Judgment aggregation and the problem of truth-tracking

Gabriella Pigozzi
Computer Science and Communications
University of Luxembourg
L–1359 Luxembourg
email:gabriella.pigozzi@uni.lu

Stephan Hartmann
Center for Logic and Philosophy of Science
Tilburg University
NL–5000 LE Tilburg
email: S.Hartmann@uvt.nl

Abstract

The problem of the aggregation of consistent individual judgments on logically interconnected propositions into a collective judgment on the same propositions has recently drawn much attention. The difficulty lies in the fact that a seemingly reasonable aggregation procedure, such as propositionwise majority voting, cannot ensure an equally consistent collective outcome. The literature on judgment aggregation refers to such dilemmas as the *doctrinal paradox*. Three procedures have been proposed in order to overcome the paradox: the premise-based and conclusion-based procedures on the one hand, and the fusion approach on the other hand. In this paper we assume that the decision which the group is trying to reach is factually right or wrong. Hence, the question is how good the fusion approach is in tracking the truth, and how it compares with the premise-based and conclusion-based procedures. We address these questions in a probabilistic framework and show that belief fusion does especially well for individuals with a middling competence of hitting the truth of a proposition.

1   Introduction

In many situations (such as expert panels, boards, councils, societies) individuals are required to express their opinions on several propositions. Then, the individual judgments need to be combined to form a collective decision. In this paper we will consider a recent method to aggregate conflicting individual judgments into a consistent group outcome. This method is called the *fusion procedure* [9]. The fusion procedure comes from computer science, where the problem of combining information from equally reliable sources arises in several contexts.

We will assume that the resulting collective judgment is factually right or wrong. Thus, the question that naturally arises is how good the fusion procedure is in tracking the truth, and how it compares with the other approaches. We are especially interested in how the fusion procedure compares with the so-called *premise-based procedure* and the so-called *conclusion-based procedure* that were recently studied by Bovens and Rabinowicz [1].

The combination of finite sets of logically interconnected propositions has been investigated in the emerging field of *judgment aggregation*.[1] Judgment aggregation studies how consistent individual judgments on logically interconnected propositions are combined into a collective judgment on the same propositions.[2] A *judgment* is an assignment of yes/no to a proposition. The difficulty lies in the fact that a seemingly reasonable aggregation procedure, such as propositionwise majority voting, cannot ensure a consistent collective outcome.

Here is an illustration [5]. A court has to make a decision on whether a person is liable of breaching a contract (represented by a proposition R, also referred to as the conclusion). The judges have to reach a verdict following the legal doctrine. This states that a person is liable if and only if she did a certain action X (represented by a proposition P, also referred to as the first premise) and had contractual obligation not to do X (represented by a proposition Q, also referred to as the second premise). The legal doctrine can be formally expressed as the rule $(P \wedge Q) \leftrightarrow R$. Each member of the court expresses her judgment on P, Q and R such that the rule $(P \wedge Q) \leftrightarrow R$ is satisfied. Suppose now that the court has seven members who make their judgments according to Table 1. We see that, although each judge expresses a consistent opinion, propositionwise majority voting (consisting in the separate aggregation of the votes for each of the propositions P, Q and R via the majority rule) results in a majority for P and a majority for

---

[1]See [7] for a comprehensive bibliography documenting the growing literature on judgment aggregation.

[2]The relations between the aggregation of individual judgments and the aggregation of individual preferences as traditionally studied in social choice theory have been explored in [8, 2].

	P	Q	R
Members 1, 2, 3	Yes	Yes	Yes
Members 4, 5	Yes	No	No
Members 6, 7	No	Yes	No
Majority	Yes	Yes	No

Table 1: An illustration of the doctrinal paradox

	K_1, K_2, K_3	K_4, K_5	K_6, K_7	$\mathcal{F}_{IC}(E)$
(1,1,1)	0	2	2	8
(1,0,0)	2	0	2	10
(0,1,0)	2	2	0	10
(0,0,0)	3	1	1	13

Table 2: The fusion operator applied to the doctrinal paradox

Q, but in a majority for $\neg R$. This is clearly an inconsistent collective result as it violates the rule $(P \wedge Q) \leftrightarrow R$. The paradox (called the *doctrinal paradox*) lies in the fact that majority voting can lead a group of rational agents to endorse an irrational collective judgment. Clearly, the relevance of such aggregation problems goes beyond the specific court example and it applies to all situations in which individual binary evaluations need to be combined into a group decision.

The first two escape-routes that have been suggested are the *premise-based procedure* (PBP) and the *conclusion-based procedure* (CBP). According to PBP, each judge votes only on the premises. The conclusion is then inferred from the rule $(P \wedge Q) \leftrightarrow R$ applied to the judgment of the group on P and Q obtained from applying the majority rule. According to CBP, the judges decide privately on P and Q and only express their opinions on R publicly. The judgement of the group is then inferred from applying the majority rule to the individual judgments on R.

In [9] it has been argued that PBP and CBP are not satisfactory methods for group decision-making and proposed that the fusion procedure is superior from a theoretical point of view. However, theoretical superiority is one thing, and truth another. The court example shows that there are situations where the propositions in question are factually true or false. What we then want from an aggregation procedure is that it is a good truth tracker, i.e. that it comes to the conclusion that R if and only if R is the case, and not, if not. Bovens and Rabinowicz in [1] and List in [6] explored how well PBP and CBP do in terms of truth-tracking. In this paper, we follow their account and ask how well fusion does compared to the other two procedures.

2 The Fusion Procedure

As shown in [9], the application of a fusion operator originally defined in [4] to judgment aggregation problems allows to define consistent group decisions and to avoid paradoxical outcomes. This section summarizes the approach and the results of [9]. The reader is referred to that paper for more details.

One of the key points stressed in the literature on information fusion is that the aggregation of consistent knowledge bases does not guarantee a consistent collective outcome.

To overcome this problem, *integrity constraints* (IC) are imposed on the final base. Let's consider a finite set N of individuals making their judgments on a given finite set X of propositions (called the *agenda*). A base K_i of an agent i is a consistent and complete finite set of atomic propositions and compound propositions expressing the agent's judgments. In a model-based fusion approach, each individual base is interpreted as a set of models. Given a multiset $E = \{K_1, K_2, \ldots, K_n\}$ and the integrity constraints IC, a fusion operator $\mathcal{F}$ assigns a belief base to E and IC. Let $\mathcal{F}_{IC}(E)$ denote the resulting collective base.[3]

A majority fusion operator will select the (possibly not unique) model that minimizes the *Hamming distance* to the individual bases. The Hamming distance is defined as the number of propositional letters on which two models differ. For example, the Hamming distance between $\omega = (0, 0, 0, 1)$ and $\omega' = (1, 1, 0, 0)$ is $d(\omega, \omega') = 3$. The procedure, then, consists of three steps. In step 1, one determines the distance between the models of IC and the models of each K_i in E. In step 2, one assigns a distance value to each model of IC and E. This distance is defined to be the sum of the aforementioned Hamming distances. Finally, in step 3, the model that minimizes the distance is chosen.

To illustrate how the majority fusion operator works, we apply it to the court example. In the new terminology, the agenda is $X = \{P, Q, R\}$ with $IC = \{(P \wedge Q) \leftrightarrow R\}$. The models for each belief base are:

$$\text{Mod}(K_1) = \text{Mod}(K_2) = \text{Mod}(K_3) = \{(1, 1, 1)\}$$
$$\text{Mod}(K_4) = \text{Mod}(K_5) = \{(1, 0, 0)\}$$
$$\text{Mod}(K_6) = \text{Mod}(K_7) = \{(0, 1, 0)\}$$

Table 2 shows the result of the IC majority fusion operator on $E = \{K_1, \ldots, K_7\}$. The row with a shaded background corresponds to the selected collective outcome. The possible collective outcomes are chosen among the interpretations that are models of IC. Thus, no paradox arises by using the fusion procedure. We should mention that the fusion operator does not necessarily select a unique group decision. In some cases, the operator selects a set of models, i.e. the result is a tie between some belief bases.

---

[3] See [3] for a survey on logic-based approaches to belief fusion.

3 The Framework

To find out how well PBP and CBP do in terms of truth-tracking, Bovens and Rabinowicz [1] investigate the case $(P \wedge Q) \leftrightarrow R$ in an epistemic framework. They make four assumptions:

(i) The prior probability that P and Q are true are equal (q).

(ii) All voters have the same competence to assess the truth of P and Q (p).

(iii) P and Q are (logically and probabilistically) independent.

(iv) Each individual judgment set is logically consistent.

From assumption (iv) we conclude that only four situations are possible: $S_1 = \{P, Q, R\} = (1, 1, 1)$, $S_2 = \{P, \neg Q, \neg R\} = (1, 0, 0)$, $S_3 = \{\neg P, Q, \neg R\} = (0, 1, 0)$ and $S_4 = \{\neg P, \neg Q, \neg R\} = (0, 0, 0)$.

We can now calculate the probability that fusion ranks the right judgment set first (let us denote this proposition by F) by observing that

$$\mathcal{P}(F) = \sum_{i=1}^{4} \mathcal{P}(F|S_i)\,\mathcal{P}(S_i).$$

Thus, we have to calculate the prior probabilities $\mathcal{P}(S_i)$ and the conditional probabilities $\mathcal{P}(F|S_i)$ for $i = 1, \ldots, 4$. From assumption (i), we obtain (with $\bar{q} := 1 - q$) that $\mathcal{P}(S_1) = q^2$, $\mathcal{P}(S_2) = \mathcal{P}(S_3) = q\bar{q}$ and $\mathcal{P}(S_4) = \bar{q}^2$. To calculate the conditional probabilities, suppose that S_1 is the right judgment set. Then n_i (of N) voters will vote for situation S_i, with $n_1 + n_2 + n_3 + n_4 = N$. The majority fusion operator selects the model that minimizes the distance to these situations. This means that, if S_1 is the right judgment set, then the fusion procedure tracks the truth if $d_1 \leq \min(d_1, \ldots, d_4)$. The distances d_i can be expressed in terms of the numbers n_i of voters voting for the situations S_i $(i = 1, \ldots, 4)$:

$$d_1 = 2n_2 + 2n_3 + 3n_4 \quad , \quad d_2 = 2n_1 + 2n_3 + n_4$$
$$d_3 = 2n_1 + 2n_2 + n_4 \quad , \quad d_4 = 3n_1 + n_2 + n_3$$

We can now calculate $\mathcal{P}(F|S_1)$, which is:

$$\sum_{n_1,\ldots,n_4=0}^{N} \binom{N}{n_1, \ldots, n_4} p^{2n_1} (p\bar{p})^{n_2+n_3} \bar{p}^{2n_4} \, \mathcal{C}(n_1, n_2, n_3, n_4)$$

The sum is constrained: $\mathcal{C}(n_1, n_2, n_3, n_4) = 1$ if (i) $\sum_{i=1}^{4} n_i = N$ and (ii) $d_1 \leq \min(d_1, \ldots, d_4)$. Otherwise $\mathcal{C}(n_1, n_2, n_3, n_4) = 0$. $\mathcal{P}(F|S_2)$, $\mathcal{P}(F|S_3)$ and $\mathcal{P}(F|S_4)$ can be obtained analogously. Putting everything together, we can calculate $\mathcal{P}(F)$.

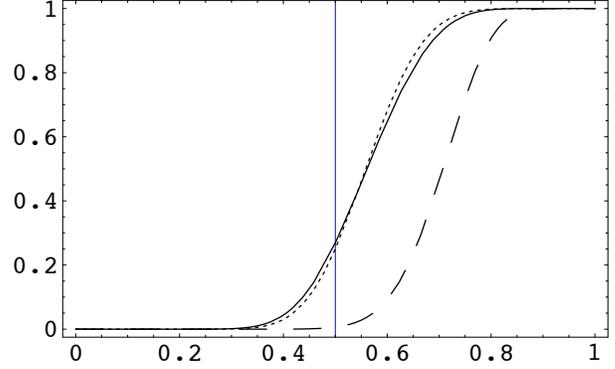With this equation in hands, we can now compare the fusion procedure (FP) with the PBP and the CBP.

Figure 1: The probability that FP (full line), PBP (dotted line), and CBP (dashed line) identify the right situation as a function of the competence of the voters p (x-axis) for $q = .5$ and $N = 21$.
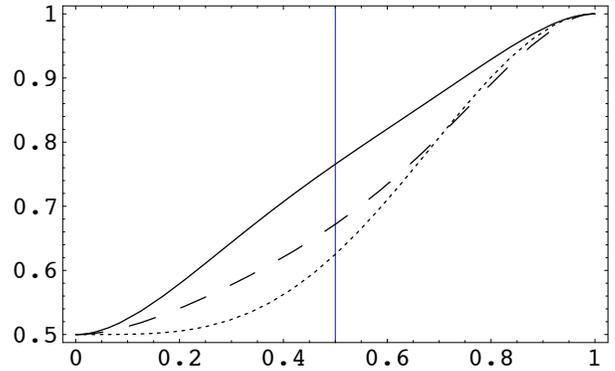

Figure 2: The probability that FP (full line), PBP (dotted line), and CBP (dashed line) identify the right result as a function of the competence of the voters p (x-axis) for $q = .5$ and $N = 3$.

4 Results

In Figure 1 we plot $\mathcal{P}(F)$ (the full line) for $q = .5$ and a medium size group with $N = 21$ and compare it with the corresponding probabilities for PBP (the doted line), and CBP (the dashed line).

Clearly, FP does significantly better than CBP. However, for high values of the competence p, PBP is slightly better than FP as a truth-tracker.

We now turn to evaluate how FP (full line) ranks a judgment set with the right result (i.e. the right conclusion, but not necessarily the right reasons) first, and we contrast this with PBP (dotted line), and CBP (dashed line). We test the procedures for $q = .5$ and for $N = 3$ (Figure 2) and $N = 31$ (Figure 3).

It turns out that FP greatly outperforms all the other aggregation procedures under investigation for small size groups. Yet, as the size of the group increases, the conclusion-based
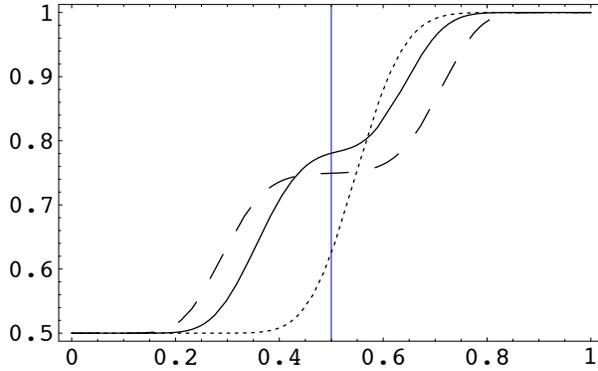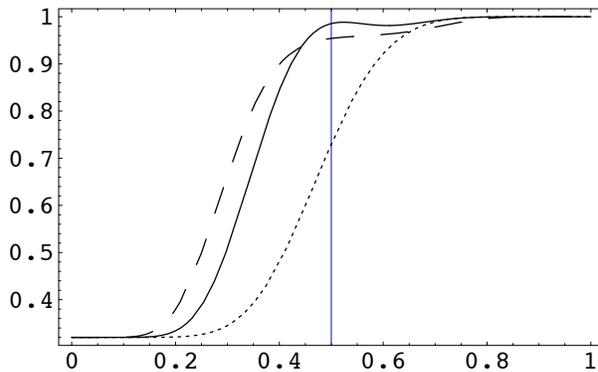
Figure 3: The probability that FP (full line), PBP (dotted line), and CBP (dashed line) identify the right result as a function of the competence of the voters p (x=axis) for $q = .5$ and $N = 31$.

Figure 4: The probability that FP (full line), PBP (dotted line), and CBP (dashed line) identify the right result as a function of the competence of the voters p (x-axis) for $q = .2$ and $N = 3$.

procedures (dashed line) does better than FP for low values of p, and PBP (dotted line) does better than FP for high values of p. But, for middling values of p, FP is always superior. We also observe that, whenever the FP is not the best procedure, it lies in-between PBP and CBP.

The next two pictures illustrate the same comparison, for a different value of the prior probability ($q = .2$). As before, the number of voters increases from $N = 3$ (Figure 4) to $N = 51$ (Figure 5).

Again, for small-sized groups, FP is the best procedure to reach the right decision. When the number of the voters increases, the conclusion-based procedure (dashed line) does better than FP, but only for low values of p. Different values of the prior do not undermine the superiority of FP for middling values of p (i.e. for $.4 \leq p \leq .6$). For higher values of p, PBP is only slightly better than FP.

We would like to stress the interesting phenomenon that, for p around .5, $\mathcal{P}(F)$ has a maximum with a value of
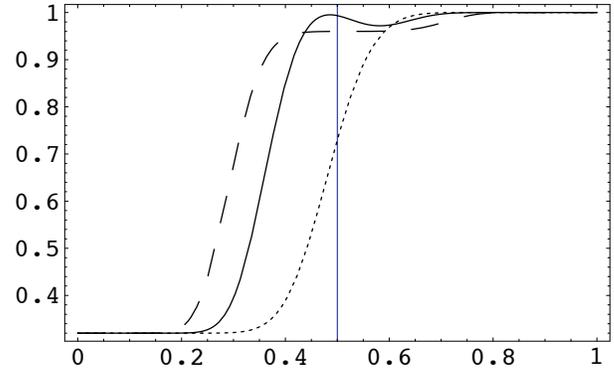

Figure 5: The probability that FP (full line), PBP (dotted line), and CBP (dashed line) identify the right result as a function of the competence of the voters p (x-axis) for $q = .2$ and $N = 51$.

almost 1. Increasing p further, $\mathcal{P}(F)$ first decreases and then, after going through a minimum, increases again and approaches 1.

In future work we plan to understand this phenomenon. Moreover we plan to systematically compare the various procedures for different values of q and for different numbers of voters. We will also test the stability of our main result – that FP does better than PBP and CBP for middling values of the competence parameter p – for different logical rules (more premises etc.). To do so, we will run Monte Carlo simulations as analytical methods as well as direct numerical evaluations of the corresponding sums are not any longer feasible.

References

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

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

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

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

[5] L. A. Kornhauser and L. G. Sager. Unpacking the court. *Yale Law Journal*, 96: 82-117, 1986.

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

[7] 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*

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

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