Judgment Aggregation and the Problem of Tracking the Truth

Stephan Hartmann,[*] Jan Sprenger[†]

May 21, 2008

Abstract

The aggregation of consistent individual judgments on logically interconnected propositions into a collective judgment on those 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 that problem as the *discursive dilemma*. In this paper, we motivate that many groups do not only want to reach a factually right conclusion, but also want to correctly evaluate the *reasons* for that conclusion. In other words, we address the problem of tracking the true situation instead of merely selecting the right outcome. We set up a probabilistic model analogous to Bovens and Rabinowicz (2006) and compare several aggregation procedures by means of theoretical results, numerical simulations and practical considerations. Among them are the premise-based, the situation-based and the distance-based procedure. Our findings confirm the conjecture in Hartmann, Pigozzi and Sprenger (2008) that the premise-based procedure is a crude, but reliable and sometimes even optimal form of judgment aggregation.

1 Introduction

Members of a group often have to express their opinions on several propositions. Examples are all kinds of committees such as expert panels, legal courts and advisory boards. The set of propositions at stake is the *agenda* of the group. 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 individually consistent judgments on logically interconnected propositions into an equally consistent group judgment has recently drawn much attention. Such problems constitute the emerging field of *judgment aggregation* (List 2007). A *judgment* 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. In this paper we compare several procedures with respect to their ability

[*]Contact information: Stephan Hartmann, Tilburg Center for Logic and Philosophy of Science, Tilburg University, 5037 AB Tilburg, S.Hartmann@uvt.nl

[†]Contact information: Jan Sprenger, Tilburg Center for Logic and Philosophy of Science, Tilburg University, 5037 AB Tilburg, jan.sprenger@gmx.net

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

Table 1: An illustration of the discursive dilemma under the constraint rule $A \wedge B \leftrightarrow C$

to *track the truth*. Thus, the resulting collective judgment should not only be factually right or wrong, but also contain a correct assessment of the *situation*. In other words, we present results concerning the ability of several methods to aggregate conflicting individual judgments into a decision that is substantiated by the correct *reasons*.

Here is an illustration. The Israeli Prime Minister asks an advisory board of military and security experts whether to engage on a military strike against an Iranian nuclear site. All board members agree that military action should be taken if and only if the Iranian nuclear program does not only serve civil, but also military purposes (represented by proposition A, referred to as the *first premise*) *and* if the Iranian Government is preparing military aggressions against Israel (represented by proposition B, referred to as the *second premise*). Taking military action is represented by proposition C, also referred to as the *conclusion*. The doctrine can be formally expressed as the formula $(A \wedge B) \leftrightarrow C$. Each member of the advisory board expresses her judgment on A, B and C such that the rule $(A \wedge B) \leftrightarrow C$ is satisfied. Suppose now that the committee has seven members who make their judgments according to table 1. We see that, although each committee member expresses a consistent opinion, propositionwise majority voting results in a majority for A and B, but in a majority for $\neg C$.[1] 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 propositionwise majority voting can lead a group of rational individuals to endorse an inconsistent collective judgment. Clearly, the relevance of such aggregation problems goes beyond the specific example: It applies to all situations where individual binary evaluations need to be combined into a judgment of the entire group.

The discursive dilemma is the point of departure for two different research programs. The first program investigates how many plausible adequacy criteria can be met by a single aggregation function. Recent results by List and Pettit (2002) and Dietrich and List (2006) indicate that Arrow's theorem on the impossibility of certain preference aggregation functions can be generalized to a theorem on the impossibility of certain judgment aggregation functions. However, we prefer to work in the second program, to take a constructive approach and to evaluate the various ways to consistently aggregate judgments from an epistemic perspective (Bovens and Rabinowicz 2006). Here, two goals have to be discerned. The most natural goal consists in selecting the factually right outcome (i.e. to take or not to take military action) regardless of whether the

[1] For reasons of simplicity, we represent propositional variables as well as realizations of those variables by capital letters in italics.

reasons for the decision are correctly discerned. Indeed, for the Israeli Prime Minister the desire to make a factually correct decision might be more important than giving the correct reasons for her decision: She has to avert a potential danger from her people and she must not expose soldiers to the risks of a futile military strike. Compared to the perils of choosing a wrong outcome, being wrong about the reasons seems to be a minor problem. In Hartmann, Pigozzi and Sprenger (2008) we pursue this route and investigate which aggregation procedure most often selects the right outcome. Sometimes, of course, right decisions are made for the wrong reasons. Suppose that Iran is actually preparing a military nuclear program, but merely with defensive intentions (e.g. in order to deter the United States from an invasion). The Isreali intelligence service erroneously comes to the opposite result – Iran has aggressive intentions, but it does not have a serious military nuclear program. The resulting decision not to take military action is factually right because the errors cancel out each other, but it would have been taken for the wrong reasons. If new evidence revealed that the Prime Minister's assessment of the situation was wrong, she would come under pressure and the trust in her competence would decline. The right decision which she took would be judged as a matter of mere luck.

This example illustrates that, although panels and committees will often merely care for factually right decisions, neglecting the reasons for a decision can have embarrassing consequences, too. A similar case already occurred: Before the Iraq war, the US Government charged Iraq with reproaches to possess and to proliferate weapons of mass destruction. When these accusations turned out to be wrong, the reputation of the US Government greatly suffered, regardless of whether the *decision* to open the war against Iraq was factually in the country's interest or not. As the reasons for the invasion of Iraq were not truthful, the relevant decision-makers were perceived either as incompetent fools or as Machiavellian liars. Therefore we sometimes need aggregation procedures that do not only select a factually right *outcome*, but also find out the right *situation*. In the judgment aggregation problem, we identify a situation with any consistent judgment on the sets of propositions at stake. The four situations that are consistent with the constraint rule $(A \wedge B) \leftrightarrow C$ are

$$S_1 = \{A, B, C\} \qquad S_2 = \{A, \neg B, \neg C\}$$
$$S_3 = \{\neg A, B, \neg C\} \qquad S_4 = \{\neg A, \neg B, \neg C\}$$

Selecting the true situation amounts to making a correct judgment on all propositions, regardless of which situation is true:

1. If S_1 were true, then S_1 would be chosen.
2. If S_2 were true, then S_2 would be chosen.
3. If S_3 were true, then S_3 would be chosen.
4. If S_4 were true, then S_4 would be chosen.

An aggregation procedure that satisfies those four conditions is said to *track the truth*. This account of truth-tracking can be brought in line with Nozick's (1981) account of knowledge and truth-tracking. According to Nozick, a method of aggregating judgments would track the truth of a situation S if and only if

Stability If S were true, then S would be chosen.

Sensitivity If S were not true, then S would not be chosen.

It is easy to see that sensitivity is satisfied whenever an aggregation procedure is stable with respect to all possible situations. Thus, the former set of conditions entails stability and sensitivity so that truth-tracking methods in our sense also track the truth of any situation in Nozick's sense.

To our mind, the problem of tracking the true situation (over and above making a factually correct decision) is important whenever institutional and individual decision-makers have to publicly justify their decisions and when they are responsible, liable or accountable for the decision which they have made. Such situations can roughly be subdivided into two different types: First, a right decision for the wrong reason often triggers costly revisions. Take the case of a job applicant who is turned down for fallacious reasons. Stating wrong reasons gives the applicant a chance to formally contest a negative decision even if she is not a suitable candidate. Similarly, in a lawsuit, a factually right decision of the court (e.g. to sentence the culprit) can be contested and revoked because the grounds for the judgment are fallacious. Such situations are *open to costly revision.*

Second, whenever decision-makers support a decision by means of their personal or institutional authority, there is the danger of *reputation loss*. In academic practice, a journal referee usually accompanies her recommendation by a list of reasons. For instance, let us assume that the referee opts for rejection because she believes the author's essential argument to be invalid. In fact, the argument is sound, but the paper has other deficits which the referee fails to notice. For example, the premises of the arguments are highly contestable or relevant literature is not taken into account. If the editor discovers that the referee's recommendation is not well substantiated (though factually correct), she might eliminate her from the journal's list of referees. The examples from politics (Israel/Iran conflict, Iraq invasion) can also be subsumed under the label of reputation loss.[2] Thus, in those two types of situations – formally contestable and reputation-intensive decisions – truth-tracking becomes an essential issue, over and above the need to make a factually right decision. Moreover, there are not only practical drawbacks – the *epistemic* evaluation of a decision is also affected: If a correct decision is generated by fallacious beliefs (e.g. because the errors cancel out each other), we will hesitate to say that this decision was *justified*. In particular, we would deny that the decision-makers *knew* which decision they had to take.

When a factually right outcome is to be selected, two major escape routes from the discursive dilemma have been suggested and compared (Bovens and Rabinowicz 2006; List 2005, 2006): the *premise-based procedure* (PBP) and the *conclusion-based procedure* (CBP). In the PBP, the committee members vote separately on each of the premises A and B. The aggregate vote on A and B, determined by simple majority voting, fixes the collective judgment on $\{A, B, C\}$, according to the constraint $(A \wedge B) \leftrightarrow C$. According to the CBP, the members decide privately on A and B and only express their opinions on C publicly. The judgment of the group is then inferred from applying the majority rule to the individual judgments on C. In the case of the example presented in table 1, the PBP recommends to conduct a military strike whereas the CBP advises the board members not to take military action. In the next section, we

[2] An overlap between both types of situations is possible, but not necessary.

construct a similar version of the discursive dilemma for the problem of tracking the truth and introduce a new procedure: the *situation-based procedure* (SBP). Then, we compare it to the distance-based procedure (Pigozzi 2006) and the PBP in terms of tracking the truth and adopt a Bayesian perspective to show the general superiority of the PBP.

2 Aggregation procedures

In the discursive dilemma, the conclusion-based procedure suppresses in general the reasons for the final decision since it only counts the votes for the conclusion. Obviously, this implies that the CBP cannot track the true situation. Due to that problem, Bovens and Rabinowicz (2006) introduce a novel procedure which we call *modified conclusion-based procedure* (MCBP). That procedure opts for situation S_i if and only if more than half of the group members support S_i (Bovens and Rabinowicz 2006, 139). In other words, situation S_i is selected if and only if an absolute majority of group members support it. Consequently, MCBP gives an indeterminate result when there is only a *relative*, but no absolute majority for a specific situation. In their calculations of the reliability of MCBP (formula (21) on page 139), Bovens and Rabinowicz treat those indeterminate cases as cases in which MCBP gives a wrong result. Since indeterminate decisions count as fallacious decisions, the reliability of MCBP is quite poor – in particular, it is always outperformed by PBP. Due to the aforementioned indeterminacy, this result is not surprising. So, instead of appraising PBP prematurely, Bovens and Rabinowicz should amend MCBP to a well-defined procedure without indeterminate cases and compare *that* procedure to PBP. Unlike the comparison in their paper, this would be a fair contest. We close that gap in this paper. Besides, Bovens and Rabinowicz do not make explicit that MCBP is *not* the original conclusion-based procedure. The original conclusion-based procedure CBP is unable to discern between S_2, S_3 and S_4 whereas MCBP can sometimes distinguish between those situations (namely, when they are supported by an absolute majority of group members). Therefore, it is misguided to compare the relative performance of the conclusion-based procedure in the right outcome problem vs. the right situation problem (Bovens and Rabinowicz 2006, 141-145) – actually, there are two conclusion-based procedures at stake and they are quite different.

For tracking the truth and making a right decision for the right reasons, the form of the logical constraint rule (so far: $(A \wedge B) \leftrightarrow C$) does not matter: When the reasons for a decision are correctly discerned, the conclusion is correctly discerned, too, regardless of the exact form of the logical constraint rule. For instance, when we have correctly assessed the truth values of A and B, we have also correctly assessed the truth value of $A \vee B$, $A \rightarrow B$, etc. Thus, aggregation procedures for tracking the truth equally apply to all conclusions that are truth-functional compounds of A and B.

Maybe the most natural way to make a judgment on the right situation consists in voting separately on the premises, as in the original case of judgment aggregation. However, the premise-based procedure has some awkward consequences. Suppose that in a panel of $2N + 1$ voters, there is a tie between S_1 ($A/B/C$) and S_4 ($\neg A/\neg B/\neg C$) – S_1 and S_4 are each supported by N voters. In other words, there is maximal disagreement between the first $2N$

group members. Now, the last voter is in a *pivotal* position – the collective judgment is sensitive to her vote. Even worse, *the group judgment just copies her judgment.* In other words: If she votes for S_1, the group endorses S_1, if she votes for S_2, the group endorses S_2, etc. It seems that this voter has obtained too much relative weight. In particular, a collective decision for S_2 can result even if this situation is merely supported by a single voter. This contradicts the intuition that the chosen situation should be backed by more than a single voter, especially in large groups. Pragmatic factors often require the panel to support the group judgment (e.g. when standing up to the public), and the premise-based procedure allows for situations where almost no one agrees with the group judgment. Moreover, even the implementation of the PBP may be difficult under some circumstances. For any kind of decisions that take a stand on more than one issue (e.g. on A and B), formal requirements can make it difficult to schedule two separate ballots (one on A and one on B, as the PBP demands). For instance, a board may have to decide between different motions that assert something about A and B, and splitting the motions may be formally impossible. This asks for a procedure where the collective judgment is gained in a single ballot, i.e. not by a ballot on the premises, but by a ballot on entire *situations*. Such procedures are *situation-based* because they are based on judgments on entire situations, not on single propositions. There, the collective judgment is a function of the number of voters that support S_1, the number of voters that support S_2, etc. Among the possible situation-based aggregation methods, we focus on a plurality-based method: the situation which receives the highest number of votes is accepted. This corresponds to a widely applied practical convention that, when different proposals compete against each other, the one with the highest number of votes goes through. We call this procedure the *situation-based procedure* (SBP). Note that the SBP extends Bovens and Rabinowicz's MCBP: It agrees with MCBP if there is an absolute majority for a specific situation, but it also yields a definite result in all other cases.

To define the SBP properly, we have to introduce some terminology. In a group of N persons, there are n_1 persons who back S_1 (i.e. they judge A, B and C to be true), n_2 persons who back S_2, and so on. Every member of the group supports exactly one of these situations (corresponding to her judgments on A and B). Hence, $n_1 + n_2 + n_3 + n_4 = N$. Now, the premise-based procedure chooses the situation where the truth value assignments to the premises are backed by a majority of voters, i.e. a decision for S_1 is made if and only if the following two inequalities are satisfied:

$$n_1 + n_2 > n_3 + n_4 \qquad\qquad n_1 + n_3 > n_2 + n_4 \tag{1}$$

The necessary and sufficient criteria for accepting another situation are analogous (see appendix). Since N is assumed to be an odd number, equality can never occur. Hence, PBP works in the same way as in the case of choosing the right outcome. By contrast, SBP opts for S_1 if the function χ_1, defined below, returns value 1. Ties are resolved by chance experiments: with equal probability, one of the situations with the highest number of votes will prevail. More refined tie-breaking rules are, of course, possible and lead to different versions

	A	B	C	Situation
Voter 1	Yes	Yes	Yes	S_1
Voter 2	Yes	Yes	Yes	S_1
Voter 3	Yes	Yes	Yes	S_1
Voter 4	No	Yes	No	S_3
Voter 5	No	Yes	No	S_3
Voter 6	No	No	No	S_4
Voter 7	No	No	No	S_4
Majority	No	Yes	No	S_1

Table 2: An illustration of the discursive dilemma when the right situation is to be selected.

of the situation-based procedure.[3]

$$\chi_1(n_1, n_2, n_3, n_4) = \begin{cases} 1 & n_1 > \max(n_2, n_3, n_4) \\ 1/2 & (n_1 = n_2) \wedge (n_1 > \max(n_3, n_4)) \\ 1/2 & (n_1 = n_3) \wedge (n_1 > \max(n_2, n_4)) \\ 1/2 & (n_1 = n_4) \wedge (n_1 > \max(n_2, n_3)) \\ 1/3 & (n_1 = n_2) \wedge (n_1 = n_3) \wedge (n_1 > n_4) \\ 1/3 & (n_1 = n_2) \wedge (n_1 = n_4) \wedge (n_1 > n_3) \\ 1/3 & (n_1 = n_3) \wedge (n_1 = n_4) \wedge (n_1 > n_2) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

It is interesting to see under which circumstances both procedures disagree. If the votes are distributed according to table 2, the PBP selects situation S_3 whereas the SBP selects S_1: S_1 is supported by three voters whereas situations S_3 and S_4 obtain only two votes each, and no one supports situation S_2. Thus, there is a simple majority for S_1, and PBP and SBP fall apart. That example creates a new version of the discursive dilemma: In the traditional problem of selecting the right outcome under the aggregation rule $A \wedge B \leftrightarrow C$, the premise-based procedure sometimes opts for the conclusion C when the conclusion-based procedure opts for $\neg C$. In a similar vein, when we try to track the truth and to select the right situation, the situation-based procedure and the premise-based procedure can equally opt for different situations, as witnessed by table 2.[4] Hence it is a non-trivial question which procedure we should adopt when

[3]Note that the SBP is strategy-proof when truth is the common good of the group members and not strategy-proof when verisimilitude (coming to the truth as close as possible) is the common good. Consider the example of table 2 – three voters vote for S_1, two for S_3 and two for S_4. The SBP will then select S_1. Now, the voters who believe S_4 to be the right situation could reason that S_3 is closer to the (allegedly) true situation S_4 than the situation S_1 which is backed by most voters. If they already knew the votes of the other group members, they could decide to conceal their real opinion and to vote for the lesser evil S_3 – S_1 is more distant from the putative truth than S_3. Note that voters 6 and 7 did not have any egoistic motivation – they voted insincerely because they wanted to promoted the common good of the group, namely verisimilitude. Only if full truth is the goal (i.e. only if the degree of error does not matter), SBP becomes strategy-proof. By contrast, PBP is always strategy-proof – no voter has a reason to conceal his true opinions if she wants the truth to be detected by the aggregate judgment.

[4]Note that in the original version of the discursive dilemma where the goal consists in

aiming at the true situation. Which aggregation procedure should the Isreali Prime Minister use? On which procedure should committee members who care for a correct decision agree? In order to respond to those questions, the next section introduces a probabilistic model which allows to assess the epistemic performance of both procedures and to compare it to the distance-based procedure (DBP) (Konieczny and Pino-Pèrez 1999; Konieczny and Grégoire 2006; Pigozzi 2006). In the example of table 2, DBP would average all individual votes into the vector (5/7 yes, 5/7 yes, 3/7 yes) which is closest to the situation S_1 = (yes, yes, yes).

3 Comparing the procedures

3.1 Preliminaries

In order to investigate the epistemic reliability of the various aggregation procedures, we adopt a probabilistic framework. In particular, we assign to every group member an individual competence $p \in (0, 1)$ to make a correct judgment on a single premise. More precisely, when a premise (either A or B) is true, the voter gives a correct report with probability p_+, and equally, if the premise is false, the voter gives a correct report with probability p_-. Then, the overall competence of the voter to evaluate the truth value of the premise A can be calculated as

$$p \quad = \quad p_+ \, \mathbb{P}(A) + p_- \, (1 - \mathbb{P}(A)) \tag{3}$$

Moreover, we assume that the voters are not biased towards the truth or the falsity of a proposition: $p := p_+ = p_-$. In other words, the probability of a false positive report on a premise equals the probability of a false negative report. Thus, the competence of an individual voter (p) is decoupled from the prior probability of the propositions at stake, and the frequency with which she makes correct or incorrect judgments does not depend on the truth values of the propositions in question.[5] Now, 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 ($N \to \infty$). The Condorcet Jury Theorem thus offers an epistemic justification to majority voting and motivates the use of the premise- and conclusion-based procedure in the judgment aggregation problem (Bovens and Rabinowicz 2006). It should be noted, though,

selecting the right outcome, the votes of table 2 would lead to *agreement* of PBP and CBP.

[5]Setting $p_+ = p_-$ also answers List's concerns (List 2006) that for a very low value of p_+ or p_-, the voters are bad at tracking the true situation since one of Nozick's subjunctive conditionals (see page 3) would be violated. This violation can occur even if the overall reliability p, as defined in (3), is high.

Note that we ascribe an individual competence only for voting on *premises*, not for voting on any truth-functional compound of A and B (such as $A \wedge B$). In many contexts it is reasonable to assign individual voting competence to 'elementary' propositions only. For instance, in the introductory example, propositions as 'Iran's nuclear program serves military purposes' or 'Iran plans a military aggression against Israel' naturally play the role of such elementary propositions, rather than 'Iran has a military nuclear program and aggressive intentions'.

that an application of the Condorcet results to judgment aggregation requires further assumptions in order to reduce computational complexity, in particular

1. A and B are logically independent.
2. All voters have the same (independent) competence to assess the truth of A and B (p). Their judgments on A and B are independent.
3. Each individual judgment set is logically consistent.

Notably, we are allowed to dismiss the original conditions of Bovens and Rabinowicz (2006) and Hartmann, Pigozzi and Sprenger (2008) who also demand that the prior probabilities of A and B are equal ($\mathbb{P}(A) = \mathbb{P}(B)$) and that A and B are probabilistically independent. The various symmetry properties of our procedures entail that for PBP, its reliability $\mathcal{R}^{\text{PBP}}$ is equal to

$$\begin{aligned} \mathcal{R}^{\text{PBP}} &= \sum_{i=1}^{4} \mathbb{P}(S_i)\, \mathbb{P}(\text{PBP opts for } S_i|S_i) \\ &= \mathbb{P}(\text{PBP opts for } S_1|S_1) \end{aligned} \tag{4}$$

and similarly for SBP (see the theorem in subsection 3.3). In other words, the reliability of PBP and SBP does not depend on which situation is actually the case and is independent of the prior probability of the various situations. The reliability of PBP and SBP is thus only a function of N and p, which greatly simplifies the calculations.

3.2 Simulations

Now we proceed to a numerical comparison of the premise-based and the situation-based procedure. First, we plot the reliability (=truth-tracking ability) of PBP and SBP as functions of p, for fixed N (see figure 1).

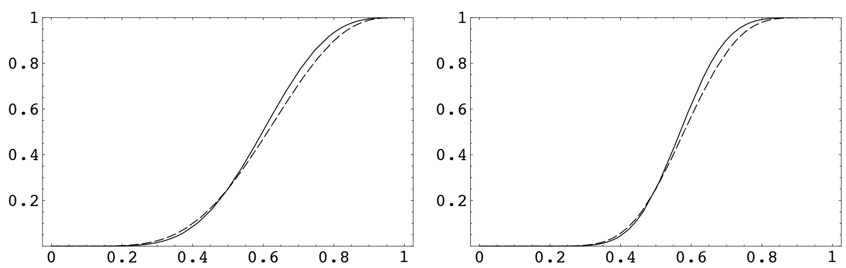

Figure 1: Reliability of the PBP (full line) and the SBP (dashed line) as a function of p, for $N = 7$ (left figure) and $N = 15$ (right figure).

Both graphs resemble Condorcet curves – for an individual competence significantly higher than 0.5 both curves rise steeply until the reliability approaches 1. Indeed, it can be proven that for increasing group size, the reliability of both procedures approaches 1 for voters that are better than randomizers:

Proposition 1. *For $N \to \infty$, the premise-based and the situation-based procedure select the right situation $\mathbb{P}$-almost surely ($\mathbb{P}$-a.s.) if and only if $p > 0.5$.*

So for very large groups of voters, the epistemic performances of PBP and SBP will resemble each other. However, this is an asymptotical property, and for small to medium-sized groups, the PBP will usually perform better, in agreement with the remarkable performance of PBP at selecting the right outcome (cf. Hartmann, Pigozzi and Sprenger 2008). Moreover, both curves intersect at $p = 0.5$. So it is tempting to conjecture that the reliability of PBP always exceeds the reliability of SBP for voters that are better than randomizers ($p > 0.5$) whereas it is the other way round for voters that are worse than randomizers ($p < 0.5$). Indeed, we later give an analytic argument for this claim. However, the performances of the two procedures are already quite close to each other for medium-scaled group size (e.g. $N = 7, 15$), as figure 1 shows. By the way, numerical simulations indicate stability if the tie-breaking rules in SBP are altered: no significant changes occur.

Now we consider a special case of the aggregation problem, namely to detect the correct situation under the logical constraint rule $(A \wedge B) \leftrightarrow C$. For this constraint rule, we compare our two procedures PBP and SBP to the distance-based procedure (DBP) known from Pigozzi (2006). The distance-based procedure selects the collective outcome that has the lowest distance to the averaged vote of the group members. Unlike PBP and SBP, DBP is sensitive to a choice of the prior probabilities of the situations. Following Bovens and Rabinowicz (2006) and Hartmann, Pigozzi and Sprenger (2008), we parametrize the set of prior distributions by $q \in [0, 1]$ where

$$\mathbb{P}(S_1) = q^2; \mathbb{P}(S_2) = \mathbb{P}(S_3) = q(1-q); \mathbb{P}(S_4) = (1-q)^2 \qquad (5)$$

For a defense of this parametrization, see the above sources. It incorporates

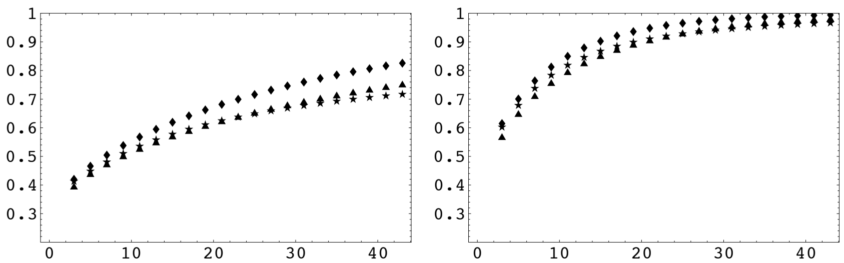

Figure 2: Reliability of the PBP (diamonds), the SBP (triangles) and the DBP (stars) as a function of N if each situation is equally likely ($q = 0.5$). The left figure shows $p = 0.6$, the right figure shows $p = 0.7$.

probabilistic independence between A and B as well as the assumption that A and B are a priori equally likely. First, we have plotted the reliability of the three procedures as a function of N for the default prior probability $q = 0.5$ (where each situation is equally likely). Besides, we have assumed standard values for the voters' individual competence, namely $p = 0.6$ and $p = 0.7$. This yields the result shown in figure 2 – DBP's performance comes close to PBP's performance for low values of N, but then it gradually deteriorates and eventually falls below the performance of SBP. For the former case ($p = 0.6$), an even stronger claim can be made: The reliability of DBP will eventually *not*

	S_1 true	S_2 true	S_3 true	S_4 true
select S_1	1	0	0	0
select S_2	0	1	0	0
select S_3	0	0	1	0
select S_4	0	0	0	1

Table 3: A standard utility matrix for the judgment aggregation problem, shown as a function of the possible actions and situations.

converge against 1 with increasing N – this is only the case for $p > (\sqrt{5}-1)/2$.[6] Apart from that, the novel plots confirm the findings from figure 1: the PBP performs somewhat better than the SBP other while the margin in favor of PBP declines with increasing individual competence p. To quantify the effect

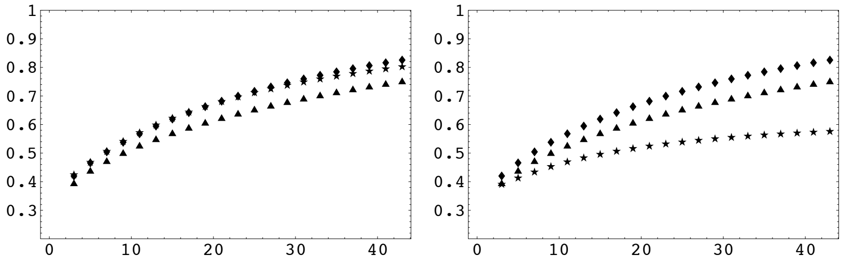

Figure 3: Reliability of the PBP (diamonds), the SBP (triangles) and the DBP (stars) for fixed $p = 0.6$ as a function of N. The left figure shows $q = 0.3$, the right figure shows $q = 0.7$.

of the prior probability q on the performance of the distance-based procedure, we have varied this parameter to $q = 0.3$ and $q = 0.7$ for fixed $p = 0.6$ (figure 3). The curves of PBP and SBP are not sensitive to a change in q (remember the remark at the end of section 2) and correspond to the curves in the left graph of figure 2. But DBP is remarkably good for $q = 0.3$ – it is roughly as reliable as PBP – and remarkably bad for $q = 0.7$, except for very small N. To explain that difference, note that DBP has a certain bias against the situation $S_1 = (A, B, C)$. The lower q, the less likely this situation and the more does the bias of DBP against S_1 improve its performance. In total, the figures teach us that for small values of q, when the board should be cautious with respect to S_1, DBP is a reasonable means of incorporating this caution. But for medium and large values of q, DBP should be rejected in favor of the premise-based procedure.

3.3 A Bayesian perspective

Let us now return to the introductory example where the Israeli Prime Minister consults an expert board. She might find comfort in our result that adopting the collective judgment that was gained by majority voting on the premises is a good strategy: The performance of PBP is reasonable for nearly all parameter values. Even if the individual competence p and the prior probabilities q are not

[6]This property of DBP is shown in proposition 2, section 3 of Hartmann, Pigozzi and Sprenger (2008).

known or hard to estimate, the PBP provides a good mechanism for tracking the true situation. But is it also the optimal way to proceed? Actually, the Prime Minister might have some surplus information, e.g. she might know the reliability of the members of the security board or know the 'objective probabilities' of A and B. This information has to affect her interpretation of the board members' judgment. The Prime Minister is then best perceived as a Bayesian updater: she updates her prior beliefs, expressed by $P(S_i)$, $i \in \{1, \ldots, 4\}$, on the votes of the group members. The judgment profiles of the board members are incoming evidence E which she uses to update her prior beliefs to a *posterior distribution* over the $(S_i)_{i \leq 4}$:

$$\mathbb{P}(S_i|E) = \frac{\mathbb{P}(S_i)\,\mathbb{P}(E|S_i)}{\sum_{i=1}^{4} \mathbb{P}(S_j)\,\mathbb{P}(E|S_j)}$$

This posterior distribution describes the rational degree of belief in the various situations, given the verdicts of the voters and their individual reliability. Now we make the simplifying assumption that each right judgment has the same utility, say 1 (i.e. correctly tracking S_1 is as valuable as correctly tracking S_2, etc.). Furthermore, each wrong judgment has the same utility, say 0 (i.e. missing the true situation S_1 in favor of S_2 is as harmful as missing S_1 in favor of S_3, etc.). These assumptions are made explicit in table 3.[7] Bayesian situation selection is now solely based on the posterior probabilities of the situations given the evidence and the utility matrix. Such decision rules are *Bayes rules*, i.e. they minimize the expected risk with regard to the prior distribution among all decision rules, see Result 1 in Berger (1985, 159). In particular, it is not possible to do better than an expected-utility-maximizing procedure based on the posterior probability distribution. For the utility matrix of table 3, a risk-minimizing decision rule has to select the situation with the highest posterior probability among all situations. This situation need not coincide with the one favored by PBP or SBP.

It is quite natural that Bayesian updating outperforms PBP, SBP and DBP since it may use additional information, e.g. the reliabilities of the board members. But what looks as a virtue can actually be a vice. Bayesian updating is in general only superior if the decision-maker's prior beliefs correspond to 'objective probabilities' in the world (e.g. the relative frequency of the occurrence of A compared to the occurrence of $\neg A$). However, in general we cannot elicit those probabilities. Then, a Bayesian approach has no advantage over the crude premise-based or situation-based procedures. Moreover, the applicability of the Bayesian approach is restricted since it takes an *external perspective*: Admittedly, a benevolent dictator will typically behave like a Bayesian: she has her own prior beliefs, updates them on the judgments of an advisory board and her estimates of the individual reliability of the board members. The final decision is then based on her posterior distribution and her personal utility matrix. But if the board has to make a decision itself, the Bayesian model does not apply. Once the group members have deliberated and agreed on a probability distribution for S_1-S_4, there is no need to update it on individual votes since all relevant information was already contained in that distribution. Still, the

[7] Our considerations are invariant under the question whether 'utility' refers to the practical or the epistemic value of an action in a specific situation. For both applications, the identity matrix seems to be the default utility matrix.

Bayesian model can serve as a *benchmark* for the other procedures. So far, we have evaluated the *relative* performance of PBP, SBP and DBP with respect to each other. Since Bayesian updating is theoretically optimal if full information (on p and q) is available, we might now check whether and when other procedures coincide with the Bayesian approach.

Here, it is remarkable that the judgments of the Bayesian procedure correspond to the judgments of PBP as long as $q = 0.5$ and $p > 0.5$: When all situations are equally likely the Bayesian procedure selects the situation which makes the actual results most likely. For $p > 0.5$, this is the situation where the judgments on the premises are backed by a majority of voters (details in the appendix). Hence, the Bayesian and the premise-based procedure agree, vindicating the theoretical optimality of PBP for this special case. Actually, this result can be generalized: PBP uniformly outperforms all aggregation procedures that are not inclined towards specific situations, regardless of the prior probabilities of the S_i.

Definition 1. *A* judgment profile J *is a* $2N$*-tuple* $(a_1, b_1, a_2, b_2, \dots, a_N, b_N)$ *where the* a_i *and* b_i *are binary variables with values in* $\{Y(es), N(o)\}$*. The set of possible judgment profiles is denoted by* $\mathcal{J} = \{Y, N\}^{2N}$*.*

Definition 2. *An* aggregation function *(aggregation procedure)* f *is a function from the sets of judgments profiles and the sets of situations to the unit interval (i.e.* $f : \mathcal{J} \times \{S_1, S_2, S_3, S_4\} \to [0, 1]$*) so that for any* $(a_i, b_i)_{1 \leq i \leq N}$*:*

$$\sum_{j=1}^{4} f(a_1, b_1, \dots, a_N, b_N; S_j) = 1.$$

We interpret the value of $f(J, S_i)$ *as the probability with which, for a judgment profile* J*, situation* S_i *is selected. The aggregation function is* random *if there is a judgment profile* J *and a situation* S_i *so that* $f(J, S_i) \notin \{0, 1\}$*.*

Definition 3. *An aggregation function* f *is* unbiased *if and only if the following equations hold:*

$$\begin{aligned} f(a_1, b_1, \dots, a_N, b_N; S_1) &= f(a_1, \neg b_1, \dots, a_N, \neg b_N; S_2) \\ f(a_1, b_1, \dots, a_N, b_N; S_1) &= f(\neg a_1, b_1, \dots, \neg a_N, b_N; S_3) \\ f(a_1, b_1, \dots, a_N, b_N; S_1) &= f(\neg a_1, \neg b_1, \dots, \neg a_N, \neg b_N; S_4) \end{aligned} \tag{6}$$

Unbiasedness means that the voters are not biased in favor of a specific situation and only listen to the evidence, or in other words, the judgments of the voters are impartial with respect to the names of the situations. In particular, a judgment profile that champions S_1 would, if all judgments were inverted, speak for S_4 to an equal degree.

Fact 1. *PBP and SBP (but not DBP) are unbiased aggregation functions.*

With these definitions at hand, we have an analytic way to demonstrate the superiority of PBP which was indicated by the numerical simulations.

Theorem. *Let the group members be individually be more competent than randomizers (i.e.* $p > 0.5$*). The premise-based procedure PBP is better at tracking* *the true situation than any other unbiased procedure, for any value of N and p, and any prior probability distribution over the possible situations.*

In other words, if the group members are not too incompetent, PBP is the best truth-tracking procedure among all aggregation procedures that are just functions of the judgment profiles and do not have an inclination towards a specific situation. In particular, it is superior to SBP and related procedures. This is the one lesson which the Bayesian perspective teaches us. The second lesson is the useful distinction between internal and external decision-making: Bayesian updating gives a model for making optimal use of the judgments of an advisory board, but not for decision-making inside such a board. The next and final section of the article summarizes the results.

Remark. *The theorem implies that for $p > 0.5$, PBP always outperforms SBP. But the conjecture that PBP is always superior to SBP is wrong. Quite to the contrary, SBP is more reliable than PBP for $p < 0.5$ which can be proven analogously to the above theorem. In other words, if the jury members are severely biased against the truth, SBP outperforms PBP.*

4 Discussion and Summary

In this paper, we have approached the problem of judgment aggregation from an epistemic perspective and focussed on the problem of tracking the truth and selecting the right situation, over and above selecting the right outcome. First we have motivated that there are relevant circumstances where truth-tracking becomes an issue, namely situations where costly revisions or loss of reputation threaten. Second, we have argued against Bovens and Rabinowicz's (2006) analysis of the truth-tracking problem. Third, we have proposed new alternatives to the premise-based procedure (PBP), among them the situation-based procedure (SBP) and compared them to the distance-based procedure (DBP) and a Bayesian approach. We have set up a probabilistic model to evaluate the various procedures.

The model adopts a Bayesian perspective to calculate and to compare the truth-tracking abilities of PBP, SBP and DBP. Numerical simulations indicate that the PBP outperforms the distance-based procedure as long as the prior probability of the premises is not too low. More importantly, we have shown that the PBP uniformly outperforms a large class of procedures that satisfy some plausible unbiasedness requirements, among them the SBP. For a specific prior distribution ($q = 0.5$), the PBP is even theoretically optimal at tracking the truth. From an externalist perspective, agents are thus justified to apply PBP because reliable truth-tracking methods justify the conclusions which they yield. Nevertheless, if more information is available, e.g. if an external decision-maker knows the individual reliabilities of the board members, she will abandon the PBP in favor of straightforward Bayesian updating. If a board is to make a decision itself, this is not applicable, however. In such cases, PBP's stunning performance suggests its application.

Hence, the surprising power of PBP found in the case of outcome selection (Hartmann, Pigozzi and Sprenger 2008) transfers to the problem of situation selection, too. On the other hand, numerical results indicate that the SBP is often only marginally less reliable than PBP. This is a substantial achievement:

First, practical reasons may speak against implementing PBP – it might be impossible to cast ballots for the premises. Second, we sometimes require that a collective judgment is supported by more than a faction of the entire group which is not guaranteed by PBP. Under those circumstances, SBP provides a decent alternative although it is not strategy-proof. From a purely epistemic perspective, PBP should always be favored over SBP; but the loss of reliability when using SBP is tolerable and sometimes even marginal.

Some questions remain open, among them the behavior of the investigated procedures when the strong assumptions of our model are relaxed. For instance, in Hartmann, Pigozzi and Sprenger (2008) we relax the independence assumptions and allow for opinion leaders as well as for correlation in the judgments on the various premises. Doing those calculations for the truth-tracking problem would go beyond the scope of this paper, but it is a fruitful and important avenue for further research.

A Details for the premise-based procedure

Let '$+S_i^{\text{PBP}}$' denote a decision for situation S_i under the premise-based procedure, $i \in \{1, 2, 3, 4\}$. Selecting a situation comes with a judgment on the truth value of the premises A and B. The premise-based procedure opts for a situation where those judgments are backed by a majority of voters, i.e.

$$\begin{aligned}
+S_1^{\text{PBP}} \quad &\Leftrightarrow \quad n_1 + n_2 > n_3 + n_4 \ \wedge \ n_1 + n_3 > n_2 + n_4 \\
+S_2^{\text{PBP}} \quad &\Leftrightarrow \quad n_1 + n_2 > n_3 + n_4 \ \wedge \ n_1 + n_3 < n_2 + n_4 \\
+S_3^{\text{PBP}} \quad &\Leftrightarrow \quad n_1 + n_2 < n_3 + n_4 \ \wedge \ n_1 + n_3 > n_2 + n_4 \\
+S_4^{\text{PBP}} \quad &\Leftrightarrow \quad n_1 + n_2 < n_3 + n_4 \ \wedge \ n_1 + n_3 < n_2 + n_4
\end{aligned}$$

We assume $N = n_1 + n_2 + n_3 + n_4$ to be odd. So equality can never occur. Now it is interesting to note that the probability to choose S_i ('$+S_i^{\text{PBP}}$') given that S_i is the case is constant in i. Given that S_i is true, PBP makes such a decision if and only if there are at least $(N+1)/2$ correct votes on premise A and there are at least $(N+1)/2$ correct votes on premise B. The number of correct votes on each premise is binomially distributed according to $B_{N,p}$ and does not depend on i. More precisely,

$$\mathbb{P}(+S_i^{\text{PBP}}|S_i) = \left(\sum_{k=(N+1)/2}^{N} \binom{N}{k} p^k (1-p)^{N-k} \right)^2 \tag{7}$$

(cf. equations (5) and (10) in Bovens and Rabinowicz (2006)). This establishes that the probability of tracking the true situation ($\mathcal{R}^{\text{PBP}}$) is independent of which situation is actually true:

$$\begin{aligned}
\mathcal{R}^{\text{PBP}} \quad &= \quad \sum_{i=1}^{4} \mathbb{P}(S_i)\mathbb{P}(+S_i^{PBP}|S_i) \\
&= \quad \left(\sum_{k=(N+1)/2}^{N} \binom{N}{k} p^k (1-p)^{N-k} \right)^2
\end{aligned}$$

In other words, the PBP's reliability is independent of the prior distribution over the S_i.

B Details for the situation-based procedure

The definition of the SBP and the proof of the theorem (see appendix C) entail that the reliability of SBP is independent of prior probabilities of the various situations. Therefore, the probability of tracking the truth $\mathcal{R}^{\text{SBP}}$ is equal to $\mathbb{P}(+S_1^{\text{SBP}}|S_1)$. Now we introduce some random variables. a is the number of votes for A and under S_1, it is distributed according to the binomial distribution $B_{N,p}$. b_1 counts the number of votes for premises A and B, and its distribution is $b_1 \sim B_{a,p}$. Finally, b_2 counts the number of votes for B that do not endorse A, and it is distributed according to $B_{N-a,p}$.[8] Clearly, $b_1 = n_1$, $a - b_1 = n_2$ so that

$$\chi(n_1, n_2, n_3, n_4) \quad = \quad \chi(b_1, a - b_1, b_2, N - a - b_2)$$

By paying attention to the definition of χ, we can now express the reliability of SBP and eliminating some superfluous terms in order to save computation time:

$$\begin{aligned}
\mathcal{R}^{\text{SBP}} &= \mathbb{P}(+S_1^{\text{SBP}}|S_1) \\
&= \sum_{a=0}^{N} \binom{N}{a} p^a (1-p)^{N-a} \sum_{b_1=0}^{a} \binom{a}{b_1} p^b (1-p)^{a-b_1} \\
&\quad \sum_{b_2=0}^{N-a} \binom{N-a}{b_2} p^b (1-p)^{N-a-b_2} \chi(b_1, a-b_1, b_2, N-a-b_2) \\
&= \sum_{a=[N/4]+1}^{N} \sum_{b_1=[a/2]}^{a} \sum_{b_2=N-a-b_1}^{b_1} \binom{N}{a} \binom{a}{b_1} \binom{N-a}{b_2} \\
&\quad p^{a+b_1+b_2} (1-p)^{2N-a-b_1-b_2} \chi(b_1, a-b_1, b_2, N-a-b_2)
\end{aligned}$$

This term is the basis of the SBP-plots in section 3 of the paper.

C Proofs

Proof of the Theorem: The proof proceeds in two steps: First, we show that the reliability of any unbiased aggregation procedure is independent of the prior probability distribution over (S_1, S_2, S_3, S_4). Second, we show with the help of a Bayesian argument that for a particular prior distribution, the PBP outperforms all other unbiased procedures. Due to the first part of the proof, this entails that PBP is the best among all unbiased procedures.

[8] Note that b_1 and b_2 are independent so that their sum is distributed as $b_1 + b_2 \sim B_{a,p} \bigstar B_{N-a,p} = B_{N,p}$, vindicating that the number of votes for B has the same distribution as the number of votes for A.

Let $ran(f)$ be the *range* of f and I be the 0-1-indicator function. Then,

$$\begin{aligned}
& \mathbb{P}(+S_1|S_1) \\
= & \sum_{x\in ran(f)} x\, \mathbb{P}(f(a_1,b_1,\ldots,a_N,b_N;S_1)=x|S_1) \\
= & \sum_{x\in ran(f)} x \sum_{(a_1,b_1,\ldots,a_N,b_N)\in\mathcal{J}} \mathbb{P}(a_1,b_1,\ldots,a_N,b_N|S_1)\, I_{f(a_1,b_1,\ldots,a_N,b_N;S_1)=x} \\
= & \sum_{x\in ran(f)} x \sum_{(a_1,b_1,\ldots,a_N,b_N)\in\mathcal{J}} \mathbb{P}(a_1,\ldots,a_N|S_1) \prod_{b_{i_l}=Y} \mathbb{P}(b_{i_l}|S_1) \prod_{b_{i_l}=N} \mathbb{P}(b_{i_l}|S_1) \\
& \cdot I_{f(a_1,b_1,\ldots,a_N,b_N;S_1)=x} \\
= & \sum_{x\in ran(f)} x \sum_{(a_1,b_1,\ldots,a_N,b_N)\in\mathcal{J}} \mathbb{P}(a_1,\ldots,a_N|S_2) \prod_{b_{i_l}=Y} \mathbb{P}(\neg b_{i_l}|S_2) \prod_{b_{i_l}=N} \mathbb{P}(\neg b_{i_l}|S_2) \\
& \cdot I_{f(a_1,b_1,\ldots,a_N,b_N;S_1)=x} \\
= & \sum_{x\in ran(f)} x \sum_{(a_1,b_1,\ldots,a_N,b_N)\in\mathcal{J}} \mathbb{P}(a_1,\ldots,a_N|S_2) \prod_{b_{i_l}=Y} \mathbb{P}(\neg b_{i_l}|S_2) \prod_{b_{i_l}=N} \mathbb{P}(\neg b_{i_l}|S_2) \\
& \cdot I_{f(a_1,\neg b_1,\ldots,a_N,\neg b_N;S_2)=x} \\
= & \sum_{x\in ran(f)} x \sum_{(a_1,b_1,\ldots,a_N,b_N)\in\mathcal{J}} \mathbb{P}(a_1,\neg b_1,\ldots,a_N,\neg b_N|S_2)\, I_{f(a_1,\neg b_1,\ldots,a_N,\neg b_N;S_2)=x} \\
= & \sum_{x\in ran(f)} x\, \mathbb{P}(f(a_1,\neg b_1,\ldots,a_N,\neg b_N;S_2)=x|S_2) \\
= & \mathbb{P}(+S_2|S_2).
\end{aligned}$$

In the fifth step, we have applied our assumptions. The same calculations can be done for S_3 and S_4, leading to

$$\mathbb{P}(+S_1|S_1) = \mathbb{P}(+S_3|S_3) \qquad \mathbb{P}(+S_1|S_1) = \mathbb{P}(+S_4|S_4)$$

which in turn entail

$$\mathbb{P}(+S_i|S_i) \quad = \quad \mathbb{P}(+S_j|S_j) \qquad \forall i,j \in \{1,2,3,4\}. \tag{8}$$

With the help of (8), it follows that the reliability $\mathcal{R}$ of any unbiased procedure equals

$$\mathcal{R} \quad = \quad \mathbb{P}(+S_1|S_1). \tag{9}$$

The right hand side of equation (9) is independent of the probability distribution over the situations S_i. It is merely a function of the parameters N, p and the aggregation function f. For comparing the reliability of unbiased aggregation functions we can thus presuppose without loss of generality that

$$\mathbb{P}(S_i) \quad = \quad \mathbb{P}(S_j) = 1/4 \qquad \forall i,j \in \{1,2,3,4\}.$$

In this case, we already know from section 3.3 that the best truth-tracking procedure is given by choosing the situation with the highest posterior probability. Since the prior distribution over (S_1, S_2, S_3, S_4) is uniform (see (10)), this amounts to selecting the situation S_i which makes the evidence most likely, i.e. the situation which has a positive log-likelihood ratio with respect to all other situations. The log-likelihood ratio of situation S_i to situation S_j as a function of the data (a, b) is denoted by $L_{ij}(a, b)$. a and b denote the number of votes for premise A and premise B and are binomially distributed random variables $(a, b \sim B_{N,p})$. Now, we show that the PBP always selects a situation S_i where $L_{ij}(a, b) > 0$ for all $j \neq i$. Without loss of generality, let $a, b > N/2$ (the other cases are analogous). Then S_1 is selected by PBP and indeed,

$$\begin{aligned}
L_{12}(a,b) &= \log\left(\frac{p}{1-p}\right)^{2b-N} > 0 \\
L_{13}(a,b) &= \log\left(\frac{p}{1-p}\right)^{2a-N} > 0 \\
L_{14}(a,b) &= \log\left(\frac{p}{1-p}\right)^{2a+2b-2N} > 0.
\end{aligned}$$

Thus, the premise-based procedure is optimal for $q = 0.5$ and it cannot be outperformed by any other unbiased procedure. $\Box$

Proof of Proposition 1: The asymptotic behavior of the PBP and DBP has already been examined – see proposition 2 and the appertaining proof in Hartmann, Pigozzi and Sprenger (2008). Hence, we have to show that the SBP eventually becomes perfectly reliable for $p > 0.5$ and $N \to \infty$. Assume that S_1 is true (this does not lead to a loss of generality due to $\mathcal{R}^{\text{SBP}} = \mathbb{P}(+S_1^{\text{SBP}}|S_1)$, see the proof of the theorem). We will not care for the tie-break rules and show the stronger claim that for increasing N, $\mathbb{P}$-a.s. $n_1 > \max(n_2, n_3, n_4)$. Again, we take a (the number of votes for premise A) and b (the number of votes for premise B) as given. Then x_1 – the number of votes for S_1 – is hypergeometrically distributed, with parameters N, b and a. With increasing N, the distribution of x_1 becomes increasingly skewed around its mean value $(ab)/N$ (proof omitted). This implies that $\mathbb{P}$-a.s. eventually $x_1 > a - x_1$ since

$$\frac{ab}{N} > a - \frac{ab}{N} \qquad \Leftrightarrow \qquad 2b > N.$$

Due to the Strong Law of Large Numbers, $2b > N$ will eventually be satisfied $\mathbb{P}$-a.s. so that $x_1 > a/2$ and S_1 will obtain more votes than S_3. The same argument can be used to show that S_1 will eventually obtain more votes than S_2. Finally, since $a/N, b/N \to p > 0.5$, the SLLN entails that there are in the limit more votes for S_1 than for S_4 ($\mathbb{P}$-a.s.).[9] $\Box$

References

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

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

[9] It is straightforward to see that for $p < 0.5$, the SBP will not become perfectly reliable.

F. Dietrich and C. List. Arrow's theorem in judgment aggregation. *Social Choice and Welfare* 29(1): 19–33, 2006.

S. Hartmann, G. Pigozzi, and J. Sprenger. Reliable methods of judgment aggregation. Under review.

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

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)*, 233-244, 1999.

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

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

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*

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

R. Nozick. *Philosophical Explanations*. Harvard: Harvard University Press, 1981.

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