Judgment Aggregation and the Problem of Tracking the Truth

Stephan Hartmann* and Jan Sprenger†

July 5, 2009

Abstract

The aggregation of consistent individual judgments on logically inter-connected 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. In this paper, we motivate that quite often, we do not only want to make a factually right decision, but also to correctly evaluate the *reasons* for that decision. In other words, we address the problem of *tracking the truth*. We set up a probabilistic model that generalizes the analysis of Bovens and Rabinowicz (2006) and use it to compare several aggregation procedures. Demanding some reasonable adequacy constraints, we demonstrate that a reasons- or premise-based aggregation procedure tracks the truth better than any other procedure. However, we also illuminate that such a procedure is not in all circumstances easy to implement, leaving actual decision-makers with a tradeoff problem.

1 Introduction

Judgment aggregation (List 2007) is an emerging research area in formal episte-mology and economics. It investigates how to aggregate individual judgments on logically related propositions into a collective judgment on those propositions. Examples of groups that need to aggregate individual judgments are all kind of committees: expert panels, legal courts, boards, and councils. The propositions are of two kinds: *premises* and a *conclusion*. The former serve as supporting reasons to derive a judgment on the latter. Consider, for example, a city council that has to make a decision on whether to build a new harbor site (represented by a proposition D, the resulting decision). This project is eligible for public funding if and only if two premises are satisfied: first, there is sufficient request for new harbor sites that cannot be met by existing harbor sites (represented by proposition A_1), and second, the nearby marine reserve is not badly affected (represented by proposition A_2). The decision rule can be formally expressed as

---

*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_1	A_2	D
Voter 1, 2, 3	True	True	True
Voter 4, 5	True	False	False
Voter 6, 7	False	True	False
Majority	True	True	False

Table 1: An illustration of the discursive dilemma under the constraint rule $A_1 \wedge A_2 \leftrightarrow D$.

the formula $(A_1 \wedge A_2) \leftrightarrow D$. Each member of the council expresses her judgment on A_1, A_2 and D such that the rule $(A_1 \wedge A_2) \leftrightarrow D$ is satisfied.

How shall we derive a group judgment given the individuals' opinions on premises and conclusion? It is assumed that each individual expresses judgments on the propositions while respecting the logical relations. If we define the group opinion as the majority view on the issues (premises and conclusion), it turns out that the group may take an inconsistent position, as shown in Table 1. The city council may face a situation where the majority thinks that the new harbor site should not be built. However, it will not be possible to provide *reasons* for this judgment as a majority of the members agrees that there is sufficient request for further harbor sites and another majority agrees that the nearby marine reserve is not badly affected. The literature on judgment aggregation refers to such a problem as the *discursive dilemma*.

We see that, although each committee member expresses a consistent opinion, propositionwise majority voting results in a majority for A_1 and A_2, but in a majority for $\neg D$.[1] This is clearly an inconsistent collective result as it violates the rule $(A_1 \wedge A_2) \leftrightarrow D$. The paradox 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 of two different research programs. The first program investigates the general scope of judgment aggregation functions. For instance, recent results demonstrate that reasonable constraints such as collective rationality (the group judgment satisfies the logical constraint rule) and non-dictatorship cannot hold jointly, if conjoined with some plausible independence constraints (List and Pettit 2002; Dietrich and List 2006). Thus, the first program methodically assimilates judgment aggregation problems to preference aggregation problems, and has a natural focus on (im)possibility theorems like Arrow's theorem. – We pursue, however, the hands-on approach of the second program. Here, various ways to consistently aggregate judgments are evaluated from an *epistemic perspective* (Bovens and Rabinowicz 2006). So the focus is rather on concretely comparing the epistemic performance of actual procedures, than on investigating what kind of procedures can, or cannot, exist.

Two epistemic goals have to be discerned. The first, and foremost goal consists in making the right *decision* (i.e. to build or not to build harbor site).

---

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

The second, more demanding goal consists in making decisions that *track the truth*: i.e. not only do we adopt the factually right decision, we also judge all propositions in the agenda correctly. We address the first goal in Hartmann, Pigozzi and Sprenger (2008), and here, we focus on the problem of substantiating the right decision by an correct overall judgment.

The paper is structured as follows. Section 2 motivates the need for substantiating decisions by procedures that track the truth, and introduces part of our formal framework. Section 3 generalizes the probabilistic model of Bovens and Rabinowicz (2006). Section 4 contains our analytical result and establishes the superiority of a reasons-based procedure for tracking the truth. Finally, section 5 discusses the practical implementation of the analyzed aggregation procedures and summarizes our insights.

2 Two Methods of Judgment Aggregation

In the discursive dilemma presented in section 1, we have identified an admissible judgment set with a consistent judgment on the sets of propositions at stake (see table 1). We now define aggregation functions over general judgment aggregation problems. The following definitions will help us to attain this goal (see also Miller and Osherson (2009)).

Definition 1. *The* agenda *is a set of propositions on which judgments are to be made.*

In our paper, we focus on agendas of the form $\mathcal{A} = \{A_1, \ldots, A_M, D\}$ where the A_i are called the *premises*, and D is called the *decision*:

Definition 2. *A logical constraint rule L on an agenda $\mathcal{A} = \{A_1, \ldots, A_M, D\}$ is a well-formed formula of propositional logic that (1) represents D as a truth-functional compound of the other elements of $\mathcal{A}$ and (2) preserves the logical independence of the A_i.*

Together, $\mathcal{A}$ and L determine a *judgment aggregation problem*. Here, we have demanded that the $A_1, \ldots, A_N$ are logically independent, but that D is equivalent to some truth-functional compound of them. This restriction has been made in order to treat the judgment aggregation problem as a decision problem, where the final decision depends in some non-trivial way on a set of premises that are assumed as independent of each other (e.g. standing for different facts of the matter).

Definition 3. *A judgment set or situation valuation S is an ordered set of true/false valuations of the propositions in $\mathcal{A}$. Let $\mathcal{J}$ denote the set of all judgment sets.*

Definition 4. *An admissible judgment set or admissible situation valuation S, given a logical constraint rule L, is an ordered set of true/false valuations of the propositions in $\mathcal{A}$ such that these valuations respect L. Let $\mathcal{J}_\bullet$ denote the set of all admissible judgment sets.*

Definition 5. *For a group of N members, an admissible judgment profile is a N-elementary vector of admissible judgment sets.*

Thus, judgment sets assign truth values to all propositions in the agenda. In other words, they evaluate all relevant aspects of a given situation – therefore the synonym situation valuation. A judgment set is admissible if and only if it conforms to the logical constraints, and a judgment profile is the vector which contains all individual judgment sets.

Now, we can define a judgment aggregation function f:

Definition 6. *For a group of N members with agenda $\mathcal{A}$ and logical constraint rule L, a function $f : \mathcal{J}_{\bullet}^{N} \rightarrow \mathcal{J}_{\bullet}$ is called a* judgment aggregation function/procedure.

The goal of our paper consists in comparing different ways of aggregating N admissible judgment sets into a single judgment set such that not only the resulting decision D, but the entire *situation* is judged correctly. Thus, our aggregation function f should aim at the following property:

> **Truth-Tracking (TT):** For any admissible situation valuation $S \in \mathcal{J}_{\bullet}$, if S were true, then S would be chosen.

An aggregation procedure that satisfies this condition is said to *track the truth*.

This account of truth-tracking can be brought in line with Nozick's (1981) account of truth-tracking in epistemology. A judgment aggregation method tracks the truth of a situation S if and only if two conditions are satisfied:

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

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

Of course, our definition of truth-tracking satisfies Stability. Neither is it difficult to see that Sensitivity is satisfied as well: If S were not true, then some other situation S' would be true, and by (TT), S' would be chosen. *A fortiori*, S would *not* be chosen, and Sensitivity is satisfied. Thus, (TT) entails Stability and Sensitivity so that truth-tracking methods in our sense also track the truth in Nozick's sense.

Of course, few aggregation procedures track the truth under all circumstances – errors are part of the game. For this reason, we believe it useful to introduce a *comparative* concept of truth-tracking: an aggregation function outperforms a rival if its (probabilistic) tendency to track the truth is higher for each situation $S \in \mathcal{J}_{\bullet}$.

Definition 7. *For two aggregation procedures f and f', f tracks the truth better than f' in a given judgment aggregation problem, if and only if, in an appropriate probability model, for any $s \in \mathcal{J}_{\bullet}$,*

$$\mathbb{P}(f(x_1,\ldots,x_N) = s \,|\, s \text{ is true}) \geq \mathbb{P}(f'(x_1,\ldots,x_N) = s \,|\, s \text{ is true}),$$

with inequality for at least one element of $\mathcal{J}_{\bullet}$.

We will construct an appropriate probability model in the next section. For the time being, note that truth-tracking is a stronger aim than just making a factually right decision. It means to make a right decision for the right reasons, based on a correct assessment of the entire situation. Not in all contexts, it is obvious that reasons are important. For example, in the initial harbor example, our desire to make a factually correct decision might be more important

4

than giving the correct reasons: Assume that we believe the harbor project to be economically pointless, but that we don't see any problems for the environment. We therefore decide not to approve of the project. Later, we learn that the harbor site would have been profitable, but that the nearby marine reserve would have been affected badly. Apparently, in that context, it does not matter whether we correctly assessed the entire situation as long as the resulting decision – not to build the harbor site – was correct.

However, the problem of tracking the truth is important in a variety of circumstances, especially when 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 be divided 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. In other words, misspecified reasons invite costly revisions.

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 main argument to be invalid. In fact, the author's 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. Similarly, a city council misjudging the reasons for an accidentally correct decision might be considered incompetent.[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, but also *epistemic* drawbacks: 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*.

Now, we investigate two different approaches to judgment aggregation. From the preceding analysis of the discursive dilemma in Bovens and Rabinowicz (2006) and List (2005, 2006), two procedures are known: the *premise-based procedure* f_P and the *conclusion-based procedure* f_C. Recall that the agenda of the version of the discursive dilemma discussed in this paper is $\mathcal{A} = \{A_1, A_2, D\}$, with the constraint rule $L = (A_1 \wedge A_2) \leftrightarrow D$. f_P determines the aggregate vote on A_1 and A_2 by simple majority voting, and fixes the collective judgment on $\mathcal{A}$ according to the constraint $(A_1 \wedge A_2) \leftrightarrow D$. In conclusion-based reasoning, however, the members decide privately on A_1 and A_2 and only express their opinions on D publicly. The judgment of the group is then inferred from applying the majority rule to the individual judgments on D. Thus, f_C only yields a decision, and no judgment on the premises, implying that it cannot track the

---

[2]An overlap between both situations is, by the way, possible, but not necessary.

true situation.

Consequently, Bovens and Rabinowicz (2006) introduce a novel procedure which we call *modified conclusion-based procedure* f_M. That procedure opts for situation $s \in \mathcal{J}_\bullet$ if and only if more than half of the group members support s (Bovens and Rabinowicz 2006, 139). In other words, situation s is selected if and only if backed by an absolute majority of group members. Consequently, f_M gives an indeterminate result when there is only a *relative*, but no absolute majority for a specific situation. In their calculations of the truth-tracking performance of f_M (formula (21) on page 139), Bovens and Rabinowicz treat those indeterminate cases as cases in which f_M gives a wrong result. Since indeterminate decisions count as fallacious decisions, it is not surprising that their analysis appraises the truth-tracking abilities of f_P over those of f_M. So, instead of appraising f_P prematurely, Bovens and Rabinowicz should amend f_M to a well-defined procedure without indeterminate cases and compare *that* procedure to f_P. Unlike the comparison in their paper, this would be a fair contest.

We close this gap in our paper. More precisely, we compare a large class of judgment aggregation procedures among which two procedures deserve special mention. The first is the premise-based procedure f_P which adopts, for each proposition $A_1, \ldots, A_N$ in the agenda, the judgment that is backed by a majority of group members. Then, the judgment on the decision D is inferred from these judgments. More, precisely, for $x_1, \ldots, x_N \in \mathcal{J}_\bullet$ and $j \leq M$:[3]

$$f_P(x_1, \ldots, x_N)^{(j)} = \begin{cases} \text{`true'} & \#\{i | x_i^{(j)} = \text{`true'}\} > \#\{i | x_i^{(j)} = \text{`false'}\} \\ \text{`false'} & \text{otherwise} \end{cases}$$

and in addition, $f_P(x_1, \ldots, x_N)^{(m+1)}$ is chosen such that $f_P(x_1, \ldots, x_N)$ is in $\mathcal{J}_\bullet$.

However, in some circumstances, f_P may be hard to implement, especially when the decision-making body is no back-room committee, but gathers in public (such as legislative bodies, public councils, etc.). Take, for instance, a motion that advances a certain resolution but also takes a stand on the reasons for the desired decision. f_P, if properly implemented, would require that the motion be split and that the final decision be inferred from the single ballots. This is twofold difficult: First, it may be impractical and time-consuming to split a complex motion. Second, how to infer the final decision? As witnessed by table 1, there can be a majority for each premise, but no majority for the corresponding decision – this was the source of the discursive dilemma. Hence, the decision-makers might end up with an inconsistent collective judgment (if they also vote on the decision), and there is, in general, no institution which could compel them to be logically consistent.

Moreover, f_P can lead to decisions that are backed by only a small minority of the group members (see table 1, and similar examples). A natural alternative is the *situation-based procedure* f_S which applies the plurality rule to selecting a peculiar judgment set or situation valuation. Such a procedure avoids the objections listed above: the group judgment corresponds to the judgment set that is backed by most voters. This is a natural generalization of the rationale underlying f_C and f_M. It is also frequently found in the practice of decision-making, e.g. whenever several draft motions are balloted. (Think of "The city

---

[3]N is, for the sake of convenience, assumed to be odd.

council approves of constructing new harbor sites because close analysis has demonstrated both the economic profitability and the ecological sustainability of the project." vs. "Although the economic profitability of new harbor sites cannot be denied, the city council does not approve of the project, because the consequences for the nearby marine reserve would be too damaging" etc.) Hence, we define $f_S(x_1, \ldots, x_N) = s \in \mathcal{J}_\bullet$ such that s satisfies either

$$\#\{i|x_i = s\} > \max_{s' \in \mathcal{J}_\bullet \setminus s} \#\{i|x_i = s'\} \tag{1}$$

or, if equality in (1) holds,

$$\min_{i \leq N}\{i|x_i = s\} < \min_{i \leq N}\{i|x_i = s'\}. \tag{2}$$

Thus, in the case of a tie between different judgment sets, the group member with the lowest number decides the day. Both f_P and f_S satisfy anonymity (with the exception of the purely conventional tie-breaking rule in the case of f_S), non-dictatorship and collective rationality, i.e. conditions that are really indispensable for any aggregation function that might be applied in democratic practice. The next section presents a formal model that allows us to compare the truth-tracking performance of f_P and f_S according to definition 7.

3　The Probabilistic Model

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 A_i is factually true, the group member submits a correct judgment with probability p, and equally, if the premise is false, the group member submits a correct report with probability p. The competence of an individual group member p is decoupled from the prior probability of the propositions at stake, and the probability of making correct or incorrect judgments does not depend on the truth values of the propositions in question.[4]

Now, the Condorcet Jury Theorem links the competence of the group members 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 group member 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 situation-based procedure in the judgment aggregation problem (Bovens and Rabinowicz 2006). It should be noted, though, that an

---

[4]If the probabilities for tracking the truth of A_i and $\neg A_i$ differed, the group members would be bad at tracking the true situation since one of Nozick's subjunctive conditionals (see page 4) would be violated. This violation can occur even if the overall reliability p is high. See also List (2006).

Note that we ascribe an individual competence only for voting on *premises* only, which play the role of 'elementary' matter-of-fact propositions from which a comprehensive judgment set is derived.

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

1. $A_1, \ldots, A_M$ are logically independent.

2. All agents have the same competence p to assess the truth of a single premise A_i. Their judgments on the premises are independent.

3. Each individual judgment set is logically consistent.

Notably, we can relax the conditions of Bovens and Rabinowicz (2006) and Hartmann, Pigozzi and Sprenger (2008) who also demand that the marginal probabilities of the premises be equal, and that the premises be probabilistically independent. Now, let $q : \mathcal{J}_\bullet \to [0, 1]$ be an assignment of prior probabilities to each judgment set, and for any $s \in \mathcal{J}_\bullet$, let $\mathbb{P}_s$ denote the probability distribution over the judgment sets that is induced by s being the correct judgment set. Evidently, for any aggregation function f, its *mean reliability* $\mathcal{R}_{p,q}(f)$ can be computed as

$$R_{p,q}(f) \quad := \quad \sum_{s \in \mathcal{J}_\bullet} q(s) \, \mathbb{P}_s(f(x_1, \ldots, x_N) = s), \tag{3}$$

and similarly for f_S.

4 The Analytical Results

With the above model at hand, we investigate the reliability and truth-tracking properties of a general class of aggregation functions.

Definition 8. *A function $g : \mathcal{J}_\bullet \to \mathcal{J}_\bullet$ is a* swap function *if and only if g satisfies for all $s \in \mathcal{J}_\bullet$ and $i, j \leq M$:*

$$s^{(i)} = s^{(j)} \qquad \to \qquad g(s)^{(i)} = g(s)^{(j)}.$$

$g(s)^{(M+1)}$ *is then inferred from the logical constraint rule L.*

Fact 1. *Any swap function is a bijection on $\mathcal{J}_\bullet$. Hence, for each $s, s' \in \mathcal{J}_\bullet$, there is a swap function g such that $s' = g(s)$, and vice versa.*

Definition 9. *An aggregation function f is* unbiased *if and only if for any swap function g and any judgment profile $(x_1, \ldots, x_N) \in \mathcal{J}_\bullet^N$:*

$$f(g(x_1), \ldots, g(x_N)) = g(f(x_1, \ldots, x_N)). \tag{4}$$

Unbiasedness means that the group members are not biased in favor of a specific situation and only listen to the evidence. In other words, the judgments of the group members are impartial with respect to the "names" of the situations. Thus, unbiasedness is complementary to anonymity – unbiased procedures don't look at the name of the proposition to be voted on, in the same way that anonymous aggregation procedures don't look at the name of the person who submitted a judgment set (Mongin 2008). This property motivates our focus on unbiased procedures in the sequel. First, it turns out that f_P and f_S are – as expected – in the class of unbiased aggregation functions:

Proposition 1. *f_P and f_S are unbiased aggregation functions.*

Remark. *Distance-based procedures as discussed in Pigozzi (2006) and Miller and Osherson (2009) are in general not unbiased.*

Second, we prove that prior probabilities do not affect the mean reliability of an unbiased procedure:

Theorem 1. *The mean reliability and truth-tracking abilities of an unbiased aggregation function f are independent of the prior probability distribution $q(\cdot)$ over the admissible judgment sets.*

In other words, for unbiased procedures, the prior probabilities of propositions or judgment sets do not matter with respect to their epistemic performance. This is a very helpful property since, if decision-making depended on the prior probabilities of a given proposition, we would leave the judgment aggregation framework in favor of a (more general) probability aggregation framework. Besides, "biased" procedures remain stuck with a bias in favor of a particular proposition, and are hard to generalize. Thus, we obtain two further reasons for focusing on unbiased procedures only.

Finally, we can demonstrate the superiority of f_P over all other unbiased procedures:

Theorem 2. *Let the group members be individually be more competent than randomizers (i.e. $p > 0.5$). The premise-based procedure f_P is better at tracking the true situation than any other unbiased procedure, for any value of N and p, and for any prior probability distribution $q(\cdot)$ over the admissible judgment sets.*

Remark. *The theorem implies that for $p > 0.5$, f_P always outperforms f_S. However, if the group members are severely biased against the truth ($p < 0.5$), f_P will be inferior to f_S, as the proof of Theorem 2 makes clear.*

In other words, if the group members are, on an individual level, not biased against the truth, (i.e. in the mean, they track the truth more often than falsehood), then f_P 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 f_S and related procedures, and it has a higher mean reliability as well.

5 Discussion and Summary

In this paper, we have approached the problem of judgment aggregation from an epistemic perspective and examined the problem of tracking the truth and selecting the right situation, over and above selecting the right decision. 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 motivated to narrow down our analysis to a particular class of aggregation functions – namely unbiased functions that are not biased in favor of a specific situation or proposition. We take this, analogous to anonymity, as a reasonable adequacy constraint on aggregation functions. Among these procedures, we have focused our atention

on two natural aggregation rules that exemplify two different heuristics, namely the premise- or reasons-based procedure f_P, and the situation-based procedure f_S. Finally, we have set up and investigated a probabilistic model to evaluate unbiased procedures with respect to their ability to track the truth.

Our analytic findings demonstrate that the premise-based procedure f_P has uniformly the best truth-tracking properties among all unbiased aggregation function. In other words, f_P *dominates* all other unbiased procedures. This result vindicates the importance of *reasons* in decision-making (see also Hartmann, Pigozzi and Sprenger (2008) for a similar result with respect to overall reliability in judgment aggregation). More precisely, our findings suggest that when a decision depends on several independent factors, it is preferable to vote on these factors separately, and to derive a decision from these individual ballots, instead of voting on all possible judgment sets. Furthermore, committees might want to discuss the premises in advance in order to achieve consensus on the reasons for a decision.

However, there are some drawbacks, too. f_P may be hard to implement in practice, and sometimes endorses judgment sets that only a tiny fraction of the group supports. If the social or political circumstances require considerable support for the entire judgment set, f_P may be inappropriate and could be replaced by f_S, for instance. Indeed, numerical results indicate that f_S is often only marginally less reliable than f_P.

To sum up, in a general judgment aggregation problem, an epistemic perspective ranks the premise-based procedure f_P always above f_S and other unbiased procedures; but these gains are often small and have to be balanced against losses in terms of feasibility. These losses may be negligible for non-public committees that are interested in tracking the truth, but severe for committees that hold their meetings in public, and where decision-making is constrained by certain formal procedures.

Acknowledgements

A   Proofs

Proof of Fact 1: All valuations of $A_1, \ldots, A_M$ are compatible with the constraint rule L, by definition 2. Thus, for any $s \in \mathcal{J}_\bullet$, also $g(s) \in \mathcal{J}_\bullet$. Moreover, a swap function swaps the valuation of one or several premises and is therefore injective. But all injective maps on a finite domain, and with equal range, are bijective as well. $\square$

Lemma 1. *Let* $g : \mathcal{J}_\bullet \to \mathcal{J}_\bullet$ *be a swap function, i.e. let g satisfy for all $s \in \mathcal{J}_\bullet$, $i, j \leq M$:*

$$s^{(i)} = s^{(j)} \qquad \to \qquad g(s)^{(i)} = g(s)^{(j)} \qquad (5)$$

Moreover, let $\mathbb{P}_s$ denote the probability distribution of an individual group member's judgments if the situation valuation s is the correct one. Then, for any

10

$s, s' \in \mathcal{J}_\bullet$:

$$\mathbb{P}_s(s') = \mathbb{P}_{g(s)}(g(s'))$$

Proof of Lemma 1: By a straightforward calculation:

$$
\begin{aligned}
\mathbb{P}_s(s') &= p^{\#\{j|s^{(j)}=s'^{(j)}\}} (1-p)^{M-\#\{j|s^{(j)}=s'^{(j)}\}} \\
&= p^{\#\{j|g(s)^{(j)}=g(s')^{(j)}\}} (1-p)^{M-\#\{j|g(s)^{(j)}=g(s')^{(j)}\}} \\
&= \mathbb{P}_{g(s)}(g(s')) \qquad \square
\end{aligned}
$$

Proof of Proposition 1: The unbiasedness of f_P follows directly from (5). For f_S, let without loss of generality $x_1 = f_S(x_1, \ldots, x_N)$ (the finally chosen judgment set must be among the submitted ones). Hence, $f_S(g(x_1), \ldots, g(x_N)) = g(x_1)$, by simple majority reasoning. Combining both equations yields

$$g(f_S(x_1, \ldots, x_N)) = f_S(g(x_1), \ldots, g(x_N)),$$

as required. $\square$

Proof of Theorem 1: We have to show that in our probabilistic model, the reliability of any unbiased aggregation procedure is independent of the prior probability distribution over the judgment sets. Take $s, s' \in \mathcal{J}_\bullet$, and let $g : \mathcal{J}_\bullet \to \mathcal{J}_\bullet$ be a swap function with $s' = g(s)$. Let f be an unbiased aggregation procedure, and let $\mathbb{P}_s$ be the probability distribution over $\mathcal{J}_\bullet^N$ in case s is the correct situation. Moreover, let I_C be the 0-1-indicator function that returns 1 if C is satisfied, and 0 otherwise. Then,

$$
\begin{aligned}
\mathbb{P}_s(f(x_1, \ldots, x_N) = s) &= \sum_{(x_1,\ldots,x_N)\in\mathcal{J}_\bullet^N} I_{f(x_1,\ldots,x_N)=s} \, \mathbb{P}_s(x_1, \ldots, x_N) \\
&= \sum_{(x_1,\ldots,x_N)\in\mathcal{J}_\bullet^N} I_{g(f(x_1,\ldots,x_N))=g(s)} \prod_{k=1}^{N} \mathbb{P}_s(g^{-1}(g(x_k))) \\
&= \sum_{(x_1,\ldots,x_N)\in\mathcal{J}_\bullet^N} I_{f(g(x_1),\ldots,g(x_N))=s'} \prod_{k=1}^{N} \mathbb{P}_{s'}(g(x_k)) \\
&= \sum_{(x_1,\ldots,x_N)\in\mathcal{J}_\bullet^N} I_{f(x_1,\ldots,x_N)=s'} \prod_{k=1}^{N} \mathbb{P}_{s'}(x_k) \\
&= \mathbb{P}_{s'}(f(x_1, \ldots, x_N) = s')
\end{aligned}
$$

In the second line, we have made use of the independence of the group members' judgments and the bijectivity of g. The third line has used the unbiasedness of f and Lemma 1. The rest was straightforward, making use of the bijectivity of g once more (the x_i are just random variables).

Hence, $c := \mathbb{P}_s(f(x_1, \ldots, x_N) = s)$ is independent of $s \in \mathcal{J}_\bullet$. Therefore,

$$
\begin{aligned}
R_{p,q}(f) &= \sum_{s\in\mathcal{J}_\bullet} q(s) \, \mathbb{P}_s(f(x_1, \ldots, x_N) = s) \\
&= \sum_{s\in\mathcal{J}_\bullet} q(s) \, c \\
&= c,
\end{aligned}
$$

proving the theorem, since c is, of course, independent of $q(\cdot)$. $\square$

Proof of Theorem 2: To avoid unneccessary technicalities, we only sketch the proof idea. We show with the help of a Bayesian argument that for a particular prior distribution, f_P is an epistemically *optimal* aggregation function. Due to Theorem 1, this entails that f_P is uniformly the best among all unbiased procedures for all prior distributions $q(\cdot)$.

Let $q(s) = 1/|\mathcal{J}_\bullet|$ be a constant. Now, all judgment sets are equally likely at the outset. This implies that not knowing which situation is the right one, the best thing we can do is to treat the group members' judgments as *evidence* for a particular judgment set, and to go for the judgment set with the highest posterior probability. In this situation of indifference over prior beliefs, the judgment set with the highest posterior corresponds to the *best confirmed* judgment set (as long as $p > 0.5$). Due to the independence assumptions, the best confirmed judgment set is the one where the valuations of each A_i are backed by more than the half of voters. (For the sake of convenience, we may assume N to be odd here.)

Now, observe that f_P picks precisely the judgment set where each valuation of A_i is in agreement with the majority vote. Hence, for group members that are more competent than pure randomizers ($p > 0.5$), the premise-based procedure f_P outperforms all other unbiased procedures. Since this effect was independent of $q(\cdot)$ and since, by Theorem 1, for any unbiased aggregation procedure f, $\mathbb{P}_s(f(x_1, \ldots, x_N) = s') = \mathbb{P}_{s'}(f(x_1, \ldots, x_N) = s')$, f_p tracks the truth better than any other unbiased aggregation procedure. $\square$

References

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

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, 2009.

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.

M. K. Miller and D. Osherson. Methods for distance-based judgment aggregation. *Social Choice and Welfare* 32: 575–601, 2009.

P. Mongin. Factoring out the impossibility of logical aggregation. *Journal of Economic Theory* 141: 100–113, 2008.

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.