CORE
Metadata, citation and similar papers at core.ac.uk
Provided by SAS-SPACE

Reliable Methods of Judgment Aggregation

Stephan Hartmann[†], Gabriella Pigozzi[‡], Jan Sprenger[§]

Abstract

The aggregation of consistent individual judgments on logically interconnected propositions into a collective judgment on the same propositions has recently drawn much attention. Seemingly reasonable aggregation procedures, such as propositionwise majority voting, cannot ensure an equally consistent collective conclusion. The literature on judgment aggregation refers to such a problem as the *discursive dilemma*. In this paper we assume that the decision which the group is trying to reach is factually right or wrong. Hence, the question we address in this paper is how good the various approaches are at selecting the right conclusion. We focus on two approaches: distance-based procedures and Bayesian analysis. Under the former we also subsume the conclusion- and premise-based procedures discussed in the literature. Whereas we believe the Bayesian analysis to be theoretically optimal, the distance-based approaches have more parsimonious presuppositions and are therefore easier to apply.

1 Introduction

Members of a group often have to express their opinions on several propositions. Examples are expert panels, legal courts, boards, and councils. Once the members have stated their views on the issues in the agenda, the individual judgments need to be combined to form a collective decision. The aggregation of individual consistent judgments on logically interconnected propositions into an equally consistent group judgment on the same propositions has recently drawn much attention. The difficulty lies in the fact that there is no general agreement upon which procedure to use.

In this paper we evaluate and compare the methods proposed so far in the literature with a Bayesian approach to judgment aggregation. The assessment criterion we employ is reliability. We assume that the resulting collective judgment is factually right or wrong, and we compare the procedures in terms of how reliable they are at selecting the right social decision. In particular, we present results concerning the reliability of several methods to aggregate conflicting individual judgments into a consistent group conclusion. The first group of methods are *distance-based procedures*, among them the majority fusion operator [11]. Fusion originates from computer science, where the problem of combining information from equally reliable sources arises in several contexts [5]. The second method is a full *Bayesian analysis* of the underlying decision problem.

The combination of finite sets of logically interconnected propositions has been recently investigated in the emerging field of *judgment aggregation* [9]. A *judg-*

[†]Contact information: Stephan Hartmann, Tilburg Center for Logic and Philosophy of Science, Tilburg University, NL-5037 AB Tilburg, S.Hartmann@uvt.nl
[‡]Contact information: Gabriella Pigozzi, Computer Science and Communications Department, University of Luxembourg, L-1358 Luxembourg
[§]Contact information: Jan Sprenger, Tilburg Center for Logic and Philosophy of Science, Tilburg University, NL-5037 AB Tilburg, J.Sprenger@uvt.nl.

	A	B	C
Judges 1, 2, 3	Yes	Yes	Yes
Judges 4, 5	Yes	No	No
Judges 6, 7	No	Yes	No
Majority	Yes	Yes	No

Table 1: An illustration of the discursive dilemma

ment is an assignment of yes/no to a proposition. The problem is that a seemingly reasonable aggregation procedure, such as propositionwise majority voting, cannot ensure a consistent collective conclusion.

Here is an illustration. A court has to make a decision on whether a person is liable of breaching a contract (represented by a proposition C, also referred to as the conclusion). The judges have to reach a verdict following the legal doctrine. This states that a person is liable if and only if she did an action X (represented by proposition A, also referred to as the first premise) and had contractual obligation not to do X (represented by proposition B, also referred to as the second premise). The legal doctrine can be formally expressed as the formula $(A \wedge B) \leftrightarrow C$. Each member of the court expresses her judgment on A, B and C such that the rule $(A \wedge B) \leftrightarrow C$ is satisfied. Suppose now that the court has seven members making their judgments according to Table 1. We see that, although each judge expresses a consistent opinion, propositionwise majority voting (consisting in the separate aggregation of the votes for each of the propositions A, B and C via the majority rule) results in a majority for A and a majority for B, but in a majority for $\neg C$. This is clearly an inconsistent collective result as it violates the rule $(A \wedge B) \leftrightarrow C$. The paradox (called the *discursive dilemma*) rests with the fact that majority voting can lead a group of rational individuals to endorse an irrational collective judgment. Clearly, the relevance of such aggregation problems goes beyond the specific court example and it applies to all situations in which individual binary evaluations need to be combined into a group decision.

Two escape-routes to the discursive dilemma have been suggested: the *premise-based procedure* (PBP) and the *conclusion-based procedure* (CBP). According to PBP, each judge votes on each premise. The conclusion is then inferred from the rule $(A \wedge B) \leftrightarrow C$ and from the judgment of the majority of the group on A and B. In case the judges of the example followed the PBP, the defendant would be declared liable of breaching the contract. On the other hand, according to the CBP, the judges decide privately on A and B and only express their opinions on C publicly. The judgement of the group is then inferred from applying the majority rule to the individual judgments on C. In the example, contrary to the PBP, the application of the CBP would free the defendant. Moreover, no reasons for the court decision could be supplied.

In this paper, we study further properties of the information fusion procedure which takes a middle position between PBP and CBP. Above all, we address the question how good an aggregation procedure is at selecting the right conclusion.[1] The behavior of the fusion procedure will be contrasted with the PBP and the CBP that were studied by Bovens and Rabinowicz [2] and by List [7, 8]. Furthermore, we apply Bayesian conditionalization to the group decision problem.

[1] See [4] for an investigation of aggregation procedures in terms of reliability in selecting the right situation, i.e. premises and conclusion.

It is shown that the Bayesian approach enjoys theoretical optimality and high flexibility. In particular, we can combine it with any set of prior distributions and utility matrices. On the other hand, it requires a costly computation of a posterior distribution where both the prior distribution and the competence of the voters have to be made explicit. These requirements may be hard to meet in many practical applications. Finally, we compare the distance-based procedures to the Bayesian analysis.

The paper is structured as follows: In Section 2, we describe the fusion procedure and show that it is an element of a continuum of distance-based procedures which also contains PBP and CBP. Section 3 compares these three approaches in terms of their reliabilities at selecting the right conclusion. A full Bayesian analysis of a group decision problem is provided in Section 4. Section 5 concludes and sketches further open questions. Finally, the appendix contains the proofs and the calculation details.

2 Distance-based procedures

2.1 Introduction

As shown in [11], the application of a fusion operator to judgment aggregation problems allows to define consistent group decisions and to avoid paradoxical outcomes without having to choose between two possibly conflicting procedures like the PBP and CBP. This subsection summarizes the approach and the results of [11]. The reader is referred to that paper for more details.

One of the key points in the literature on information fusion is that the aggregation of finite sets of propositions satisfying some constraints does not guarantee a collective judgment satisfying the same constraints. One way to overcome this problem is to restrict the space of the possible solutions to the set of the *admissible situations* only, i.e. to those sets of premises and conclusion that satisfy the constraints. Then, the fusion operator selects one of these consistent situations, namely the (possibly not unique) element that minimizes the distance to the actual individual inputs.

To illustrate how the majority fusion operator works, we apply it to the court example. We have to form a judgment on the set of propositions $X = \{A, B, C\}$ with the constraint $(A \wedge B) \leftrightarrow C$. Hence, there are only four consistent situations:

$$\begin{aligned} S_1 &= \{A, B, C\} = (1,1,1) \\ S_2 &= \{A, \neg B, \neg C\} = (1,0,0) \\ S_3 &= \{\neg A, B, \neg C\} = (0,1,0) \\ S_4 &= \{\neg A, \neg B, \neg C\} = (0,0,0) \end{aligned} \tag{1}$$

In this terminology, A is identified with a 1 and $\neg A$ with a 0. In a group of N persons, there are n_1 persons endorsing the situation S_1 (i.e. they judge A, B and C to be true), n_2 persons endorsing S_2, and so on. Hence, $n_1+n_2+n_3+n_4 = N$. On pain of individual irrationality, every member of the group has to endorse exactly one of these situations. In principle, the equations in (1) involve an abuse of notation because the situations refer to states of the world as well as to elements in a vector space that bear distance relations to the submitted

	A	B	C	Total
Judges 1,2,3	1	1	1	S_1
Judges 4,5	1	0	0	S_2
Judges 6,7	0	1	0	S_3
Average	5/7	5/7	3/7	—
Hamming distance to S_1 (componentwise)	2/7	2/7	4/7	8/7
Hamming distance to S_2 (componentwise)	2/7	5/7	3/7	10/7
Hamming distance to S_3 (componentwise)	5/7	2/7	3/7	10/7
Hamming distance to S_4 (componentwise)	5/7	5/7	3/7	13/7

Table 2: The distance-based fusion operator in the original example of table 1.

judgments. Nonetheless, the intended meaning of "S_1" will always be evident from the context.

To apply the fusion operator, the four situations must be weighed with the number of persons that endorsed that situation. In other words, we look at

$$\overline{S} := \frac{1}{N}\sum_{i=1}^{4} n_i S_i \tag{2}$$

Fusion opts for the situation in $\{S_1, S_2, S_3, S_4\}$ which has the lowest distance to $\overline{S}$. In other words, if "$+S_i$" denotes a decision for S_i as the aggregated collective judgment set and "$-S_i$" a decision against S_i then

$$+S_i \iff \|S_i - \overline{S}\| \leq \min_{j \neq i} \|S_j - \overline{S}\|$$

If we define $d_i := \|S_i - \overline{S}\|$, fusion ranks the situation S_i first if and only if

$$d_i < \min_{j \neq i} d_j$$

Note that nothing hinges on the choice of a particular norm as a distance function because all norms in finite-dimensional spaces are (ordinally) equivalent. In particular, the S_i and $\overline{S}$ are all members of $\mathbb{R}^3$ so that the fusion procedure is invariant under the choice of a norm: only the ordering of the distances matters for the decision. In order to simplify calculations we recommend to use the 1-norm (i.e. to sum the absolute values of the three components) which corresponds to the Hamming distance.

The principle behind the distance minimization is the selection of a situation that is closest to the average. Table 2 illustrates that, in the court example, the situation selected by the fusion operator is $S_1 = \{1, 1, 1\}$. This is because S_1 is — among the possible situations — the closest to the collective average.

Since $d_1 = 2/7 + 2/7 + 4/7$ is smaller than any other distance to a situation, the fusion operator selects S_1. We also see that, with regard to a decision on S_1, we can apply all three procedures – PBP, CBP and fusion – to conjunctive aggregation rules ($A \wedge B \leftrightarrow C$) as well as to disjunctive rules ($A \vee B \leftrightarrow C$) because the latter are representable as $((\neg A \wedge \neg B) \leftrightarrow \neg C)$. The judgment aggregation mechanism is absolutely isomorphic – accepting $\neg A \wedge \neg B$ amounts to accepting $A \vee B$ and vice versa.

2.2 Representation results

There is also an intuitively understandable representation of the majority fusion procedure (henceforth FP). Assume that a voters judge A to be true, b voters judge B to be true and c voters judge $C \leftrightarrow (A \wedge B)$ to be true. (Of course, a, b, and c can be calculated from the n_i and vice versa.) Then several facts can be shown:[2]

Fact 1 *The following claims hold:*

1. $\min(a, b) \geq c \geq a + b - N$
2. $+S_1 \Leftrightarrow (a + c > N) \wedge (b + c > N)$.
3. *If* $d_1 < d_2, d_3$*, then also* $d_1 < d_4$.

Among these claims, we would like to draw special attention to the second one:

$$+S_1 \Leftrightarrow (a + c > N) \wedge (b + c > N) \tag{3}$$

In order to satisfy equation (3) and to accept S_1 in the aggregated judgment, a sufficient number of people have to endorse the conclusion C individually. In particular, a two-third majority on each of the premises is sufficient to guarantee that fusion selects the situation S_1.

From (3), we can also derive the following fact:

Fact 2 *Let* $+S_i(X)$ *denote a decision for the situation* S_i *under procedure* X*. Then*

- $+S_1(CBP) \rightarrow +S_1(FP) \rightarrow +S_1(PBP)$.
- $-S_1(PBP) \rightarrow -S_1(FP) \rightarrow -S_1(CBP)$.

Hence, if the CBP opts for S_1, so does fusion. And if fusion opts for S_1, so does the PBP. However, if the result from the CBP is negative, then fusion is more cautious than the premise-based procedure. To recall, the PBP suffers from a high false positive rate, i.e. it often endorses $C \leftrightarrow (A \wedge B)$ when it is in fact *false* (cf. [7, 8]). Fusion is less vulnerable to this mistake, as fact 2 shows. We will expand on this point in the subsequent section.

It might be suspected that for an increasing number of premises, fusion more and more resembles the premise-based procedure because the contribution of the premises outweighs the contribution from the conclusions. Nonetheless this is only true in the trivial sense that all three approaches make it increasingly hard to endorse the conclusion even when it is true.

Fact 3 *Let* $a_1, \ldots, a_m$ *denote the number of votes for each of the* m *premises. Then,*

1. $\min a_i \geq c \geq \sum_{i=1}^{m} a_i - (m-1)N$

[2] All proofs are given in the appendix.

2. $+S_1 \equiv a_i + c > N \, \forall i \in \{1, \ldots, m\}$

This entails that for constant p and large m, c/N (i.e. the fraction of people voting for S_1) is probably small. This is so because the large number of premises which have to be affirmed raises the hurdles for endorsing S_1. Even if there is a majority for each of the premises: the greater m, the higher the sampling variance in the a_i, so that the additional condition $a_i + c > N$ becomes increasingly hard to satisfy. Hence, for a large number of premises, FP will resemble CBP and set the standards for an endorsement of S_1 substantially higher than PBP.[3]

All this suggests that fusion is intimately related to CBP and PBP. Indeed, we can represent the two latter procedures as distance-based procedures when we parametrize the situation S_1 by means of $S_1^t := (1, 1, t)$ with $t \in [0, \infty]$. (For any t, S_1^t refers to the same real world situation as the original S_1 – both premises and the conclusion are true. Merely the distance between this situation and the submitted set of judgment is now measured differently, in particular $\overline{S}^t := (a/N, b/N, tc/N)$.) Again, the situation which minimizes the distance to $\overline{S}^t$, the average of the submitted judgment sets, is chosen. This elucidates the connection between fusion, CBP and PBP:

Proposition 1 *Let $S_1^t = (1, 1, t)$. Choosing the situation with the minimum distance to $\overline{S}^t$ is equivalent to PBP for $t = 0$, yields FP for $t = 1$ and converges to CBP for $t \to \infty$.*

In other words, for $t \to 0$, the distance-based operator converges against the PBP which is attained for $t = 0$.[4] On the other hand, when $t \to \infty$, the distance-based operator converges against the CBP. Finally, for $t = 1$, we obtain the conventional fusion operator.[5] From now on, we intend the term "distance-based procedure" to refer to all aggregation procedures that correspond to a specific value of t, including $t = \infty$. This gives us a continuum of distance-based approaches, ranging from the premise-based to the conclusion-based operator, with fusion having a middle position. We now turn to an evaluation of the procedures.

3 Comparing the distance-based procedures

3.1 Preliminaries

In order to investigate the epistemic reliability of the fusion procedure, we adopt a probabilistic framework. In particular, we assign to every voter an individual competence $p \in (0, 1)$ to make a correct judgment about a single premise. This means that when a premise (either A or B) is true, the voter gives a correct report with probability $p_1 = p$, and equally, if the premise is false, the voter gives

[3]This does not contradict the obvious fact that, for fixed prior probabilities, the probability of correctly detecting S_1 approaches 0 as m goes to infinity for all three approaches.

[4]We can think of that case as a projection of S_1^t onto the hyperplane spanned by the other three situations.

[5]The value $t = 1$ is special because it is the only value where the yes=1/no=0 assignment scheme introduced in the previous subsection applies to premises *and* conclusion.

a correct report with probability $p_2 = p$.[6] It would, of course, also be possible to assign two different competences to the voter, one for correctly discerning A and one for correctly discerning $\neg A$. But that would couple the competence of the voter to the prior probabilities of the various situations. To see this, note that

$$p = p_1\, \mathbb{P}(A) + p_2\, (1 - \mathbb{P}(A)) \tag{4}$$

We would often like to say that the reliability of the voters is independent of the prior probabilities over the four situations. The only possible way to ensure this independence is to assume that the probability of a false positive report on a premise equals the probability of a false negative report, in other words, $p_1 = p_2$.[7] Then, the Condorcet Jury Theorem links the competence of the voters to the reliability of majority voting: Assume that the individual votes on a proposition A are independent of each other, conditional on the truth or falsity of that proposition. If the chance that an individual voter correctly judges the truth or falsity of A is greater than fifty percent (in other words, $p > 0.5$), then majority voting eventually yields the right collective judgment on A with increasing size of the group. Therefore, the Condorcet Jury Theorem offers an epistemic justification to majority voting and motivates the use of the PBP and CBP in the judgment aggregation problem ([2]).

It should be noted, though, that an application of the Condorcet results to judgment aggregation requires further assumptions which we now make explicit. They are also required to avoid computational complexity and are formulated as in [2]:

(i) The prior probabilities that A and B are true are equal ($\mathbb{P}(A) = \mathbb{P}(B)$).

(ii) A and B are (logically and probabilistically) independent.

(ii) All voters have the same (independent) competence to assess the truth of A and B (p). Their judgments on A and B are independent.

(iv) Each individual judgment set is logically consistent.

Assumption (iv) entails that only four situations are possible:

$$\begin{array}{ll} S_1^t = \{A, B, C\} = (1, 1, t) & \quad S_2^t = \{A, \neg B, \neg C\} = (1, 0, 0) \\ S_3^t = \{\neg A, B, \neg C\} = (0, 1, 0) & \quad S_4^t = \{\neg A, \neg B, \neg C\} = (0, 0, 0) \end{array}$$

Moreover, assumption (i) and independence claim (ii) entail that we can parametrize the set of prior distributions by a single parameter $q := \mathbb{P}(A) = \mathbb{P}(B)$. From the independence assumptions we then obtain

[6]We ascribe an individual competence only for voting on *premises*, not for voting on any proposition (such as $A \wedge B$). Indeed, it follows that given an individual voting competence p on A and B, the voting competence on $A \wedge B$ is $p^2 \neq p$ ([8]). However, in many contexts it is reasonable to assign individual voting competence to only a certain kind of propositions. E. g. in a legal case this would be propositions as "P had contractual obligation not to do X" or "P actually did X", but not on propositions as "P should go to jail".

[7]Setting $p_1 = p_2$ also answers List's concerns ([8]) that for a very low value of p_1 or p_2, the voters are bad at tracking the true situation although the overall reliability p, as defined in (4), can still be high. Regardless of whether this point is really convincing, setting $p_1 = p_2$ kills two birds with one stone: we circumvent List's objection and we decouple overall reliability and prior probabilities.

$$\mathbb{P}(S_1^t) = q^2; \;\; \mathbb{P}(S_2^t) = \mathbb{P}(S_3^t) = q(1-q); \;\; \mathbb{P}(S_4^t) = (1-q)^2$$

The probability that a distance-based procedure chooses the right conclusion can be calculated via

$$\begin{aligned} \mathbb{P}(G) &:= \mathbb{P}(\text{A distance-based procedure selects the right conclusion}) \\ &= \mathbb{P}(S_1^t)\,\mathbb{P}(+S_1^t|S_1^t) + \sum_{i=2}^{4} \mathbb{P}(S_i^t)\,\mathbb{P}(-S_1^t|S_i^t) \end{aligned}$$

where "$+S_1^t$" denotes a collective judgment that selects the situation S_1^t and the $\mathbb{P}(S_i^t)$-terms can be replaced by the corresponding q-terms.

3.2 Results and generalizations

With the above equations in hand, we can now compare the fusion procedure (FP) to the PBP and the CBP. Bovens and Rabinowicz ([2]) show that the PBP is always better at identifying the correct situation, while the CBP is sometimes better at selecting the right conclusion. This means either to accept or to reject S_1^t as the correct situation, and, in case of a rejection, to be silent on whether S_2^t, S_3^t or S_4^t is true. Indeed, in a variety of real aggregation problems, it is most urgent to come to a verdict with regard to S_1^t and it is less important to discern between S_2^t, S_3^t and S_4^t (e.g. because they have the same practical consequences). However, that does *not* mean that the aggregation procedures neglect the *reasons* for either accepting or rejecting S_1^t: The number of votes for each premise plays a substantial part in all distance-based approaches to judgment aggregation, with the obvious exception of CBP. The complementary problem of situation selection is covered in detail in a sequel paper ([4]).

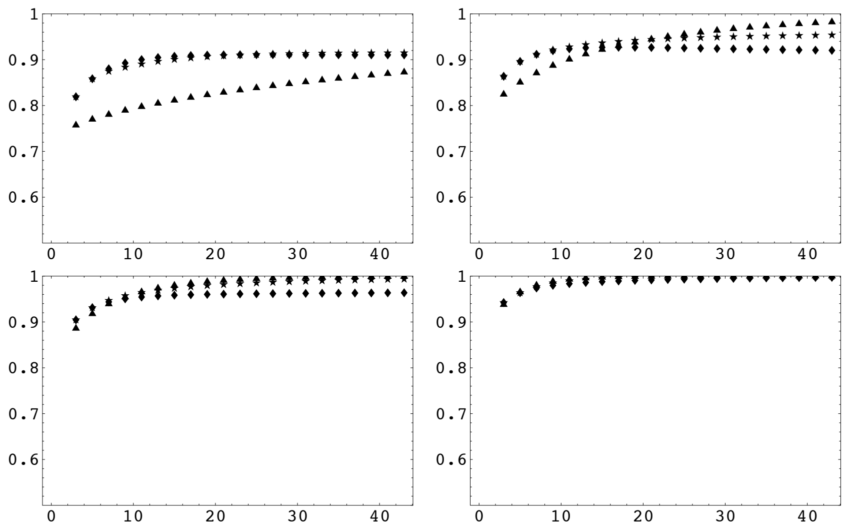

Figure 1: Reliability of PBP (triangles), FP (stars) and CBP (diamonds) as a function of N, for various values of p and a fixed value of $q = 0.3$. Upper left figure: $p = 0.56$. Upper right figure: $p = 0.64$. Lower left figure: $p = 0.72$. Lower right figure $p = 0.8$.

Figures 1-3 depict the reliability of PBP, CBP and FP for various values of p, q and odd values of N. First we would like to discuss figure 1. It turns out that

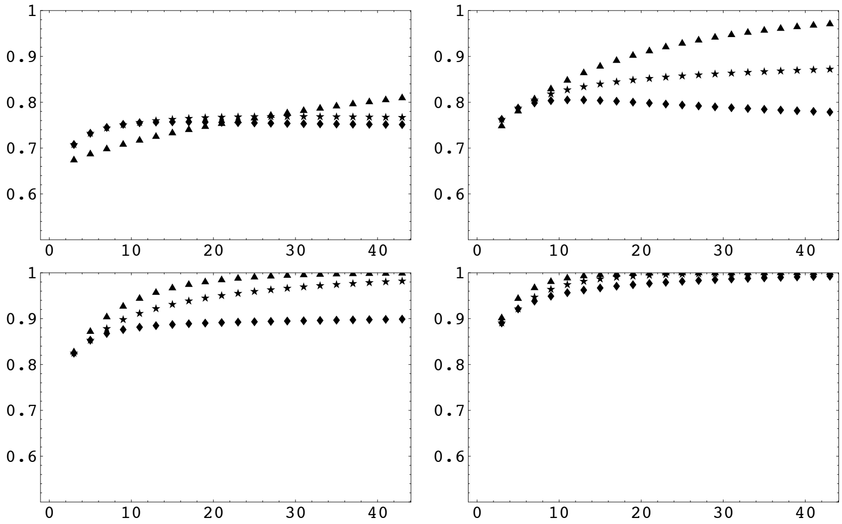

Figure 2: Reliability of PBP (triangles), FP (stars) and CBP (diamonds) as a function of N, for various values of p and a fixed value of $q = 0.5$. Upper left figure: $p = 0.56$. Upper right figure: $p = 0.64$. Lower left figure: $p = 0.72$. Lower right figure $p = 0.8$.

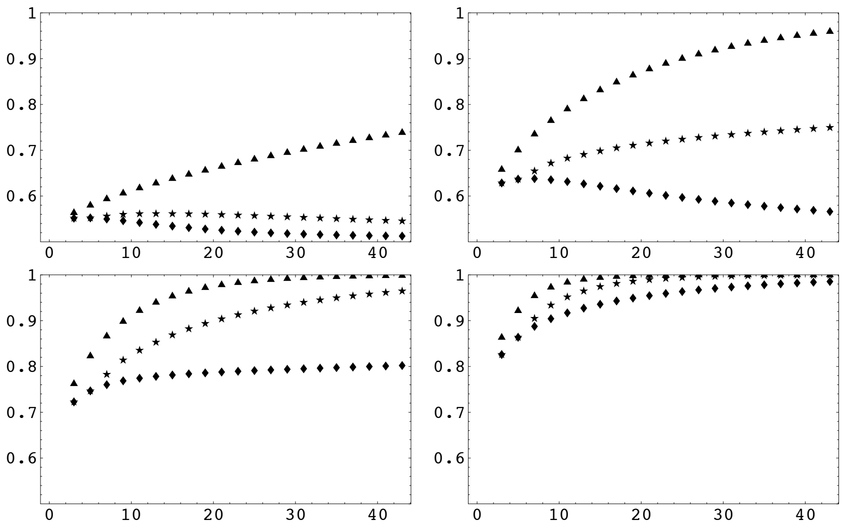

Figure 3: Reliability of PBP (triangles), FP (stars) and CBP (diamonds) as a function of N, for various values of p and a fixed value of $q = 0.7$. Upper left figure: $p = 0.56$. Upper right figure: $p = 0.64$. Lower left figure: $p = 0.72$. Lower right figure $p = 0.8$.

for relatively small values of p ($p = 0.56$, 0.64), the premise-based procedure too often erroneously endorses S_1^t, and especially so for small values of N. In this case, a majority for a premise can emerge by mere random sampling effects although the premise is actually not satisfied. Therefore PBP is inferior to both FP and CBP in such circumstances. For higher values of p, however, the three procedures nearly coincide and do not differ much. This is especially salient for $p = 0.8$. Figure 2 confirms the local failure of PBP for a modest p ($p = 0.56$). However, we also see that for intermediate values of p ($p = 0.64$), PBP clearly dominates the two other approaches whereas there is again no significant difference between PBP and the rest for high values of p.

The superiority of PBP is most pronounced in Figure 3 where $q = 0.7$, i.e. S_1^t is the most probable situation. For any value of p smaller than 0.8, PBP clearly outperforms the two other procedures. That is not surprising: the greater q, the more important is it to avoid erroneous rejection of S_1^t, just because S_1^t occurs more often. Fact 2 has established that, among the three scrutinized procedures, PBP is most inclined towards accepting S_1^t, as already noted by ([2], [8]). This "optimism" towards S_1^t naturally pays off in terms of overall reliability when q is quite large.[8] On the other hand, we see that CBP fails to benefit from the greater stability in the data which accompanies the increasing number of voters. Especially, we see that CBP performs quite poorly for large values of N in comparison to the other procedures. Besides we see again that all three procedures are almost equally reliable for $p = 0.8$ because the high individual reliability guarantees that any procedure is well protected against error.

The concrete observations for large N in the above examples can be generalized. We perform an asymptotic analysis of the distance-based procedures in a general framework, building on the parametrization already used in proposition 1. Consider first the case that S_1^t is true.

Proposition 2 *Assume that $S_1^t = (1, 1, t)$ is the true situation. Then $\mathbb{P}$-almost surely ($\mathbb{P}$-a.s.) for $N \to \infty$:*

$$+S_1^t \iff d_1^t < \min_{j \neq 1} d_j^t \iff p > p_t := \frac{\sqrt{2t^2 + 2t + 1} - 1}{2t}$$

In particular, this translates as $p > 0.5$ for PBP, $p > (\sqrt{5} - 1)/2$ for FP and $p > 1/\sqrt{2}$ for CBP.

The following corollary asserts that $p > p_t$ is both necessary and sufficient in order to ensure the $\mathbb{P}$-a.s. correct conclusion selection for increasing group size:

Corollary 1 *For the group size going to infinity ($N \to \infty$), the distance-based procedures select the right conclusion $\mathbb{P}$-a.s. if and only if $p > p_t$.*

Put another way, $\mathbb{P}(G) \overset{N\to\infty}{\to} 1$ if and only if p is larger than the specified threshold. Hence, PBP has better asymptotic properties than fusion because for a large number of voters, it eventually becomes perfectly reliable for $p > 0.5$ whereas fusion requires the stronger $p > (\sqrt{5} - 1)/2$. CBP requires the even higher $p > 1/\sqrt{2}$. This superiority of PBP for large voting groups is exemplified in all three figures. Moreover, the asymptotic results explain why CBP is not

[8]We even conjecture that there is a threshold for q (dependent on p) so that for any N, PBP is more reliable than any other distance-based procedure ($t \in (0, \infty)$). We would like to prove such a result in future work.

monotonously increasing as a function of N for $p = 0.56$ or $p = 0.64$. The same holds for FP with regard to $p = 0.56$. However, the figures also teach us that for large values of p, the asymptotic results carry little importance for the actual reliability because all three procedures tend to agree quickly. Furthermore, the asymptotic properties are not always correlated with the performance in small voting groups: For small to moderate values of p, q and N, FP and CBP outperform PBP – see the upper graphs in figure 1 and 2. As already mentioned, we believe that this is due to random sampling effects which occur in small voting groups.

We can summarize the results as follows: For high values of p (approximately $p > 0.75$), all three examined procedures are very reasonable. Choosing an aggregation method among the infinity of distance-based procedures does not make much of a difference. Only for moderate values of p ($p \in [0.5, 0.75]$), there is a real difference between the aggregation procedures. It turns out that the prior probability of S_1^t, q^2, plays a crucial role here. Roughly, we can say that the higher q and the higher N, the more should we be inclined towards PBP, whereas for small groups and modest q, FP or even CBP can be the better choice. In comparison to CBP, FP has the virtue of not performing too badly for large samples and medium values of p. For potential applications, it might be interesting to note that in a lot of jury and panel decisions, the number of voters is quite small, typically $N \in \{5, 7, \ldots, 15\}$. Hence, especially when we have some reasons not to fully trust the voters competence (take, for instance, a laymen jury in a criminal trial), we have a rationale for applying the fusion operator. For such cases, we also suggest further calibrations of t in order to combine the power of PBP with the conservativeness of FP, e.g. $t = 0.5$. By contrast, when we face a large number of voters, for instance in a plebiscite, recommending PBP is the safest option due to the asymptotic superiority. Such calibrations can be further refined by considering the relative severity of a decision error. E.g. if erroneous acceptance of S_1^t were in a specific situation much worse than erroneous rejection of S_1^t, we would tend to set t to a higher value than if the opposite were true.

4 The Bayesian Approach

4.1 General Remarks

The probabilistic framework which we used for the evaluation of PBP, CBP and FP can be transferred to a full Bayesian approach, too. In a Bayesian approach, we have a prior probability distribution over the situations S_1 to S_4, given again by $(q^2, q(1-q), q(1-q), (1-q)^2)$. We treat the judgments of the voters (call them V) as incoming evidence which we use to update the prior probabilities to a *posterior distribution* over S_1 to S_4:

$$\mathbb{P}(S_i|V) \quad = \quad \frac{\mathbb{P}(S_i)\mathbb{P}(S_i|V)}{\mathbb{P}(V)}$$

This posterior distribution describes our rational degree of belief in the various situations, given the verdicts of the voters and their individual reliability. Then we base our decision exclusively on that posterior distribution and the utility matrix which describes the actual decision problem.

	S_1 is true	S_1 is false
accept S_1 ("$+S_1$")	1	0
reject S_1 ("$-S_1$")	0	1

Table 3: The utility matrix that corresponds to using $P(G)$ as a benchmark for the performance of the aggregation procedures, shown as a function of the possible actions and states of the world.

We are now interested in the average probability that the right conclusion (S_1 or $\neg S_1$) is selected. In other words, we want to calculate $\mathbb{P}(G)$ and use it as a the benchmark for the various aggregation procedures. Keeping in mind that $\mathbb{P}(G) = \mathbb{P}(S_1)\,\mathbb{P}(+S_1|S_1) + \mathbb{P}(\neg S_1)\,\mathbb{P}(-S_1|\neg S_1)$, this corresponds to a decision problem where utility 1 is assigned to a correct conclusion selection and 0 to a wrong conclusion selection (see table 3). [9]

The *Conditional Bayes Principle* ([1], p. 8) tells us that, relative to a given (posterior) distribution, we ought to take the action that maximizes the expected utility. Faced with the above utility matrix, we will opt for S_1 if and only if $\mathbb{P}(S_1|V) > 0.5$.

We can generalize that principle to the following decision rule: for any set of judgments that will be observed, take the action that maximizes expected utility relative to the posterior distribution which we obtain by updating on the voters' judgments. Indeed, it can be shown that such a decision rule is a *Bayes rule*, i.e. a decision rule that minimizes the expected risk with regard to the prior distribution among all decision rules.[10] Hence we see that the decision rule given by the Bayesian approach is optimal in the risk-minimizing (or utility-maximizing) sense. We can base all decisions solely on the posterior distribution and the problem-specific utility matrix.

This has a number of implications. First, all data that do not affect the posterior distribution of the S_i can be neglected. The posterior distribution is uniquely determined by the number of votes for A and B, in other words, the statistics a and b. All further information is irrelevant, given the values of those functions. Technically spoken, a and b are *sufficient statistics*.[11] Once we know the values of a and b, we can totally neglect how many people endorsed the conclusion. This vindicates an intuition underlying the premise-based procedure: complete information about the premises is all we need to make a reliable decision. Indeed, the *Rao-Blackwell Theorem* ([1], p. 41) guarantees that there is no information that can improve the decision rule beyond what is contained in the sufficient statistics. More precisely, any decision rule can be improved in a way that it is only a function of the sufficient statistics.

This also implies that both the premise- and the conclusion-based approach cannot be optimal: Both are based on 0-1 statistics that measure whether there are majorities for A, B or C. But those statistics are neither sufficient nor jointly sufficient. Too much information gets lost when only checking the majorities. A similar result holds for fusion, where the decision on the right conclusion is only based on the statistics $(a+c)$ and $(b+c)$ which are neither (jointly) sufficient. Hence, none of the these three approaches can be optimal. Contrarily, by the

[9] In other words, erroneously opting for S_1 is equally devastating as erroneously opting for $\neg S_1$. A similar matrix can be found for the situation selection problem, see again [4].

[10] Cf. Result 1 in [1], p. 159.

[11] A statistic T is *sufficient* with regard to an unknown parameter Θ when $\mathbb{P}(\Theta = \vartheta|T = t) = \mathbb{P}(\Theta = \vartheta|X = x)$ where X denotes the full data.

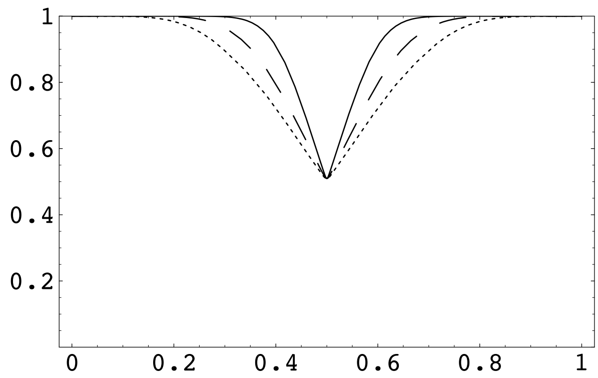

Figure 4: The probability that the Bayesian procedure identifies the right conclusion as a function of the competence of the voters p for $q = .7$ and $N = 11$ (dotted line), $N = 21$ (dashed line) and $N = 51$ (full line).

aforementioned theorem, the Bayesian approach provides an upper bound for the reliability of all decision rules. We now turn to the performance of the Bayesian approach and compare its reliability to the procedures discussed in the previous section.

4.2 Results and Discussion

In Figure 4.2, we have plotted the reliability of the Bayesian aggregation procedure as a function of the individual reliability for $q = 0.7$ and various group sizes ($N = 11, 21, 51$). We see that the Bayesian procedure is almost perfectly reliable when p is far from 0.5. From the above arguments it is clear that the Bayesian reliability constitutes an upper bound for all other procedures. Also, we see that the reliability is monotonously increasing in N, and it approaches 1 (pointwise) for all values of p except for a small neighbourhood of 0.5. It can even be shown that for all values of p except for $p = 0.5$, the Bayesian procedure eventually selects the right conclusion.

Proposition 3 *For any $p \neq 0.5$, the Bayesian decision rule eventually selects the right conclusion $\mathbb{P}$-a.s. when $N \to \infty$.*

This draws our attention to another feature of the Bayesian analysis: It is symmetric as a function of p. So Bayesian aggregation is perfectly reliable even when the individual voters are very unreliable. This sheds light on a substantial premise of the Bayesian approach, namely that knowledge of p and q is required to update the prior distribution. In other words, both p and q have to be *transparent* to the decision maker. Then, it is no more surprising that the Bayesian procedure is highly reliable for $p \approx 0$: When the aggregators *know* that the voters nearly always submit wrong judgments, they will just replace the individual judgments on the premises by their negations. So, when a highly unreliable voter submits $(1, 0, 0)$, this amounts to the submission of $(0, 1, 0)$ by a highly reliable voter. Therefore the Bayesian procedure works perfectly fine for low values of p.

On the other hand, it is questionable whether knowledge of p and q can really be presupposed. Such transparency is hardly realistic and not applicable in a wide class of cases. For example, any kind of assigning prior probabilities is frowned upon in legal decision making. Even more disturbing, our estimate of the individual voting competence may be grossly mistaken. For example, assume that we erroneously claim that $p = 0.5$, i.e. we believe the voters to be randomizers. For the Bayesian, this means that the results of the voting process have no impact on the posterior distribution: it equals the prior distribution. Even if the voters unanimously endorse both premises and the conclusion, this has absolutely no impact. In particular, when $\mathbb{P}(S_1) \leq 0.5$, even an unanimous endorsement of S_1 does not lead to the acceptance of S_1. This is a result which we intuitively find absurd: Not only does the Bayesian recommendation conflict with the principle of unanimity, the data also suggest to revise our estimate of p because unanimity is much less likely for randomizing voters than for competent voters. For such an extreme set of submitted judgments, we would rather tend to accept S_1 as it is recommended by all discussed distance-based procedures. The data seem to falsify our previous estimate $p = 0.5$. Such a revision is, however, not possible in a genuine Bayesian framework because this would amount to double-counting the data: once for eliciting an estimate of p and afterwards for updating the prior distribution. Assigning a prior distribution over p and averaging the results over the prior distribution of p ("Bayesian model averaging") would mitigate, but not eliminate the effect. We leave it open to future research whether the Bayesian aggregation procedure can be protected against severe estimation errors. In any case, the Bayesian approach carries a considerable estimation risk when p is hard to elicit.

In several contexts it is, of course, not awkward to assume that p is more or less transparent, e.g. when we have N independent measuring instruments in a scientific experiment. In such a case we should always use Bayesian updating and make a decision on the basis of the posterior distribution. But in a variety of cases where human judgments are aggregated, e.g. when a jury has to decide upon the liability of the defendant or to award or deny tenure to a faculty member, the competence of the jury may be hard to estimate. In particular, there may be no relative success frequency as a basis for an estimate of the voting competence. Moreover, the Bayesian procedure will sometimes, namely for low values of p, recommend to do exactly the opposite of what the voters are thinking. This prevents the Bayesian procedure from being applied in many practical cases. It fares best when the decision-maker has a sensible estimate of p and when he does not need the consent of the voters to take an action.[12]

To summarize: The Bayesian approach has the advantage that the decision procedure can be flexibly adapted to the particular problem by changing the utility matrix. Various judgment aggregation problems are characterized by different utility matrices to which the Bayesian procedure can be flexibly transferred. E.g. sending someone to jail erroneously has a much higher associated loss than setting free a guilty person. But denying tenure to an outstanding researcher and teacher might be worse than giving tenure to a person who is a capable researcher but only a mediocre teacher. In other words, the utilities tell us which kind of error we have to avoid. The Bayesian approach naturally incorporates the variation in the losses associated with a wrong decision of a certain kind whereas distance-based approaches as CBP, PBP and FP are not sensitive to the severity of a particular kind of error. The Bayesian framework can be naturally extended to connection rules different from the conjunction of the premises, too. On the other hand, the distance-based procedures do not

[12] So far, we have been silent on the possibility of strategic voting which can occur in the Bayesian model as well as in the distance-based approaches. Therefore we assume that the voters have no interest in a specific conclusion – otherwise no voter would submit the judgment set S_2 or S_3.

involve a complicated calculation of posterior probabilities and they are easier to apply to real decision problems.

5 Conclusions

By means of the t-parametrization in section 2, we have integrated PBP, CBP and FP into a continuum of distance-based procedures, with FP taking a middle position between the two extremes PBP and CBP. Thereby we achieve a conceptual unification with regard to the traditional methods of judgment aggregation. With regard to selecting the right conclusion, the distance-based procedures are theoretically inferior to the Bayesian decision procedure which also enjoys high flexibility. Nevertheless, the Bayesian approach can only be applied when the individual competence p and the prior probability q of each premise can be elicited. Often, estimating p might be associated with a unpredictable estimation risk, deteriorating the performance of the Bayesian approach. This was illustrated in the toy example of the previous section. In practice, working with an estimate of p might also be prohibited by external, pragmatic constraints. The more we are uncertain about p, the more we should be inclined to apply a distance-based approach which performs reasonably well for various values of p. The most suitable value of t then depends on the actual number of voters and the estimated prior probabilities. Distance-based approaches, although not strictly optimal from a theoretical point of view, might be a reasonable compromise between two ends: to neglect an estimate of individual competence in the actual decision-making and to have a procedure that reliably selects the right conclusion. Among these procedures, those which resemble the premise-based procedure ($t \ll 1$) will usually perform best. The precise calibration of t in an actual application is sensitive to the specific social decision problem and will be investigated in further work. Moreover, our sequel paper [4] examines the performance of the discussed procedures at tracking the right situation.

A Calculational details

A.1 Properties of the fusion operator

We examine the case $A \wedge B \leftrightarrow C$. There are N voters and we assume that N is an odd number ≥ 3. n_1 voters vote for $S_1 = (1,1,1)$, n_2 for situation $S_2 = (1,0,0)$, n_3 for situation $S_3 = (0,1,0)$ and n_4 for situation $S_4 = (0,0,0)$. Obviously, $N = n_1 + n_2 + n_3 + n_4$. Fusion chooses the model which has the lowest distance to the average of submitted judgment sets, $\overline{S} := \frac{1}{N}\sum_{i=1}^{4} n_i S_i$ (cf. equation (2)). For reasons of convenience, we work with the Hamming distance which corresponds to the 1-Norm in real Euclidean vector spaces. (As mentioned in the main text, all norms are ordinally equivalent). Let a denote the number of voters that vote for premise A and b the number of voters that vote for B. Accordingly, let c denote the number of voters that vote for the conclusion C. Let $d_i := \|S_i - \overline{S}\|_1$. Hence, fusion ranks model S_i first if and only if $d_i < \min_{j \neq i} d_j$.

Proof of Fact 1: The first claim is trivial. For the rest, note that

$$\begin{aligned}
N(d_2 - d_1) &= (N-a) + b + c - (N-a) - (N-b) - (N-c) = 2b + 2c - 2N \\
N(d_3 - d_1) &= a + (N-b) + c - (N-a) - (N-b) - (N-c) = 2a + 2c - 2N \\
N(d_4 - d_1) &= a + b + c - (N-a) - (N-b) - (N-c) = 2a + 2b + 2c - 3N
\end{aligned}$$

The first two equations yield $d_1 < d_2 \equiv b + c > N$ and $d_1 < d_3 \equiv a + c > N$. If these two conditions are satisfied, it follows that

$$N(d_4 - d_1) > a + b - N > 0 \tag{5}$$

because $a + c > N$ also implies $a > N/2$ and analogously for b. Thus, also $d_1 < d_4$, showing the second and the third part of the fact. $\square$

Proof of Fact 2: We only prove the first claim of the fact, the second follows by contraposition. A positive report in the conclusion-based procedure ("$+S_1$(CBP)") occurs if and only if $c > N/2$. Since $a, b \geq c$, this also implies $a + c > N$ and $b + c > N$. So fusion opts for S_1, too, in virtue of fact 1. The two latter inequalities imply that $a, b > N/2$ (again, because $a, b \geq c$). Hence, if fusion opts for S_1, so does the premise-based procedure. $\square$

Proof of Proposition 1:

$$d_1^t = \frac{1}{N}(N - a + N - b + t(N - c)) \qquad d_2^t = \frac{1}{N}(N - a + b + tc)$$

$$d_3^t = \frac{1}{N}(a + N - b + tc) \qquad d_4^t = \frac{1}{N}(a + b + tc)$$

A decision for S_1^t is made if and only if $d_1^t < \min_{i \neq 1} d_i^t$. For $a \leq N/2$ or $b \leq N/2$, we also get $c \leq N/2$ and $t(N - c) \geq tc$ for all values of t. Hence either $d_1^t \geq d_2^t$ or $d_1^t \geq d_3^t$ so that S_1^t is rejected independent of the value of t.
Therefore assume that $a, b > N/2$. Then a decision for the conclusion S_1^t amounts to $d_1^t < \min_{i \neq 1} d_i^t$ which can be written as

$$\begin{aligned}
2a - N + t(2c - N) &> 0 \\
2b - N + t(2c - N) &> 0 \\
2a - N + 2b - N + t(2c - N) &> 0
\end{aligned}$$

Obviously, the first two inequalities entail the third one.
For t=1, this leads to the usual conditions $a + c > N$ and $b + c > N$, as shown in fact 1.
For t=0, the inequalities are satisfied if and only if $a > N/2$ and $b > N/2$, i.e. when there is a majority for each of the premises. Hence we obtain PBP.
Finally, for very large t, the $2a - N$ and $2b - N$ terms drop out of the picture. Hence we accept S_1^t if and only if $c \geq N/2$. CBP – accept the conclusion if and only if it is endorsed by a genuine majority – is obtained in the limit $t \to \infty$. $\square$

Remark: It is a simple fact of linear algebra that the relative ordering between the distances d_2^t, d_3^t and d_4^t is not affected by the value of t. So, the proof would go through even if we aimed at choosing the right situation instead of the right conclusion.

Proof of Fact 3: The first inequality is trivial. For the second one, the maximal set of voters who do *not* accept the conclusion is $\sum_{i=1}^m (N - a_i)$. Hence, $c \geq N - \sum_{i=1}^m (N - a_i) = \sum_{i=1}^m a_i - (m-1)N$.

For the second claim, first assume that there is an i so that $a_i + c \leq N \equiv N - c \geq a_i$. Then, by measuring distance by the supremum norm, we see that

$$\begin{aligned} & \sup(N-a_1, N-a_2, \ldots, N-a_m, N-c) \\ = \; & N-c \\ \geq \; & \sup(N-a_1, N-a_2, \ldots, N-a_{i-1}, a_i, N-a_{i+1}, \ldots, c) \end{aligned}$$

so that $S_i := (1,1,\ldots,1,0,1,\ldots,0)$ is closer to $\overline{S}$ than $S_1 = (1,1,\ldots,1)$. Hence S_1 is rejected.
Now assume that $\forall i : a_i + c \geq N$. Then S_1 is chosen if and only if $N - c < \min(a_1, a_2, \ldots, a_m)$ which is true by hypothesis. Hence we have established necessity and sufficiency of the proposed condition. $\square$

A.2 Asymptotical behaviour of the distance-based procedures

Proof of Proposition 2: Under S_1^t, a is $B_{N,p}$-distributed. Similarly, $b \sim B_{N,p}$ and $c \sim B_{N,p^2}$. All three variables are sums of N independent and identically distributed random variables so that the Strong Laws of Large Numbers applies. It follows that $\mathbb{P}$-a.s. a/N, $b/N \to p$, $c/N \to p^2$. In particular, with the exception of a set of measure zero,

$$\forall \varepsilon \geq 0 \; \exists N_0 \; \forall N \geq N_0 : \; \frac{c}{N} \in (p^2 - \varepsilon, p^2 + \varepsilon), \; \frac{a}{N}, \frac{b}{N} \in (p - \varepsilon, p + \varepsilon) \qquad (6)$$

Choose $(3+3t)\varepsilon := \left|2tp^2 + 2p - (1+t)\right|$. A simple computation yields that $2tp^2 + 2p - (1+t) > 0$ if and only if $p > p_t := (1/2t)(\sqrt{2t^2+2t+1} - 1)$ for $t \neq 0$ and $p > 0.5$ for $t = 0$. Recall that S_1^t is chosen ("$+S_1^t$") if and only if

$$\begin{aligned} 2\frac{a}{N} - 1 + t\left(2\frac{c}{N} - 1\right) > 0 \\ 2\frac{b}{N} - 1 + t\left(2\frac{c}{N} - 1\right) > 0 \end{aligned} \qquad (7)$$

(cf. the proof of proposition 1). And indeed, for $p > p_t$ and $N \geq N_0$,

$$\begin{aligned} & 2\frac{a}{N} + t\left(2\frac{c}{N} - 1\right) \\ > \; & 2p - 2\varepsilon - 1 + t(2p^2 - 2\varepsilon - 1) \\ = \; & (3+3t)\varepsilon - 2\varepsilon - 2t\varepsilon \\ > \; & 0 \end{aligned}$$

Exactly the same computation can be done for the other equation in (7). Thus, if $p > p_t$, $\mathbb{P}$-almost surely the right conclusion is eventually chosen if S_1^t is true.
Assume on the other hand that $p \leq p_t$. Then, in virtue of (6), for $N \leq N_0$,

$$\begin{aligned} & 2\frac{a}{N} + t\left(2\frac{c}{N} - 1\right) \\ < \; & 2p + 2\varepsilon - 1 + t(2p^2 + 2\varepsilon - 1) \\ = \; & (-3-3t)\varepsilon + 2\varepsilon + 2t\varepsilon \\ < \; & 0 \end{aligned}$$

and eventually selects the wrong conclusion is $\mathbb{P}$-a.s. selected if S_1^t is true.
For $t \to \infty$ (CBP), a simple asymptotic analysis yields the condition $p > 1/\sqrt{2}$,

whereas $t = 0$ (PBP) amounts to $p > 0.5$. Finally, in the case $t = 1$ (FP), we get the threshold $p > (\sqrt{5} - 1)/2$. $\square$

Proof of Corollary 1: Proposition 2 has already proved that, if S_1^t is true, the right conclusion is eventually chosen $\mathbb{P}$-a.s. if and only if $p > p_t$. Three remaining cases are to examine.
(a) S_2^t is true. If $p > 0.5$ we will $\mathbb{P}$-a.s. get $b < N/2$ for increasing N and thus reject S_1^t, independent of the value of t.
(b) S_3^t is true. The same as (a) due to symmetry.
(c) S_4^t is true. If $p > 0.5$ we will $\mathbb{P}$-a.s. get $a, b < N/2$ for increasing N. Hence S_1^t is rejected, independent of the value of t.
Since $p > p_t$ entails $p > 0.5$, all distance-based procedures eventually choose the right conclusion $\mathbb{P}$-a.s. if and only if $p > p_t$. $\square$

A.3 Bayesian posteriors

We calculate the posterior probability of S_1 conditional on the observed variables a and b, the number of votes for premise A and B, respectively. Calculations are completely analogous for the rivalling models. Let $L_{ij}(a, b)$ be the likelihood ratio of situation S_i to situation S_j as a function of the data x. Then, we have

$$
\begin{aligned}
L_{21}(a,b) &= \left(\frac{1-p}{p}\right)^{2b-N} & L_{31}(a,b) &= \left(\frac{1-p}{p}\right)^{2a-N} \\
L_{41}(a,b) &= \left(\frac{1-p}{p}\right)^{2a+2b-2N} & L_{32}(a,b) &= \left(\frac{1-p}{p}\right)^{2a-2b} \\
L_{42}(a,b) &= \left(\frac{1-p}{p}\right)^{2a-N} & L_{43}(a,b) &= \left(\frac{1-p}{p}\right)^{2b-N}
\end{aligned}
$$

Note that $L_{ij}(a, b) = L_{ji}^{-1}(a, b)$. We can now calculate the posterior distribution if a votes for premise A and b votes for premise B are submitted:

$$
\begin{aligned}
\mathbb{P}(S_1 \mid a, b) \quad &= \quad \frac{\mathbb{P}(S_1)\,\mathbb{P}(a, b \mid S_1)}{\mathbb{P}(a, b)} \quad = \quad \frac{\mathbb{P}(S_1)\,\mathbb{P}(a, b \mid S_1)}{\sum_{i=1}^{4} \mathbb{P}(S_i)\,\mathbb{P}(a, b \mid S_i)} \\
&= \quad \left[1 + \sum_{i=2}^{4} \frac{\mathbb{P}(S_i)}{\mathbb{P}(S_1)} L_{i1}(a, b)\right]^{-1} \\
&= \quad \left[1 + \frac{1-q}{q} L_{21} + \frac{1-q}{q} L_{31} + \left(\frac{1-q}{q}\right)^2 L_{41}\right]^{-1} \qquad =: \qquad M^{-1}
\end{aligned}
$$

In other words, we denote the term inside the square brackets (without the exponent) by M. For any event E, due to independence,

$$\mathbb{P}(E \mid S_1) = \sum_{a=1}^{N}\sum_{b=1}^{N} \binom{N}{a} p^a (1-p)^{N-a} \binom{N}{b} p^b (1-p)^{N-b} 1_E$$

$$\mathbb{P}(E \mid S_2) = \sum_{a=1}^{N}\sum_{b=1}^{N} \binom{N}{a} p^a (1-p)^{N-a} \binom{N}{b} (1-p)^b p^{N-b} 1_E$$

$$\mathbb{P}(E \mid S_1) = \sum_{a=1}^{N}\sum_{b=1}^{N} \binom{N}{a} (1-p)^a p^{N-a} \binom{N}{b} p^b (1-p)^{N-b} 1_E$$

$$\mathbb{P}(E \mid S_4) = \sum_{a=1}^{N}\sum_{b=1}^{N} \binom{N}{a} (1-p)^a p^{N-a} \binom{N}{b} (1-p)^b p^{N-b} 1_E$$

1_E denotes the 0-1 indicator function of the event E. Assume now that S_1 is true. When we want to calculate the probability that the Bayesian decision procedure gets the *conclusion* right, we compute the probability that M is smaller than 2, given the prior distribution:

$$\mathbb{P}(+S_1 \mid S_1) = \mathbb{P}(M < 2 \mid S_1)$$

which can be calculated numerically according to the probability densities given above. Analogously,

$$\begin{aligned}
& \mathbb{P}(-S_1 \mid \neg S_1) \\
= \; & \mathbb{P}(-S_1 \mid S_2 \vee S_3 \vee S_4) \\
= \; & \frac{1}{1-q^2}\left[q(1-q)\mathbb{P}(-S_1|S_2) + q(1-q)\mathbb{P}(-S_1|S_3) + (1-q)^2\mathbb{P}(-S_1|S_4)\right] \\
= \; & \frac{1}{1-q^2}\left[2q(1-q)\mathbb{P}(M \geq 2|S_2) + (1-q)^2\mathbb{P}(M \geq 2|S_4)\right]
\end{aligned}$$

so that we can compute $\mathbb{P}(G)$.

A.4 Bayesian asymptotics

Proof of Proposition 3: It is clear that for $p = 0.5$, the reliability of the Bayesian decision rule is constant in N, because the judgments of the voters do not affect the posterior distribution. Two cases remain to examine.
(a) $p < 0.5$. Assume S_1 is true. Then $(1-p)/p > 1$ and in the long run, almost certainly $a, b < N/2$ which entails $L_{21}(a,b) = ((1-p)/p)^{2b-N} \to 0$. Similar considerations hold for L_{31} and L_{41} as well as for the case that another situation is true.
(b) $p > 0.5$. The proof is similar: Assume S_1 is true. Then $(1-p)/p < 1$ and in the long run, almost certainly $a, b > N/2$ which entails $L_{21}(a,b) = ((1-p)/p)^{2b-N} \to 0$, etc. $\square$

References

[1] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Second Edition. Springer, New York, 1985.

[2] L. Bovens and W. Rabinowicz. Democratic answers to complex questions. An epistemic perspective, *Synthese*, 150: 131–153, 2006.

[3] F. Dietrich and C. List. Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, 2006.

[4] S. Hartmann, G Pigozzi, and J. Sprenger. Judgment aggregation and the problem of truth-tracking. In preparation.

[5] S. Konieczny and E. Grégoire. Logic-based approaches to information fusion. *Information Fusion*, 7: 4–18, 2006.

[6] S. Konieczny and R. Pino-Pérez. Merging with integrity constraints. In *Fifth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'99)*, pages 233-244, 1999.

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

[8] C. List. The discursive dilemma and public reason. *Ethics*, 116(2): 362–402, 2006.

[9] C. List. Judgment aggregation - A bibliography on the discursive dilemma, the doctrinal paradox and decisions on multiple propositions. 2007 *http://personal.lse.ac.uk/LIST/doctrinalparadox.htm*

[10] C. List and P. Pettit. Aggregating sets of judgments: An impossibility result. *Economics and Philosophy*, 18: 89-110, 2002.

[11] G. Pigozzi. Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. *Synthese*, 152(2): 285-298, 2006.