General representation of epistemically optimal procedures

Franz Dietrich[1]
forthcoming in *Social Choice and Welfare*

Abstract

Assuming that votes are independent, the epistemically optimal procedure in a binary collective choice problem is known to be a weighted supermajority rule with weights given by personal log-likelihood-ratios. It is shown here that an analogous result holds in a much more general model. Firstly, the result follows from a more basic principle than expected-utility maximisation, namely from an axiom ("Epistemic Monotonicity") which requires neither utilities nor prior probabilities of the 'correctness' of alternatives. Secondly, a person's input need not be a vote for an alternative, it may be any type of input, for instance a subjective degree of belief or probability of the correctness of one of the alternatives. The case of a profile of subjective degrees of belief is particularly appealing, since here no parameters such as competence parameters need to be known.

1 Introduction and overview

Throughout, we consider a group of n persons, labelled $i = 1, ..., n$, which faces a binary choice problem. One option (it does not matter which) is called 0, the other one 1. By assumption, exactly one of these options is the 'correct' alternative; everybody would prefer making a correct choice, but there is uncertainty as to which is the correct alternative.[2]

Let us take a purely *epistemic* approach by caring only of the correctness of the decision regardless of the fairness of the decision procedure.[3] Within this paradigm it is intuitively clear that more competent persons should have more say. It is known

[1] I wish to express my thanks to various people, including Luc Bovens, Bernhard Grofman, James Hawthorne, Christian List and Josh Snyder. I also thank the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research, and the Program for the Investment in the Future (ZIP) of the German Government, for supporting this research. Address for correspondence: Center For Junior Research Fellows, University of Konstanz, 78457 Konstanz (Germany). Telephone: ++49 (0)7531 88-4733. Email: franz.dietrich@uni-konstanz.de.

[2] This assumption may, for instance, be justified if the choice is between convicting or acquitting a defendant, since conviction is correct if and only if the defendant is guilty. In other situations, this assumption is more critical, such as in choices between two (political) candidates.

[3] The epistemic approach assumes a procedure-independent standard of correctness of decisions and takes the only aim to be a decision in favour of that correct alternative. By contrast, the procedural approach aims at reaching decisions that are a fair or democratic reflection of the profile. For a discussion of epistemic and procedural justifications for procedures, see for instance Cohen (1986), Dahl (1979), Coleman and Ferejohn (1986), Estlund (1993, 1997) and List and Goodin (2001).

(see the below literature review) that this intuition can be turned into a precise statement that specifies 'how much' more say should be given to more competent persons: Under independent voting, expected-utility maximisation requires that

$$\textit{the procedure decides according to a threshold for } \sum_{i=1}^{n} \log \frac{P(x_i|H_1)}{P(x_i|H_0)} \tag{1}$$

(see Definition 1 for "deciding according to a threshold"). Here, H_0 (H_1) is the hypothesis that alternative 0 (1) is correct, and $P(x_i|H_0)$ ($P(x_i|H_1)$) is the probability that person i votes x_i given H_0 (H_1). Under this rule, voting power increases with competence on a specific logarithmic scale. Indeed, the more competent a person i is, the more the likelihood-ratio $\frac{P(x_i|H_1)}{P(x_i|H_0)}$ differs from 1, and hence the larger the term $\log \frac{P(x_i|H_1)}{P(x_i|H_0)}$ becomes in absolute value.[4] An extreme case is that person 3 (say) is so much more competent than all others that the term $\log \frac{P(x_3|H_1)}{P(x_3|H_0)}$ dominates the entire sum and the procedure reduces to an expert rule with person 3 as the expert. The other extreme is that the persons are nearly identically competent, so that the terms $\log \frac{P(x_3|H_1)}{P(x_3|H_0)}$ have very similar magnitudes and the procedure reduces to a (non-weighted) supermajority rule.

Before mentioning the literature, let me give a brief overview of the two contributions of this paper, namely the criterion of Epistemic Monotonicity and the generalisation to arbitrary types of personal inputs.

Epistemic Monotonicity. This paper has a slightly less economical and more social-choice-theoretic focus than some of the below-cited literature. Our main objective is not to derive the single optimal procedure (except in Section 10), rather we formulate a basic principle of epistemic social choice and derive the class of those procedures satisfying it. Our principle replaces the heavier concept of expected-utility maximisation and might be acceptable for groups who reject the latter or cannot agree on the values of utilities and/or prior probabilities of the correctness of alternatives. Although quite elementary, our principle will take us nearly as far as expected-utility maximisation.

To motivate Epistemic Monotonicity, notice that in (1) the individual log-likelihood-ratios depend neither on prior probabilities (of H_0 and H_1) nor on any utility values – prior probabilities and utilities become relevant only when specifying the threshold to which the sum is to be compared. This suggests that a more basic principle than expected-utility maximisation underlies the fact that (under independence) procedures should have the form (1), where this basic principle should not depend on particular utilities or prior probabilities.

A first remark is that (1) also follows if, instead of maximising the expected utility one maximises the probability of a correct decision, which does not require utilities. However, this still involves prior probabilities, and moreover it makes an implicit

[4] Assume for instance that person i is highly competent. Then a vote of $x_i = 0$ is highly probable given H_0 and little probable given H_1, implying that $\frac{P(x_i|H_1)}{P(x_i|H_0)}$ is close zero, and hence $\log \frac{P(x_i|H_1)}{P(x_i|H_0)}$ is 'close' $-\infty$. On the other hand, a vote of $x_i = 1$ is highly probable given H_1 and little probable given H_0, implying that $\frac{P(x_i|H_1)}{P(x_i|H_0)}$ is 'close' ∞, and hence $\log \frac{P(x_i|H_1)}{P(x_i|H_0)}$ is also 'close' ∞.

assumption about utilities since it is equivalent to expected-utility maximisation in the special case in which the utilities satisfy a quite particular relation.

So, what then is the basic principle entailing (1) without making (explicit or implicit) assumptions about priors or utilities? As we will show, it is Epistemic Monotonicity, which states roughly that if a profile that yields decision $y \in \{0,1\}$ is modified so as to make y more probably correct, then the modified profile should still yield decision y. More precisely, Epistemic Monotonicity requires that,

for any two profiles $x\,, x'$*, if* $P(H_1|x) \le P(H_1|x')$ *then* $f(x) \le f(x')$.

This does not involve utilities, and despite the appearance not even prior probabilities, as will be shown. The generality of this principle comes at the expense of a unique characterisation of the procedure: The threshold for the weighted sum does not follow from this axiom.

Arbitrary inputs. The other contribution of this paper is a full generalisation of the type of profile submitted: Epistemic Monotonicity implies (1) regardless of the type of information people submit. They may submit simple votes for either option (as in the below-cited literature), or votes with abstention allowed, or subjective probabilities of the correctness of alternative 1, or statements chosen from the set of statements {'I am sure of guilt', 'I am rather sure of guilt', 'I don't know', 'I am rather sure of innocence', 'I am sure of innocence'}, and so on. We also allow the situation in which different people are asked to submit different types of inputs; for example, 'experts' might be asked to submit more informative inputs than others.

The case in which people submit their subjective probabilities of H_1 (the hypothesis that 1 is correct) has a distinguishing property: Under independence and an additional assumption, epistemically monotonic procedures can be devised without recourse to any (unknown) parameters – in contrast, for instance, to the case of simple votes analysed by the literature, in which knowledge of competence parameters is required. More precisely, we shall show that a procedure aggregating subjective probabilities $(x_1, ..., x_n)$ of H_1 is epistemically monotonic if and only if

$$\textit{the procedure decides according to some threshold for } \sum_{i=1}^{n} \log \frac{x_i}{1-x_i}. \qquad (2)$$

The only quantities involved in this sum are the known submitted subjective probabilities $x_1, ..., x_n$ of H_1. Note the formal similarity of (2) to the corresponding statement for the case of simple votes: One only needs to replace person i's submitted probability x_i in (2) by his or her competence p_i in the simple-vote case and put a minus in front of those summands in which the person votes for 0. But this analogy seems to be a coincidence, since (2) holds for quite different reasons.

Literature review on epistemically optimal procedures. Classical Condorcet jury models stand between the epistemic and the procedural approach[3]: On the one hand, simple majority voting is adopted on procedural grounds, and, on the other hand, the epistemic (truth-tracking) performance of this procedure is analysed.[5] Nitzan and

[5] For a collection of results in the case of independent voting, see Grofman, Owen and Feld (1983).

Paroush (1982) and Shapley and Grofman (1984) have taken a 'purely epistemic' perspective by choosing the procedure so as to optimise the truth-tracking performance, at the expense of giving up the procedural concern of 'one man, one vote'. Assuming independent voting, they derive optimal procedures of the type (1). The result is successively generalised by Nitzan and Paroush (1984a, 1984b) and Ben-Yashar and Nitzan (1997).[6] Gradstein and Nitzan (1986) analyse conditions under which, in small groups, the epistemically optimal weighted rule reduces to common rules such as simple majority rule or expert rule. The important and complex question of the effect of dependence between votes on the performance of group judgments has been studied for non-weighted majority rules (e.g. Berg (1993), Boland, Proschan, and Tong (1989), and Ladha (1993, 1995)), but the same question for weighted majority rules has been little addressed; see, however, Shapley and Grofman (1984) and Berry (1994). The problem that competence levels are little known in practice has been tackled mainly by treating a person's competence as a random variable with a certain (known) distribution; see Nitzan and Paroush (1984c), Gradstein and Nitzan (1986), Berend and Harmse (1993), Sapir (1998), and Berend and Sapir (2002).

Structure of the paper. After defining the model (Sections 2 and 3), I introduce and discuss Epistemic Monotonicity (Section 4). I then prove representation theorems for the class of epistemically monotonic procedures, first in the general case (Section 5) and then in the case of independence (Section 6). To illustrate these results, I then discuss three examples: In Section 7, I consider the case that people submit simple votes, as assumed in the above-cited literature; in Section 8, I discuss the case in which people submit their degrees of belief; and in Section 9, I consider the case of a mixed profile, where some individuals submit simple votes and others (e.g. the 'experts') submit their degrees of belief. In Section 10, I restate all of our theorems for the case in which the criterion is not Epistemic Monotonicity but the more classical criterion of expected-utility maximisation, thereby generalising the main results of the literature to arbitrary inputs. Section 11 contains a conclusion. In Appendix A, we discuss a certain generalisation. Proofs are given in Appendix B.

Notation. We denote random variables by capital letters $(X, X_1, X_2, ...)$, and particular realisations of them by small letters $(x, x_1, x_2, ...)$. The probability of the event $X = x$, $P(X = x)$, is simply written $P(x)$; and analogously for all other random variables $(X_1, X_2, ...)$. This notation is naturally extended to conditional and joint probabilities: $P(x|H_0)$ means $P(X = x|H_0)$, $P(H_0|x)$ means $P(H_0|X = x)$, $P(x_1, ..., x_n|H_1)$ means $P(X_1 = x_1, ..., X_n = x_n|H_1)$, etc. Sets are written in calligraphic $(\mathcal{X}_1, \mathcal{X}, \mathcal{F}, ...)$.

[6] Nitzan and Paroush (1984a, 1984b) allow both alternatives to have different prior probabilities of correctness, and allow the four possible outcomes (correct/incorrect decision in favour of alternative 0/1) to have four different utility values. Later, Ben-Yashar and Nitzan (1997) also allow non-symmetric competence, that is, they allow the case in which a person's probability of making the correct choice given that 0 is the correct choice differs from that given that 1 is the correct choice.

2 A binary choice problem with an arbitrary type of profile

Recall that we are considering a group of n persons, labelled $i = 1, ..., n$, facing any binary decision problem. By assumption, one of the two options, 0 and 1, is the 'correct' alternative and the other option is the 'incorrect' alternative. H_0 (H_1) is the hypothesis that option 0 (1) is correct. Note that H_0 is equivalent to $\sim H_1$, and H_1 to $\sim H_0$.

We allow complete generality regarding the type of profile collected from the group. For each person i, let $\mathcal{X}_i$ be the set of all inputs x_i that person i may submit. This set could be the same for all persons ($\mathcal{X}_1 = \mathcal{X}_2 = ... = \mathcal{X}_n$) or differ across persons (e.g. the 'experts' are asked to provide more informative inputs than the others). For instance, recalling the examples of the introduction, we have

(a) $\mathcal{X}_i = \{0, 1\}$ for simple-vote submission,

(b) $\mathcal{X}_i = \{0, 1, \text{'abstention'}\}$ if abstention is also allowed,

(c) $\mathcal{X}_i = \{1\%, 2\%, 3\%, .., 98\%, 99\%\}$ as an example of submission of a subjective probability of the hypothesis H_1 (that 1 is the correct option),

(d) $\mathcal{X}_i = \{$'I am sure of guilt', 'I am rather sure of guilt', 'I don't know', 'I am rather sure of innocence', 'I am sure of innocence'$\}$ as an example of 'ticking a statement'.

For technical simplicity, we assume that each person's $\mathcal{X}_i$ is a (non-empty) *finite* set, i.e. that each person chooses between only finitely many possible inputs.[7] We impose no restrictions on the submitted profiles, and so the set of allowed profiles is the Cartesian product

$$\mathcal{X} := \mathcal{X}_1 \times ... \times \mathcal{X}_n = \{(x_1, ..., x_n) | x_1 \in \mathcal{X}_1 \& ... \& x_n \in \mathcal{X}_n\} \quad \text{(universal domain)}.$$

Let $\mathcal{F}$ be the set of all decisive procedures defined on the universal domain $\mathcal{X}$, i.e. of all functions $f : \mathcal{X} \mapsto \{0, 1\}$. Of course, $f(x)$ is 0 (1) if for profile x the decision is 0 (1).

3 Likelihoods of profiles

We assume that each profile x has a certain likelihood of occurring given H_0 and a certain likelihood of occurring given H_1 – an assumption we have in common with the above-cited literature. More precisely, for each person i, the input x_i is seen as the realisation of a random variable X_i taking values in $\mathcal{X}_i$, and hence the profile $x = (x_1, ..., x_n)$ is a realisation of the random vector $X := (X_1, ..., X_n)$ taking values in the Cartesian product $\mathcal{X} = \mathcal{X}_1 \times ... \times \mathcal{X}_n$. For each profile $x = (x_1, ..., x_n)$ there is

- a probability of this profile given H_0: $P(x|H_0) = P(x_1, ..., x_n|H_0)$, and
- a probability of this profile given H_1: $P(x|H_1) = P(x_1, ..., x_n|H_1)$.

These probabilities are often called "likelihoods", as opposed to the "(prior) probabilities" $P(H_0)$ and $P(H_1)$ and the "(posterior) probabilities" $P(H_0|x)$ and $P(H_1|x)$.

[7] Our discussion could be extended to infinite sets $\mathcal{X}_i$. In the case of uncountably infinite sets $\mathcal{X}_i$ (e.g. $\mathcal{X}_i = (0, 1)$), one would have to talk not about probabilities of different $x_i \in \mathcal{X}_i$, but of a probability density function defined over $\mathcal{X}_i$, and the likelihood-ratio would be a ratio not of probabilities of x_i, but of densities in x_i.

For instance, in the simple-vote case $\mathcal{X}_i = \{0,1\}$ of the cited literature, the likelihood functions $P(x|H_0)$ and $P(x|H_1)$ can, under independence, be expressed in terms of competence parameters. But, of course, likelihood functions may in principle be specified for *any* type of profile (such as for the types (a)-(d) in Section 2), and moreover independence is not a necessary assumption.

We further suppose that $P(x|H_0)$ and $P(x|H_1)$ are non-zero for all profiles x. So each profile x must have a positive probability of occurring both under H_0 and H_1. This assumption might be too restrictive, and in Appendix A we see that our results essentially remain if the assumption is relaxed.

4 Epistemic Monotonicity

To be able to state the principle of Epistemic Monotonicity independently of prior information, we need a brief preparation. For two profiles x, x', consider this question: Which profile makes the hypothesis H_1 (that 1 is the correct choice) more probable, i.e. which of $P(H_1|x)$ and $P(H_1|x')$ is larger? Let us see why the answer to this question does not depend on the size of the prior probability $P(H_1)$, i.e. on the available information prior to observing any votes. We have to show that the relation $P(H_1|x) \leq P(H_1|x')$ holds for *some* specification of the prior $r = P(H_1) \in (0,1)$ if and only if it holds for *any* specification of $P(H_1)$. By Bayes' Theorem,

$$P(H_1|x) = \frac{rP(x|H_1)}{rP(x|H_1) + (1-r)P(x|H_0)}.$$

By dividing numerator and denominator by $P(x|H_1)$, we obtain

$$P(H_1|x) = \frac{r}{r + (1-r)\{LR(x)\}^{-1}}, \tag{3}$$

where $LR(x)$ denotes the

$$\textit{likelihood-ratio } LR(x) := \frac{P(x|H_1)}{P(x|H_0)} \quad \text{(for all } x \in \mathcal{X}\text{)}. \tag{4}$$

(3) shows that $P(H_1|x)$ is a strictly increasing function of the likelihood-ratio $LR(x)$, regardless of the value of the prior $r = P(H_1)$. In other words, $P(H_1|x) \leq P(H_1|x')$ if and only if $LR(x) \leq LR(x')$, whatever the value r.

Epistemic Monotonicity (EM). *For every two profiles $x, x' \in \mathcal{X}$, if $P(H_1|x) \leq P(H_1|x')$ (for some arbitrary and hence every specification of the prior probability $r = P(H_1) \in (0,1)$), then $f(x) \leq f(x')$.*

Note that the inequality $f(x) \leq f(x')$ means that the pair $(f(x), f(x'))$ is either (0,0), (0,1), or (1,1), but not (1,0). To check whether a procedure f satisfies (EM) one does not need to know the true prior probability of H_1 or any utilities. But one does need to know the likelihood functions $P(x|H_0)$ and $P(x|H_1)$; whether f is epistemically monotonic crucially depends on the likelihood functions.

Why impose (EM)? The main reason is the general social choice theoretic search for weak requirements. Indeed, (EM) seems a minimal condition of epistemic consistency or soundness, which might be acceptable for groups that reject expected-utility maximisation or cannot agree on the values of utilities and/or priors.

5 Which procedures are epistemically monotonic?

It is relatively easy to prove that Epistemic Monotonicity requires that procedures decide according to some threshold for the posterior $P(H_1|x)$, or, equivalently, according to some threshold for the likelihood-ratio $LR(x)$. First, let us clarify our language.

Definition 1 *For every procedure $f \in \mathcal{F}$ and every function $h(x)$ mapping $\mathcal{X} \mapsto \mathbf{R}$, we say that f decides "according to (the) threshold h^* for $h(x)$" (where $h^* \in \mathbf{R}$ is a given number) if*

$$f(x) = \begin{cases} 1 & \text{if } h(x) > h^*, \\ 0 & \text{if } h(x) \leq h^*, \end{cases} \quad \text{for all } x \in \mathcal{X};$$

and we say that f decides "according to some threshold for $h(x)$" if there exists an $h^* \in \mathbf{R}$ such that f decides according to threshold h^* for $h(x)$.

For instance, f may decide according to some threshold for the posterior probability $h(x) := P(H_1|x)$, or according to some threshold for the likelihood-ratio $h(x) := LR(x)$, or, if f is a weighted supermajority rule, according to some threshold for the weighted sum of votes. Note that the value of the threshold h^* is not uniquely defined.[8]

A general characterisation of epistemically monotonic procedures is given by

Theorem 1 *For a procedure $f \in \mathcal{F}$, the following statements are equivalent:*
(i) *Epistemic Monotonicity (EM).*
(ii) *For some specification of the prior probability $r = P(H_1) \in (0,1)$, f decides according to some threshold for $P(H_1|x)$.*
(iii) *For every specification of the prior probability $r = P(H_1) \in (0,1)$, f decides according to some threshold for $P(H_1|x)$.*
(iv) *f decides according to some threshold for $LR(x)$.*

The proof is given in Appendix B. What should be relatively clear from our discussion prior to introducing (EM) is that each of (ii), (iii), (iv) implies (i). Further, (ii) is a weaker statement than (iii), and my only reason for including (ii) is to point out that (ii) is already sufficient for Epistemic Monotonicity.

In summary, an epistemically monotonic procedure can be characterised either by a threshold for the likelihood-ratio $LR(x)$, or by a threshold for the posterior $P(H_1|x)$ once a prior is specified. Just as there are many possible thresholds for $P(H_1|x)$ resp. $LR(x)$, there are many epistemically monotonic procedures. By Theorem 1, each threshold for $P(H_1|x)$ resp. $LR(x)$ yields an epistemically monotonic procedure; conversely, each epistemically monotonic procedure has the form of deciding according

[8] Indeed, for a procedure f that decides according to some threshold for $h(x)$ the corresponding threshold h^* may be any value in the interval $\max\{h(x)|x \in \mathcal{X} \& f(x) = 0\} \leq h^* < \min\{h(x)|x \in \mathcal{X} \& f(x) = 1\}$ (where this interval of equivalent thresholds has positive length since $\mathcal{X}$ is by assumption finite). If h^* is chosen *strictly* between both interval bounds (rather than equal to the lower bound) then $h(x) \neq h^*$ for all $x \in \mathcal{X}$. This shows that, in the definition of "deciding according to some threshold for $h(x)$", it is not essential that we defined the decision as 0 if $h(x)$ equals the threshold.

to some threshold for $P(H_1|x)$ resp. $LR(x)$. In Section 10, I provide an analogous theorem for expected-utility maximisation; the latter condition will this time determine precise thresholds for $P(H_1|x)$ and $LR(x)$, thereby uniquely determining the procedure f.

6 Which procedures are epistemically monotonic under independence?

Theorem 1 characterises epistemically monotonic procedures f in full generality. Now let us make the significant restriction of

Independence (I). *The inputs are independent given the true hypothesis. Explicitly, for all inputs $x_1 \in \mathcal{X}_i, ..., x_n \in \mathcal{X}_n$ and each alternative $a \in \{0,1\}$,*

$$P(x_1, ..., x_n|H_a) = P(x_1|H_a) \times ... \times P(x_n|H_a).$$

This assumption is, in particular, made by Condorcet's jury model and by most of the cited literature on optimal procedures. An advantage of having (I) is that it is sufficient to specify the individual likelihoods $P(x_i|H_a)$ ($i \in \{1,...,n\}$, $a \in \{0,1\}$), because then the joint likelihood $P(x|H_a)$ can be obtained by taking the product of all of the individual likelihoods. With regard to the class of epistemically monotonic procedures, we will show that (I) entails that such procedures must be weighted supermajority rules, in the following sense of weighted supermajority rules, generalised to allow arbitrary types of input.

Definition 2 *A procedure $f \in \mathcal{F}$ is called a "weighted supermajority rule" or just a "weighted rule" if it decides according to some threshold for $w(x) := \sum_{i=1}^{n} w_i(x_i)$, where, for each person $i \in \{1,...,n\}$, $w_i(x_i)$ is some function mapping $\mathcal{X}_i \mapsto \mathbf{R}$. The function $w_i(x_i)$ is called (person i's) "weight function", since it assigns "weights" $w_i(x_i)$ to inputs $x_i \in \mathcal{X}_i$.*

By Theorem 1 (iv), f is epistemically monotonic if and only if f decides according to some threshold for $LR(x)$. By Independence, $LR(x)$ factorises into the product of all individual likelihood-ratios:

$$LR(x) = \frac{P(x|H_1)}{P(x|H_0)} = \frac{P(x_1|H_1) \times ... \times P(x_n|H_1)}{P(x_1|H_0) \times ... \times P(x_n|H_0)} = LR(x_1) \times ... \times LR(x_n),$$

where $LR(x_i)$ is person i's

$$\textit{individual likelihood-ratio } LR(x_i) := \frac{P(x_i|H_1)}{P(x_i|H_0)}, \quad \text{for all } x \in \mathcal{X}.$$

Now, f decides according to threshold h^* for $LR(x)$ if and only if f decides according to threshold $\log h^*$ for

$$\log LR(x) = \log\{LR(x_1) \times ... \times LR(x_n)\} = \sum_{i=1}^{n} \log LR(x_i) = \sum_{i=1}^{n} w_i(x_i),$$

where $w_i(x_i)$ is defined as person i's

$$\textit{individual log-likelihood-ratio } w_i(x_i) := \log LR(x_i) = \log \frac{P(x_i|H_1)}{P(x_i|H_0)}, \text{ for all } x \in \mathcal{X}. \tag{5}$$

This proves

Theorem 2 *Suppose Independence (I). A procedure $f \in \mathcal{F}$ satisfies Epistemic Monotonicity (EM) if and only if f is a weighted rule with weight functions given by the individual log-likelihood-ratios* (5).

Again, note that Theorem 2 leaves the threshold of the weighted rule open. There are many thresholds, and hence many epistemically monotonic procedures. The analogue of Theorem 2 for expected-utility maximisation is provided in Section 10.

7 Example 1: Aggregating simple votes

The simplest type of profile is given when each x_i is simply a vote for either alternative, i.e. when $\mathcal{X}_i = \{0, 1\}$ for all i. In this situation, Theorem 2 resembles results in the literature (e.g. Nitzan and Paroush (1984a), p. 214), apart from the different optimality criterion. Expressed in terms of person i's competence parameters $p_i^0 := P(X_i = 0|H_0)$ and $p_i^1 := P(X_i = 1|H_1)$, the individual likelihood-ratios are given by

$$LR(x_i) = \frac{P(x_i|H_1)}{P(x_i|H_0)} = \begin{cases} \frac{p_i^1}{1-p_i^0}, & \text{if } x_i = 1, \\ \frac{1-p_i^1}{p_i^0}, & \text{if } x_i = 0. \end{cases}$$

So, under Independence, a procedure $f \in \mathcal{F}$ is epistemically monotonic if and only if f is a weighted rule with weight functions given by

$$w_i(x_i) := \log LR(x_i) = \begin{cases} \log \frac{p_i^1}{1-p_i^0}, & \text{if } x_i = 1, \\ \log \frac{1-p_i^1}{p_i^0} = -\log \frac{p_i^0}{1-p_i^1}, & \text{if } x_i = 0. \end{cases}$$

If, for simplicity, we assume symmetric competence $p_i^1 = p_i^0 =: p_i$, $w_i(x_i)$ can be expressed in a condensed way as

$$w_i(x_i) = (-1)^{x_i+1} \log \frac{p_i}{1-p_i}, \quad \text{for all } x_i \in \mathcal{X}_i.$$

In the case that the first three persons vote $x_i = 1$ and all others vote $x_i = 0$, option 1 is chosen according to some threshold for the sum

$$\sum_{i=1}^{n} w_i(x_i) = \sum_{i=1}^{n} (-1)^{x_i+1} \log \frac{p_i}{1-p_i} = \sum_{i=1}^{3} \log \frac{p_i}{1-p_i} - \sum_{i=4}^{n} \log \frac{p_i}{1-p_i}. \tag{6}$$

If all jurors i are 'competent' ($p_i > 1/2$), votes for 1 increase the chances of 1 ($w_i(1) > 0$) and votes for 0 increase the chances of 0 ($w_i(0) < 0$), as expected. The more competent a person i is (higher p_i), the more strongly the vote x_i is weighted (higher absolute value $|w_i(x_i)| = \log \frac{p_i}{1-p_i}$). If person i has competence $p_i = 1/2$, the vote has no influence since $w_i(1) = w_i(0) = \log 1 = 0$. If person i has competence $p_i < 1/2$, the vote is counted negatively ($w_i(1) < 0$ and $w_i(0) > 0$), so that a vote for one alternative increases the chances of the other alternative. This strong discrimination depending on competence reflects the purely epistemic orientation of Epistemic Monotonicity, which neglects all procedural concerns.

8 Example 2: Aggregating subjective probabilities

The reason for choosing the aggregation of subjective probabilities as the second example is that this is perhaps the only example in which epistemically monotonic procedures can be devised without specifying the likelihood functions $P(x|H_1)$ and $P(x|H_0)$. Indeed, we shall see that, under a special assumption, the weight function $w_i(x_i) = LR(x_i)$ equals $x_i/(1-x_i)$ (up to an additive constant) – independently of any parameters coming from the way likelihoods are specified! This contrasts with the case of simple votes, in which the competence specification is crucial for determining the weights.

Since people are submitting subjective probabilities, the set $\mathcal{X}_i$ of person i's allowed inputs x_i is some subset of $(0,1)$;[9] For instance, each $\mathcal{X}_i$ could be equal $\{1\%, 2\%, 3\%, ..., 98\%, 99\%\}$.

Consider the posterior probability of H_1 conditional just on person i's submission x_i. Using Bayes' Theorem and then dividing numerator and denominator by $P(x_i|H_1)$,

$$P(H_1|x_i) = \frac{rP(x_i|H_1)}{rP(x_i|H_1) + (1-r)P(x_i|H_0)} = \frac{r}{r + (1-r)\{LR(x_i)\}^{-1}}.$$

Solving this for the individual likelihood-ratio $LR(x_i)$, we obtain

$$LR(x_i) = \frac{P(H_1|x_i)}{1 - P(H_1|x_i)} \times \frac{1-r}{r}. \quad (7)$$

This expresses $LR(x_i)$ in terms of the prior r and the posterior given x_i. There is a natural way to get hold of this posterior. Suppose that person i's individual probabilities are "well-calibrated" in the sense that the person neither tends to exaggerate, nor to understate probabilities. Then we may assume that the posterior probability of H_1 *given* just the information that person i gives H_1 a probability of x_i is precisely x_i, i.e. $P(H_1|x_i) = x_i$ – a property for which a referee kindly suggested the term "calibration". Formally:[10]

Calibration (C). *For all persons $i \in \{1, ..., n\}$ and inputs $x_i \in \mathcal{X}_i$, $P(H_1|x_i) = x_i$.*

Before discussing Calibration, let us see what it entails. Calibration is a truly convenient assumption, since now the individual likelihood-ratio (7) becomes

$$LR(x_i) = \frac{x_i}{1-x_i} \times \frac{1-r}{r}. \quad (8)$$

[9] For the following reason $\mathcal{X}_i$ cannot contain 0 or 1, i.e. we have to exclude the submission of $x_i = 0$ or $x_i = 1$. If $x_i = 0$ then $P(x_i|H_1) = 0$ by (8), which was excluded earlier – see the end of Section 3, and see Appendix A where this restriction is removed. Similarly, if $x_i = 1$ then $P(x_i|H_0) = 0$ by (8), which was also excluded.

[10] My condition of Calibration (C) is related to, but not identical with the concept of *well-calibrated* probability assignments in the literature (e.g. Dawid (1982)).

By Theorem 2, under Independence, a procedure f is epistemically monotonic if and only if f decides according to some threshold for

$$\begin{aligned}\sum_{i=1}^{n}\log LR(x_i) &= \sum_{i=1}^{n}\log\left(\frac{x_i}{1-x_i}\times\frac{1-r}{r}\right)\\ &= \sum_{i=1}^{n}\left(\log\frac{x_i}{1-x_i}+\log\frac{1-r}{r}\right)\\ &= \sum_{i=1}^{n}\log\frac{x_i}{1-x_i}+n\log\frac{1-r}{r}.\end{aligned}$$

Since $n\log\frac{1-r}{r}$ is just a constant (i.e. does not depend on x), saying that f decides according to some threshold for $\sum_{i=1}^{n}\log LR(x_i)$ is equivalent with saying that f decides according to some threshold for $\sum_{i=1}^{n}\log\frac{x_i}{1-x_i}$ (the new threshold being shifted by the amount $n\log\frac{1-r}{r}$). So, Theorem 2 becomes

Theorem 3 *Suppose Independence (I) and Calibration (C). A procedure $f\in\mathcal{F}$ satisfies Epistemic Monotonicity (EM) if and only if f is a weighted rule with weight functions $w_i(x_i):=\log\frac{x_i}{1-x_i}$ ($i\in\{1,...,n\}$).*

Since these weight functions do not depend on any parameters, one can construct epistemically monotonic procedures aggregating subjective probabilities without even specifying the likelihood functions $P(x_i|H_1)$ and $P(x_i|H_0)$, provided that one accepts the two tough assumptions of Independence and Calibration. This is surprising since (EM) is based on posterior probabilities and hence on likelihoods. Again, the many epistemically monotonic rules are given by the many possible thresholds for $\sum_{i=1}^{n}w_i(x_i)$.

Justification of Calibration (C). While Theorems 1 and 2 are entirely general regarding the informational content profiles, Theorem 3 can only apply to profiles of subjective probabilities, because only then can assumption (C) *possibly* be justified. But what is this justification?

First, there may be good reasons to reject (C). For instance, if person i determines the submitted x_i by tossing a coin or by any other procedure unrelated to the truth of H_1, then x_i provides no evidence about H_1, and hence $P(H_1|x_i)=P(H_1)$ instead of $P(H_1|x_i)=x_i$. Or, if person i is a notorious liar who submits x_is that are the higher, the lower his or her genuine belief about H_1, then x_i provides evidence in the opposite direction: $P(H_1|x_i)$ should be a decreasing function of x_i, not $P(H_1|x_i)=x_i$. Finally and more realistically, person i might be sincere but unable to form rational beliefs; for instance, if person i notoriously underestimates the probability of H_1, then $P(H_1|x_i)>x_i$ instead of $P(H_1|x_i)=x_i$; and if person i notoriously overestimates the probability of H_1, then $P(H_1|x_i)<x_i$.

However, in groups where people submit individual probabilities of high quality, (C) could be defended along the following lines. Roughly, the idea is that there is, on the one hand, *shared* knowledge which all group members have in common, and, on the other hand, *private* knowledge held by individual group members. The probability function $P(.)$ represents the shared knowledge of the group, and so the prior probability $P(H_1)$ reflects only shared knowledge. Person i has more knowledge than the shared knowledge, since he or she also has private knowledge, and so person i's probability of H_1, namely x_i, is built on more knowledge than the prior $P(H_1)$. As a consequence, someone who starts with the shared knowledge and then learns that person i (who knows more) assigns a probability of x_i to H_1 should revise his or her probability of H_1 from $P(H_1)$ to x_i; formally, $P(H_1|x_i) = x_i$. More precisely, the relation $P(H_1|x_i) = x_i$ is based on the assumption that person i is a rational Bayesian agent[11] and is sincere in his or her submission, which guarantees that the submitted probability x_i of H_1 was derived by person i by applying Bayes' rule in the light of private information. Under this interpretation of x_i, it is indeed true that the probability of H_1 given the submission x_i is $P(H_1|x_i) = x_i$.[12] Deviation from this would mean that it is wrong *either* that person i is a rational Bayesian agent[11] (i.e. has obtained x_i by Bayesian updating), *or* that person i represents the shared knowledge (his or her basis for updating) by the probability function $P(.)$, *or* that person i is sincere (i.e. tells his or her true probability of H_1).

Note the analogy between Calibration and van Fraassen's Reflection Principle.[13]

9 Example 3: Mixing simple votes and subjective probabilities

By combining Examples 1 and 2, we can derive epistemically monotonic procedures for the case that some persons submit simple votes and others (the 'experts') submit their subjective probabilities of H_1. Denoting by $\mathcal{I}_{\mathrm{v}}$ the set of persons i submitting simple votes ($\mathcal{X}_i = \{0, 1\}$), and by $\mathcal{I}_{\mathrm{p}}$ the set of persons i submitting subjective probabilities ($\mathcal{X}_i \subset (0, 1)$), we define weight functions by

$$w_i(x_i) := \begin{cases} (-1)^{x_i+1} \log \frac{p_i}{1-p_i}, & \text{if } i \in \mathcal{I}_{\mathrm{v}}, \\ \log \frac{x_i}{1-x_i}, & \text{if } i \in \mathcal{I}_{\mathrm{p}}. \end{cases} \tag{9}$$

[11] A rational Bayesian agent updates his or her beliefs in light of new information according to Bayes' rule.

[12] More precisely, assume that person i's private information is the observation of the value s_i taken by a random variable S_i (for instance, person i might have observed that the random variable "weather on the day of the crime" took the value "sunshine"). Then person i, as a rational Bayesian agent[11], assigns to H_1 a (posterior) probability of $x_i = P(H_1|S_i = s_i)$. Under this assumption about person i's belief formation, (C) actually follows as a theorem. Let me sketch the simple proof. For each value s_i of S_i, we denote by $d(s_i) = P(H_1|S_i = s_i)$ the corresponding value of x_i. So $X_i = d(S_i)$, i.e. the vote is a function of the signal S_i. If the function $d(.)$ is one-to-one, conditionalising on the value x_i of X_i is equivalent to conditionalising on the corresponding value $s_i := d^{-1}(x_i)$ of S_i, which implies that $P(H_1|X_i = x_i) = P(H_1|S_i = s_i) = x_i$. These relations, in fact, hold even if $d(.)$ is many-to-one, with the only difference being that now s_i is one of possibly many inverse images of x_i under $d(.)$.

[13] While Calibration compares the shared knowledge of the group with the (larger) knowledge of person i, the Reflection Principle compares the knowledge of a person at a time t with the (larger) knowledge of the same person at a later time t'. The Reflection Principle states that, if the person at time t (hypothetically) learns that at time t' he or she will assign a probability of p to an event, then the person should revise his or her present probability of that event to p. In short, if you are told that in a month time you will believe that party X will win the election, then you should now believe that party X will win the election. The justification is, of course, that future beliefs are based on more knowledge. A more remote analogy may be drawn from Calibration (C) to Lewis' Principal Principle (1980).

Here, for simplicity, we have again assumed symmetric competence $p_i := P(X_i = 1|H_1) = P(X_i = 0|H_0)$ for persons $i \in \mathcal{I}_\text{v}$. Supposing Independence, as well an axiom analogous to Calibration but restricted to persons $i \in \mathcal{I}_\text{p}$, a procedure is epistemically monotonic if and only if it is a weighted rule with weight functions given by (9). In other words, an epistemically monotonic procedure decides according to some threshold for

$$\sum_{i \in \mathcal{I}_\text{v}} (-1)^{x_i+1} \log \frac{p_i}{1-p_i} + \sum_{i \in \mathcal{I}_\text{p}} \log \frac{x_i}{1-x_i}.$$

The latter expression highlights that it is easier to incorporate a submitted degree of belief x_i into the decision than a submitted vote x_i, because the term $\log \frac{x_i}{1-x_i}$ does not depend on parameters, while the term $(-1)^{x_i+1} \log \frac{p_i}{1-p_i}$ depends on the competence p_i.

10 The analogous results based on expected-utility maximisation

It is worthwhile to restate our theorems for the case in which the criterion is expected-utility maximisation instead of Epistemic Monotonicity, thereby connecting to the literature.[14] This leads to precise thresholds. To indicate the analogies to our earlier theorems, let us use the same labels as before, appended with a star "*".

Unlike before, assume now that there are fixed prior probabilities of correctness and utilities of outcomes. More precisely, there is a prior probability $r := P(H_1) \in (0,1)$ of 1 being the correct alternative, and there is a utility $u_{ya} := u(y,a)$ assigned to each of the four possible outcomes $(y,a) \in \{0,1\}^2$, where $y \in \{0,1\}$ stands for the collective decision and $a \in \{0,1\}$ for the correct alternative. The values u_{11} and u_{00} are the utilities of choosing 1 when 1 is correct resp. 0 when 0 is correct. The values u_{10} and u_{01} are the utilities of the two types of incorrect decisions. We make the reasonable assumption that $u_{00} > u_{10}$ and $u_{11} > u_{01}$, reflecting that correct decisions are always better than incorrect decisions.

Under procedure $f \in \mathcal{F}$, utility is a random variable $U_f := u(f(X), A)$, which depends on the random profile X and on the random correct alternative A. The expected utility can be calculated by summing the utilities of all four outcomes $(y,a) \in \{0,1\}^2$ weighted by their probabilities:

$$E(U_f) = \sum_{(y,a) \in \{0,1\}^2} u_{ya} P(f(X) = y \& H_a) = \sum_{(y,a) \in \{0,1\}^2} u_{ya} P(H_a) P(f(X) = y|H_a).$$

Since the event $f(X) = y$ is the event that the random profile X belongs to the set of profiles $x \in \mathcal{X}$ for which $f(x) = y$, the probability $P(f(X) = y|H_a)$ can be written as the sum of probabilities $\sum_{x \in \mathcal{X} \& f(x)=y} P(x|H_a)$. So,

$$\begin{aligned} E(U_f) &= \sum_{(y,a) \in \{0,1\}^2} u_{ya} P(H_a) \sum_{x \in \mathcal{X} \& f(x)=y} P(x|H_a) \\ &= r \sum_{y \in \{0,1\}} u_{y1} \sum_{x \in \mathcal{X} \& f(x)=y} P(x|H_1) + (1-r) \sum_{y \in \{0,1\}} u_{y0} \sum_{x \in \mathcal{X} \& f(x)=y} P(x|H_0), \end{aligned}$$

[14] See Ben-Yashar and Nitzan (1997, Theorem 3.1) for the most general result, although resticted to the simple-vote case.

which illustrates how the expected utility $E(U_f)$ depends on the three components: the prior $r = P(H_1)$, utilities u_{ya}, and likelihoods $P(x|H_0)$ and $P(x|H_1)$.

In principle, there can be more than one procedure f with maximal expected utility $E(U_f)$. To achieve uniqueness (and stick to decisive procedures), we will impose a requirement to the effect that for those profiles x where the decision $f(x)$ does not affect the expected utility $E(U_f)$, this decision $f(x)$ will always be alternative 0. Let us say that a procedure $f \in \mathcal{F}$ is "more favourable to option 0" than another procedure $f_1 \in \mathcal{F} \backslash \{f\}$ in case $f(x) \leq f_1(x)$ for all $x \in \mathcal{X}$. In other words, f is more favourable to 0 than f_1 in case f decides for 0 whenever f_1 does (but might decide for 0 when f_1 decides for 1). We impose

Expected-Utility Maximisation (UM). *f has maximal expected utility (i.e. $E(U_f) = \sup_{f' \in \mathcal{F}} E(U_{f'})$) and is more favourable to option 0 than any other procedure in $\mathcal{F}$ with maximal expected utility.*

This axiom requires that among all procedures maximising the expected utility, we pick the one most favourable to 0.[15] Often there is just one procedure with maximal $E(U_f)$, in which case the clause "and is more favourable..." is superfluous.

In analogy to Theorem 1, we have

Theorem 1* *For a procedure $f \in \mathcal{F}$, the following statements are equivalent:*
(i*) *Expected-Utility Maximisation (UM).*
(ii*) *f decides according to the threshold $\frac{1}{1+\frac{u_{11}-u_{01}}{u_{00}-u_{10}}}$ for $P(H_1|x)$.*
(iv*) *f decides according to the threshold $\frac{u_{00}-u_{10}}{u_{11}-u_{01}} \times \frac{1-r}{r}$ for $LR(x)$.*

Note that in (iv*) the threshold for $LR(x)$ is a decreasing function of the prior $r = P(H_1)$. So, the higher the prior r, the easier for $LR(x)$ to exceed the threshold, and hence the more likely a decision in favour of 1, which is intuitively plausible. The threshold in (ii*) does not depend on the prior r, intuitively because the prior is already contained in $P(H_1|x)$.

Comparing Theorems 1 and 1*, it is immediately clear that (ii*) implies (ii) and that (iv*) implies (iv). So, (i*) implies (i), i.e. we have

Corollary 1 *If $f \in \mathcal{F}$ satisfies Expected-Utility Maximisation (UM), then f satisfies Epistemic Monotonicity (EM).*

Now assume Independence. In exactly the same way as Theorem 1 implies Theorem 2, Theorem 1* implies

Theorem 2* *Suppose Independence (I). A procedure $f \in \mathcal{F}$ satisfies Expected-Utility Maximisation (UM) if and only if f is the weighted rule with weight functions given by the individual log-likelihood-ratios* (5) *and threshold* $\log \frac{u_{00}-u_{10}}{u_{11}-u_{01}} - \log \frac{r}{1-r}$.

[15] If instead of (UM) one prefers to impose that the procedure be maximally favourable to *alternative 1* among all procedures f with maximal $E(U_f)$, then one should impose (UM) after swapping the names of both alternatives. Or, alternatively, one should redefine "deciding according to threshold h^* for $h(x)$" (Definition 1) such that the decision is 1 (not 0) if $h(x) = h^*$; the results of this section would then stay true for this new definition and the modified (UM).

Finally, regarding the aggregation of subjective probabilities, the analogue of Theorem 3 is

Theorem 3* *Suppose Independence (I) and Calibration (C). A procedure $f \in \mathcal{F}$ satisfies Expected-Utility Maximisation (UM) if and only if f is the weighted rule with weight functions $w_i(x_i) := \log \frac{x_i}{1-x_i}$ ($i \in \{1, ..., n\}$) and threshold $\log \frac{u_{00}-u_{10}}{u_{11}-u_{01}} + (n-1) \log \frac{r}{1-r}$.*

Note that here the threshold is an increasing function of the prior $r = P(H_1)$, unlike in Theorems 1*(iv*) and 2*. At first sight this is surprising since more prior support for H_1 should make it easier to decide for option 1 – so how can there be a higher threshold? Intuitively, the explanation is that higher prior support for H_1 increases the chances of 1 *despite* the higher threshold, because more prior support for H_1 leads to larger subjective beliefs x_i of H_1 (see footnote 12). On the other hand, if we hold the submissions x_i fixed and increase the prior $r = P(H_1)$, then the decreasing chances of 1 (the increasing threshold) have the following intuitive justification. The more probable H_1 is *a priori* (higher r), the smaller the submitted beliefs x_i of H_1 are compared to r, and hence the more it must be the case that people's private information (which is not part of the prior information) supports H_0. But if people's private information gives more support for H_0, it becomes more appropriate to choose option 0, which justifies the higher threshold.

11 Conclusion

We have proven that, from an epistemic perspective, procedures should decide according to some threshold for the likelihood-ratio $LR(x)$, whatever the type of profile x (Theorem 1). In the special case of Independence, this is equivalent to deciding according to some threshold for the sum of personal log-likelihood-ratios $\sum_{i=1}^{n} \log LR(x_i)$, which we call a *weighted (supermajority) rule* with *weight functions* $w_i(x_i) := \log LR(x_i)$. Perhaps the most interesting application is the aggregation of subjective probabilities. Here, under Calibration, epistemically monotonic procedures decide according to some threshold for $\sum_{i=1}^{n} \log \frac{x_i}{1-x_i}$, which does not involve any unknown parameters such as competence (Theorem 3).

The entire analysis of optimal procedures may be done either based on Epistemic Monotonicity (see Sections 2-9), or based on Expected-Utility Maximisation (Section 10). Both approaches yield the same optimal procedures, with the only difference that Epistemic Monotonicity leaves the threshold open, while Expected-Utility Maximisation provides precise thresholds as a function of the prior probability of H_1 and utilities. This is highlighted by the analogy between Theorems 1, 2, 3 and Theorems 1*, 2*, 3*.

It should be emphasised that the assumption of Independence is quite radical and limits the practical use of Theorems 2 and 3. Independence requires that the individually perceived evidences that caused the persons' inputs be probabilistically independent (and hence in particular have no overlap), which involves more than just causal independence between voters. The more realistic case, probabilistic dependence, is covered by Theorem 1, but here the problem is that the n-dimensional likelihoods are hard to specify in practice.

One might generally object that Epistemic Monotonicity does not yield a well-defined procedure and that, in order to choose a specific threshold, at least some sort of utility and prior-probability considerations are inevitable. This is a fair point, to which one might give two answers. First, what our analysis shows is that in situations where priors and utilities are unavailable or controversial, the procedure choice does not become arbitrary from an epistemic perspective but that Epistemic Monotonicity requires a very particular type of procedure. The threshold might then be chosen intuitively, in the worst case arbitrarily. Second, a person or group that rejects the principle of Expected-Utility Maximisation might still accept Epistemic Monotonicity as a minimal requirement of epistemic consistency. Since, as we have shown, Epistemic Monotonicity leads to essentially the same procedure as Expected-Utility Maximisation, apart from the open threshold, this person or group might want to rethink the rejection of Expected-Utility Maximisation. In this sense, there is no conflict between Epistemic Monotonicity and Expected-Utility Maximisation; rather, the former makes a good case for the latter.

A Allowing zero likelihoods of profiles

The assumption that $P(x|H_1) \neq 0$ and $P(x|H_0) \neq 0$ for all profiles x (see Section 3) might be too restrictive. Indeed, perhaps the occurrence of certain profiles x is impossible under H_1 or under H_0. As an example in which one might prefer to let $P(x|H_0) = 0$, consider a convict-or-acquit problem, and assume person i submits the input $x_i =$ 'I saw the defendant commit the murder'. Suppose that person i is surely not a liar and that his or her memory and eye sight are infallible. It is impossible that person i submits x_i given innocence (H_0), and so a profile x containing this x_i should have probability zero given innocence: $P(x|H_0) = 0$. A (less realistic) example arises if simple vote are aggregated (see Section 7) and some person i has competence $p_i = 1$: Then, profiles x in which person i's vote x_i is incorrect have probability 0. For a more realistic example, consider the case of aggregating subjective probabilities of H_1 and assume Calibration (see Section 8). Here, profiles can have zero likelihood if for some person i the set $\mathcal{X}_i$ contains 0 and/or 1; see footnote 9. This is why we had to assume $0, 1 \notin \mathcal{X}_i$ in Section 8. But this restriction might be undesirable, because it prevents people from submitting the information that they are certain of the correctness of one of the alternatives.

We now allow $P(x|H_0)$ and $P(x|H_1)$ to be zero. The first remark is that if *both* $P(x|H_1)$ and $P(x|H_0)$ are zero, the unconditional probability of x is zero, too:

$$P(x) = P(x|H_1)P(H_1) + P(x|H_0)P(H_0) = 0. \quad (10)$$

But then the posterior probability $P(H_1|x)$ is undefined[16], which is a problem because (EM) is based on posterior probabilities. To avoid this problem, we have to consider those (decisive) procedures that are defined on a *subset* of the universal domain $\mathcal{X}_1 \times ... \times \mathcal{X}_n$, namely on the set of those profiles x for which this problem does not occur. So, we redefine the domain of profiles $\mathcal{X}$ as

$$\mathcal{X} := \{x \in \mathcal{X}_1 \times ... \times \mathcal{X}_n | P(x|H_1) \neq 0 \text{ or } P(x|H_0) \neq 0\} \quad \text{(restricted domain).}$$

[16] A conditional probability $P(A|B)$ is, per definition, the ratio $P(A\&B)/P(B)$, which is well-defined only if $P(B) \neq 0$.

A procedure f defined on $\mathcal{X}$ does not yield a decision when $x \notin \mathcal{X}$. This restriction does not seem strong, because by (10) profiles $x \notin \mathcal{X}$ occur with probability zero. We redefine the set $\mathcal{F}$ as the set of (decisive) procedures with (non-universal) domain $\mathcal{X}$, i.e. as the set of functions $f : \mathcal{X} \mapsto \{0, 1\}$.

As an example, in the aggregation of subjective probabilities we may now put $\mathcal{X}_i := \{0\%, 1\%, ..., 99\%, 100\%\}$, thus allowing the submission of $x_i = 0\%$ (certainty of H_0) or of $x_i = 100\%$ (certainty of H_1). Note, however, that the restricted definition of $\mathcal{X}$ prevents profiles $x = (x_1, ..., x_n) \in \mathcal{X}$ from containing *both* 0%s *and* 100%s, because for such profiles $P(x|H_0) = P(x|H_1) = 0$, under Independence and Calibration.[17] Again, this restriction does not seem strong, since it should be impossible that among rational agents some are certain of H_0 and others are certain of H_1.

All of our axioms ((EM), (I), (C) and (UM)), definitions (Definitions 1 and 2) and results (Theorems 1, 2, 3, 1*, 2*, 3*, Proposition **??** and Corollary 1) remain true in the present context. None of the axioms or results has to be reformulated, and the results can be proven analogously. But the definitions and the meaning of certain operations have to be extended in a sensible way so as to apply to new special cases. These necessary generalisations are now discussed:

- We put $c/0 := \infty$ and $c/\infty := 0$, for all $c > 0$. This applies to the likelihood-ratio $LR(x) = P(x|H_1)/P(x|H_0)$ when $P(x|H_0) = 0$. Also, $\log 0 := -\infty$ and $\log \infty = \infty$, which is needed when computing $\log LR(x)$.

- Definition 1 of deciding *according to some threshold for* $h(x)$ has to be extended to accommodate the case in which $h(x)$ can take infinite values (needed when $h(x) = LR(x)$ or when $h(x) = \log LR(x)$). We also need to allow the threshold h^* for $h(x)$ to take infinite values $\pm\infty$ and even to take the 'value' $-\infty - 1$. Here, '$-\infty - 1$' denotes an artificial number smaller than $-\infty$, which is needed for the (obviously epistemically monotonic) procedure that *always* chooses 1, even when $LR(x) = 0$, i.e. when $h(x) = \log LR(x) = -\infty$; here, the threshold of $h^* = -\infty - 1$ is needed since $h^* = -\infty$ would lead to choosing 0 when $h(x) = -\infty$. Thus, the new definition formally states: *For any procedure $f \in \mathcal{F}$ and any function $h(x)$ mapping $\mathcal{X} \mapsto R \cup \{-\infty, +\infty\}$, f decides "according to some threshold for $h(x)$" if there exists an $h^* \in R \cup \{-\infty - 1, -\infty, +\infty\}$ ('threshold') such that*

$$f(x) = \begin{cases} 1 & \text{if } h(x) > h^*, \\ 0 & \text{if } h(x) \leq h^*, \end{cases} \quad \text{for all } x \in \mathcal{X}.$$

- Definition 2 of a *weighted rule* also has to be restated so as to allow the weight functions $w_i(x_i)$ to take on infinite values. Since $(+\infty) + (-\infty)$ is undefined, it has to be excluded that for some $x \in \mathcal{X}$ among $w_1(x_1), ..., w_n(x_n)$ there can be infinite values of different signs: *A procedure $f \in \mathcal{F}$ is called a "weighted supermajority rule" or just a "weighted rule" if it decides according to some threshold for $w(x) := \sum_{i=1}^{n} w_i(x_i)$, where for all $i \in \{1, ..., n\}$ $w_i(x_i)$ is some function mapping $\mathcal{X}_i \mapsto \mathbf{R} \cup \{-\infty, +\infty\}$, such that for no $x \in \mathcal{X}$ both $-\infty$ and $+\infty$ occur among $w_1(x_1), ..., w_n(x_n)$.* Crucially, the latter condition imposed on weight functions is satisfied by the weight functions

[17] To see why, we use (8). If there is a 0 among $x_1, ..., x_n$, by (8) there is a 0 among $P(x_1|H_1), ..., P(x_n|H_1)$, implying that $P(x|H_1) = P(x_1|H_1) \times ... \times P(x_n|H_1) = 0$. Similarly, if there is a 1 among $x_1, ..., x_n$, by (8) there is a 0 among $P(x_1|H_0), ..., P(x_n|H_0)$, implying that $P(x|H_0) = P(x_1|H_0) \times ... \times P(x_n|H_0) = 0$.

$w_i(x_i)$ of Theorem 2, which are defined as $\log LR(x_i)$ (and should be defined arbitrarily if $LR(x_i)$ is undefined, i.e. if $P(x_i|H_0) = P(x_i|H_1) = 0$).[18] Analogously, the condition on weight functions is satisfied by the weight functions of Theorem 3, defined as $w_i(x_i) := \log \frac{x_i}{1-x_i}$, because among $x_1, ..., x_n$ there cannot be both 0s and 1s, as noted earlier.

B Proofs

Theorem 1. We prove the equivalences by showing the implications (i)→(iii), (iii)→(ii), (ii)→(iv), and (iv)→(i).

(iii)→(ii). Trivial.

(iv)→(i). Assume (iv), i.e. f decides according to some threshold h^* for $LR(x)$. Let $x, x' \in \mathcal{X}$ be any profiles such that $P(H_1|x) \leq P(H_1|x')$. We have to show that $f(x) \leq f(x')$, i.e. that the pair $(f(x), f(x'))$ is not the pair $(1, 0)$. We assume $f(x') = 0$ and have to show that $f(x) = 0$. By $f(x') = 0$ and (iv), $LR(x') \leq h^*$. By the discussion before introducing (EM), we have $LR(x) \leq LR(x')$, which implies $LR(x) \leq h^*$. By (iv) it follows that $f(x) = 0$.

(ii)→(iv). Consider a prior $r = P(H_1) \in (0, 1)$ such that f decides according to some threshold h^* for $P(H_1|x)$. By (3),

$$P(H_1|x) = \frac{r}{r + (1-r)\{LR(x)\}^{-1}},$$

which can be solved with respect to $LR(x)$ to give

$$LR(x) = \frac{(1-r)}{r(\{P(H_1|x)\}^{-1} - 1)}.$$

Since $LR(x)$ is an increasing function of $P(H_1|x)$, we have

$$P(H_1|x) > h^* \qquad \text{if and only if} \qquad LR(x) > \frac{(1-r)}{r(\{h^*\}^{-1} - 1)} =: h^{**}.$$

So, f decides according to the threshold h^{**} for $LR(x)$.

(i)→(iii). Assume (EM). If $f(x) = 1$ for all $x \in \mathcal{X}$, (iii) is trivially satisfied. Now assume that $f(x) = 0$ for some $x \in \mathcal{X}$. Consider any specification of the prior probability $r = P(H_1) \in (0, 1)$, and let us show that f decides according to the threshold

$$h^* := \sup\{P(H_1|x) | x \in \mathcal{X} \text{ and } f(x) = 0\}$$

for $P(H_1|x)$. By definition of h^*, it is trivial that $f(x) = 1$ whenever $P(H_1|x) > h^*$. Now assume an $x \in \mathcal{X}$ such that $P(H_1|x) \leq h^*$, and let us show that $f(x) = 0$. Since each set $\mathcal{X}_i$ is assumed finite, $\mathcal{X}$ is also finite, and so h^* is the supremum over a finite set. Hence $h^* = P(H_1|x')$ for some $x' \in \mathcal{X}$ such that $f(x') = 0$. Since

[18] The reason is that if $\log LR(x_i) = \infty$ and $\log LR(x_j) = -\infty$, then $P(x_i|H_0) = 0$ and $P(x_i|H_1) = 0$ and hence $P(x|H_0) = 0$ and $P(x|H_1) = 0$, implying that $x \notin \mathcal{X}$.

$P(H_1|x) \leq P(H_1|x')$, by (EM) $f(x) \leq f(x')$. So, since $f(x') = 0$, we have $f(x) = 0$. **QED**.

Theorem 1*. We first prove the equivalence of (i*) and (ii*). Since

$$E(U_f) = \sum_{x \in \mathcal{X}} P(x)\{u_{f(x),0} P(H_0|x) + u_{f(x),1} P(H_1|x)\},$$

f satisfies (UM) if and only if for all $x \in \mathcal{X}$ the decision $y = f(x)$ maximises the term in curly brackets $u_y := u_{y0}P(H_0|x) + u_{y1}P(H_1|x)$ and is 0 if both decisions y maximise it, i.e. if $u_1 = u_0$; in other words, the decision is $y = 1$ if and only if $u_1 > u_0$. The latter holds if and only if

$$u_{10}P(H_0|x) + u_{11}P(H_1|x) > u_{00}P(H_0|x) + u_{01}P(H_1|x),$$

i.e., by $P(H_0|x) = 1 - P(H_1|x)$, if and only if

$$P(H_1|x) > \frac{u_{00} - u_{10}}{u_{00} - u_{10} + u_{11} - u_{01}} = \frac{1}{1 + (u_{11} - u_{01})/(u_{00} - u_{10})}. \tag{11}$$

This proves the equivalence of (i*) and (ii*). To prove the equivalence with (iv*), one needs only to write $P(H_1|x)$ as $\frac{r}{r+(1-r)\{LR(x)\}^{-1}}$ and then solve inequality (11) for $LR(x)$. **QED**.

Theorem 3*. By Theorem 2*, f satisfies (UM) if and only if f decides according to the threshold $\log \frac{u_{00}-u_{10}}{u_{11}-u_{01}} - \log \frac{r}{1-r}$ for $\sum_{i=1}^n \log LR(x_i)$. By (8),

$$\log LR(x_i) = \log\left(\frac{x_i}{1-x_i} \times \frac{1-r}{r}\right) = \log \frac{x_i}{1-x_i} + \log \frac{1-r}{r}.$$

So f satisfies (UM) if and only f decides according to the threshold $\log \frac{u_{00}-u_{10}}{u_{11}-u_{01}} - \log \frac{r}{1-r}$ for

$$\sum_{i=1}^n \log LR(x_i) = \sum_{i=1}^n \log \frac{x_i}{1-x_i} + n \log \frac{1-r}{r} = \sum_{i=1}^n \log \frac{x_i}{1-x_i} - n \log \frac{r}{1-r},$$

which is equivalent with deciding according to the threshold $\log \frac{u_{00}-u_{10}}{u_{11}-u_{01}} + (n-1) \log \frac{r}{1-r}$ for $\sum_{i=1}^n \log \frac{x_i}{1-x_i}$. **QED**.

References

Ben-Yashar, R and S Nitzan (1997). The optimal decision rule for fixed-size committees in dichotomous choice situations: The general result. Int Econ Rev 38, p. 175-186.

Berend, D and JE Harmse (1993). Expert rule versus majority rule under partial information. Theory and Decision 35, p. 179-197.

Berend, D and L Sapir (2002). Expert rule versus majority rule under partial information, II. Journal of Applied Mathematics and Decision Sciences 6 (2), p. 79-99.

Berg, S (1993). Condorcet's jury theorem, dependence among voters. Social Choice and Welfare 10, p. 87-95.

Berry, S (1994). Evaluation of some weighted majority decision rules under dependent voting. Mathematical Social Sciences 28, p. 71-83.

Boland, P, F. Proschan, and Y. Tong (1989). Modelling dependence in simple and indirect majority systems. Journal of Applied Probability 26, p. 81-88.

Cohen, J (1986). An epistemic conception of democracy. Ethics 97 (1), p. 26-38.

Coleman, J and J Ferejohn (1986). Democracy and social choice. Ethics 97 (1), p. 6-25.

Dahl, RA (1979). Procedural democracy. In: Laslett, P and J Fishkin (eds) Ethics, Politics and Society, 5th series, Oxford (Blackwell) p. 97-133.

Dawid, PA (1982). The Well-Calibrated Bayesian, Journal of the American Statistical Association, p. 605-610.

Estlund, D (1993). Making truth safe for democracy. In: Copp, DJ Hampton and JE Roemer (eds). The idea of democracy, New York, Cambridge University Press, p. 71-100.

Estlund, D (1997). Beyond fairness and deliberation: The epistemic dimension of democracy. In: Bohman, J and W Rehg (eds). Deliberative democracy: essays on reason and politics, Cambridge, MA (MIT Press) p. 173-204.

Gradstein, M and S Nitzan (1986). Performance evaluation of some special classes of weighted majority rules. Mathematical Social Sciences 12, p. 31-46.

Grofman, B, G Owen and S Feld (1983). Thirteen theorems in search of the truth. Theory and Decision 15, p. 261-278.

Ladha, K (1993). Condorcet's jury theorem in light of Finetti's theorem: majority-rule voting with correlated votes. Social Choice and Welfare 10, p. 69-86.

Ladha, K (1995). Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes. Journal of Economic Behavior and Organization 26, p. 353-372.

Lewis, D (1980). A subjectivist's guide to objective chance. In: Jeffrey, RC (ed.) (1980). Studies in inductive logic and probability, vol. 2, Berkeley, Berkeley University Press.

List, C and RE Goodin (2001). Epistemic democracy: Generalizing the Condorcet Jury Theorem. Journal of Political Philosophy 9, p. 277-306.

Nitzan, S and J Paroush (1982). Optimal decision rules in uncertain dichotomous choice situations. International Economic Review 23, p. 289-297.

Nitzan, S and J Paroush (1984a). A general theorem and eight corollaries in search of a correct decision. Theory and Decision 17, p. 211-220.

Nitzan, S and J Paroush (1984b). Are qualified majority rules special? Public Choice 42, p. 257-272.

Nitzan, S and J Paroush (1984c). Partial information on decisional competencies and the desirability of the expert rule in uncertain dichotomous choice situations. Theory and Decision 17, p. 275-286.

Sapir, L (1998). The optimality of the expert and majority rules under exponentially distributed competence. Theory and Decision 45(1), p. 19-35.

Shapley, L and B Grofman (1984). Optimizing group judgmental accuracy in the presence of interdependencies. Public Choice 43, p. 329-343.

van Fraassen, BC (1984). Belief and the will. Journal of Philosophy 81, p. 235-256.