  = “defendant breached the contract”, where voting  $\neg P$  forces to vote  $\neg Q$ .

In these situations, some more data is needed, namely, the competence of each voter on one of the premises alone, and on the other premise conditioned to have guessed correctly the first, and conditioned to have guessed it incorrectly. To wit, suppose that  $\theta_P$  is the competence of a voter on premise  $P$ , that  $\theta_{Q|P}$  is her competence on  $Q$  if she guesses correctly on  $P$ , and that  $\theta_{Q|\bar{P}}$  is her competence on  $Q$  assuming she does not guess correctly on  $P$ . Then, the first row in the table of Proposition A.4 would read

	$P_x$	$P_y$	$P_z$	$P_t$
$P \wedge Q$	$\theta_P \theta_{Q P}$	$\theta_P(1 - \theta_{Q P})$	$(1 - \theta_P)\theta_{Q \bar{P}}$	$(1 - \theta_P)(1 - \theta_{Q \bar{P}})$

and similarly for the other true states of nature. Interconnection between issues has been considered recently by Bozbay (2019), in a case isomorphic to the extreme one just mentioned, but when the voters have private conflicting partial information that could lead to inconsistent conclusions depending on the aggregation rule. Bozbay also introduces the possibility of abstention in an issue to obtain efficient aggregation rules in the sense of Nash equilibrium in this situation.

- *Competence depending on the true state.* The probability to guess the truth may depend on the truth itself, i.e  $\theta = (\theta_P, \theta_{\neg P}, \theta_Q, \theta_{\neg Q})$  could be four different parameters associated to the committee members, giving the probabilities of guessing the true state of nature when  $P$  is true,  $P$  is false,  $Q$  is true and  $Q$  is false, respectively. And these probabilities can be different for each individual, of course.
- *The Bayesian approach.* It is immediate to cast our ROC analysis into a Bayesian setting. Only the definition of false positive rate has to be changed. In Sect. 4, we defined the FPR as the worst case probability of making the error, the situation that corresponds to one of the premises being true whereas the other is false. If we hypothesise the existence of a priori probabilities  $\pi_{P \wedge Q}, \pi_{P \wedge \neg Q}, \pi_{\neg P \wedge Q}$  and  $\pi_{\neg P \wedge \neg Q}$  on the set of states of nature, then the different probabilities  $\mathbb{P}$  become the different conditional versions of only one probability  $\mathbb{P}$ . In that situation, Definition 4.2 would read

$$\text{FPR}(n, \theta, R) := \frac{\mathbb{P}\{R = 1 \mid P \wedge \neg Q\} \pi_{P \wedge \neg Q} + \mathbb{P}\{R = 1 \mid \neg P \wedge Q\} \pi_{\neg P \wedge Q} + \mathbb{P}\{R = 1 \mid \neg P \wedge \neg Q\} \pi_{\neg P \wedge \neg Q}}{1 - \pi_{P \wedge Q}}$$

where  $\{R = 1\}$  is a simplified notation for  $\{R(X, Y, T, Z) = 1\}$ . The definition of TPR does not change. The probabilistic settings in Bozbay (2019) and Terzopoulou and Endriss (2019) follow the Bayesian paradigm.

- *More than two premises.* It is not difficult to extend the setting to more than two premises when the truth of the conclusion is equivalent to the truth of all and every premise. If the premises are represented by  $P_1, \dots, P_s$ , and  $n$  is the com-

mittee size, then the total number of individual voting profiles is  $2^s$ , and a voting table is 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 extension of the concept of admissible rule is straightforward. The probability of a false negative is computed, obviously, under the state of nature  $P_1 \wedge \dots \wedge P_s$ , and that of false positive must be computed, according to the logic explained in Sect. 4, under the state  $P_1 \wedge \dots \wedge P_{s-1} \wedge \neg P_s$ .

All these extensions can be combined together. The key to compute FPR, FNR, and consequently the values of AOT and  $WAOT_w$  of any given rule is the ability to compute the probability of appearance of all possible voting tables; and this is possible for all the extensions of the list above. It is not easy to establish general theorems of comparison between rules, but a computer software can evaluate and compare rules for any specific value of all the parameters involved. Notice that it is not even necessary to fix  $n$  and  $\theta$ . Two differently formed committees, adhering to the same or to different decision rules, can be compared applying the same ideas, using the symmetric or the weighted area of the triangle. In this paper, we have stuck to the simplest of the situations in our exposition, to better highlight the methodology, and to obtain some specific theoretical results. We have also selected three particular rules, two of them classical and founded in well understood principles, corresponding to the comprehensive deliberative and the minimal liberal approaches to decision making (Dietrich and List 2007); but the methodology can be applied to compare any given subset of general binary rules.

The extension of the model to other truth-functional agendas can be more involved, although in principle all computations should be possible. For instance, assume that the conclusion is true if and only if the three-premisses formula  $(P_1 \vee P_2) \wedge (P_1 \vee P_3)$  is true, and that the state of nature is “all premisses are true”. If a voter competence is  $\theta$  for each premise, then their probability to get the conclusion right is  $\theta^3 + 3\theta^2(1 - \theta) + \theta(1 - \theta)^2$ , corresponding to get  $P_1$  right (true), or wrong (false) and the other two right (true). The probability to get the conclusion wrong is the complement  $2\theta(1 - \theta)^2 + (1 - \theta)^3$ . Note that a voter may get the conclusion right even when failing on all premisses; for instance, if the true state of nature is  $\neg P_1 \wedge P_2 \wedge P_3$ , then voting  $P_1 \wedge \neg P_2 \wedge \neg P_3$  would lead to the correct conclusion.

In ROC analysis one usually relies in sample training data for the classifier. Here we have postulated the existence of an exact competence parameter  $\theta$ . This parameter can be assigned on subjective grounds, but of course past data, if available, can be used to estimate its value. Furthermore, if after a new experiment it is possible to assess the quality of a voter’s decision, this value could be readjusted in a Bayesian manner. As examples of possible practical relevance, we mention: In court cases, the level of a court, and the proportion of cases that have been successfully appealed to higher instances, can measure the competence of individual judges; in some professional sport competitions, referees are ranked according to their performance in past events, and it is usually easy to determine a posteriori the proportion of their correct decisions in a given event; in simultaneous medical tests, the “competence” or reliability of each one is usually known to some extent, and they can be combined to offer the best diagnostic decision, taking possibly into account the risks of false positives or false negatives through the weight parameter  $w$ .

It may be argued that the appearance of the discrepancy between different decision rules is rare in practice, but this depends on the competence parameter  $\theta$ , and the committee size  $n$ . In fact, the probability to obtain different outcomes in different decision rules can be computed explicitly. For instance, List (2005, Proposition 2) computes the probability of the occurrence of the paradox from our formula (14), for the two classical rules premise-based ( $Prem, R_1$ ) and conclusion-based ( $Conc, R_3$ ).

However, the potential appearance of the paradox cannot be avoided, except for the trivial one-member committee. The interest must then be focused in the choice of the “best decision rule” among a catalogue of rules. The question then becomes to define a criterion to evaluate decision rules. In this work, we have considered a family of criteria, which are completely objective, once fixed the subjective weight of the two competing risks, the false positive and the false negative, in assessing if the conjunction  $P \wedge Q$  is true. This is in contrast with the classical theory of Hypothesis Testing, but in line with Statistical Decision Theory.

Once the weight  $w$  is given, the WAOT criterion to choose the rule is an definite way to arrive to a collective decision. To be honest, though, we must mention again that the definition of the false positive rate (Definition 4.2), although logical, and justified by Proposition A.6 in the Appendix, it is to some extent arbitrary.

Let us finish by commenting on possible extensions and open questions:

Instead of using the majority principle, another “qualified majority” or *quota* can be employed, and the analysis of the modified rules can be done similarly. Still another possibility is to use *score versions* of the rules: Instead of 0/1 outputs, one can use more general mappings from the set of tables into the set of real numbers. These mappings are called *scores*. The natural scores to associate to rules  $R_1, R_2, R_3$  are, respectively

$$S_1 = x + y \wedge z - m, \quad S_2 = x + \lfloor \frac{y \wedge z}{2} \rfloor - m, \quad S_3 = x - m,$$

and clearly  $S_3 \leq S_2 \leq S_1$ . ROC analysis can be done the same with scores, where the goal is to find the best score to fix as a boundary between  $P \wedge Q$  and  $\neg(P \wedge Q)$  from the point of view of the “area under the curve”. Admissible scores can be defined similarly to admissible rules: they must be non-decreasing functions with respect to the partial order on  $\mathbb{T}$  defined in Sect. 2. See Fawcett (2006) for a good short introduction to score rules and ROC curves. Note that this notion is different from the “judgement aggregation scoring rules” introduced by Dietrich (2014).

We have compared three particular rules, two of them traditional, and a reasonable third one that lies in between. They are easily expressed in terms of the entries of the voting table. But there are much more admissible rules; a total of 36 in the case  $n = 3$ , and they are not easy to enumerate systematically in general. Hence, the question of enumerating all admissible rules and choosing the best according to some criterion is open.

## Appendix

We first prove the equivalence between the expressions defining the three rules in Section 2.

**Proposition A.1** *Rules  $R_1, R_2, R_3$  can be defined by*

$$R_1 : \text{Decide } P \wedge Q \text{ if and only if } x > m - y \wedge z \quad (11)$$

$$R_2 : \text{Decide } P \wedge Q \text{ if and only if } x > m - \left\lfloor \frac{y\wedge z}{2} \right\rfloor \quad (12)$$

$$R_3 : \text{Decide } P \wedge Q \text{ if and only if } x > m \quad (13)$$

**Proof** For rule  $R_1$ , we must show the equivalence

$$(x + y > z + t, x + z > y + t) \iff x > m - y \wedge z.$$

Indeed,

$$\begin{aligned} x + y > z + t &\iff x + y > n - x - y \iff 2x + 2y > n = 2m + 1 \\ &\iff x + y > m + \frac{1}{2} \iff x > m - y, \end{aligned}$$

taking into account that we are dealing with integer numbers. The equivalence  $x + z > y + t \iff x > m - z$  is proved in an identical way and (11) follows.

For rule  $R_2$ , we must prove the equivalence in (12), that is,

$$(x > z + t, x > y + t) \iff x > m - \left\lfloor \frac{y\wedge z}{2} \right\rfloor.$$

We have

$$x > z + t \iff x > n - x - y \iff 2x > 2m + 1 - y \iff x > m - \frac{y-1}{2}.$$

If  $y$  is odd, say  $y = 2q + 1$ , then both  $\frac{y-1}{2}$  and  $\left\lfloor \frac{y}{2} \right\rfloor$  are equal to  $q$ , and we get the equivalence with  $x > m - \left\lfloor \frac{y}{2} \right\rfloor$ . If  $y = 2q$ ,

$$x > m - \frac{y-1}{2} = m - q + \frac{1}{2} \iff x > m - q = m - \left\lfloor \frac{y}{2} \right\rfloor.$$

Analogously, we obtain  $x > y + t \iff x > m - \left\lfloor \frac{z}{2} \right\rfloor$ , and (12) is proved.

Finally, for rule  $R_3$ , we have to prove the equivalence in (13):

$$x > y + z + t \iff x > m.$$

Indeed,

$$\begin{aligned} x > y + z + t &\iff x > n - x \iff 2x > 2m + 1 \\ &\iff x > m + \frac{1}{2} \iff x > m, \end{aligned}$$

and we are done. □



The relations among the rules  $R_1, R_2, R_3$  can be stated as follows:

**Proposition A.2** *Rules  $R_1, R_2, R_3$  of Definition 2.1 are admissible and, as functions  $\mathbb{T} \rightarrow \{0, 1\}$ , they satisfy  $R_3 \leq R_2 \leq R_1$ .*

*Moreover, we have  $R_3 = R_2 < R_1$  for  $n = 3, 5$ , and  $R_3 < R_2 < R_1$  for  $n \geq 7$ .*

**Proof** The admissibility of the three rules is completely obvious from the defining inequalities (1), (2) and (3). The order  $R_3 \leq R_2 \leq R_1$  as functions  $\mathbb{T} \rightarrow \{0, 1\}$  is also clear, so we focus on the equalities and strict inequalities.

First, we will prove that  $R_2 = R_3$  for  $n = 3$  (i.e.  $m = 1$ ). When  $n = 3, y + z \leq 3$  and this implies  $\lfloor \frac{y\wedge z}{2} \rfloor = 0$ . According to Proposition A.1, rules  $R_2$  and  $R_3$  are the same.

Secondly, for  $n = 5$  ( $m = 2$ ), we have  $y + z \leq 5$  and then  $\lfloor \frac{y\wedge z}{2} \rfloor \leq 1$ . On the points  $(x, y, z, t)$  for which  $\lfloor \frac{y\wedge z}{2} \rfloor = 0$ , the two rules  $R_2$  and  $R_3$  clearly coincide, by Proposition A.1. Assume  $\lfloor \frac{y\wedge z}{2} \rfloor = 1$ . That means  $y \geq 2, z \geq 2$ , and therefore  $x \leq 1$ . For such points, the value of both rules is zero. Hence, they coincide everywhere.

To see that  $R_1 > R_2$  for all  $n \geq 3$ , recall that this inequality means that  $R_1 \geq R_2$  and the two rules are not identical. Consider the point  $(x, y, z, t) = (1, m, m, 0)$ . On this point, condition (1) is satisfied, but (2) fails. Therefore,  $R_1(1, m, m, 0) = 1$  and  $R_2(1, m, m, 0) = 0$ .

To show that  $R_2 > R_3$  for all  $n \geq 7$  (implying that  $m \geq 3$ ), we distinguish two cases, depending on whether  $m$  is even or odd. For an even  $m \geq 4$ , take  $(x, y, z, t) = (m, \frac{m}{2}, \frac{m}{2}, 1)$ ; for  $m \geq 3$  odd, take  $(x, y, z, t) = (m, \frac{m+1}{2}, \frac{m+1}{2}, 0)$ . In both cases condition (2) is satisfied and (3) fails. □

**Remark A.3** We could have started by defining an equivalence relation  $\sim$  on  $\mathbb{T}$ , identifying tables

$$\begin{bmatrix} x & y \\ z & t \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x & z \\ y & t \end{bmatrix}$$

The partial order would be shorter to define, and an admissible rule considered on the quotient set  $\mathbb{T} / \sim$  would not need condition 1 of the definition. Figure 1 implicitly uses this equivalence. Although elegant, this setup would introduce more complications in the discussions of the next sections.

**Proposition A.4** *Under conditions (C1)–(C4) of Sect. 3, the joint distribution of  $(X, Y, Z, T)$  is multinomial  $M(n, p_x, p_y, p_z, p_t)$ , with parameters depending on the true state of nature, according to the following table:*

	$p_x$	$p_y$	$p_z$	$p_t$
$P \wedge Q$	$\theta^2$	$\theta(1 - \theta)$	$\theta(1 - \theta)$	$(1 - \theta)^2$
$P \wedge \neg Q$	$\theta(1 - \theta)$	$\theta^2$	$(1 - \theta)^2$	$\theta(1 - \theta)$
$\neg P \wedge Q$	$\theta(1 - \theta)$	$(1 - \theta)^2$	$\theta^2$	$\theta(1 - \theta)$
$\neg P \wedge \neg Q$	$(1 - \theta)^2$	$\theta(1 - \theta)$	$\theta(1 - \theta)$	$\theta^2$

**Proof** For each voter  $k = 1, \dots, n$ , let  $V_k$  be the random variable taking values in one of the four possible final decisions of the voter, and let  $p_x, p_y, p_z, p_t$  the probability of each of them. Since these  $n$  variables are mutually independent and identically distributed, the law of the counts  $(X, Y, Z, T)$  is multinomial with parameters  $(n, p_x, p_y, p_z, p_t)$ . The independence of the decisions concerning the two premises, gives immediately the particular parameters of the table, once given the true state of nature.

Recall that the multinomial probability function is given by

$$\mathbb{P}\{X = k, Y = j, Z = i, T = \ell\} = \binom{n}{k, j, i, \ell} p_x^k \cdot p_y^j \cdot p_z^i \cdot p_t^\ell, \tag{14}$$

where  $k, j, i, \ell$  are non-negative integers such that  $n = k + j + i + \ell$ , and  $\binom{n}{k, j, i, \ell}$  means the quotient of factorials  $\frac{n!}{k! \cdot j! \cdot i! \cdot \ell!}$ . □

**Proposition A.5** *Under the assumptions (C1)–(C4), the true positive rates defined in (7) for rules  $R_1, R_2, R_3$  are:*

$$\text{TPR}(n, \theta, R_1) = \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m-i \wedge j+1}^{n-i-j} \binom{n}{k, j, i, \ell} \theta^{i+j+2k} (1 - \theta)^{2n-i-j-2k}, \tag{15}$$

$$\text{TPR}(n, \theta, R_2) = \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m-\lfloor \frac{i \wedge j}{2} \rfloor + 1}^{n-i-j} \binom{n}{k, j, i, \ell} \theta^{i+j+2k} (1 - \theta)^{2n-i-j-2k}, \tag{16}$$

$$\text{TPR}(n, \theta, R_3) = \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m+1}^{n-i-j} \binom{n}{k, j, i, \ell} \theta^{i+j+2k} (1 - \theta)^{2n-i-j-2k}, \tag{17}$$

where  $\ell = n - i - j - k$ . □

**Proof** The probability of deciding  $P \wedge Q$  for any decision rule can be computed summing up the multinomial probability function (14) over the set where the rule takes the value 1. That means, taking into account Proposition A.1,

$$\begin{aligned} \mathbb{P}\{R_1(X, Y, Z, T) = 1\} &= \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m-i \wedge j+1}^{n-i-j} \frac{n!}{k!j!i!(n-i-j-k)!} p_x^k p_y^j p_z^i p_t^{n-i-j-k} . \\ \mathbb{P}\{R_2(X, Y, Z, T) = 1\} &= \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m-\lfloor \frac{1}{2}i \wedge j \rfloor + 1}^{n-i-j} \frac{n!}{k!j!i!(n-i-j-k)!} p_x^k p_y^j p_z^i p_t^{n-i-j-k} . \\ \mathbb{P}\{R_3(X, Y, Z, T) = 1\} &= \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m+1}^{n-i-j} \frac{n!}{k!j!i!(n-i-j-k)!} p_x^k p_y^j p_z^i p_t^{n-i-j-k} . \end{aligned}$$

Now using the values of Proposition A.4 for the case  $P \wedge Q$ , we obtain the stated true positive rates. □

**Proposition A.6** *Let  $R$  be any one of the rules  $R_1, R_2$  or  $R_3$ . Under assumptions (C1)–(C4) of Sect. 3, we have*

$$\begin{aligned} \mathbb{P}_{P \wedge Q}\{R(X, Y, Z, T) = 1\} &= \mathbb{P}_{\neg P \wedge Q}\{R(X, Y, Z, T) = 1\} > \\ \mathbb{P}_{\neg P \wedge \neg Q}\{R(X, Y, Z, T) = 1\} . \end{aligned} \tag{18}$$

**Proof** To check the equality in (18), first recall that every admissible decision rule satisfies  $R(x, y, z, t) = R(x, y, z, 1-t)$ . Moreover, the law of the random vector  $(X, Y, Z, T)$  under  $P \wedge \neg Q$ , which is multinomial  $M(n, \theta(1-\theta), \theta^2, (1-\theta)^2, \theta(1-\theta))$ , coincides with the law of  $(X, Z, Y, T)$  under  $\neg P \wedge Q$ . Therefore,

$$\begin{aligned} \mathbb{P}_{P \wedge \neg Q}\{R(X, Y, Z, T) = 1\} &= \mathbb{P}_{\neg P \wedge Q}\{R(X, Z, Y, T) = 1\} \\ &= \mathbb{P}_{\neg P \wedge Q}\{R(X, Y, Z, T) = 1\} . \end{aligned}$$

For the inequality in the statement, we have, from Proposition A.4,

$$\begin{aligned} \mathbb{P}_{P \wedge \neg Q}\{R_1(X, Y, Z, T) = 1\} &= \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m-i \wedge j+1}^{n-i-j} \binom{n}{k, j, i, \ell} \theta^{n-i+j} (1-\theta)^{n+i-j} , \\ \mathbb{P}_{\neg P \wedge \neg Q}\{R_1(X, Y, Z, T) = 1\} &= \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m-i \wedge j+1}^{n-i-j} \binom{n}{k, j, i, \ell} \theta^{2n-i-j-2k} (1-\theta)^{i+j+2k} , \end{aligned}$$

where  $\ell = n - i - j - k$ . Each term in the first line is greater than the corresponding one in the second line:  $\theta^{n-i+j} (1-\theta)^{n+i-j} > \theta^{2n-i-j-2k} (1-\theta)^{i+j+2k}$  is equivalent to  $\theta^{2k+2j-n} > (1-\theta)^{2k+2j-n}$ , and this inequality, since  $\theta > \frac{1}{2}$ , is true provided that  $2k + 2j - n > 0$ . Indeed,

$$k \geq m - i \wedge j + 1 \geq m - j + 1 \Rightarrow 2k + 2j \geq 2m + 2 \Rightarrow 2k + 2j - n \geq 1 .$$

The same argument is valid for  $R_2$  and  $R_3$ , since  $m - i \wedge j + 1 \leq m - \lfloor \frac{i \wedge j}{2} \rfloor + 1 \leq m + 1$ . This completes the proof.

In fact, one can prove, using a probabilistic coupling argument, that the inequality  $\mathbb{P}_{P \wedge \neg Q}\{R = 1\} \geq \mathbb{P}_{\neg P \wedge \neg Q}\{R = 1\}$  holds true not only for  $R_1, R_2, R_3$ , but for any admissible rule as defined in 2.3. For simplicity, we have restricted ourselves here to prove the statement as it is. □

Using the formulae of Proposition A.4, one can write the analogue of Proposition A.5 for FPR. The computations are completely analogous to those of Proposition A.5.

**Proposition A.7** *Under the assumptions (C1)–(C4), the false positive rate defined in (8) for each fixed rule is:*

$$FPR(n, \theta, R_1) = \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m-i \wedge j+1}^{n-i-j} \binom{n}{k, j, i, \ell} \theta^{n-i+j} (1 - \theta)^{n+i-j} , \tag{19}$$

$$FPR(n, \theta, R_2) = \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m-\lfloor \frac{i \wedge j}{2} \rfloor+1}^{n-i-j} \binom{n}{k, j, i, \ell} \theta^{n-i+j} (1 - \theta)^{n+i-j} , \tag{20}$$

$$FPR(n, \theta, R_3) = \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m+1}^{n-i-j} \binom{n}{k, j, i, \ell} \theta^{n-i+j} (1 - \theta)^{n+i-j} , \tag{21}$$

where  $\ell = n - i - j - k$ .

**Proof of Theorem 5.2:** We start with the second inequality, which is equivalent, by (4), to  $TPR(n, \theta, R_2) - FPR(n, \theta, R_2) < TPR(n, \theta, R_1) - FPR(n, \theta, R_1)$ , and therefore to

$$FPR(n, \theta, R_1) - FPR(n, \theta, R_2) < TPR(n, \theta, R_1) - TPR(n, \theta, R_2) . \tag{22}$$

In other words, it is equivalent to say that the increase in true positives when changing from  $R_2$  to  $R_1$  more than compensates the increase in false positives.

Using formulae (15), (16), (19) and (20), it is clearly enough to prove that, for all  $0 \leq i \leq n$ , for all  $0 \leq j \leq n - i$ , and for all  $m - i \wedge j + 1 \leq k \leq m - \lfloor \frac{i \wedge j}{2} \rfloor$ ,

$$\theta^{n-i+j} (1 - \theta)^{n+i-j} < \theta^{i+j+2k} (1 - \theta)^{2n-i-j-2k}$$

or, equivalently, that

$$(1 - \theta)^{2i+2k-n} < \theta^{2i+2k-n} ,$$

which can be easily checked taking into account that  $0 < 1 - \theta < \theta$  and  $2i + 2k - n > 0$ : Indeed,  $n = 2m + 1$  and  $k \geq m - i \wedge j + 1$  imply that  $2i + 2k - n \geq 2i + 2(m - i \wedge j + 1) - (2m + 1) = 2i - 2i \wedge j + 1 \geq 1$ . Thus, the second inequality is proved.

A similar argument proves the first one, using (16), (17), (20) and (21), and noticing that  $k \geq m - i \wedge j + 1$  also holds. The inequality is strict, unless  $n \leq 5$ , in which case  $R_2$  and  $R_3$  are equal (see Proposition A.2). □

**Proof of Corollary 5.3:** The proof of Theorem 5.2 is based in checking the inequality (22). The analogue for the weighted area is  $w(\text{FPR}(n, \theta, R_1) - \text{FPR}(n, \theta, R_2)) < (1 - w)(\text{TPR}(n, \theta, R_1) - \text{TPR}(n, \theta, R_2))$ , which trivially follows from (22) when  $w < \frac{1}{2}$ . □

Theorem 5.4 will follow from two preliminary lemmas. In the proofs, we only treat in detail the claims relating  $R_1$  and  $R_2$ , those relating  $R_2$  and  $R_3$  being analogous, with only slight changes that will be noted. Condition  $n \geq 7$  in the second case is needed since otherwise  $R_2$  and  $R_3$  coincide, as we have seen in Sect. 2.

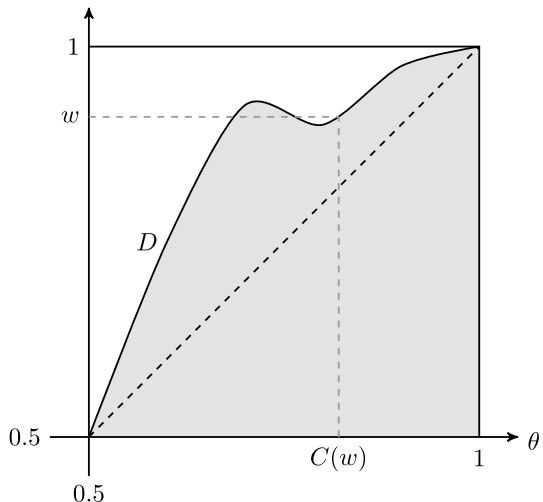
The first lemma is interesting in itself in that it establishes a dichotomy when we look at  $\theta$  as fixed and let  $w$  vary. The second lemma states some technical properties of the functions introduced in the first. Refer to Fig. 7 for a graphical clue of the situation presented in theorem and lemmas.

**Lemma A.8** Fix  $n \geq 3$ . For every competence  $\frac{1}{2} < \theta < 1$ , there exists a constant  $D_1(\theta)$ , greater than  $\theta$  (except that  $D_1(\theta) = \theta$  if  $n = 3$ ), such that

$$w < D_1(\theta) \Rightarrow \text{WAOT}_w(n, \theta, R_1) > \text{WAOT}_w(n, \theta, R_2)$$

$$w > D_1(\theta) \Rightarrow \text{WAOT}_w(n, \theta, R_1) < \text{WAOT}_w(n, \theta, R_2).$$

**Fig. 7** An example graph of a function  $D(\theta)$  with the properties of Lemmas A.8 and A.9. For competence values  $\theta$  greater than  $C(w)$ , rule  $R_1$  is better than rule  $R_2$  in the sense of the weighted area of the triangle



Fix  $n \geq 7$ . For every competence  $\frac{1}{2} < \theta < 1$ , there exists a constant  $D_2(\theta)$ , greater than  $\theta$ , such that

$$\begin{aligned} w < D_2(\theta) &\Rightarrow \text{WAOT}_w(n, \theta, R_2) > \text{WAOT}_w(n, \theta, R_3) \\ w > D_2(\theta) &\Rightarrow \text{WAOT}_w(n, \theta, R_2) < \text{WAOT}_w(n, \theta, R_3). \end{aligned}$$

**Lemma A.9** For  $i = 1, 2$ , the functions  $\theta \mapsto D_i(\theta)$  in Lemma A.8 satisfy:

1. For any  $\frac{1}{2} < \theta < 1$ , we have also  $\frac{1}{2} < D_i(\theta) < 1$ .
2.  $D_i$  is continuous in the interval  $(\frac{1}{2}, 1)$ .
3.  $\lim_{\theta \searrow \frac{1}{2}} D_i(\theta) = \frac{1}{2}$  and  $\lim_{\theta \nearrow 1} D_i(\theta) = 1$ .

**Proof of Lemma A.8:** We prove the first part of the lemma, the other one being analogous, and write simply  $D$  instead of  $D_1$ .

Fix  $n \geq 3$ , and let  $\frac{1}{2} < \theta < 1$ . From the definitions of WAOT, FPR, FNR and TPR, the inequality  $\text{WAOT}_w(n, \theta, R_2) < \text{WAOT}_w(n, \theta, R_1)$  is equivalent to

$$w(\text{FPR}(n, \theta, R_1) - \text{FPR}(n, \theta, R_2)) < (1 - w)(\text{TPR}(n, \theta, R_1) - \text{TPR}(n, \theta, R_2)),$$

or

$$(1 - w)(\text{TPR}(n, \theta, R_1) - \text{TPR}(n, \theta, R_2)) - w(\text{FPR}(n, \theta, R_1) - \text{FPR}(n, \theta, R_2)) > 0. \tag{23}$$

We know that inequality (23) is true for  $w \leq \frac{1}{2}$ , for all  $n \geq 3$  and all  $\theta > \frac{1}{2}$ , by Theorem 5.2 and Corollary 5.3, whereas for  $w = 1$  is manifestly false, since clearly  $\text{FPR}(n, \theta, R_1) - \text{FPR}(n, \theta, R_2) > 0$ , from (19)–(20).

Since the left-hand side is a linear function of  $w$ , there must be a unique point  $D(\theta)$  such that the equality holds in (23), and such that for every  $w < D(\theta)$  rule  $R_1$  yields a greater weighted area than  $R_2$ , whereas for every  $w > D(\theta)$  it is the other way round.

Let us now prove that  $D(\theta) > \theta$ : Consider the particular weight  $w = \theta$ . Clearly, we only need to check that (23) is satisfied for this  $w$ . Using (15–16) and (19–20), inequality (23) is equivalent, for this particular value, to

$$\begin{aligned} &\sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m-i/j+1}^{\lfloor \frac{i+j}{2} \rfloor \wedge (n-i-j)} \binom{n}{k, j, i, \ell} \\ &\left[ \theta^{i+j+2k} (1 - \theta)^{1+2n-i-j-2k} - \theta^{1+n-i+j} (1 - \theta)^{n+i-j} \right] > 0. \end{aligned} \tag{24}$$

The expression in square brackets is non-negative if and only if

$$\left( \frac{\theta}{1 - \theta} \right)^{2(i+k) - (n+1)} \geq 1,$$

which is true because  $\theta > \frac{1}{2}$ , and  $2(i + k) \geq n + 1$  in this range of indices. Therefore, the sum (24) is non-negative. Furthermore, one can easily check that there is at least one positive term in the sum for  $n > 3$ . We conclude that the critical point  $D(\theta)$  is

greater than  $\theta$ . In the special case  $n = 3$ , there is only one term in the sum and it is equal to zero, hence  $D(\theta) = \theta$ .

The claim on  $R_2$  and  $R_3$  can be proved in the same way, with the only difference that the index  $k$  in the sum (24) ranges from  $m - \lfloor \frac{i \wedge j}{2} \rfloor + 1$  to  $m \wedge (n - i - j)$ . If  $n = 7$ , the only term in the sum (24) is already positive, so there is no need to consider this case separately.

**Proof of Lemma A.9:** Again, we prove the claims first for  $D_1$  and we call it simply  $D$ . Fix  $n \geq 3$ . The point  $D(\theta)$  is easily computed:  $D(\theta) = A(\theta)/(A(\theta) + B(\theta))$ , with

$$A(\theta) := \text{TPR}(n, \theta, R_1) - \text{TPR}(n, \theta, R_2), \quad B(\theta) := \text{FPR}(n, \theta, R_1) - \text{FPR}(n, \theta, R_2).$$

Using (15–16) and (19–20), their explicit expressions are:

$$A(\theta) = \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m-i \wedge j+1}^{(m-\lfloor \frac{i \wedge j}{2} \rfloor) \wedge (n-i-j)} \binom{n}{k, j, i, \ell} \theta^{i+j+2k} (1-\theta)^{2n-i-j-2k}$$

$$B(\theta) = \sum_{i=0}^n \sum_{j=0}^{n-i} \sum_{k=m-i \wedge j+1}^{(m-\lfloor \frac{i \wedge j}{2} \rfloor) \wedge (n-i-j)} \binom{n}{k, j, i, \ell} \theta^{n-i+j} (1-\theta)^{n+i-j}.$$

Notice first that the function  $D(\theta)$  is the quotient of two polynomials in  $\theta$  that never vanish because both the true and the false positive rates are greater for rule  $R_1$  than for  $R_2$ , for all  $\theta > \frac{1}{2}$  and  $n \geq 3$  (see again (15)–(16) and (19)–(20)). Hence, it is clear that  $D$  is continuous and less than 1 in its domain. Moreover, Lemma A.8 ensures that  $\frac{1}{2} < \theta \leq D(\theta)$ . In particular,  $\frac{1}{2} < D(\theta) < 1$ . For  $\theta = \frac{1}{2}$ , both  $A$  and  $B$  are well defined and  $A(\frac{1}{2}) = B(\frac{1}{2})$ , giving  $\lim_{\theta \searrow \frac{1}{2}} D(\theta) = \frac{1}{2}$ . Finally,  $\theta \leq D(\theta) < 1$ , implies that  $\lim_{\theta \nearrow 1} D(\theta) = 1$ , and the three claims of the Lemma are proved for the function  $D_1$ .

The proof for  $D_2$  is identical, using formulae (16–17) and (20–21) and the hypothesis  $n \geq 7$ . □

**Proof of Theorem 5.4:** Denote, as before, by  $D$  and  $C$  the functions  $D_1$  and  $C_1$ , respectively. The proof of the second part of the theorem is identical, with  $D_2$  and  $C_2$ .

From Lemmas A.8 and A.9, we know that  $D(\theta)$  defines a continuous function that maps  $[\frac{1}{2}, 1]$  onto  $[\frac{1}{2}, 1]$  (extending it by continuity at the endpoints), and that the curve  $(\theta, D(\theta))$  lies above the diagonal, as depicted in Fig. 7. The shaded region below the curve is the set of pairs  $(\theta, w)$  for which  $\text{WAOT}_w(n, \theta, R_1) > \text{WAOT}_w(n, \theta, R_2)$ , and the inequality is reversed above the curve.

Regardless whether  $D$  is and increasing function or not, and because it is a quotient of polynomials, for any fixed  $w$  in  $(\frac{1}{2}, 1)$  there is a largest value  $C(w) \in (\frac{1}{2}, 1)$  that solves for  $\theta$  the equation  $D(\theta) = w$ . And it is clear from the figure that for all  $\theta > C(w)$ , all points of the segment  $\{(\theta, w) : \theta > C(w)\}$  lie below the curve, hence  $\text{WAOT}_w(n, \theta, R_1) > \text{WAOT}_w(n, \theta, R_2)$  for these values.

Finally, since  $D(\theta) > \theta$  for all  $\theta$  if  $n > 3$ , we have in particular  $w = D(C(w)) > C(w)$ . For  $n = 3$ ,  $D(\theta) = \theta$ , and we get  $w = C(w)$ . This finishes the proof. □

In Theorem 5.4 we have not assumed that the function  $\theta \rightarrow D(\theta)$  is increasing. We conjecture that it always is, but we do not have yet a formal proof. If the conjecture is true, or in a particular case in which it is checked to be true, then one could strengthen the theorem by claiming that for every weight  $\frac{1}{2} < w < 1$ , there exists a value  $C(w)$ , smaller than  $w$ , such that

$$\begin{aligned}\theta > C(w) &\Rightarrow \text{WAOT}_w(n, \theta, R_1) > \text{WAOT}_w(n, \theta, R_2) \\ \theta < C(w) &\Rightarrow \text{WAOT}_w(n, \theta, R_1) < \text{WAOT}_w(n, \theta, R_2),\end{aligned}$$

and similarly for  $R_2$  in relation with  $R_3$ .

### R session info.

Computations have been done in R with the following setup:

R version 4.0.2 (2020-06-22), x86\_64-w64-mingw32.

Base packages: base, datasets, graphics, grDevices, methods, stats, utils.

Other packages: knitr 1.30, xtable 1.8-4, compiler 4.0.2, digest 0.6.25, evaluate 0.14, magrittr 2.0.1, stringi 1.5.3, stringr 1.4.0, tools 4.0.2, xfun 0.18.

**Acknowledgements** This work has been partially supported by grant numbers MTM2014-59179-C2-1-P from the Ministry of Economy and Competitiveness of Spain, and 2017-SGR-1094 from the Ministry of Business and Knowledge of Catalonia. The authors sincerely thank the work of the anonymous reviewer for their very helpful comments, literature recommendations, and presentation suggestions. We also thank the Associate Editor in charge, for their additional comments and contributions to improve the manuscript.

**Funding** Open Access Funding provided by Universitat Autònoma de Barcelona.

**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

- Ahn DS, Oliveros S (2014) The Condorcet jur(ies) theorem. *J Econ Theory* 150:841–851
- Boland PJ (1989) Majority systems and the Condorcet Jury Theorem. *J R Stat Soc Ser D (Stat)* 38(3):181–189
- Bonnefon J.-F (2010) Behavioral evidence for framing effects in the resolution of the doctrinal paradox. *Soc Choice Welf* 34(4):631–641
- Bovens L, Rabinowicz W (2004) Voting procedures for complex collective decisions: an epistemic perspective. *Ratio Juris* 17(2):241–258
- Bovens L, Rabinowicz W (2006) Democratic answers to complex questions—an epistemic perspective. *Synthese* 150(1):131–153
- Bozbay I (2019) Truth-tracking judgment aggregation over interconnected issues. *Soc Choice Welf* 53(2):337–370
- Camps R, Mora X, Saumell L (2012) A general method for deciding about logically constrained issues. *Ann Math Artif Intell* 64(1):39–72
- Condorcet, MJAN Caritat de (1785) *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie royale, Paris



- de Clippel G, Eliaz K (2015) Premise-based versus outcome-based information aggregation. *Games Econ Behav* 89:34–42
- Dietrich F (2007) A generalised model of judgment aggregation. *Soc Choice Welf* 28(4):529–565
- Dietrich F (2014) Scoring rules for judgment aggregation. *Soc Choice Welf* 42(4):873–911
- Dietrich F, List C (2004) A model of jury decisions where all jurors have the same evidence. *Synthese* 142(2):175–202
- Dietrich F, List C (2007) Arrow's theorem in judgment aggregation. *Soc Choice Welf* 29(1):19–33
- Dietrich F, List C (2007) Strategy-proof judgment aggregation. *Econ Philos* 23(3):269–300
- 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
- Egan JP (1975) Signal detection theory and ROC-analysis. Academic Press
- Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27(8):861–874
- Hand DJ, Till RJ (2001) A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach Learn* 45(2):171–186
- Karotkin D, Paroush J (2003) Optimum committee size: quality-versus-quantity dilemma. *Soc Choice Welf* 20(3):429–441
- Kornhauser LA (1992) Modeling collegial courts I: path-dependence. *Int Rev Law Econ* 12(2):169–185
- Kornhauser LA, Sager LG (1993) The one and the many: adjudication in collegial courts. *Calif Law Rev* 81(1)
- Ladha KK (1992) The Condorcet jury theorem, free speech, and correlated voters. *Am J Polit Sci* 36:617–634
- Ladha KK (1993) Condorcet's jury theorem in light of de Finetti's theorem. *Soc Choice Welf* 10(1):69–85
- Ladha KK (1995) Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes. *J Econ Behav Org* 26(3):353–372
- List C (2005) The probability of inconsistencies in complex collective decisions. *Soc Choice Welf* 24(1):3–32
- List C (2006) The discursive dilemma and public reason. *Ethics* 116(2):362–402
- 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, Puppe C (2009) Judgment aggregation: a survey. In: Anand P, Pattanaik P K, Puppe C (eds), *Handbook of rational and social choice*, Oxford University Press, Oxford
- Nehring K, Pivato M (2011) Incoherent majorities: the McGarvey problem in judgement aggregation. *Discret Appl Math* 159(15):1488–1507
- Peleg B, Zamir S (2012) Extending the Condorcet jury theorem to a general dependent jury. *Soc Choice Welf* 39(1):91–125
- Pivato M (2017) Epistemic democracy with correlated voters. *J Math Econ* 72:51–69
- Sapir L (1998) The optimality of the expert and majority rules under exponentially distributed competence. *Theor Decis* 45(1):19–36
- 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)*
- Terzopoulou Z, Endriss U (2019) Strategyproof judgment aggregation under partial information. *Soc Choice Welf* 53(3):415–442