Merging judgments and the problem of truth-tracking

Gabriella Pigozzi and Stephan Hartmann

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 *discursive paradox*. So far, three procedures have been proposed to overcome the paradox: the premise-based and conclusion-based procedures on the one hand, and the merging 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 merging approach is in tracking the truth, and how it compares with the premise-based and conclusion-based procedures.

1 Introduction

The problem of judgment aggregation was first identified by the Law professors Lewis Kornhauser and Larry Sager [10, 11]. In their example, a court has to make a decision on whether a person is liable of breaching a contract (proposition R, or 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 (first premise P) and had contractual obligation not to do X (second premise Q). The legal doctrine can be formally expressed as the rule $(P \wedge Q) \leftrightarrow R$. Each member of the court expresses her judgment (in the form of yes/no) on the propositions P, Q and R such that the rule $(P \wedge Q) \leftrightarrow R$ is satisfied.

Suppose now that the seven members of the court make their judgments according to the following table:

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

Each judge expresses a consistent opinion, i.e. she says yes to R if and only if she says yes to both P and Q. However, propositionwise majority voting (consisting in the separate aggregation of the votes for each proposition P, Q and R via majority rule) results in a majority for P and Q and yet a majority for $\neg R$. This is clearly an inconsistent collective result. The paradox lies in the fact that majority voting can lead a group of rational agents to endorse an irrational collective judgment. The literature on judgment aggregation refers to such dilemma as the *discursive paradox* (or *doctrinal paradox*).

The first two escape-routes that have been suggested are the *premise-based procedure* and the *conclusion-based procedure* [14, 4, 13]. The first procedure is to let each member vote on each premise and to declare the defendant liable only if a majority of the court believes that she did the action X and that she was under contract obligation not to do X. The second procedure requires the judges to privately decide about P and Q and to publicly express their opinions on R only. The defendant will be declared liable if and only if a majority of the judges actually believes that she is liable. Clearly, in the conclusion-based procedure nothing can be said about the reasons supporting the final decision.

In [15] it has been argued that the two above suggested escape-routes from the paradox are not satisfactory methods for group decision-making. The premise-based procedure is problematic because it does not univocally identify what a premise is. To see why, suppose that a group of individuals make their judgments on the propositions A, B and C according to the decision rule $((A \wedge B) \vee (\neg A \wedge \neg B)) \leftrightarrow C$. It is easy to construct examples where premise-based procedure gives two divergent results depending on what we take to be the premises (the atomic propositions A, B and C or the disjuncts $A \wedge B$ and $\neg A \wedge \neg B$). This problem was first noticed by Bovens and Rabinowicz [3] who referred to it as the *instability* of the premise-based procedure. On the other hand, the conclusion-based procedure avoids the paradox at the price of incomplete collective judgments. In all those situations in which a group has to reach a conclusion, but also needs to provide reasons for that decision (as in the original formulation of the doctrinal paradox), the conclusion-based cannot serve as proper aggregation method.

Therefore, a new aggregation procedure, providing a collective decision as well as the reasons for that decision, was introduced in [15]. This approach (that we will call merging — or fusion — procedure) was inspired by a family of operators defined in artificial intelligence [7, 6] in order to merge finite sets of propositions. Not only complex collective decisions are paradox-free when the inconsistent collective judgments are ruled out from the set of possible solutions. Also, an outcome in the merging approach is a complete collective judgment on the premises and on the conclusion

However, situations like the Kornhauser and Sager' court example do not only require that consistent individual opinions are aggregated into a rational group judgment, but also that the group makes the *right* decision. The defendant factually is (or is not) guilty: There exists an objective truth that the court is trying to reach. Therefore, a natural question is: In addition to guarantee consistent group outcomes, does the merging procedure also select the correct decision? The present paper addresses this question.

An epistemic perspective on judgment aggregation and, in particular, on the premise-based and conclusion-based procedure, was discussed by Bovens and Rabinowicz in [3]. Following their work, and making various independence assumptions as in the Condorcet Jury Theorem, we will introduce our framework in order to test how good the fusion procedure is in tracking the truth. Finally, we will illustrate the results obtained by computer simulation and compare them with the results for premise-based and conclusion-based procedure.

Let us first start by briefly recalling the merging approach.

2 The merging procedure

The fusion procedure is inspired by an aggregation operator defined in artificial intelligence in order to combine several finite sets of propositions (*bases*) [7, 6]. In fact, one of the major problems that artificial intelligence needs to address is the combination of different and potentially conflicting sources of information. Examples are multi-sensor fusion, database integration and expert systems development.[1]

Clearly, belief fusion and judgment aggregation share a similar problem, viz. the definition of operators that produce collective opinion from individual bases. The discursive dilemma rests upon the fact that, when the individual judgments on atomic propositions conform to some logical constraints on those propositions, this does not ensure to obtain a consistent (i.e. obeying the same logical constraints) collective judgment set. On the other hand, one of the key points in the literature of belief fusion is precisely that the aggregation of consistent knowledge bases does not guarantee a consistent collective outcome. To overcome this problem, domain-specific restrictions (*integrity constraints*) are imposed on the final base. This ensures that unwanted solutions are ruled out from the set of possible group outcomes.

Let $N = \{1, 2, ..., n\}$ ($n \geq 2$) be a set of individuals making their judgments on a given finite set X of propositions (*agenda*). Let $\mathcal{L}$ be a finitary propositional language, built up from a finite set $\mathcal{P}$ of propositional letters and the usual logical connectives ($\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$). The belief base K_i of an agent i is a consistent and complete finite set of atomic propositions and compound propositions (this corresponds to the individual judgment set).

An *interpretation* is a function $\mathcal{P} \rightarrow \{0, 1\}$ and it is represented as the list of the binary evaluations. For example, given three propositional variables P, Q and R, the vector (0,1,0) stands for the interpretation in which P and R are false and Q is true. Let $\mathcal{W} = \{0, 1\}^{\mathcal{P}}$ be the set of all interpretations. An interpretation is a *model* of a propositional formula if and only if it makes the formula true in the usual truth functional way.

[1]See [5] for a survey on logic-based approaches to information fusion.

IC is the belief base whose elements are the integrity constraints. These are extra conditions imposed on the result of the merging operator. Given a multi-set $E = \{K_1, K_2, \ldots, K_n\}$ and IC, a merging operator $\mathcal{F}$ is a function that assigns a belief base to E and IC. By borrowing the term from judgment aggregation, we call E a *profile*. Let $\mathcal{F}_{IC}(E)$ denote the collective belief base resulting from the IC merging on E. In a model-based merging operator the only possible collective outcomes are the models of IC. A majority fusion operator will the select the (eventually more than one) model that minimizes the *distance* to the profiles.

The most widely used distance in the literature is the Hamming distance. This is defined as the number of propositional letters on which two interpretations differ. For example, the Hamming distance between $\omega = (1, 0, 0, 1)$ and $\omega' = (0, 1, 0, 1)$ is $d(\omega, \omega') = 2$.

The first step is to determine the Hamming distance between those interpretations that are models of IC and the models of each base K_i in the profile E. The next step is to assign a distance value to each model of IC and a profile E. This is defined by the sum of the Hamming distances defined before.

To illustrate how the majority belief fusion operator works, we apply it to our initial 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 the following:

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

The table below shows the result of the IC majority fusion operator on $E = \{K_1, \ldots, K_7\}$. The row with a shaded background correspond to the selected collective outcome.

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

Because $\mathcal{F}_{IC}(E)$ is an IC merging operator, the possible collective outcomes are chosen among the interpretations that are models of IC. Thus, no paradox arises by using this fusion operator. 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.

The question we want to address now is whether the fusion approach not only prevents the discursive dilemma, but also is a good truth-tracker. Hence, whether a group that applies the merging procedure can not only keep away from irrational decisions, but has also a good chance to make the right decision. Using the Condorcet Jury Theorem, Bovens and Rabinowicz have explored how good truth-trackers the premise-based and the conclusion-based procedures are.

Our framework is introduced in the next section following [3] and making various independence assumptions as in the Condorcet Jury Theorem. We will then present some results about the fusion procedure and, finally, we will compare the performance of the fusion operator with the performance of the premise-based and the conclusion-based approaches described in [3].

3 The framework

The Condorcet Jury Theorem provides a justification for the majority rule in epistemic terms. It states that if the chance that an individual correctly judges the truth or falsity of a proposition is greater than fifty percent (her *competence*), then the chance that the majority of the group will come to the right decision will increase with the size of the group. In other words, individual probabilities turn into a group probability that is greater. More precisely, the Condorcet Jury Theorem can be formulated as follows:

> Suppose there is a group of N individuals (with N odd and greater than 1). Assume also that each group member has a chance $0.5 < p < 1$ of correctly assessing the truth or falsity of a proposition, and this chance does not depend on the other group member's judgments. Then, the probability that the group's majority judgment on that proposition is correct is greater than p and converges to 1 as the number of voters increases to infinity.

The Condorcet Jury Theorem requires that the number of voters is odd, that the voters are equally competent and independent. In order to avoid computational complexity, we need to make additional assumptions. These are as in [3]:

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

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

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

As [3], we will model the merging procedure for $P \wedge Q \leftrightarrow R$. Both the literature on judgment aggregation and the fusion approach assume that each individual judgment set is logically consistent. Hence, for $P \wedge Q \leftrightarrow R$ only four situations are possible (their corresponding models are also annotated):

$S_1 = \{P, Q, R\} = (1, 1, 1)$
$S_2 = \{P, \neg Q, \neg R\} = (1, 0, 0)$
$S_3 = \{\neg P, Q, \neg R\} = (0, 1, 0)$
$S_4 = \{\neg P, \neg Q, \neg R\} = (0, 0, 0)$

From (a), we derive that the prior probabilities of the four possible situations are (with $\bar{x} := 1 - x$):

$$\mathcal{P}(S_1) = q^2;\ \ \mathcal{P}(S_2) = \mathcal{P}(S_3) = q\bar{q};\ \ \mathcal{P}(S_4) = \bar{q}^2$$

We now want to calculate the probability that fusion ranks the right judgment set first (let us denote this proposition with $\mathcal{P}(F)$). Note that $\mathcal{P}(F) = \sum_{i=1}^{4} \mathcal{P}(F|S_i) \cdot \mathcal{P}(S_i)$. Thus, we have to calculate the conditional probabilities $\mathcal{P}(F|S_i)$ for $i = 1, \ldots, 4$. To see how it works, suppose that S_1 is the right judgment set. Then n_i (of N) voters will vote for profile S_i, with $n_1 + n_2 + n_3 + n_4 = N$.

We have seen that the majority merging operator selects the (eventually more than one) model that minimizes the distance to the profiles. This means that — if S_1 is the right judgment set — fusion is a good truth-tracker 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 for the situations S_i $(i = 1, \ldots, 4)$:

$$\begin{aligned} d_1 &= 2n_2 + 2n_3 + 3n_4 \quad &; \quad d_2 &= 2n_1 + 2n_3 + n_4 \\ d_3 &= 2n_1 + 2n_2 + n_4 \quad &; \quad d_4 &= 3n_1 + n_2 + n_3 \end{aligned}$$

For example, d_1 is obtained by summing the distances between S_1 and S_2, S_3 and S_4 times the number of voters for each S_i. The Hamming distance between S_1 and S_2 is 2. Hence, this value is multiplied by the number of voters for S_2 (which is n_2). The values $2n_3$ and $3n_4$ are obtained with the same procedure with respect to S_3 and S_4.

Finally, we can calculate the probability that fusion selects S_1 provided that S_1 is the right judgment set :

$$\mathcal{P}(F|S_1) = \sum_{n_1,\ldots,n_4=0}^{N} \binom{N}{n_1, \ldots, n_4} p^{2n_1} (p\bar{p})^{n_2+n_3} \bar{p}^{2n_4}\, \mathcal{C}(n_1, \ldots, n_4)$$

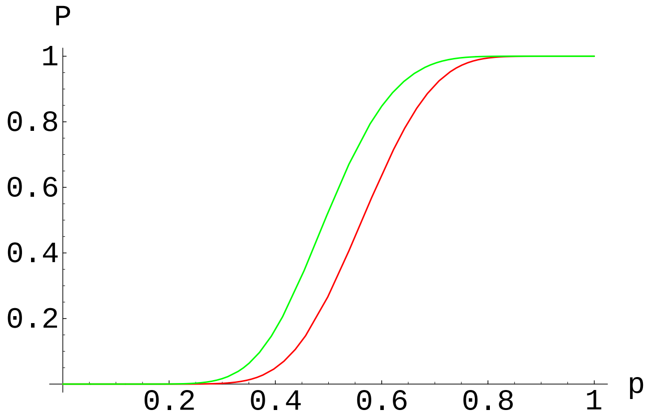
The sum is constrained: $\mathcal{C}(n_1, \ldots, n_4) = 1$ if (i) $\sum_{i=1}^{4} n_i = N$ and (ii) $d_1 \leq \min(d_1, \ldots, d_4)$. Otherwise $\mathcal{C}(n_1, \ldots, n_4) = 0$.

We can now present some results about how good in selecting the right judgment set the merging operator is. We will then turn to some figures showing the behavior of the fusion approach compared to the premise-based and the conclusion-based procedures.

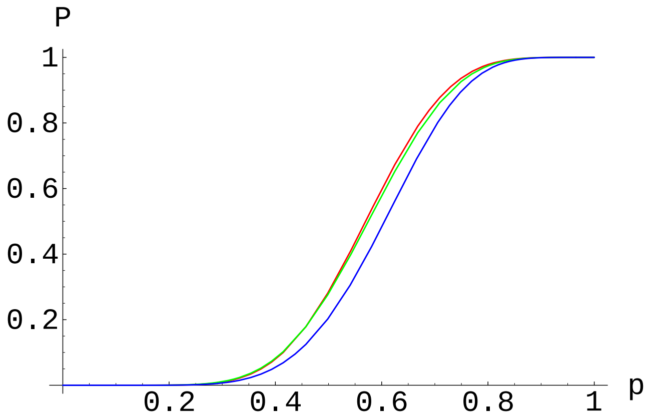
4 Results

4.1 Testing the merging procedure

In Section 2, we have seen that the notion of distance used in the fusion approach defines a pre-order on the possible outcomes. Thus, our first question is how good is belief fusion in selecting the correct judgment set as the first element in the ranking. The figure below shows how fusion ranks the right profile first (the red curve — *abbr.* R) or second (the green curve — *abbr.* G) for $N = 19$ and $q = .5$

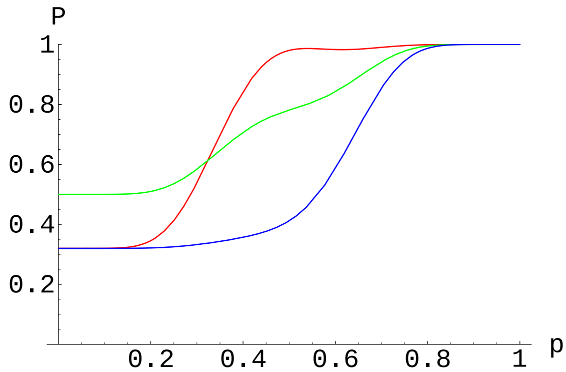


We now want to compare how the probability that fusion ranks the right profile first depends on different values of q. The plot is for $N = 11$ and three values of q: $q = .2$ (R), $q = .5$ (G), $q = .8$ (the blue line — *abbr.* B)



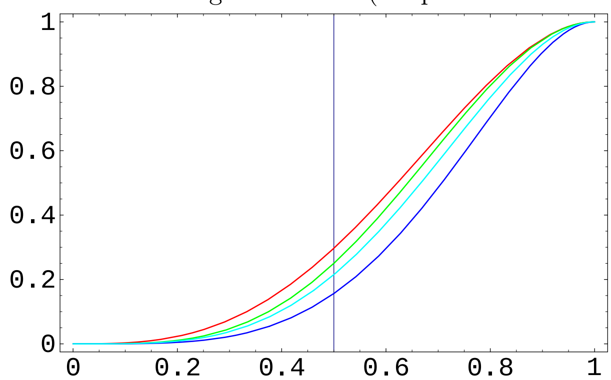
It turns out that the probability P is quite independent on the priors q.

However, different values of the priors q matter when we look at how good fusion is in ranking first a judgment set with the right decision (but not necessarily the correct reasons for that decision). The figure below shows the results for $N = 17$ and $q = .2$ (R), $q = .5$ (G), $q = .8$ (B)



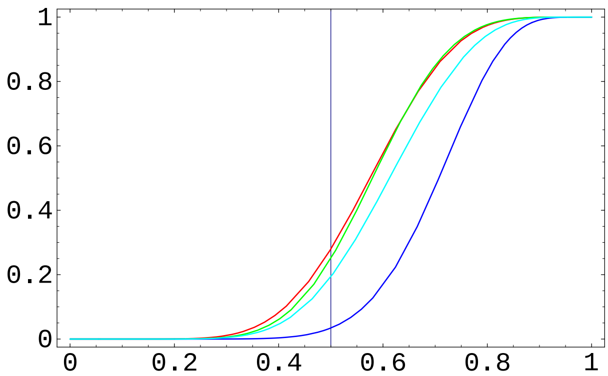
4.2 The merging approach compared to the premise-based and the conclusion-based procedures

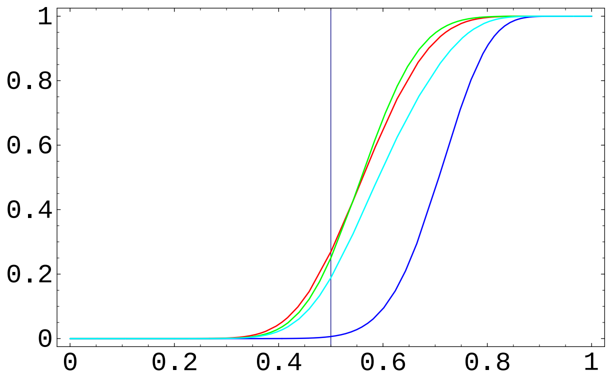
The second set of our results compare the merging approach with the premise-based and the conclusion-based procedures. We start with a small number of voters ($N = 3$) and $q = .5$. The first figure below shows how fusion ranks the right profile first (R) compared with premise-based procedure (G), conclusion-based procedure (B) and the conclusion-based procedure with the right reasons (turquoise curve — *abbr.* T).



The merging operator outperform all the other procedures. However, it is no surprise that the second best procedure is the premise-based one. In fact, from the Bovens and Rabinowicz's findings, we know that if we aim at reaching the right decision for the right reasons, we should prefer the premise-based procedure to the conclusion-based.

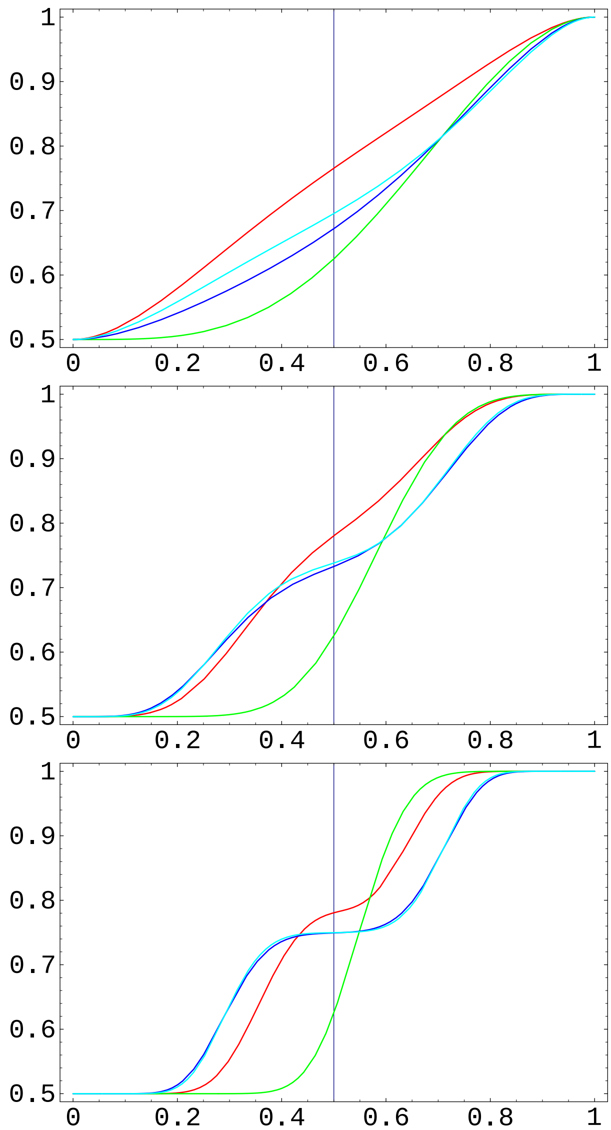
The next two figures illustrate the behavior of the fusion operator compared to the contender procedures when the number of voters increases ($N = 11$ and $N = 21$ respectively):





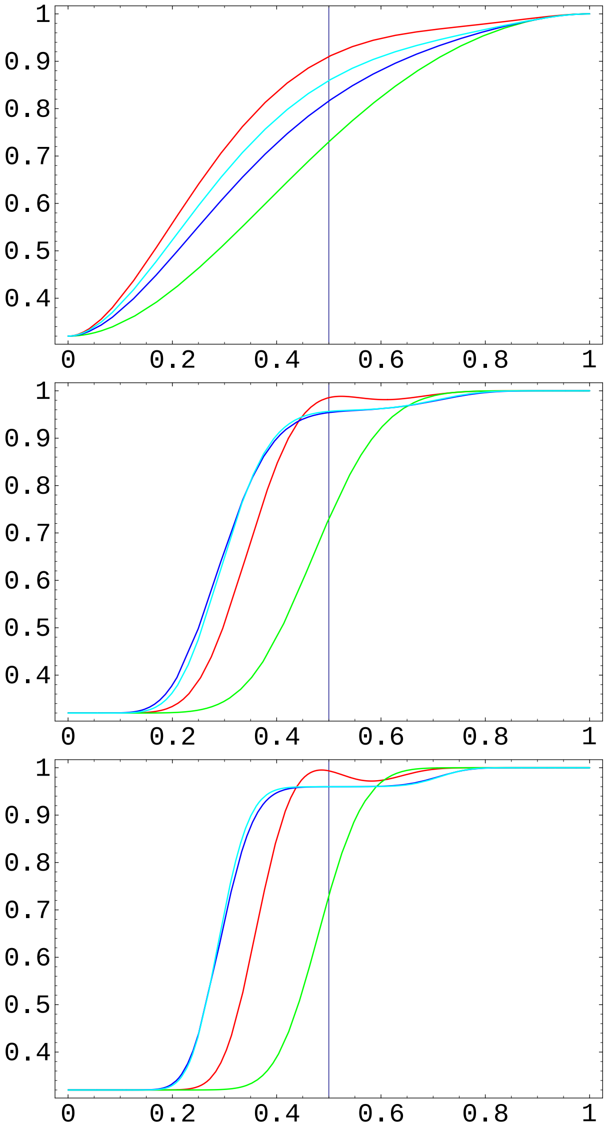
Clearly, the fusion approach (R) does significantly better than the conclusion-based (for the right reasons or not — the turquoise and blue lines). However, for high values of competence p, the premise-based procedure (G) is slightly better as a truth-tracker.

We now turn to evaluate how the fusion approach (R) ranks a judgment set with the right result (but not necessarily the right reasons) first, and we contrast this with the premise-based (G), the conclusion-based (B) and the conclusion-based for the right reasons (T) procedures. As before, we test the procedures for $q = .5$ and for increasing number of voters ($N = 3$, $N = 11$ and $N = 31$ respectively):



It turns out that fusion greatly outperforms all the other aggregation procedures under investigation for small size groups. Yet, as the size of the group increases, both the conclusion based procedures (B and T lines) do better than the fusion operator for low values in competence, and the premise-based procedure (G) does better than fusion for high values of p. But, for the middle values of p, merging is always superior. We can also observe that, whenever the fusion is not the best procedure, it lies in-between the premise-based and the conclusion-based procedures.

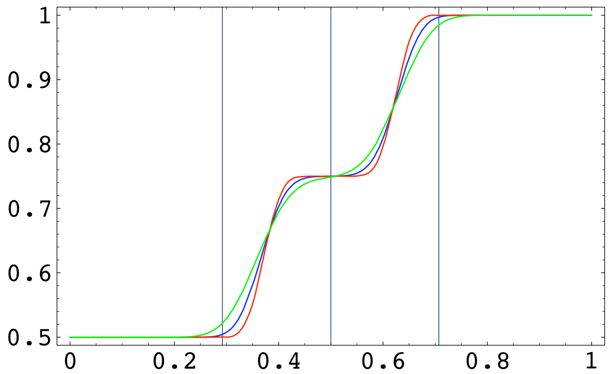
The next three pictures illustrate the same comparison, for a different value of prior ($q = .2$). As before, the number of voters increases, from $N = 3$ (first plot) to $N = 21$ (second) and $N = 51$ (third plot).



Again, for small-sized groups, fusion is the best procedure to reach the right decision. When the number of the voters increase, the conclusion-based procedures (B and T curves) do better than fusion, but only for low competence values. Different values of priors do not undermine the superiority of the fusion approach in the middle values of p ($.4 \leq p \leq .6$). More interesting, for p around .5, the probability that fusion selects the right decision is almost 1! Finally, for higher values of p, premise-based procedure is only slightly better than the merging operator.

The last figure shows that the trend of fusion for $N =?$ voters (G) is very close to the curve obtained for increasingly higher number of voters: $N =?$ (B) and $N =?$ (R).

Say something about the two values (vertical lines) of BR06.



Summarizing, our computer simulations show that the fusion approach does especially well for middling values of p. Nevertheless, for other values of p, the fusion operator is often in between the premise-based and the conclusion-based procedures (whichever is better in the case at hand).

Hypothesis: Fusion works best for realistic cases ($p \approx .5$) and takes the best of both worlds, i.e. PBP and CBP.

5 Conclusion and future plans

- Belief merging as a valuable tool to aggregate individual judgment sets:
  - no paradox
  - ranking on all possible social outcomes
  - no instability problem
  - propositions can be give different interpretation $\Rightarrow$ different fusion operators?
- We examined how good a truth-tracker the fusion approach is.
- In future work, we will:
  - work with a larger number of voters,
  - a larger number of premises,
  - examine the disjunctive case, and
  - use other distance measures.
- We will also explore the political and philosophical significance of the fusion approach.

References

[1] K. J. Arrow (1963) *Social Choice and Individual Values*, Wiley, New York, second edition.

[2] K. J. Arrow, A. K. Sen, and K. Suzumura (eds.). *Handbook of Social Choice and Welfare*, Vol.1, Elsevier, 2002.

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

[4] B. Chapman. Rational Aggregation. *Politics, Philosophy and Economics*, 1(3): 337-354, 2002.

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

[6] S. Konieczny. *Sur la Logique du Changement: Révision et Fusion de Bases de Connaissance*, Ph.D. dissertation, University of Lille, France, 1999.

[7] S. Konieczny and R. Pino-Pérez. On the logic of merging. In *Proceedings of KR'98*, Morgan Kaufmann, pages 488–498, 1998.

[8] S. Konieczny and R. Pino-Pérez. Propositional belief base merging or how to merge beliefs/goals coming from several sources and some links with social choice theory. *European Journal of Operational Research*, 160(3): 785–802. 2005.

[9] L. A. Kornhauser. Modeling collegial courts II. Legal doctrine. *Journal of Law, Economics and Organization*, 8: 441-470, 1992.

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

[11] L. A. Kornhauser and L. G. Sager. The one and the many: Adjudication in collegial courts. *California Law Review*, 81: 1-51, 1993.

[12] C. List. Judgment Aggregation: a Bibliography on the Discursive Dilemma, the Doctrinal Paradox and Decisions on Multiple Propositions, 2006.
http://personal.lse.ac.uk/LIST/doctrinalparadox.htm

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

[14] P. Pettit. Deliberative democracy and the discursive dilemma. *Philosophical Issues*, 11: 268–299, 2001.

[15] G. Pigozzi. Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. Forthcoming in *Synthese*, 2007.

Gabriella Pigozzi
ILIAS, Computer Science and Communications
University of Luxembourg
6, rue Richard Coudenhove - Kalergi
L-1359 Luxembourg
Email: gabriella@pigozzi.org

Stephan Hartmann
Department of Philosophy, Logic and Scientific Method
London School of Economics and Political Science
Houghton Street
London WC2A 2AE - UK
Email: S.Hartmann@lse.ac.uk