Hierarchical fusion of expert opinion in the Transferable Belief Model, application on climate sensitivity

Minh Ha-Duong

Hierarchical fusion of expert opinions in the Transferable Belief Model, application to climate sensitivity*

Minh Ha-Duong†

May 19, 2008

Abstract

This paper examines the fusion of conflicting and not independent expert opinion in the Transferable Belief Model. A hierarchical fusion procedure based on the partition of experts into schools of thought is introduced, justified by the sociology of science concepts of epistemic communities and competing theories. Within groups, consonant beliefs are aggregated using the cautious conjunction operator, to pool together distinct streams of evidence without assuming that experts are independent. Across groups, the non-interactive disjunction is used, assuming that when several scientific theories compete, they can not be all true at the same time, but at least one will remain. This procedure balances points of view better than averaging: the number of experts holding a view is not essential.

This approach is illustrated with a 16 expert real-world dataset on climate sensitivity obtained in 1995. Climate sensitivity is a key parameter to assess the severity of the global warming issue. Comparing our findings with recent results suggests that the plausibility that sensitivity is small (below 1.5C) has decreased since 1995, while the plausibility that it is above 4.5C remains high.

Résumé:

Ce texte propose examine la fusion des opinions d'experts en situation de controverse scientifique, à l'aide du Modèle des Croyances Transférables.

Parmi les procédures qui combinent les experts symmétriquement, nous constatons que lorsque les croyances sont bayésiennes (une modélisation classique s'appuyant sur les probabilités), l'opérateur de disjonction non-interactif donne de meilleurs résultats que les autres (conjonction prudente, la conjonction non-interactive, règle de Dempster).

Puis nous proposons une procédure de fusion hiérarchique. En premier lieu, une partition des experts en écoles de pensée est réalisée à l'aide des méthodes de sociologie des sciences. Puis les croyances sont aggrégées à l'intérieur des groupes avec l'opérateur de conjonction prudente: on suppose que tous les experts sont fiables, mais pas qu'ils constituent des sources d'information indépendantes entre elles. Enfin les groupes sont combinés entre eux par l'opérateur de disjonction non-interactive: on suppose qu'au moins l'une des écoles de pensée s'imposera, sans dire laquelle. Cette procédure offre un meilleur équilibre des points de vue que la simple moyenne, en particulier elle ne pondère pas les opinions par le nombre d'experts qui y souscrivent.

La méthode est illustrée avec un jeu de données de 1995 obtenu en interrogeant 16 experts à propos de la sensibilité climatique (le paramètre clé exprimant la gravité du problème du réchauffement global). La comparaison de nos résultats avec la littérature récente montre que la plausiblité que ce paramètre soit relativement faible (moins que 1.5C) a diminué depuis 1995, alors que la plausibilité qu'il soit au delà de 4.5C n'a pas décru.

*Submitted to the International Journal of Approximate Reasoning

†Chargé de recherche au CIRED-CNRS. Mail to haduong@centre-cired.fr. This research was supported by the Centre National de la Recherche Scientifique, France and by the Center for Integrated Assessment of Human Dimensions of Global Change, Pittsburgh PA, created through a cooperative agreement between the National Science Foundation (SBR-9521914) and Carnegie Mellon University. We gratefully thank David Keith for providing access to this survey data. We acknowledge the precious comments of an anonymous referee, of T. Denœux and CIRED colleagues.

1 Introduction

Is there a single all-purpose aggregation method for expert opinions ? According to Ouchi [2004], the answer is negative. Indeed, there are at least three different ways to represent mathematically an expert opinion. One is probabilistic risk analysis [Cooke and Goossens, 1999]. Another approach is to use the fuzzy numbers theory to combine opinions represented as possibility distributions [Sandri et al., 1995]. We are interested here in a third approach: Dempster-Shafer theory of evidence [Shafer, 1990].

We will use a variant of the theory of evidence named the Transferable Belief Model, and more specifically examine new operators for information fusion recently proposed by Denœux [2008]. We study the applicability of these operators for the aggregation of expert opinion, using a real-world dataset from Morgan and Keith [1995].

This dataset illustrates four challenges for mathematical aggregation methods. First, it cannot be assumed that opinions are, statistically speaking, independent: that would overestimate the precision of the actual information in the field. Second, there is complete contradiction among experts: aggregation methods that take somehow the intersection of the opinions can not work when the intersection is empty. Third, the disagreement between experts is not a balanced opposition, but rather a dissent minority situation. Some aggregation methods, like averaging, give more weight to views held by a larger number of experts, but this is arguably unbalanced because scientific theories should be evaluated only on their own merits, not by the number of proponents. And fourth, there is no proxy available to calibrate the reliability of experts, so we can't assume that some experts are less reliable than others.

Section 2 describes the mathematical theory for information fusion in the Transferable Belief Model. It defines three ways to combine opinions, namely the non-interactive conjunction, non-interactive disjunction and the cautious conjunction. Section 3 discusses theoretically these operators, along with the well known averaging and Dempster's rules.

We argue that none of these ways to combine expert opinions adequately addresses the four challenges defined above. To this end, we propose a hierarchical method for the fusion of expert opinion. Experts are not combined symmetrically, but grouped into schools of thought. Within groups, beliefs are combined using the cautious conjunction rule, whereas across groups the non-interactive disjunction is used.

These approaches are numerically applied and compared in Section 4. The data used in this study represents the opinion of 16 experts on climate sensitivity, a key parameter of the climate change issue. We examine which fusion method works best, showing that the answer is not the same for Bayesian beliefs and consonant beliefs. The discussion Section 5 analyzes sensitivity of the results, comparing them with the more recent literature, and points to existing social science concepts that could be used with the proposed hierarchical approach. Section 6 concludes.

2 Operators of the Transferable Belief Model

2.1 Basic Belief Assignments

The Transferable Belief Model is an elaboration of the Dempster-Shafer mathematical theory of evidence [Dempster, 1967, Shafer, 1976], a theory that represents and combines uncertain beliefs. This section briefly reminds the parts of this model that are relevant for expert aggregation. The reader may refer to Denœux [2006], Smets [2000] for a more complete exposition including the mathematical demonstrations.

As usual, let us denote by Ω a frame of reference, that is, a set of mutually exclusive states of the world. This paper assumes a finite number of states of the world. Classical probability theory represents uncertainty by allocating a unit mass of belief among states of the world, that is a function $p : \Omega \to [0, 1]$ such that $\sum_{\omega \in \Omega} p(\omega) = 1$.

Dempster-Shafer theory of evidence represents uncertainty by allocating the unit mass of belief among *subsets* of the frame of reference Ω. Formally, let 2^Ω denote the power set of Ω, that is the set of all its subsets. Elements of 2^Ω will be denoted with upper case letters such as $A \subseteq \Omega$ or $X \subseteq \Omega$. The empty subset will be denoted $\emptyset$. A basic belief assignment (BBA) is a function $m : 2^\Omega \to [0, 1]$ such that:

$$\sum_{A \subseteq \Omega} m(A) = 1 \tag{1}$$

The mass $m(A)$ is the portion of the total belief supporting A which do not support more precisely any specific subset of A. Any subset $A \subset \Omega$ such that $m(A) > 0$ is called a *focal set* of m.

As a classical example, consider a drawing from an urn containing white, black, and red marbles ($\Omega = \{\text{white}, \text{black}, \text{red}\}$). Knowing only that there is 1/3 of white marbles would lead to the basic belief assignment defined as: $m(\{\text{white}\}) = 1/3$, $m(\{\text{black}, \text{red}\}) = 2/3$. This is not the same as drawing from an urn known to have 1/3 of each color, which would be represented with the basic belief assignment defined as: $m(\{\text{white}\}) = m(\{\text{black}) = m(\{\text{red}\}) = 1/3$.

For any subset $A \subseteq \Omega$, the BBA that represents the certain belief that the state of the world is in A is the indicator function $\mathbf{1}_A \ : 2^\Omega \rightarrow [0, 1]$ defined by:

$$\begin{cases} \mathbf{1}_A(A) = 1 \\ \mathbf{1}_A(X) = 0 \quad \text{if } X \neq \ A \end{cases} \tag{2}$$

The basic belief assignment $\mathbf{1}_\Omega$ is called the *vacuous* BBA. It allocates all belief to Ω itself, and represents the absence of information. Following up the urn example above, the vacuous BBA is defined by $m(\{\text{white}, \text{black}, \text{red}\}) = 1$, $m(X) = 0$ otherwise. Again, this is not the same as the equidistribution. We will call $m(\Omega)$ the weight of ignorance.

In Shafer's original theory, in addition to Equation 1, a BBA must verify the axiom $m(\emptyset) = 0$. The Transferable Belief Model drops this constraint: it allows non-zero belief mass to the empty set, and considers that renormalization, defined as follows, should not be applied systematically. *Renormalizing* a BBA m means replacing it by the BBA m^* defined as:

$$\begin{cases} m^*(\emptyset) = 0 \\ m^*(A) = \frac{m(A)}{1 - m(\emptyset)} \quad \text{if } A \neq \ \emptyset \end{cases} \tag{3}$$

Smets [1992] discusses two reasons for using unnormalized BBAs: incompleteness and conflict. Incompleteness means that $m(\emptyset)$ measures the belief that something out of Ω happens. For example, if $\Omega = \{\text{Head}, \text{Tail}\}$ models a coin toss, then $m(\emptyset)$ is the extend of the belief that the coin could fall sideways, break or otherwise disappear. In what follows, we assume that the states of the world are collectively exhaustive, disregarding incompleteness.

Therefore in this context, $m(\emptyset)$ relates to conflict only. The number $m(\emptyset)$, called weight of conflict, is a measure of internal contradiction which arises when forming belief from information sources pointing in different directions. The extreme case $\mathbf{1}_\emptyset$ represents being confounded by completely contradictory information sources. As opposed to the vacuous BBA $\mathbf{1}_\Omega$ which can be adopted when one has no information at all, the complete contradiction BBA $\mathbf{1}_\emptyset$ represents a situation of confusion arising from too much information inconsistency.

2.2 Non-interactive fusion operators

The two basic combination rules of the transferable belief model will be denoted $\textcircled{\cap}$ and $\textcircled{\cup}$. They provide a way to compute the "intersection" or the "union" of two experts' opinions.

Before turning to the formal definitions, these rules will be illustrated on a special case: the fusion of two experts holding certain beliefs. Expert 1 views that the state of the world is in $A \subseteq \Omega$, and expert 2 views that the state of the world is in $B \subseteq \Omega$. Their beliefs are represented, respectively, by $\mathbf{1}_A$ and $\mathbf{1}_B$.

To start with $\textcircled{\cup}$, consider what the result of the fusion should be when one thinks that either expert 1 or expert 2 is a reliable information source. In this case, one is led to believe that the state of the world is in A or B, that is in $A \cup B$. The $\textcircled{\cup}$ combination rule is precisely such that $\mathbf{1}_A \textcircled{\cup} \mathbf{1}_B = \mathbf{1}_{A \cup B}$. It is called the non-interactive disjunction rule. This operator can be qualified as a "gullible" rule, which means it accepts all that it is told.

The non-interactive conjunction rule $\textcircled{\cap}$ is meant to be used when one thinks that both expert 1 and expert 2 are reliable information sources. Apparently, there are two cases. When $A \cap B$ is non-empty, the fusion of the two opinions should be the belief that the state of the world is in $A \cap B$. When the experts have no common ground, that is $A \cap B = \emptyset$, then we have a contradiction problem. However, in the transferable belief model this is not a problem, this state of affairs is represented with $\mathbf{1}_\emptyset$. So actually in both cases, the operator should be such that $\mathbf{1}_A \textcircled{\cap} \mathbf{1}_B = \mathbf{1}_{A \cap B}$. This operator can be qualified as a “consensus” rule, to mean that all parties accept the result.

For reasons that will become apparent with Equation 8, we define below these two combination rules with slightly more general functions than BBAs. Let use the Greek letter μ to denote a real-valued subset function $\mu : 2^\Omega \to \Re$ which verifies Equation 1, but may or may not be a basic belief assignment, that is, take values in [0,1] or not. The non-interactive conjunction of μ_1 and μ_2 is defined as the subset function $\mu_1 \circledcap \mu_2 : 2^\Omega \to \Re$ such that, for any subset X:

$$(\mu_1 \circledcap \mu_2)(X) = \sum_{\substack{A \subseteq \Omega \\ B \subseteq \Omega \\ A \cap B = X}} \mu_1(A) \times \mu_2(B) \tag{4}$$

In the same way, $\circledcup$ is defined by:

$$(\mu_1 \circledcup \mu_2)(X) = \sum_{\substack{A \subseteq \Omega \\ B \subseteq \Omega \\ A \cup B = X}} \mu_1(A) \times \mu_2(B) \tag{5}$$

These operators are commutative, associative and if μ_1 and μ_2 are two BBAs then the result is also a BBA. These properties allow to treat the experts symmetrically when combining their opinions. Vacuous beliefs $\mathbf{1}_\Omega$ is an absorbing element for disjunction and a neutral element for conjunction. Conversely, contradiction $\mathbf{1}_\emptyset$ is absorbing for conjunction and neutral for disjunction.

As an example, consider $\Omega = \{a, b\}$, and the BBA m defined by $m(\{a\}) = m(\{b\}) = 1/2$. Then $(m \circledcap m)(\{a\}) = (m \circledcap m)(\{b\}) = 1/4$, and $(m \circledcap m)(\emptyset) = 1/2$. Such a large weight of conflict in the result may seem surprising. One way out is to systematically renormalize, as described by Equation 3. The non-interactive conjunction $\circledcap$ followed by normalization is known as Dempster’s combination rule, usually denoted $\oplus$ in the literature :

$$m_1 \oplus m_2 = (m_1 \circledcap m_2)^* \tag{6}$$

However, in some situations the surprising result is the correct one, and renormalization should not be used. It depends on what is being modeled. Consider for example a setting in which two scientists simultaneously replicate a large number of fair coin tosses. Both conclude that $p(\text{Head}) = p(\text{Tail}) = 1/2$ in the long run. But if the experiments are independent, then results of the coin tosses were in conflict half the time. This suggests that the non-interactive conjunction $\circledcap$ is relevant to combine information sources only when some kind of independence relation can be assumed between information sources. It justifies why this operator is called *non-interactive*.

2.3 Factorization and cautious conjunction

The non-interactive combination rules should not be used to combine experts who share pieces of evidence. To perform information fusion in this kind of situations, Denœux [2008] introduced an operator called *cautious conjunction*. To define it mathematically, it is necessary to introduce first the factorization of BBAs.

For any proper subset $A \subset \Omega$ and any real number s, we denote A^s the function $\mu : 2^\Omega \to \Re$ such that:

$$\begin{cases} \mu(\Omega) = e^{-s} \\ \mu(A) = 1 - e^{-s} \\ \mu(X) = 0 & \text{if } X \neq A \text{ and } X \neq \Omega \end{cases} \tag{7}$$

The letter s stands for “Shafer’s weight of evidence”. This value was previously denoted w by Shafer [1976, Chapter 5]. But the recent literature [Denœux, 2006] uses the letter w to denote the “weight of evidence” defined by $w = e^{-s}$.

Regarding the interpretation of A^s, when $s \geq 0$ the function A^s is a BBA, but when $s < 0$ it is not, so A^s can generally not be interpreted as a state of belief. Smets [1995] has shown that for any BBA m such that $m(\Omega) > 0$ there is a unique function $\mathrm{s} : 2^\Omega \setminus \Omega \to \Re$ such that:

$$m = \bigcircledcap_{A \subset \Omega} A^{\mathrm{s}(A)} \tag{8}$$

Any BBA m such that $m(\Omega) > 0$ is the non-interactive conjunction of elementary pieces of the form $A^{\mathrm{s}(A)}$. The weights of evidence function s may take negative values, in which case the BBA is not *separable* according to Shafer [1976], who did not consider negative weights of evidence.

This unique factorization theorem allows us to come back to the interpretation of A^s. It can be seen as the change in one's beliefs realized when integrating with weight s a piece of evidence stating that the state of the world is in A. Positive infinity for s represents a perfectly convincing proof that the state of the world is in A. This remains excluded in the above definition, for reasons discussed further below. Negative weights $s < 0$ have an algebraic justification similar to that of negative numbers: considering A with weight s exactly counterbalances considering A with weight $-s$, to produce vacuous beliefs $\mathbf{1}_\Omega$. It is more difficult to achieve an intuitive understanding of negative information. Smets [1995] suggested that A^s, for a negative value of s, represents a 'good reason not to believe' that the state of the world is in A.

Let us denote $|X|$ the number of elements (cardinality) of a subset $X \subseteq \Omega$. The weights can be computed as follows, introducing the function q called the commonality function :

$$q(X) = (m \circledcirc \mathbf{1}_X)(X) = \sum_{A \supseteq X} m(A) \tag{9}$$

For any $X \subset \Omega$, note that $q(X) \geq m(\Omega)$, therefore $m(\Omega) > 0$ implies $q(X) > 0$, so the logarithm is well defined in the following:

$$s(X) = \sum_{A \supseteq X} (-1)^{|X|-|A|} \ln\big(q(A)\big) \tag{10}$$

Using equations 4 and 8, along with commutativity and associativity, it is straightforward to verify that, if two BBA m_1 and m_2 admit corresponding weight functions s_1 and s_2, their non-interactive conjunction can be computed simply by adding those:

$$m_1 \circledcirc m_2 = \bigcircledcirc_{A \subset \Omega} A^{s_1(A)+s_2(A)} \tag{11}$$

This property allows us to clarify the intuition behind the $\circledcirc$ operator. The non-interactive conjunction adds up distinct pieces of evidence. For example, when combining two experts who point exactly in the same direction A with the same weight s, the result is $A^s \circledcirc A^s = A^{2s}$. Once again, it is correct to argue that a stream of evidence pointing out in the same direction leads to stronger beliefs only when they are distinct.

To combine experts that share evidence, Denœux [2006, 2008] defined the cautious conjunction operator, denoted by $\owedge$. It combines any two BBA such that $m_1(\Omega) > 0$ and $m_2(\Omega) > 0$ by taking the maximum of their weight functions as follows:

$$m_1 \owedge m_2 = \bigcircledcirc_{A \subset \Omega} A^{\max(s_1(A)\,,\, s_2(A))} \tag{12}$$

It can be shown that if m_1 and m_2 are BBAs, then $m_1 \owedge m_2$ is also a BBA (this is immediate only when m_1 and m_2 are separable). This combination rule $\owedge$ is also commutative and associative, it treats experts symmetrically. It is also idempotent, that is $m \owedge m = m$, and distributes over the noninteractive rule $(m_1 \circledcirc m_2) \owedge (m_1 \circledcirc m_3) = m_1 \circledcirc (m_2 \owedge m_3)$.

Distributivity has an interesting interpretation related to the fusion of beliefs. Consider two experts in the following scenario. Expert 1's belief results from the noninteractive conjunction of two pieces of evidence, $m_1 = A^s \circledcirc B^t$. Expert 2 shares one piece of evidence with expert 1, and has an independent piece, so that $m_2 = A^s \circledcirc C^u$. Then distributivity implies that in the fusion, the shared evidence A^s is not counted twice $m_1 \owedge m_2 = A^s \circledcirc (B^t \owedge C^u)$.

2.4 Discounting Beliefs

A BBA m that verifies $m(\Omega) = 0$ cannot be factorized as described above. Equation 4 implies that $(\mu_1 \circledcirc \mu_2)(\Omega) = \mu_1(\Omega) \times \mu_2(\Omega)$, and we defined A^s in Equation 7 such that $A^s(\Omega) > 0$ always holds. Therefore the right hand side of Equation 8 cannot be BBA such that $m(\Omega) = 0$.

Various reasons justify to take basic beliefs assignments such that $m(\Omega) = 0$ with a grain of salt:

- No information source is 100% reliable, especially human ones.
- Many philosophers consider that fundamentally, scientific knowledge can never be absolute and definitive. On the contrary, it is necessarily based on a possibly large but finite number of human observations, and is always open to revision in front of new experimental evidence.
- The elicitation of expert's opinions, for example by asking them probability density functions, is necessarily coarse. Experts who allo-

cated no significant probability weight to extreme outcomes might have agreed that there was a very small possibility.

Shafer [1976, p. 255] proposed a simple way to add doubt to a basic belief assignment, called *discounting*. Let r be a number in [0, 1] called a *reliability factor*. Discounting the BBA m means replacing it by the BBA defined as:

$$\text{discount}(m, r) = rm + (1 - r)\mathbf{1}_{\Omega}. \tag{13}$$

Discounting allows beliefs to be factorized, and therefore combined using the cautious operator. Admittedly, discounting expert beliefs is deliberately blurring the data, a practice to be considered with extreme care if used at all. However, the reasons above justify using reliability factors, provided they are close enough to unity. The theoretical literature suggests that the fusion operators can be extended by continuity to deal with $m(\Omega) = 0$, and the sensitivity analysis will allow to check that results don't change much when r varies from 0.99 to 0.9999.

To sum up, that section defined a mathematical object used to represent an expert's opinion, denoted m and called a basic belief assignment (BBA). Four operators were defined to combine the opinions of two experts. The cautious conjunction operator $\owedge$ is meant to be used when experts share data. Otherwise, the non-interactive disjunction $\ocup$ takes the union of expert beliefs, while non-interactive conjunction $\ocap$ takes their intersection. Dempster's rule $\oplus$ is the renormalized non-interactive conjunction.

3 Fusion in the Transferable Belief Model

Having defined the mathematical framework and the binary fusion operators, we discuss now the complete procedures involving pooling the opinions of experts. Experts opinions are typically called for in situations in which there is not enough statistical evidence to support precise probabilities. This motivates our interest in an imprecise probability theory, such as the Transferable Belief Model, to model and combine beliefs. But imprecision has implications along the whole analytical process, not just the fusion of beliefs.

First, we discuss the implications of imprecision for the process' ultimate aim, to facilitate decision making. In our view, it implies to take a step back from the standard expected utility-maximization methodology implicit in probabilistic risk analysis. Second, we discuss the elicitation of opinions, a necessary step before the fusion, and question the validity of asking experts for probability density functions when more imprecise communication instruments can be used. Third, we discuss theoretically alternative ways to fusion beliefs, and fourth, we introduce a hierarchical approach to set the stage for the numerical application that will follow.

3.1 Decision Making and Uncertainty Communication

A reason why decision analysis processes involving the fusion of opinion is important is that when decisions involve different parties and scientific experts are not unanimous, policymakers will tend to break the symmetry of the elicitation process by myopically focusing on the results best supporting their interest. Another risk is that the press and other media outlets tend to paint issues in black and wite and to present two sides on everything. Organizations seeking a balanced point of view would overemphasize the most extreme positions in the group, even when they are actually a minority not representative of the experts' general opinion.

Smets [2005] offers a way to find a balanced point of view for decision making in the Transferable Belief Model. He points out that any BBA m such that $m(\emptyset) \neq 1$ defines a probability function $BetP$, that he calls the pignistic probability function of m, by:

$$BetP(\omega) = \frac{1}{1 - m(\emptyset)} \sum_{X \ni \omega} \frac{m(X)}{|X|} \tag{14}$$

Smets then argues that when beliefs are described by m, a decision-maker should choose actions that maximize the expected utility, where expectation is computed using the probability distribution $BetP$. However, other decision making rules can be used. For example, Cobb and Shenoy [2006] point out that the justification of $BetP$ is an argument of symmetry, which fundamentally contradicts the semantics of ignorance underlying the use of BBAs. These authors suggest instead to use another way to transform a BBA m into a probability distribution PlP, by renormalizing the plausibility of singletons:

$$PlP(\omega) = \frac{1}{K} \sum_{A \ni \omega} m(A) \tag{15}$$

where K is chosen so that $\sum_{\omega \in \Omega} PlP(\omega) = 1$.

But offering a single precise probability distribution from which expected utility maximization can provide an optimal answer to all policy issues is problematic. This position has been put forward by Morgan and Keith [1995], who argued that while expert aggregation can help decision making by presenting a simpler picture of the multiplicity of opinions on a given subject, in many cases presenting an aggregate probability is an oversimplification and it is better to leave with the decision-maker the task of the combining the judgment of all experts. Keith [1996] discusses in more detail why combining experts is rarely appropriate, and suggests instead to use alternative analysis framework such as seeking robust adaptive strategies or using scenario analysis to bound the problem.

Such an alternative framework could be provided by imprecise probabilities, where one uses sets of probabilities as basic uncertainty representation. Mathematically, it is straightforward to view a BBA as implicitly defining upper and lower bounds on admissible probabilities, using Equations 17 and 16. But there are significant semantic and technical difficulties with this view. The combination operators of the Transferable Belief Model, especially Dempster's rule, do not correspond directly with the combination operators of the imprecise probability theory.

Today there is no consensus in the scientific literature on precautionary decision making. The core agreement is that when beliefs are Bayesian, the standard approach is expected utility maximization. But in the more general case several rules have been proposed. Some reject what has been historically the first axiom in the field: that there is a total ordering between decisions. This leads to an analysis that recommends a set of *maximal* or *E-admissible* actions [Troffaes, 2007]. The set can be large, and results do not prescribe further which action should be selected in that set. These incomplete ordering approaches provide less guidance for decision-making than other rules. While this can be seen as a fatal limitation, rejecting the total ordering axiom follows the intuition that when there is a multiplicity of opinions, it is not possible to determine precisely and objectively an optimal answer to the policy issue.

In any case, communication of the results obtained by information fusion in the Transferable Belief Model does not have to put forward a single probability distribution. Instead, it can involve the measures of *belief* and *plausibility* associated with a BBA m. The value of the belief function for an event $X \subseteq \Omega$, denoted $bel(X)$, measures the strength of conviction that X must happen. The value of the plausibility function, denoted $pl(X)$, relates to the strength of conviction that X could happen. With the special case $bel(\emptyset) = pl(\emptyset) = 0$, these functions are defined when $X \neq \emptyset$ as :

$$bel(X) = \sum_{\substack{A \subseteq X \\ A \neq \emptyset}} m(A) \tag{16}$$

$$pl(X) = \sum_{\substack{A \subseteq \Omega \\ A \cap X \neq \emptyset}} m(A) \tag{17}$$

An intuitive interpretation of the theory of evidence sees $m(X)$ as a mass of belief that can flow to any subset of X. In this view, $bel(X)$ represents the minimal amount of belief that is constrained to stay within X, while $q(X)$ represents the amount of belief that can flow to every point of X, and $pl(X)$ the maximal amount of belief that could flow into X.

These functions can be used with the 'calibrated vocabulary' approach to communicate qualitatively about uncertainty. For example, if the analytical result is $bel(X) > 0.90$, it could be said that X is correct with *very high confidence.* If $pl(X) < 0.33$, it could be said that X is *unlikely.* No calibrated uncertainty vocabulary (probabilistic or otherwise) is universally accepted, and presumably it would depend upon the readers' language and culture.

Calibrated vocabularies have often been defined using a probability scale [IPCC, 2005, Weiss, 2003, Wallsten et al., 1986]. Such scales have to be revised, if one wishes to account for the multi-dimensionality of uncertainty: a basic belief assignment m allows to express levels of belief, plausibility, and contradiction.

3.2 Elicitation: Bayesian or Consonant BBAs?

We now turn to the methods for expert elicition, upstream of the information fusion itself. Approaches in experts elicitation include:

1. Expert's opinion elicitation in the tradition of risk assessment: asking the experts about probabilities, obtaining subjective probability density functions.
2. Expert's knowledge elicitation in the tradition of fuzzy logic: collecting opinions in natural language, modeling them with fuzzy numbers or possibility distributions [Zadeh, 1978].
3. Qualitative methods: asking the experts to make hypothetical choices. Opinions can then be deducted from elicited preferences, using the assumption that choices follow rationally from beliefs. This approach was applied to belief functions by Yaghlane et al. [2006]

Formally, information fusion in the Transferable Belief Model can deal with these three approaches. We will focus on the first two, because qualitative methods, which could potentially be used to elicit directly BBAs, are also less well developed. There is a natural embedding of probability distributions in the set of BBAs and a natural embedding of possibility functions in the set of BBAs.

Any probability function $p : \Omega \to [0, 1]$ naturally defines a BBA m by:

$$\begin{cases} m(\{\omega\}) = p(\omega) & \text{for any } \omega \in \Omega \\ m(X) = 0 & \text{if } |X| \neq 1 \end{cases} \tag{18}$$

A BBA m that naturally corresponds with a probability p by the above equation is said to be *Bayesian*. A BBA is Bayesian when its only focal sets are singletons.

By definition, a normalized possibility distribution is a function $\pi : \Omega \to [0, 1]$ such that $\max_{\omega\in\Omega} \pi(\omega) = 1$. Given such π, a BBA m naturally associated with π can be computed via its commonality function as follows:

$$q(A) = \min_{\omega\in A} \pi(\omega) \tag{19}$$

$$m(A) = \sum_{B\supseteq A} (-1)^{|B|-|A|} q(B) \tag{20}$$

In the numerical application (section 4 and following), we will use a dataset where opinions are given as probabilities, using Equation 18 to transform them into Bayesian belief functions when needed.

We will also explore information fusion when beliefs are more imprecise. This is necessary theoretically because probabilities are a very special kind of basic belief assignments, in which all belief mass is supported by singletons, and the information fusion methods need to be tested in a more general case.

There is also a potential practical interest to explore information fusion without assuming that beliefs are Bayesian. While in this specific dataset, as in many other, opinions are specified with probabilities, other elicitation exercises might use different approaches. These include providing judgements using natural language, probability bounds or possibility estimates. We argued that presenting a single probability distribution was not justified when statistical data is insufficient, even considering the whole pool of experts. This data scarcity argument is even stronger at the individual level, since each expert holds only a fraction of the data.

In order to compare better the fusion of Bayesian and non-Bayesian beliefs, we will re-use the same dataset and transform each expert's elicited distribution into a corresponding consonant belief function. This transformation problem was already discussed by Sandri et al. [1995] in a possibilistic context, which is not surprising given that most of the existing available datasets are probabilistic.

There are many ways to transform a probability p into a BBA m, starting with the natural injection defined Equation 18. But if we relax the assumption that beliefs in the mind of experts are necessarily Bayesian, a principle of least commitment (or maximal uncertainty) can be used to compute which m an expert could have held, knowing that it has answered the probability distribution p. The principle is applied as follows. Given p, consider the set M of belief functions consistent with p, for some definition of consistency. Then select m as the member of M which has the most uncertainty in it, having defined an uncertainty-related order relation that admits a single maximum in M.

Dubois et al. [2008] suggested to select for M the set of all BBAs m such that $BetP = p$, where $BetP$ is defined by Equation 14. This set is never empty because it contains the BBA naturally corresponding to p itself. This amounts to argue that even if the elicitation procedure does not explicitly use bets, experts, when asked to provide probabilities, actually provided pignistic probabilities ($BetP$ defined Equation 14), that is, probabilities they would use if they were asked to bet.

Following the least commitment principle, one then computes the least committed belief functions compatible with these pignistic probabilities. The uncertainty order relation is defined as follows: for any two BBAs m_1 and m_2 with respective commonality function q_1 and q_2, if $q_1(A) \geq q_2(A)$ for all $A \subset \Omega$, we write that $m_1 \sqsubseteq_q m_2$.

Dubois et al. [2008] states that there is an unique maximum in M with respect to $\sqsubseteq_q$, which can be computed as follows. Order the states of the world from most to least probable, that is $p(\omega_{n_1}) > \cdots > p(\omega_{n_{|\Omega|}})$. Consider the sets $A_k = \{\omega_{n_1}, \ldots, \omega_{n_k}\}$ and assign to A_k the belief mass:

$$m(A_k) = |A| \times \big(p(\omega_{n_k}) - p(\omega_{n_{k+1}})\big) \qquad (21)$$

with the convention that $p_{n_{|\Omega|+1}} = 0$. The procedure is illustrated on Figure 3.2, which demonstrates graphically that m is indeed a basic belief assignment, it adds up to unity. Note that the focal sets A_k are nested, that is $A_k \subset A_{k+1}$ for all k. In this case, it is said that m is *consonant*. It can be shown that the result m is naturally associated with a possibility distribution (via Equation 20).

For each expert i, we have a method to transform the Bayesian belief function (corresponding to the elicited probability distribution p_i) into a consonant belief function (corresponding to a possibility distribution that we will denote π_i).

3.3 Symmetric Fusion of Expert Opinions

Having discussed opinion elicitation and decision making, we now turn to the fusion of opinions. The literature offers many rules to combine beliefs, see Smets [2007] for a survey. This section examines systematically ten ways to combine opinions symmetrically: five operators defined above, each used with or without discounting.

We will explore two discounting options. The high reliability factor, $r = 0.999$, amounts to prac-

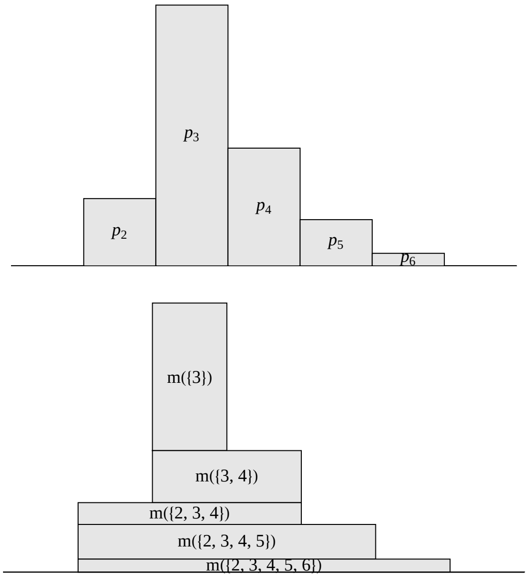

Figure 1: From Bayesian to consonant beliefs. Top, Bayesian beliefs (from expert 1 in the dataset). Assuming that the width of each rectangle is 1, and its height is proportional to the probability, the area of the rectangle denoted p_3 is p_3, and the sum of all rectangles' area is 1. Bottom, the corresponding consonant belief function. The area of the rectangle denoted $m(\{3, 4\})$ is the belief mass going to the focal set $\{3, 4\}$. The slices are cut horizontally, but the outline remains the same. The total area remains 1, meaning that m is a BBA.

tically no discounting at all, but is technically necessary to ensure that beliefs can be factorized and combined using the cautious operator. A medium reliability factor, $r = 0.8$, can be justified as in 2.4. The five operators are: the non-interactive conjunction and disjunction, the cautious conjunction, Dempster's combination rule, and averaging.

Theoretical analysis allows us to disregard seven of the ten different ways to fusion opinions, because they can be expected to give mathematically degenerate or otherwise uninteresting results in the context of expert opinion fusion.

Consider first averaging, also called the linear opinion pool. It is mathematically equivalent to discount the opinions before averaging, or to discount after averaging. But there is no reason to discount the average opinion, once it is computed. That only adds unjustified imprecision to the result. This explains why we will only check averaging with $r = .999$ in the next section. More precisely, denoting m_i the BBA associated with expert i and denoting n the number of experts, we will compute :

$$m_{average} = \frac{1}{n} \sum_{i=1 \ldots n} \text{discount}(m_i, 0.999) \qquad (22)$$

On the contrary, using Dempster's combination rule $\oplus$ without discounting can give counter-intuitive results [Zadeh, 1979]. Consider, for example, three states of the world, $\Omega = \{A, B, C\}$, and the problem of combining Bayesian beliefs corresponding to the two probability distributions p_1 and p_2, defined respectively by $p_1(A) = 0.9$, $p_1(B) = 0$, $p_1(C) = 0.1$, and $p_2(A) = 0$, $p_2(B) = 0.9$, $p_2(C) = 0.1$. The result according to Dempster's rule has a belief weight 0.85 to the state of the world C, which is paradoxical since both information sources agree that this is the least probable outcome. In the same example, if opinions are taken with a reliability factor $r = 0.8$ before combination, the weight going to state of the world C is only 0.105, which is much more intuitive. This is why we will only examine Dempster's rule with the medium reliability factor:

$$m_{dDempster} = \bigoplus_{i=1 \ldots n} \text{discount}(m_i, 0.8) \qquad (23)$$

Turning now to the non interactive disjunction $\circledcirc$, this operator tends to produce very uninformative beliefs. Adding imprecision to the input by discounting leads even faster to a vacuous result $\mathbf{1}_\Omega$. This goes against the purpose of information fusion, so we will only consider the fusion with almost no discount:

$$m_{niDisjunction} = \bigcirc\!\!\!\!\cup_{i=1 \ldots n} \text{discount}(m_i, 0.999) \qquad (24)$$

The non interactive conjunction operator $\circledcirc$ and the cautious operator $\circledwedge$ produce a trivial result when the information sources conflict completely. In this case, the fusion falls into pure contradiction $\mathbf{1}_\emptyset$. As with Dempster's rule, discounting could be used to decrease conflict before the fusion. This would technically allow to recover more informative results. But discounting is not justified for these operators, since in the transferable belief model $\mathbf{1}_\emptyset$ is accepted as a theoretically correct result. Worse, the non-interactive conjunction finds conflict when combining a Bayesian belief with itself. As seen previously, when combining the fifty-fifty probability with itself, the belief mass of $\emptyset$ is 0.5.

The introduction enumerated four challenges for mathematical aggregation methods: non-independence, complete contradiction, minority views, and discounting. Contradiction between experts rules out conjunction operators, but is not a problem for the remaining three approaches. None of these, however, completely answers the other challenges. Dempster's rule needs discounting, but there is little evidence to determine reliability factors. Contrary to the cautious conjunction, Dempster's rule and the non-interactive operators assume that experts are independent. This can lead to artificially over-precise results, by counting the same pieces of evidence more than once.

With averaging and Dempster's rule, the weight of an opinion increases with the number of experts holding it. This can be seen as a problem, as scientific arguments should be evaluated on their own merits, not by *argumentum ad populum* (Latin: "appeal to the people"). It is only at the social decision-making stage that the quality and number of people behind each view should matter. Groupthink and bandwagon effects are known dangers when pooling opinions. Thus, all other things being equal, a fusion method that gives equal attention to the minority and the majority views is preferable.

3.4 A Hierarchical Approach

The difficulties of symmetric fusion methods to aggregate conflicting beliefs have led researchers to suggest *adaptive* fusion rules [Schubert, 1995, Ayoun and Smets, 2001, Destercke et al., 2006]. The general idea is to merge conjunctively subgroups of coherent sources, before disjunctively merging the different results. We propose a hierarchical fusion procedure based on this idea. This procedure aims to be relevant when science is not yet stabilized, and the notion of 'competing theories' can be used. Sociology of science suggests that at some moments in the progress of science, in front of a big unexplained problem, scientists tend to group into schools of thought, which correspond to alternative candidate theories [Kuhn, 1962]. Within each group, experts share an explanation of the way the world works. But only time can tell which theory will emerge, and only one will be adopted in the end.

This suggests to use different operators across and within groups. Across groups, we will use a non-interactive disjunction operator, assuming that at least one theory, but not all theories, is a reliable information source. This deals with the challenge of representing equally minority views because all theories are treated equally, regardless of the number of experts in the group.

Within groups, beliefs will be combined using a cautious conjunction operator. This assumes that experts are all reliable but not independent information sources. Discounting is needed if the beliefs verify $m_i(\Omega) = 0$, but this is only a technical operation; as the reliability factor can be as close to 1 as desired, we will use $r = 0.999$. This method deals with contradiction as far as the degree of conflict remains low between experts within groups. Denoting $G_1, \ldots, G_N$ the groups of experts, we will compute:

$$m_{Hierarchical} = \underset{k=1\ldots N}{\bigcirc\!\!\!\!\cup} \; \underset{i \in G_k}{\bigcirc\!\!\!\!\wedge} \operatorname{discount}(m_i, 0.999) \tag{25}$$

In that equation, using the $\bigcirc\!\!\!\!\cup$ operator is tantamount to assuming that schools of thought are non-interactive, that is somewhat independent. This assumption could be discussed, but the disjunctive combination rule corresponding to the cautious conjunction has been published by Denœux [2008] too recently to be examined here.

At this point, we have defined four ways to combine beliefs: the simple linear opinion pool (Equation 22), the discounted Dempster's combination rule (eq. 23), the non-interactive disjunction (eq. 24), and a hierarchical disjunctive-cautious fusion based on the notion of competing theories (eq. 25). These four methods will be applied both to the elicited Bayesian beliefs (eq. 18), and to the consonant beliefs (eq. 21). This defines theoretically eight distinct ways to perform opinion fusion in the transferable belief model. The next section examines how they perform on a real-world dataset.

4 Application to Climate Sensitivity

4.1 Data used

Climate sensitivity is a proxy for the severity of the climate change problem. It is denoted $\Delta T_{2\times}$, and defined as the equilibrium global mean surface temperature change following a doubling of atmospheric CO_2 concentration, compared to preindustrial levels [Randall et al., 2007]. Over the last two decades, climate sensitivity has become one of the main communication anchors between the scientists and policymakers to quantify the seriousness of the climate change issue, as discussed by van der Sluijs [1997], Boa [2003].

The value of this parameter is not known precisely. For a long time, the [1.5°C, 4.5°C] interval has been regarded as the canonical uncertainty range for $\Delta T_{2\times}$ [National Research Council, 1979]. Yet knowing better climate sensitivity is critical for climate policy. According to current trends, humankind is well on track to double the CO_2 concentration in the Earth's atmosphere, not to mention other greenhouse gases. IPCC [2001a] estimated that 2°C of global warming raises serious concerns such as risks to many unique and threatened ecosystems (for example coral reefs or the arctic ice sheet), plus a large increase in the frequency and magnitude of extreme climate events (like heatwaves, droughts and storms).

If climate sensitivity were around 1.5°C, one could argue that doubling the CO_2 concentration would not lead immediately to a dangerous interference with the climate system. But if climate sensitivity were at the upper end of the canonical uncertainty range, 4.5°C, then doubling the CO_2 concentration would certainly be a very dangerous interference with the climate system.

Morgan and Keith [1995] conducted structured interviews using expert elicitation methods drawn from decision analysis with 16 leading U.S. climate scientists. The authors obtained quantitative, probabilistic judgments about a number of key climate variables, including the climate sensitivity parameter.

This dataset received a significant interest in the climate change literature, as in the late nineties there were very few other estimates for this parameter's probability distribution. For example Webster and Sokolov [2000, 4.1] derived a climate sensitivity probability distribution by taking the median (across the 16 experts) of each of the fractiles (0.05, 0.25, 0.5, 0.75, 0.95), and using the median fractile values to fit a beta distribution. According to this distribution, $p(\Delta T_{2\times} \leq 1.5°\text{C}) = 0.24$, $p(1.5°\text{C} \leq \Delta T_{2\times} \leq 4.5°\text{C}) = 0.67$, $p(\Delta T_{2\times} \geq 4.5°\text{C}) = 0.09$.

But Reichert and Keith [2001] raised theoretical issues against combining these opinions into a single judgment on climate sensitivity like Webster and Sokolov [2000] did. The 16 experts are not independent, they are part of a research community regularly sharing data, models and ideas. And yet opinions on climate sensitivity are widely different in qualitative terms. The authors confirmed that there is an interest in finding an aggregation technique where the combined probability distribution does not necessarily narrow as the number of experts increases, and which is more robust with respect to extreme experts judgments than previously published techniques.

4.2 Implementation

The Transferable Belief Model was implemented in *Mathematica* version 6 using matrix calculus as described by Smets [2002]. The whole notebook file used to create the published figures and tables is available as an electronic supplement to this manuscript.

In the dataset, no probability is allocated to climate sensitivity lower than −6°C, or larger than 12°C. For the sake of numerical tractability, this range was subdivided in seven ranges:

$$\begin{aligned}\Omega &= \{\omega_1, \ldots, \omega_7\} \\ &= \{[-6,0], [0,1.5], [1.5,2.5], [2.5,3.5], [3.5,4.5], [4.5,6], [6,12]\}\end{aligned}$$

Each expert's probability distribution on Ω was computed from the elicited probability density function P_i:

$$\begin{aligned}p_i(\omega_1) &= P_i\left(-6 \leq \Delta T_{2\times \mathrm{CO}_2} < 0\right) \\ p_i(\omega_2) &= P_i\left(0 \leq \Delta T_{2\times \mathrm{CO}_2} < 1.5\right) \\ &\ldots\end{aligned}$$

The procedure described in Section 3.2 (see Equation 21 and Figure 3.2) was used to transform the Bayesian beliefs into consonant beliefs. We computed an implicit possibility distribution π_i associated with each expert's probability distribution p_i. Figure 2 represents p_i and π_i for the 16 experts.

Four qualitatively different groups of distributions can be identified. The widest distributions come from experts 2, 3 and 6, they allow a positive probability both to cooling and to climate sensitivity well above 6°C. Distributions from experts 4, 7, 8, 9 do not give weight to cooling, but have an upper bound above 8°C. Experts {1,10–16} disallow extreme cases, the width of the range supporting their probability distributions is between 4.2 and 5.5°C. Expert's 5 probability distribution lies in the range $\omega_2 = [0°\text{C}, 1.5°\text{C}]$.

The 0.80 reliability factor used for Dempster's rule is arbitrary. Discounting is also necessary to compute the cautious conjunction, as all experts except {2,3,6} give a zero probability to some outcomes. We used a reliability factor 0.999. Since results will be shown only to 2 digits, that is presumably close enough to 1, an assumption that will be tested in the sensitivity analysis.

We used the four qualitative groups outlined above for the hierarchical approach: $G_1 = \{2,3,6\}$, $G_2 = \{4,7,8,9\}$, $G_3 = \{1,10,11,12,13,14,15,16\}$, $G_4 = \{5\}$. Better ways to group experts together will be discussed in section 5.3, but this heuristic is sufficient to illustrate the method.

We further assumed that within a school of thought, all experts are reliable but not inde-

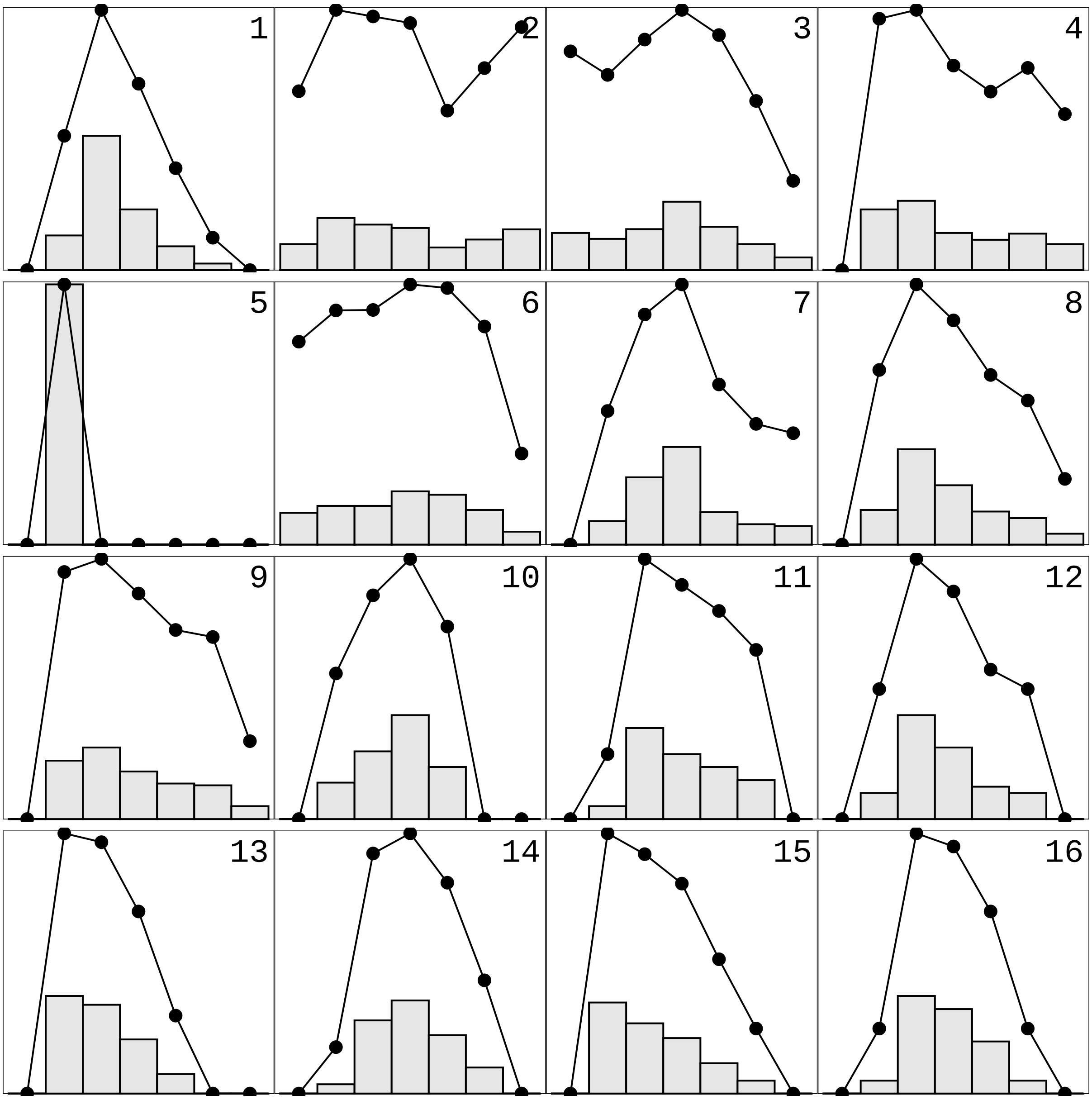

Figure 2: The probability (grey histograms) and implicit possibility (dotted lines) for the 16 experts in [Morgan and Keith, 1995]. The vertical axis goes from 0 to 1. The horizontal axis discretizes the [-6°C, 12°C] climate sensitivity range into seven intervals using a non-uniform subdivision at -6, 0, 1.5, 2.5, 3.5, 4.5, 6 and 12°C. Four qualitatively different groups of distributions can be seen: Experts 2,3,6 allow cooling, 4,7,8,9 allow high outcomes but no cooling, 1,10–16 disallow extreme cases, and 5 is concentrated on [0°C,1°C]. Data are given numerically in Table 6.

pendent information sources. Their beliefs were combined using a cautious conjunction operator: $m_{G_k} = \owedge_{i \in G_k} m_i$. The second stage combined the four groups together using the non-interactive disjunction operator.

4.3 Results

Figure 3 and Table 1 present the results, in two different ways. The figure shows the results obtained by combining Bayesian beliefs in the left column, and those obtained with consonant beliefs in the right column. Correspondingly, the table is divided in a top half showing the fusion of Bayesian beliefs, whereas the bottom half is devoted to the consonant beliefs. In each half, we compare the results obtained using the four ways to combine opinions: averaging (Equation 22), discounted Dempster's rule (Equation 23), non-interactive disjunction (Equation 24) and the hierarchical approach (Equation 25). Numbers are shown with two significant digits.

On each plot in Figure 3, the vertical axis goes from 0 to 1, and horizontally the numbers (from 1 to 7) denote the states of the world ω_1 to ω_7. The legend at the bottom defines these states of the world in terms of climate sensitivity. Finally, there are three series of points on each plot. The top one is labelled *pl*, while the middle one is labelled p and the bottom *bel*. They display, respectively, the plausibility $pl(\omega_i)$, the pignistic probability $BetP(\omega_i)$, and the belief $bel(\omega_i)$. Labels are sometimes superposed. The lines are drawn for readability, but it does not mean that we plot continuous densities.

Showing these three functions only on the ω_i does not represent completely the results, except when the result is Bayesian. Since there are 7 states of the world, a general BBA m is defined with $2^7 = 128$ numbers. As an example of what the full results look like, the basic belief assignment resulting from the hierarchical fusion of the consonant beliefs is completely tabulated in Table 5 (see Annex) with 5 decimals. It has 18 focal sets.

In Table 1, each line describes aspects of the BBA obtained using a different fusion method. Lines 1 to 4 show the combination of Bayesian beliefs, lines 5 to 8 of consonant beliefs. There are five columns. The first column shows the degree of conflict $m(\emptyset)$, while the second column shows $m(\Omega)$.

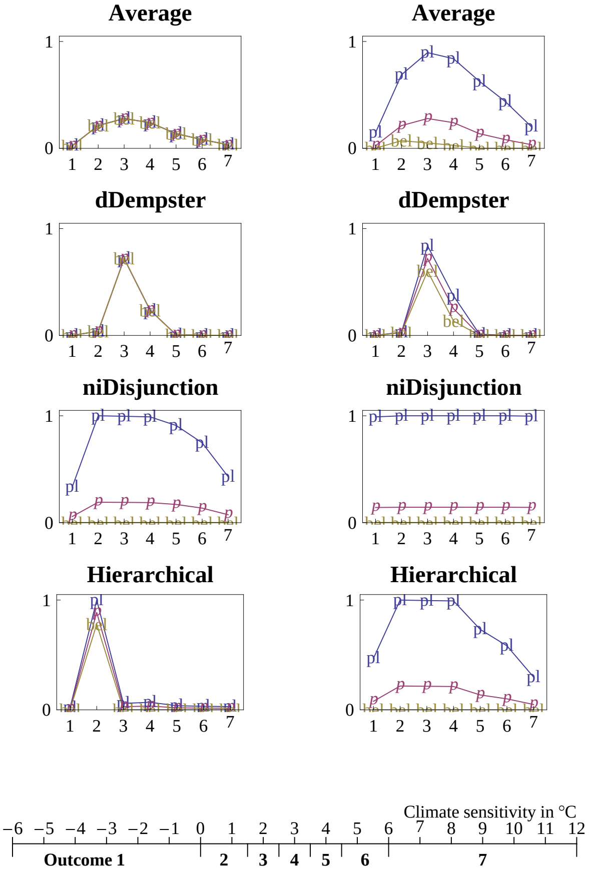

Figure 3: Results of the fusion, using different operators. Left column using Bayesian beliefs, right column using consonant beliefs.

	Conflict $m(\oslash)$	Ignorance $m(\Omega)$	Below 1.5°C *bel*–*pl*	In range *bel*–*pl*	Above 4.5°C *bel*–*pl*
$m_{average}$	0.	0.00	0.23–0.24	0.65–0.65	0.11–0.11
$m_{dDempster}$	0.	0.	0.04–0.04	0.96–0.96	0.–0.
$m_{niDisjunction}$	0.	0.08	0.–1.	0.–1.	0.–0.86
$m_{Hierarchical}$	0.	0.00	0.79–1.	0.–0.16	0.–0.06
	Conflict $m(\oslash)$	Ignorance $m(\Omega)$	Below 1.5°C *bel*–*pl*	In range *bel*–*pl*	Above 4.5°C *bel*–*pl*
$m_{average}$	0.	0.08	0.07–0.69	0.27–0.93	0.–0.45
$m_{dDempster}$	0.	0.	0.02–0.03	0.97–0.98	0.–0.
$m_{niDisjunction}$	0.	0.99	0.–1.	0.–1.	0.–1.
$m_{Hierarchical}$	0.	0.18	0.–1.	0.–1.	0.–0.61

Table 1: The fusion of expert opinion on climate sensitivity. Top table using Bayesian beliefs, bottom table using consonant beliefs.

Heuristically, smaller numbers in these columns are better, since they correspond to intuitively more interesting or informative results. The last three columns show the values of the belief and plausibility functions. They refer to a coarsened frame of reference: states of the world have been grouped into three policy relevant cases. The less worrying case is $\{\omega_1, \omega_2\}$, that is sensitivity below 1.5°C. The historical canonical range is represented by $\{\omega_3, \omega_4, \omega_5\}$. The worst case groups together outcomes for which climate sensitivity is above 4.5°C, that is $\{\omega_6, \omega_7\}$.

We now discuss each operator successively. Results obtained by averaging are shown in Figure 3 on the top row. Top left, the three curves *bel*, *p* and *pl* are superposed: when all beliefs are Bayesian, the average is also Bayesian. The top right graphic represents the average of consonant beliefs. The plausibility and belief curves are now very different. For a decision-maker only focused on pignistic probabilities (curve labelled *p*), the left and right results would seem very close. But from an evidence theory perspective, the left plot under-represents scientific controversies and the need for precautionary decision-making.

Consider for example what the results say about $\Delta T_{2\times} < 1.5$. As shown in Table 1, averaging in the Bayesian case leads to the conclusion that $bel(\{\omega_1, \omega_2\}) = 0.23$ and $pl(\{\omega_1, \omega_2\}) = 0.24$ (the small difference between belief and plausibility levels is explained by the reliability factor 0.999 we introduced.). Yet the bottom half of the table shows that averaging in the consonant case leads to $bel(\{\omega_1, \omega_2\}) = 0.07$ and $pl(\{\omega_1, \omega_2\}) = 0.7$.

These results are qualitatively different. The former could be stated as 'There is a low confidence that $\Delta T_{2\times} < 1.5$', to mean a probability around 0.2. But the latter result could be stated as '$\Delta T_{2\times} < 1.5$°C has a low degree of belief but a significant level of plausibility'. Such a more imprecise statement describes more accurately the state of scientific controversies.

The graphics in the second row reveal that Dempster's rule with discounting produces very focused beliefs around ω_3. This rule considerably reduces the plausibility of states of the world that are outside the canonical range. The issues with discounting, shared evidence and the bandwagon effect discussed in section 3.3 suggest that much of this precision is unwarranted.

Results with the non-interactive disjunction are shown on the third row. With consonant beliefs, the result is almost completely uninformative: line 7, column 2 in Table 1 shows indeed that $m(\Omega) = 0.99$. With Bayesian beliefs, the figure shows that the results are well shaped. In that case, while the levels of belief are close to zero, the levels of plausibility do have lower values for the extreme cases. This suggests empirically that the non-interactive disjunction rule produces degenerate results when used to combine consonant beliefs, but works better when combining Bayesian beliefs.

Lastly, let us consider the results of the hierarchical fusion. In the case of Bayesian beliefs, almost all the weight goes to ω_2, and the plausibility of

$\{\omega_3, \omega_4, \omega_5\}$ is only 0.16. This can be explained by looking at the cautious conjunction within groups (Table 7 in the annex). At this stage, the degree of conflict is high, respectively 0.86, 0.86 and 1 within G_1, G_2 and G_3.

Thus it appears that the hierarchical fusion method is useless, or at least highly unstable, when applied to subjective probabilities that are represented by Bayesian BBA's. Bayesian BBA's tend to be conflicting, and their conjunction leads to a large mass on the empty set. Thus, groups of multiple experts tend to eliminate themselves. This is the opposite issue of averaging, where the majority got a larger weight than minority opinions.

That contradiction problem does not arise when combining consonant beliefs: the degree of conflict within groups is only 0.01, 0.03 and 0.14. The non-interactive disjunction rule across groups gives a more balanced image of the opinion pool.

Figure 3 shows that some combinations of operators and input data produce degenerate results, while other give more interesting BBAs. Designing a study to formally elicit and aggregate information using expert opinion involves several heuristic choices: which experts to interview, how to elicit their opinions and how to fusion them. To avoid degenerate results, one has to balance these choices.

Conjunction operators move the belief masses 'down', into smaller focal sets, but too much precision is a problem given conflicting opinions. Disjunction operators move the belief masses 'up', to larger focal sets, but too much generality produces useless results. To avoid extreme results, a proper heuristic may be to apply a precision increasing operator (conjunction) to imprecise data (consonant beliefs), or conversely an imprecision-increasing operator (disjunction) to precise data (Bayesian beliefs). Averaging is neutral here.

This may contribute to explain the qualitative results in Figure 3. For example, when opinions are represented with Bayesian beliefs, it is interesting to combine them with the non-interactive disjunction rule: this operator produces beliefs that are more imprecise than its inputs. Conversely, this operator gives the trivial vacuous beliefs when opinions are represented using consonant beliefs, because it adds imprecision to a pool of already imprecise data.

5 Discussion

5.1 Sensitivity Analysis

Tables 2 and 3 show the result of the fusion under alternative operators (also represented in the annex, figures 5 and 6). First, we examine the sensitivity of the discounted Dempster's rule to the reliability factor. Decreasing the reliability factor means adding doubt to the beliefs to be combined. This spreads around the belief weights, so the result becomes less focused compared to the previous case with $r = 0.8$. The magnitude of change in the results can be significant. Consider for example the 'below 1.5°C' case when combining Bayesian beliefs. Between $r = 0.8$ (Table 1, line 2, column 3) and $r = 0.5$ (Table 2, line 1, column 3), its probability increases by a factor 4.

The other four lines in Table 2 illustrate the problem of contradiction. In the non-interactive conjunction and in the cautious conjunction of all experts, the degree of conflict $m(\emptyset)$ is very high. We check that $\owedge$ is less sensitive to conflict than $\circledcap$, and that adding doubt, either by discounting or by transforming Bayesian into consonant beliefs, decreases conflict. This only confirms the theoretical reasons why in section 3.3 we disqualified these operators.

Table 3 presents variants of the hierarchical approach. Using a reliability factor $r = 0.99$ does not change the results much compared to the case with $r = 0.999$. Adding more doubt to the input data mechanically increases $m(\Omega)$ in the output, which in turn increases plausibility levels. Moving to $r = 0.9999$, the results do not change visibly, as the display is rounded to 2 digits.

Merging G_1 and G_2 together allows to check that results are significantly sensitive to the clustering of experts: the plausibility of the 'above 4.5°C' case, with consonant beliefs, drops from 0.61 to 0.15 (Table 1, line 7, column 5). Finally, we examined a hierarchic fusion where the first step is averaging, rather than the cautious conjunction. The plausibility function levels are generally greater with averaging, as the extreme cases get more plausible.

	Conflict $m(\oslash)$	Ignorance $m(\Omega)$	Below 1.5°C *bel–pl*	In range *bel–pl*	Above 4.5°C *bel–pl*
Dempster r=0.5	0.	0.01	0.16–0.17	0.8–0.81	0.03–0.04
cautious	1.	0.	0.–0.	0.–0.	0.–0.
cautious r=0.8	0.96	0.00	0.01–0.02	0.01–0.02	0.00–0.01
niConj.	1.	0.	0.–0.	0.–0.	0.–0.
niConj. r=0.8	1.	0.	0.–0.	0.–0.	0.–0.
	Conflict $m(\oslash)$	Ignorance $m(\Omega)$	Below 1.5°C *bel–pl*	In range *bel–pl*	Above 4.5°C *bel–pl*
Dempster r=0.5	0.	0.	0.06–0.12	0.88–0.94	0.–0.01
cautious	0.95	0.	0.05–0.05	0.–0.	0.–0.
cautious r=0.8	0.71	0.	0.1–0.13	0.16–0.19	0.–0.01
niConj.	1.	0.	0.–0.	0.–0.	0.–0.
niConj. r=0.8	0.87	0.	0.00–0.00	0.13–0.13	0.–0.

Table 2: Sensitivity analysis: the fusion of expert opinions using alternative symmetric operators. Top half, using Bayesian beliefs; bottom half using the corresponding consonant beliefs.

	Conflict $m(\oslash)$	Ignorance $m(\Omega)$	Below 1.5°C *bel–pl*	In range *bel–pl*	Above 4.5°C *bel–pl*
Hierarchical r=.99	0.	0.01	0.74–1.	0.–0.2	0.–0.08
Hierarchical r=.9999	0.	0.00	0.79–1.	0.–0.16	0.–0.06
Hierarchical 3-way	0.	0.00	1.–1.	0.–0.00	0.–0.00
average within	0.	0.00	0.01–1.	0.–0.96	0.–0.39
	Conflict $m(\oslash)$	Ignorance $m(\Omega)$	Below 1.5°C *bel–pl*	In range *bel–pl*	Above 4.5°C *bel–pl*
Hierarchical r=.99	0.	0.2	0.–1.	0.–1.	0.–0.62
Hierarchical r=.9999	0.	0.18	0.–1.	0.–1.	0.–0.61
Hierarchical 3-way	0.	0.00	0.01–1.	0.–0.99	0.–0.15
average within	0.	0.58	0.–1.	0.–1.	0.–0.95

Table 3: Sensitivity analysis: the fusion of expert opinions using alternative hierarchic procedures. Top half, using Bayesian beliefs; bottom half using the corresponding consonant beliefs.

5.2 Existing Results on Climate Sensitivity

In its third assessment published in 2001, the Intergovernmental Panel on Climate Change [IPCC, 2001b, Technical Summary F.3] stated that "climate sensitivity is likely to be in the range of 1.5 to 4.5°C. This estimate is unchanged from the first IPCC Assessment Report in 1990". This estimate can be traced back even earlier [National Research Council, 1979]. The [1.5, 4.5°C] was not offered as a 90% confidence interval, but as a "likely" range. The word "likely" had a formally defined meaning, it was used to indicate a judgmental estimate of confidence of 66 to 90% chance. Since this report, several studies have estimated probability density functions for climate sensitivity based on models and observations.

Hall et al. [2007] combined a set of 7 such distributions in an imprecise probability framework. The result, given as upper probability bounds, suggests that $p(\Delta T_{2\times} \leq 1.5) \leq 0.10$ and $p(\Delta T_{2\times} \geq 4.5) \leq 0.60$ (determined graphically from Figure 5 in Hall et al. [2007]).

Kriegler [2005, section 3.2.3] conducted a deeper analysis of the combination of these distributions with imprecise probabilities. Four of the six estimates examined show a 90% confidence interval in the range between 1.3 and 6.3°C. In the other two studies, these ranges are [1.4, 7.7] and [2.2, 9.3]. The author then estimated a prior imprecise distribution based on the literature, and then updated it using a climate model and observational data for 1870–2002. Updating was done using both Dempster's rule and the Generalized Bayes Rule, but only Dempster's rule produced meaningful results. Table 4 summarizes them. For example, the posterior results suggest that the probability of climate sensitivity being less than 1.5°C is very small (0.00 meaning less than 1 per thousand). In the posterior, the probability that climate sensitivity falls in the [1.5, 4.5] range is between 0.53 and 0.99.

According to these results, there is a large possibility that climate sensitivity lies above 4.5°C. The relatively high upper bound (10°C) has been contested by Hegerl et al. [2006, Figure 3], who recently estimated that the 5–95 per cent confidence range of climate sensitivity was about 1.5–6.2°C. Still, this does not refute the idea that the 90% confidence interval has its upper bound above 4.5°C.

Hegerl et al. [2007, 718–727] offers a comprehensive assessment of the literature on climate sensitivity. In this more recently published Fourth Assessment Report, IPCC continues to formulate uncertainty statement literally, with an explicit correspondence on a probability scale [IPCC, 2005]. The conclusion is that, in spite of new research, the result is not changed much since the previous report: the likely range is [2, 4.5], where "likely" means a probability between 66 and 90 percent. IPCC also states that it is "very unlikely" that climate sensitivity lies below 1.5, meaning a less than 10% probability.

Andronova et al. [2007, Figure 1.1a] also published an historical perspective on climate sensitivity. They conclude that recent studies based on observations indicate that there is more than a 50% likelihood that $\Delta T_{2\times}$ lies outside the canonical range of 1.5°C to 4.5°C, with disquietingly large values not being precluded. They combined the 16 experts opinions in terms of their mean estimation and variance into a single cumulative density function, under the assumption that each of the 16 estimations is normally distributed, but this was mostly for historical comparison.

Results presented Table 1 can be compared to this more recent literature. Only the non-degenerate cases in lines 1, 3, 5 and 8 need to be considered. The plausibility that climate sensitivity lies below 1.5 appears to be low in the recent literature. But it is high in our results (respectively 1, 0.7 and 1 in lines 3, 5 and 8). In the linear pooling case line 1, the probability is 0.23 which can also be seen as rather significant. The fusion results are not in line with the more recent literature here.

Given that the dataset included one opinion certain that $\Delta T_{2\times} \leq 1.5$°C, this discrepancy can hardly be seen as a mathematical artefact. A more intuitive explanation is that the scientific consensus has evolved since 1995, to revise downward the likelihood of that event. The increase in the IPCC lower bound from 1.5°C to 2°C can be taken as a sign of this change.

Consider now the last column in Table 1, related to the case in which climate sensitivity lies above 4.5°C. The recent literature finds that this case is rather plausible. Lines 3, 5 and 8, this event's plausibility is respectively 0.88, 0.45 and 0.62. Line 1,

	$T_{2\times} \in [0°C, 1.5°C]$	$[1.5°C, 4.5°C]$	$[4.5°C, 10°C]$
†Prior	[0, 0.07]	[0.31, 0.98]	[0.02, 0.62]
†Posterior	[0, 0.00]	[0.53, 0.99]	[0.01, 0.47]
‡Prior	[0, 0.08]	[0.12, 1.0]	[0, 0.80]

Table 4: Probability bounds on climate sensitivity $\Delta T_{2\times}$. Source: Top two rows (†) from Kriegler [2005, table 4.2] and Bottom row (‡) from Kriegler and Held [2005]. The prior summarizes the literature, the posterior is updated by Dempster's rule, using the results of a simulation model.

the probability is 0.11. Thus, the fusion results are in better agreement with the more recent findings here. If the visibility given to the higher than 4.5°C case has increased in the recent publications, its subjective weight was already present in experts' minds back in 1995.

5.3 Remarks on the hierarchical approach

Clemen and Winkler [1999] dichotomize ways to summarize the opinion of a variety of experts in two classes: behavioral approaches and mathematical methods. In behavioral approaches, also called interactive expert aggregation methods, experts exchange information with each other. In mathematical approaches, each expert is interviewed separately in a first phase, and then opinions are combined afterward according to some algorithmic aggregation method.

Behavioral approaches have many interesting advantages over algorithmic methods. The group judgment is more legitimate since it comes from the experts themselves and collective deliberation is a natural social process. The way scientific panels such as the Intergovernmental Panel on Climate Change (IPCC) write their reports is an interactive expert aggregation method. However, behavioral approaches also have drawbacks. Any group of experts is subject to the social dynamics inherent in any group of humans. There are known biases towards conservatism and overconfidence in groupthinking. More importantly, managing all the interactions between the experts is complicated, time consuming and thus costly.

Mathematical methods aim at simplifying and rationalizing the procedure by separating in time the expert opinion elicitation step from the aggregation step, and performing the later without the experts. The simplest aggregation method we have seen is linear pooling, that is averaging. It works with Bayesian as well as with consonant beliefs. As an alternative, we have seen that the non-interactive disjunction produces meaningful and non-trivial results, when beliefs are Bayesian. Finally, we examined a hierarchical approach. Using consonant beliefs, it gave results comparable to those obtained with the non-interactive disjunction.

The main limitations of our work are the following. Firstly, when beliefs are Bayesian, the hierarchical fusion works poorly. This is because, in that case, the degree of conflict within groups is too high. Secondly, we used a probability-possibility transformation, which is an abductive reasoning, an inference to the best explanation. But had we used a possibilistic dataset from the start, we would also have had to use a possibility-probability transformation in order to compare fusion methods in the Bayesian and the non-Bayesian cases. Taken together, these two limitations suggest that hierarchical fusion as presented here is more appropriate when beliefs are elicited as possibility distributions.

Thirdly, the hierarchical approach is based on a partition of experts into a small number of schools of thought. Contrary to symmetric fusion operators, it requires to structure the pool of experts. Thus, it requires to put back some sociology aspects in a mathematical aggregation framework. While in this paper we determined the groups from the elicited probability distribution, social sciences offer much better procedures:

- The network of experts can be analyzed through publications. Experts who have published together have seen the same data, they are more likely to share evidence. Newman [2004] shows that bibliometrics can help determine the patterns of scientific collaborations.
- Expert elicitation techniques involve semi-

structured interviews. That material is prime experimental data for social scientists. Working from transcripts is a classical method to analyze how a group of people is organized. Such analysis is usually conducted without mathematical tools. There are more formal content analysis methods, often based on the written rather than oral production of the subjects.

- The experts themselves know their community. They can help to discover how it is organized, and they can validate the results of the sociological analysis.

Note that the expert selection step, in an elicitation exercise, has to make sure that no major point of view is omitted. This shows that sociological considerations on the population of experts cannot be avoided, even in a mathematically oriented study. When it is clear from the start what the different schools of thought are, one can select a single expert to represent each position, and then pool the opinions symmetrically. Otherwise, it is only after analyzing the interviews transcripts that the population of experts can be organized around a small number of archetypes.

Representing the diversity of viewpoints by a small numbers of schools of though is admittedly a strong simplification of complex social reality. But it is less simplistic than treating all experts symmetrically. Finding out the detailed structure of epistemic communities, and explaining the differences between theories can be very informative in itself. Bringing forward that qualitative analysis is a valuable advantage of the hierarchical approach.

6 Conclusion

This paper compared several procedures to aggregate expert opinion in the Transferable Belief Model. We considered both Bayesian beliefs and consonant beliefs. The former correspond naturally with probabilities, the latter with possibilities. Regarding the procedures that combine opinions symmetrically, results show that:

- Taking either the non-interactive conjunction or the cautious conjunction of all opinions produces degenerate results, indicating only that experts contradict each other.

- Dempster's rule of combination, even after discounting, led to excessively narrow results (overconfidence).

- Averaging always produces non-degenerate results, but there are two problems with that method. First, when beliefs are Bayesian, the result is Bayesian too. In the Dempster-Shafer theory of evidence, Bayesian beliefs under-represent scientific controversies. Second, averaging is essentially a way to allocate more weight to views held by a larger number of experts. This is a problem because scientific theories should be assessed only on their own merit.

- The non-interactive disjunction rule produces a degenerate (uninformative) result when beliefs are consonant. The intuition is that consonant beliefs are vague to start with, and the result of the disjunction is more imprecise than its inputs. The non-interactive disjunction of Bayesian beliefs represents more appropriately scientific controversies than their average.

Then a hierarchical fusion procedure was assessed. This procedure is built around a simple model of experts' social relations: it divides them into schools of thought. Social science methods are available to determine the fine structure of epistemic communities, and knowing this structure may be as interesting as knowing an aggregate opinion. Within each school, beliefs are aggregated using the cautious conjunction operator. Across the groups, beliefs are combined using the non-interactive disjunction rule. Hierarchical fusion in the Transferable Belief Model offers a solution to several theoretical problems regarding opinion aggregation:

- It allows to represent the issue of precautionary decision making due to scientific controversies in ways that purely probabilistic methods are not able to, beyond standard expected utility maximization.

- Disjunction allows coping with complete contradiction among opinions without falling into degenerate results or paradoxes. When several scientific theories compete to explain the same observations, it should not be assumed that both are true at the same time (conjunction), but that at least one will remain (disjunction).

- Within groups, cautious conjunction does pool together distinct streams of evidence to make beliefs firmer. But it is not assumed that opinions are independent: this would overestimate the precision of actual information.

- Pooling opinions across schools of thoughts, rather than across individual experts, is arguably a more balanced procedure. Contrary to averaging, where the number of experts holding a view is essential, minority views are equally taken into account in hierarchical fusion.

- This hierarchical fusion procedure uses only a technical approach to discounting. It applies the same very high reliability factor to all experts. This avoids the two issues of discounting: adding lots of doubt to experts opinions, or saying that some experts are less qualified than others.

This study was conducted using a real-world dataset on climate sensitivity, published in 1995. The fusion of expert opinion was compared to the more recent stochastic results on climate sensitivity, some of them based on model simulations. That comparison suggest that since 1995, the plausibility that climate sensitivity will remain below 1.5°C has decreased. The plausibility that climate sensitivity is above 4.5°C was significant in the community's opinion in 1995. It remains so today.

References

Natalia Andronova, Michael E. Schlesinger, Suraje Dessai, Mike Hulme, and Bin Li. The concept of climate sensitivity: history and development. In *Human-Induced Climate Change. An inter-disciplinary assessment*, chapter 1. Cambridge University Press, 2007. ISBN 978-0521-86603-3. URL `http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=9780521866033`.

André Ayoun and Philippe Smets. Data association in multi-target detection using the transferable belief model. *International Journal of Intelligent Systems*, 16(10):1167–1182, 2001. doi: 10.1002/int.1054. URL `http://citeseer.ist.psu.edu/ayoun00data.html`.

Estimating Climate Sensitivity: Report of a Workshop, 2003. Board on Atmospheric Sciences and Climate, Division of Earth and Life Studies, National Research Council of the National Academies, The National Academies Press. ISBN 0-309-09056-3 (Book). URL `http://www.nap.edu/catalog.php?record_id=10787#toc`.

Robert T. Clemen and Robert L. Winkler. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2):187–203, 1999. doi: 10.1023/A:1006917509560. URL `http://faculty.fuqua.duke.edu/~clemen/bio`.

Barry R. Cobb and Prakash P. Shenoy. On the plausibility transformation method for translating belief function models to probability models. *International Journal of Approximate Reasoning*, 41(3):314–330, 2006. doi: 10.1016/j.ijar.2005.06.008.

R. M. Cooke and L. J. H. Goossens. Procedures guide for structured experts judgment. Technical Report EUR 18820 EN, European Commission, Nuclear Science and Technology (EURATOM), 1999. URL `ftp://ftp.cordis.europa.eu/pub/fp5-euratom/docs/eur18820_en.pdf`. Project report produced by the Delft University of Technology.

A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38(2):325–339, 1967. URL `http://projecteuclid.org/euclid.aoms/1177698950`.

Thierry Denœux. The cautious rule of combination for belief functions and some extensions. In *Proceedings of FUSION'2006*, Florence, Italy, July 2006. URL `http://www.hds.utc.fr/~tdenoeux/papers.htm#conf`.

Thierry Denœux. Conjunctive and disjunctive combination of belief functions induced by non distinct bodies of evidence. *Artificial Intelligence*, 172(2–3):234–264, 2008. URL `http://www.hds.utc.fr/~tdenoeux/perso`.

Sebastien Destercke, Didier Dubois, and Eric Chojnacki. Fusion d'opinions d'experts et théories de l'incertain. In *LFA'06. Rencontres Francophones sur la Logique Floue et Applications*, October 19–20 2006.

Didier Dubois, Henri Prade, and Philippe Smets. A definition of subjective possibility. *International Journal of Approximate Reasoning*, In Press, Corrected Proof, 2008. doi: 10.1016/j.ijar.2007.01.005.

Jim W. Hall, Guangtao Fu, and Jonathan Lawry. Imprecise probabilities of climate change: aggregation of fuzzy scenarios and model uncertainties. *Climatic Change*, 81(3–4):265–281, 2007. doi: 10.1007/s10584-006-9175-6.

Gabriele C. Hegerl, Thomas J. Crowley, William T. Hyde, and David J. Frame. Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature*, 440:1029–1032, 20 April 2006. doi: 10.1038/nature04679.

Gabriele C. Hegerl, Francis W. Zwiers, Pascale Braconnot, Nathan P. Gillett, Yong Luo, Jose A. Marengo Orsini, Neville Nicholls, Joyce E. Penner, and Peter A. Stott. Understanding and attributing climate change. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]*, chapter 9. Cambridge University Press, 2007. URL `http://ipcc-wg1.ucar.edu/wg1/wg1-report.html`.

IPCC. *Climate Change 2001: Impacts, Adaptation, and Vulnerability*. Published for the Intergovernmental Panel on Climate Change by GRID Arendal, 2001a. ISBN 0521-01500-6. URL `http://www.ipcc.ch/pub/tar/wg2/index.htm`. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change.

IPCC. *Climate Change 2001: the scientific basis*. Cambridge University Press, 2001b. ISBN 0521-01495-6. URL `http://www.ipcc.ch/`. Contribution of the Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change.

IPCC. Guidance notes for lead authors of the IPCC fourth assessment report on addressing uncertainties, July 2005. URL `http://ipcc-wg1.ucar.edu/wg1/Report/AR4_UncertaintyGuidanceNote.pdf`. Accessed online 2007-07-04.

David W. Keith. When is it appropriate to combine expert judgments? *Climatic Change*, 33:139–144, 1996. doi: 10.1007/BF00140244.

Elmar Kriegler. *Imprecise probability analysis for integrated assessment of climate change*. Doctoral thesis, University of Potsdam (Germany), January 2005. URL `http://pub.ub.uni-potsdam.de/volltexte/2005/561/`.

Elmar Kriegler and Hermann Held. Utilizing belief functions for the estimation of future climate change. *International Journal of Approximate Reasoning*, 39(2–3):185–209, June 2005.

Thomas Samuel Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, 3rd, november 1996 edition, 1962. ISBN 0-226-45808-3.

M. Granger Morgan and David W. Keith. Subjective judgments by climate experts. *Environmental Science and Technology*, 29(10):468A–476A, October 1995. URL `http://www.andrew.cmu.edu/user/dk3p/`.

National Research Council. *Carbon Dioxide and Climate: A Scientific Assessment*. National Academy Press, 1979. Ad Hoc Study Group on Carbon Dioxide and Climate. Also known as 'The Charney committee report'.

M. E. J. Newman. Coauthorship networks and patterns of scientific collaboration. *PNAS, Proceedings of the National Academy of Sciences of the United States of America*, 101 (1):5200–5205, 2004. doi: 10.1073/pnas.0307545100. URL `http://www.pnas.org/cgi/content/full/101/suppl_1/5200`.

Fumika Ouchi. A literature review on the use of expert opinion in probabilistic risk analysis. Technical Report 3201, World Bank, February 2004. URL `http://econ.worldbank.org/files/32873_wps3201.pdf`.

David A. Randall, Richard A. Wood, Sandrine Bony, Robert Colman, Thierry Fichefet, John Fyfe, Vladimir Kattsov, Andrew Pitman, Jagadish Shukla, Jayaraman Srinivasan, Ronald J. Stouffer, Akimasa Sumi, and Karl E. Taylor. Climate models and their evaluation. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller (eds.)]*, chapter 8. Cambridge University Press, 2007. URL `http://ipcc-wg1.ucar.edu/wg1/wg1-report.html`.

Peter Reichert and David W. Keith. Bayesian combination of expert judgments: reinterpretation of the meaning and dependence of elicited distributions, 2001.

Sandra A. Sandri, Didier Dubois, and H. W. Kalfsbeek. Elicitation, assessment, and pooling of expert judgments using possibility theory. *IEEE Transactions on Fuzzy Systems*, 3(3):313–335, 1995. doi: 10.1109/91.413236. Corrections were published in the next issue 3(4) p. 479.

Johan Schubert. Cluster-based specification techniques in dempster-shafer theory. In *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 395–404, 1995. URL `http://citeseer.ist.psu.edu/schubert94clusterbased.html`.

Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton (NJ), 1976. ISBN 0-691-10042-X (hardback), 0-608-02508-9 (reprint).

Glenn Shafer. Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, 4:323–362, 1990. doi: 10.1016/0888-613X(90)90012-Q. URL `http://www.glennshafer.com/assets/downloads/articles/article40.pdf`.

Philippe Smets. The nature of the unnormalized beliefs encountered in the Transferable Belief Model. In D. Dubois, M. P. Wellman, B. D'Ambrosio, and Philippe Smets, editors, *Uncertainty in Artificial Intelligence 92*, pages 292–297. Morgan Kaufman, San Mateo, Ca, 1992.

Philippe Smets. The canonical decomposition of a weighted belief. In *International Joint Conference on AI*, pages 1896–1901, 1995.

Philippe Smets. Quantified epistemic possibility theory seen as an hyper cautious Transferable Belief Model. In *Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2000), La Rochelle, France*, pages 343–353, Toulouse, France, 2000. Cepadues-Editions. URL `http://www.iridia.ulb.ac.be/~psmets/`.

Philippe Smets. Matrix calculus for belief functions. *International Journal of Approximate Reasoning*, 31(1):1–30, October 2002. doi: 10.1016/S0888-613X(02)00066-X. URL `http://iridia.ulb.ac.be/~psmets/MatrixRepresentation.pdf`.

Philippe Smets. Decision making in the tbm: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38(2):133–147, May 2005. doi: 10.1016/j.ijar.2004.05.003. URL `http://iridia.ulb.ac.be/~smets`.

Philippe Smets. Analyzing the combination of conflicting belief functions. *Information fusion*, 8(4):387–412, October 2007. doi: 10.1016/j.inffus.2006.04.003. URL `http://iridia.ulb.ac.be/~psmets`.

Matthias C. M. Troffaes. Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45(1):17–29, May 2007. doi: 10.1016/j.ijar.2006.06.001.

Jeroen P. van der Sluijs. *Anchoring amid uncertainty. On the management of uncertainties in risk assessment of anthropogenic climate change.* PhD thesis, Universieit Utrecht, 21 June 1997. URL `http://www.chem.uu.nl/nws/www/publica/PhDThesisJeroenvanderSluijs1997.pdf`.

Thomas S. Wallsten, David V. Budescu, Amon Rapoport, Rami Swick, and Barbara Forsyth. Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, 115(4):348–365, 1986. doi: 10.1037//0096-3445.115.4.348.

Mort D. Webster and A. P. Sokolov. A methodology for quantifying uncertainty in climate projections. *Climatic Change*, 46(4):417–446, 2000. doi: 10.1023/A:1005685317358. URL `http://web.mit.edu/globalchange/www/rpt37.html`.

Charles Weiss. Expressing scientific uncertainty. *Law, Probability and Risk*, 2:25–46, 2003. doi: 10.1093/lpr/2.1.25.

A. Ben Yaghlane, Thierry Denœux, and K. Melloui. Elicitation of expert opinions for constructing belief functions. In *Proceedings of IPMU'2006, Vol. I*, pages 403–411, Paris, France, July 2006.

L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1):3–38, 1978. doi: 10.1016/S0165-0114(99)80004-9. Reprinted in Fuzzy Sets and Systems 100:9-34, 1999.

L. A. Zadeh. On the validity of Dempster's rule of combination of evidence. Technical Report UCB/ERL Memo M79/24, 1979.

7 Annex: Additional tables and figures

S	$m(S)$
{2}	0.00014
{2, 3}	0.00765
{2, 4}	0.00340
{2, 3, 4}	0.16386
{1, 2, 3, 4}	0.00663
{2, 4, 5}	0.00114
{2, 3, 4, 5}	0.13432
{1, 2, 3, 4, 5}	0.07212
{2, 3, 4, 6}	0.02748
{1, 2, 3, 4, 6}	0.01334
{2, 3, 4, 5, 6}	0.08973
{1, 2, 3, 4, 5, 6}	0.18316
{2, 3, 4, 7}	0.02128
{2, 3, 4, 5, 7}	0.00624
{2, 3, 4, 6, 7}	0.01364
{1, 2, 3, 4, 6, 7}	0.01063
{2, 3, 4, 5, 6, 7}	0.06289
{1, 2, 3, 4, 5, 6, 7}	0.18234

Table 5: The basic belief assignment resulting of the hierarchical fusion. Cautious conjunction within groups, non interactive disjunction across, consonant beliefs.

Climate sensitivity (°C)	-6–0	0–1.5	1.5–2.5	2.5–3.5	3.5–4.5	4.5–6	6–12
Expert 1	0.	0.1333	0.5167	0.2333	0.0917	0.025	0.
Expert 2	0.1	0.2	0.175	0.1625	0.0875	0.1179	0.1571
Expert 3	0.1429	0.1203	0.1579	0.2632	0.1667	0.1002	0.049
Expert 4	0.	0.2333	0.2667	0.1429	0.1171	0.14	0.1
Expert 5	0.	1.	0.	0.	0.	0.	0.
Expert 6	0.1217	0.1488	0.1494	0.205	0.1917	0.1333	0.05
Expert 7	0.	0.0909	0.2591	0.375	0.125	0.0786	0.0714
Expert 8	0.	0.1333	0.3667	0.2286	0.127	0.1023	0.0421
Expert 9	0.	0.225	0.275	0.1833	0.1367	0.13	0.05
Expert 10	0.	0.14	0.26	0.4	0.2	0.	0.
Expert 11	0.	0.05	0.35	0.25	0.2	0.15	0.
Expert 12	0.	0.1	0.4	0.275	0.125	0.1	0.
Expert 13	0.	0.375	0.3417	0.2083	0.075	0.	0.
Expert 14	0.	0.0357	0.281	0.3583	0.225	0.1	0.
Expert 15	0.	0.35	0.27	0.2133	0.1167	0.05	0.
Expert 16	0.	0.05	0.375	0.325	0.2	0.05	0.
Climate sensitivity (°C)	-6–0	0–1.5	1.5–2.5	2.5–3.5	3.5–4.5	4.5–6	6–12
Expert 1	0.	0.5167	1.	0.7167	0.3917	0.125	0.
Expert 2	0.6875	1.	0.975	0.95	0.6125	0.7768	0.9339
Expert 3	0.8409	0.7506	0.886	1.	0.9035	0.6499	0.3427
Expert 4	0.	0.9667	1.	0.7857	0.6857	0.7771	0.6
Expert 5	0.	1.	0.	0.	0.	0.	0.
Expert 6	0.7804	0.9005	0.9022	1.	0.9867	0.8384	0.35
Expert 7	0.	0.5136	0.8841	1.	0.6159	0.4643	0.4286
Expert 8	0.	0.6714	1.	0.8619	0.6524	0.5538	0.2526
Expert 9	0.	0.95	1.	0.8667	0.7267	0.7	0.3
Expert 10	0.	0.56	0.86	1.	0.74	0.	0.
Expert 11	0.	0.25	1.	0.9	0.8	0.65	0.
Expert 12	0.	0.5	1.	0.875	0.575	0.5	0.
Expert 13	0.	1.	0.9667	0.7	0.3	0.	0.
Expert 14	0.	0.1786	0.9226	1.	0.8107	0.4357	0.
Expert 15	0.	1.	0.92	0.8067	0.5167	0.25	0.
Expert 16	0.	0.25	1.	0.95	0.7	0.25	0.

Table 6: Top half, the elicited probability distributions, corresponding to histograms in Figure 2. Bottom half, possibility distributions derived from these, represented as dotted lines in Figure 2

	Conflict $m(\emptyset)$	Ignorance $m(\Omega)$	Below 1.5°C *bel–pl*	In range *bel–pl*	Above 4.5°C *bel–pl*
2, 3, 6	0.86	0.	0.04–0.04	0.07–0.07	0.03–0.03
4, 7, 8, 9	0.86	0.	0.02–0.02	0.09–0.09	0.02–0.02
1,10, . . . , 16	1.	0.	0.–0.	0.–0.	0.–0.
5	0.	0.00	1.–1.	0.–0.00	0.–0.00
	Conflict $m(\emptyset)$	Ignorance $m(\Omega)$	Below 1.5°C *bel–pl*	In range *bel–pl*	Above 4.5°C *bel–pl*
2, 3, 6	0.01	0.14	0.02–0.75	0.24–0.97	0.–0.48
4, 7, 8, 9	0.03	0.	0.–0.44	0.52–0.97	0.–0.26
1,10, . . . , 16	0.14	0.	0.00–0.06	0.81–0.86	0.–0.
5	0.	0.00	1.–1.	0.–0.00	0.–0.00

Table 7: The cautious conjunction within groups of expert opinion on climate sensitivity. Top table using Bayesian beliefs, bottom table using consonant beliefs. See also Figure 4.

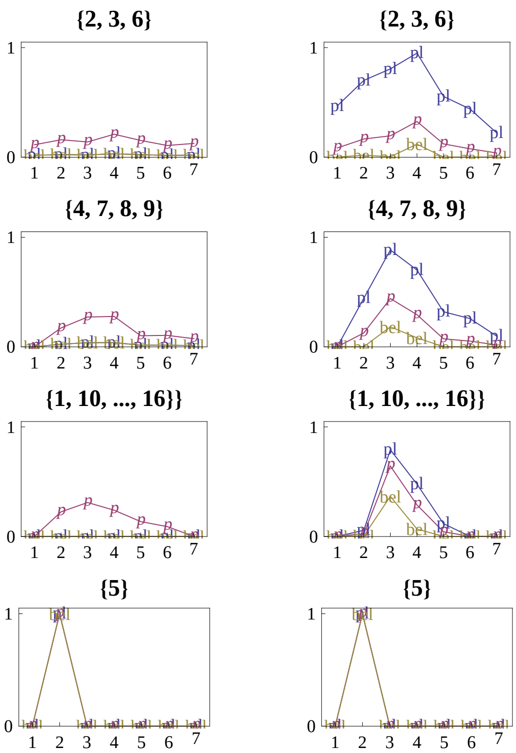

Figure 4: Results of the hierarchical fusion's first stage, the cautious conjunction within each group. Left column using Bayesian beliefs, right column using consonant beliefs. See also Table 7.

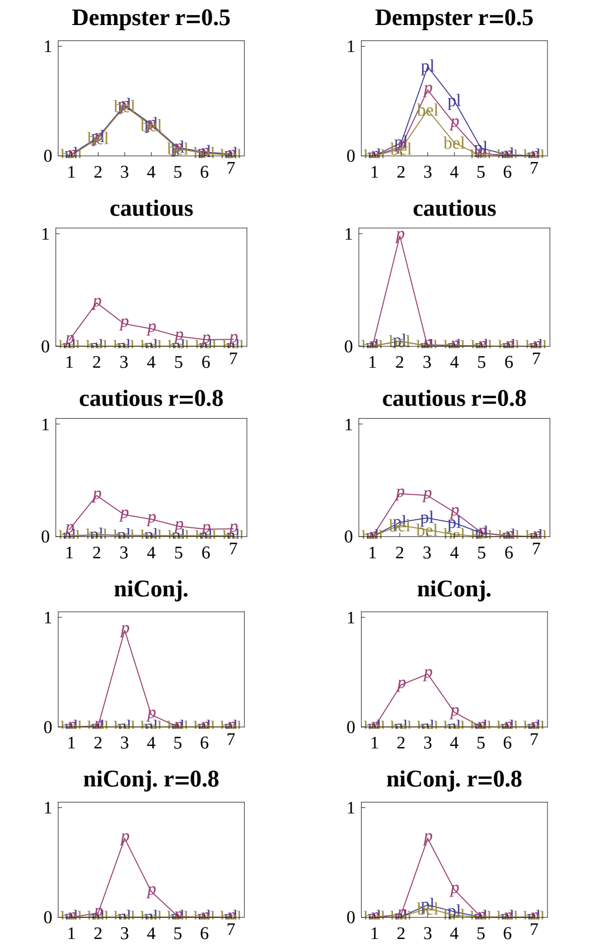

Figure 5: Sensitivity analysis, alternative symmetric fusion operators. Left column using Bayesian beliefs, right column using consonant beliefs.

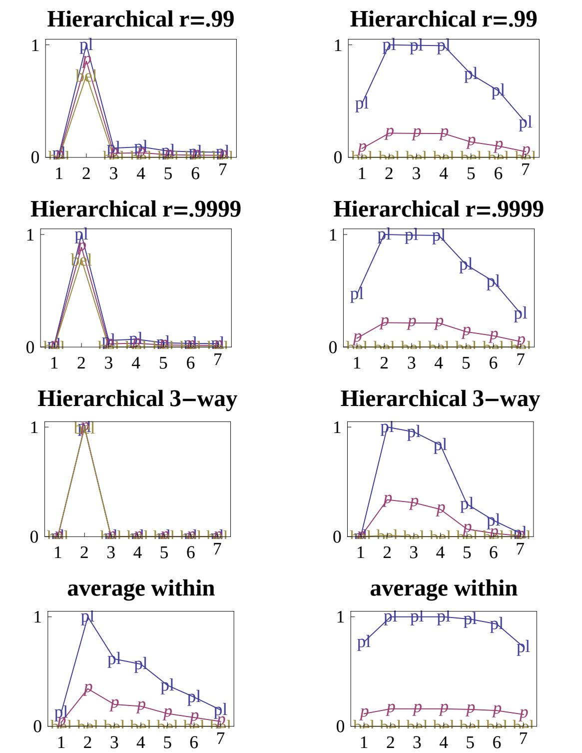

Figure 6: Sensitivity analysis, alternative hierarchic fusion. Left column using Bayesian beliefs, right column using consonant beliefs.