An Aggregation Framework Based on Coherent Lower Previsions: Application to Zadeh's Paradox and Sensor Networks

Alessio Benavoli, Alessandro Antonucci

"Dalle Molle" Institute for Artificial Intelligence (IDSIA),
Galleria 2, Via Cantonale, CH-6928 Manno-Lugano (Switzerland)

Abstract

The problem of aggregating two or more sources of information containing knowledge about a common domain is considered. We propose an aggregation framework for the case where the available information is modelled by *coherent lower previsions*, corresponding to convex sets of probability mass functions. The consistency between aggregated beliefs and sources of information is discussed. A closed formula, which specializes our rule to a particular class of models, is also derived. Two applications consisting in a possible explanation of Zadeh's paradox and an algorithm for estimation fusion in sensor networks are finally reported.

Keywords: Information fusion, coherent lower previsions, linear-vacuous mixtures, independent natural extension, natural extension, generalized Bayes rule, aggregation rule.

1. Introduction

In practical problems where modeling and handling knowledge is required, information often comes piecewise from different sources. The modeler usually wants to aggregate these pieces of information into a global model, which serves as a basis for various kinds of inference, like decision making, estimation and many others. If the available information is characterized by uncertainty, Bayesian theories can offer a suitable approach to problems of this kind. Yet, there are situations where the level of uncertainty characterizing the sources is so high that single probability measures cannot properly model the available information. This goes beyond the standard Bayesian theory, and leads to alternative models of uncertainty, like for example belief functions [1], and possibility measures [2]. As shown by Walley [3], these models represent uncertainty through sets instead of single probability measures, and they can be regarded as special cases of Walley's *coherent lower previsions* [4]. This theory, which is usually referred to as *imprecise probability*, provides a very general model of uncertain knowledge, for which some rationality criteria are also provided, these being used to identify conflicts among the different sources and check whether or not the model is self-consistent. All these features seem to be particularly suited for the aggregation of different sources of information, which can be not only uncertain and vague when considered singularly, but also conflictual or contradictory when considered jointly.

In this paper, we apply Walley's theory of coherent lower previsions to develop a general method of aggregation for uncertain information coming from different sources. In order to describe this aggregation task, let us us first formulate the problem in the Bayesian framework.

Consider n sources of information, all reporting knowledge about a variable X, whose generic value x varies in a finite set $\mathscr{X}$.[1] For each $j = 1, \ldots, n$, the knowledge associated to the j-th source is modelled by a conditional probability mass function $p_j(X|A_j = a_j)$. In this formalism, the conditioning event $A_j = a_j$ describes the actual *internal state* of each source, which is in fact modelled by a variable A_j, whose possible realizations take values a_j in a finite set $\mathscr{A}_j$. Examples of internal states of the sources can be the two states of a binary variable denoting the fact that a source is reliable or not, or a collection of measurements collected for the phenomenon under study.

The information associated to the different sources is collected by a single *information fusion center* (IFC), which aims at aggregating this information together with its prior knowledge about X, modelled as a probability mass function $p_0(X)$. This is achieved by identifying the sources' beliefs about A_j given that $X = x$ with those of the IFC, i.e., $p_0(a_j|x) := p_j(a_j|x)$, where:

$$p_j(a_j|x) = \frac{p_j(x|a_j)p_j(a_j)}{\sum_{a_j \in \mathscr{A}_j} p_j(x|a_j)p_j(a_j)}, \tag{1}$$

for each $x \in \mathscr{X}$, where $p_j(A_j)$ is the prior over the internal states of the j-th source. Thus, assuming conditional independence between variables in $A_1, \ldots, A_n$

[1] In this paper, variables are denoted by uppercase letters; the corresponding calligraphic and lowercase letters denote respectively their sets of possible values and the generic values of these sets. Accordingly, the notation $p(X)$ is used to denote a probability mass function over X, and $p(x)$ to denote the probability assigned by this mass function to a particular value $x \in \mathscr{X}$.

given X, we can aggregate the beliefs into the following joint:

$$\begin{aligned} p_0(x, a_1, \ldots, a_n) &= p_0(a_1, \ldots, a_n | x) \cdot p_0(x) \\ &= \left[\prod_{j=1}^{n} p_0(a_j | x) \right] \cdot p_0(x) \\ &= \left[\prod_{j=1}^{n} \frac{p_j(x|a_j) p_j(a_j)}{p_j(x)} \right] \cdot p_0(x), \end{aligned} \tag{2}$$

with $p_j(x) = \sum_{a_j \in \mathscr{A}_j} p_j(x|a_j) p_j(a_j)$ prior of the j-th source. Finally, from (2), the *aggregated* posterior can be written as:

$$p_0(x|\tilde{a}_1, \ldots, \tilde{a}_n) \propto \left[\prod_{j=1}^{n} \frac{p_j(x|\tilde{a}_j)}{p_j(x)} \right] \cdot p_0(x) \ , \tag{3}$$

where $\tilde{a}_j$ denotes the element of $\mathscr{A}_j$ corresponding to the observed internal state of the source. According to (3), $p_0(x|\tilde{a}_1, \ldots, \tilde{a}_n)$ is only a function of the IFC's prior $p_0(X)$, and of the sources' conditional $p_j(X|\tilde{a}_j)$ and prior $p_j(X)$, where the latter two are the only pieces of information to be shared between the sources and the IFC. Note that, while the prior over the internal states $p_j(A_j)$ is necessary to compute the conditional $p_j(A_j|x)$ from $p_j(X|a_j)$, it is not involved in (3), as we can embed it into the normalization constant.

Figure 1 depicts the sequential steps involved in the above derivation. The idea there is that each source should be regarded as an independent subject, which has inferred its conditional beliefs about X given the actual internal state of the source. As formalized in (1), each source induces a *model revision* into the IFC's beliefs. This means that, regarding the state of the source conditional on X, the IFC identifies its own beliefs with those of the source. Finally, the IFC defines a global model over all the variables by exploiting the independence among the sources as in (2).

In this architecture it has been assumed that each source processes its own information in order to compute the posterior probability $p_j(x|a_j)$, which can be regarded as a *sufficient statistical descriptor*, to be shared with the IFC together with $p_j(x)$. This is a high-level form of aggregation, since the IFC aggregates pieces of information which have already been elaborated from the sources. This is one of the most common architectures for data fusion (see for example [5, Chapter 8]). A possible alternative would be the case in which sources share directly the raw information, i.e., $p_j(a_j|x)$, with the IFC. However, this is just a sub-case of the general architecture in Figure 1 (from step (b) to step (e)).

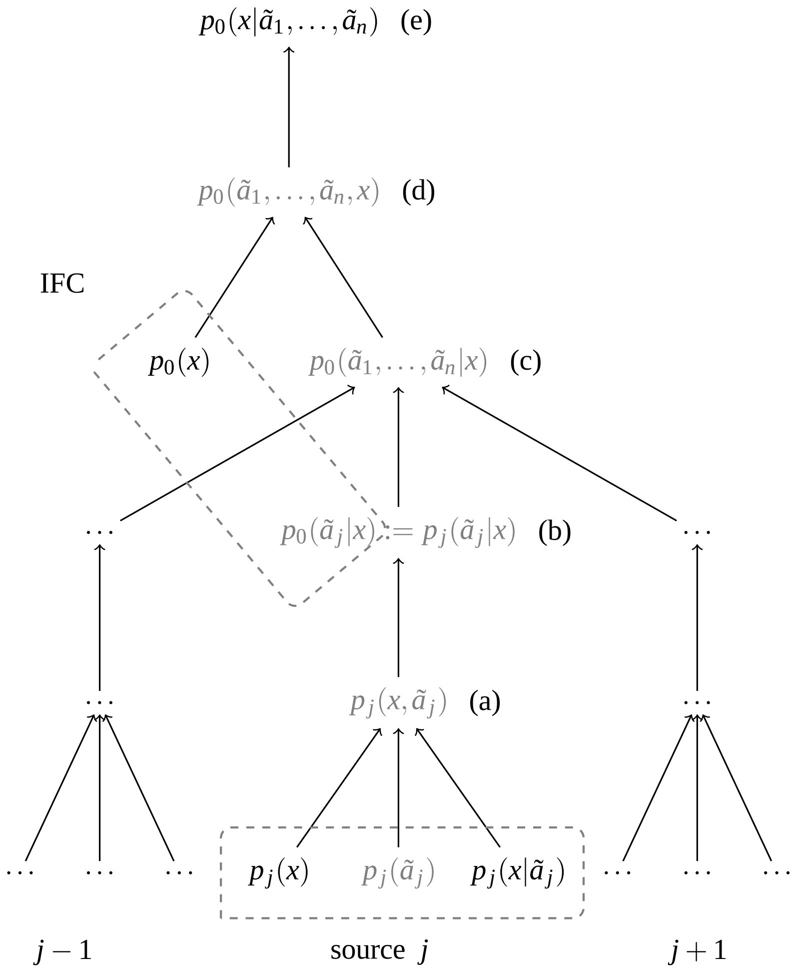

Figure 1: Aggregation of the sources of information in the Bayesian framework. The black-highlighted text describes the information used by the IFC to compute the final posterior density (still in black). The gray-highlighted text denotes the intermediate steps needed to aggregate the information. The dashed boxes are used to group the beliefs whose coherence will be checked in Section 4.

In this paper we aim at generalizing this approach to Walley's theory of imprecise probability in the general case where, instead of probability mass functions, the uncertainty about a variable is described by *coherent lower previsions*. To this end, in Section 2 we first recall the basics of the theory of coherent lower previsions. In Section 3, we detail the different steps of our derivation leading to an aggregation rule for the general case of coherent lower previsions. In particular, two cases are considered: (i) sources are assumed to be *epistemically* independent; (ii) no assumption is made about the independence among the sources. For both cases, the consistency between the obtained results and the original assessment is discussed in Section 4. In Section 5, in the case of independence, the aggregation rule is specialized for a special class of coherent lower previsions, called *linear-vacuous mixtures*. In Section 6, we show how this approach can be applied in practice for a possible explanation of Zadeh's paradox [6]. In Section 7, we show how this approach can be used to solve an estimation fusion problem in sensor networks in the case of unknown correlation among the estimates. Conclusions and outlooks for future developments are finally reported in Section 8.

2. Coherent Lower Previsions

Imprecise probability theory [4] is an extension of the Bayesian theory of subjective probability. The goal is to model a subject's uncertainty by looking at his dispositions toward taking certain actions, and imposing requirements of rationality, or consistency, on these dispositions. In order to do that, let us first recall the fundamental notion of *coherent lower prevision* (see [7] for a recent survey).

Given a variable X taking values in a set $\mathscr{X}$, we use *gambles*, i.e., bounded functions $f : \mathscr{X} \to \mathbb{R}$, in order to express a subject's uncertainty about X. For each $x \in \mathscr{X}$, the real number $f(x)$ is regarded as the (possibly negative) reward, expressed in some linear utility units that the subject receives by accepting the gamble if $X = x$. Uncertainty about the actual value of X can be modelled by the willingness to accept certain gambles and to reject others. Bayesian theory assumes that subjects are always able to provide a single fair price $P(f)$ for f, whatever information is available about X. This assumption is relaxed in the imprecise probability framework, where subjects can express two different prices, called respectively lower and upper previsions and denoted by $\underline{P}(f)$ and $\overline{P}(f)$, that correspond to the highest (lowest) buying (selling) price for the gamble f. Since selling a gamble f for a given price r is the same as buying $-f$ for the price $-r$, the conjugacy relation $\overline{P}(f) = -\underline{P}(-f)$ holds and we can therefore focus on lower previsions only. If $\mathscr{L}(\mathscr{X})$ denotes the set of all the bounded[2] gambles on $\mathscr{X}$, a lower prevision $\underline{P}$ can be regarded as a real-valued functional on $\mathscr{L}(\mathscr{X})$.

Indicator functions[3] are clearly a special class of gambles. Given set $\mathscr{X}' \subseteq \mathscr{X}$, we can consider the lower prevision for the corresponding indicator function $I_{\mathscr{X}'}$. The behavioural interpretation of $\underline{P}(I_{\mathscr{X}'})$ is the supremum rate for which the subject is disposed to bet on the event that some $x \in \mathscr{X}'$ occurs, which is the subject's *lower probability* for this event, similarly $\overline{P}(I_{\mathscr{X}'}) = 1 - \underline{P}(I_{\mathscr{X} \setminus \mathscr{X}'})$ is the *upper probability*.

Since lower previsions represent a subject's dispositions to act in certain ways, some criteria ensuring that these dispositions do not lead to irrational behaviours should be imposed. *Coherence* is the strongest requirement considered in the theory of imprecise probability. A lower prevision $\underline{P}$ is *coherent* if and only if it satisfies the following properties:

(P1) $\min_{x \in \mathscr{X}} f(x) \leq \underline{P}(f)$ [accepting sure gains],

(P2) $\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g)$ [super-additivity],

(P3) $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ [positive homogeneity],

for all $f, g \in \mathscr{L}(\mathscr{X})$ and non-negative real numbers λ. We point the reader to [4, Chapter 2] for a deep explanation of the irrational consequences of modeling beliefs by lower previsions that are not coherent. Here, we regard a *coherent lower prevision* (CLP) as the more general model of a subject's (rational) beliefs about a variable.

Let us present some examples of CLP. A *linear prevision* P on $\mathscr{L}(\mathscr{X})$ is a CLP which is also self-conjugate, i.e., $P(-f) = -P(f)$ for each $f \in \mathscr{L}(\mathscr{X})$. This property makes the prevision a linear functional, i.e., $P(\lambda(f+g)) = \lambda P(f) + \lambda P(g)$ for all $f, g \in \mathscr{L}(\mathscr{X})$ and real λ. Any linear prevision P is completely determined by its *mass function* $p(x) := P(I_{\{x\}})$, since it follows from the previous properties that for any gamble f, $P(f) = \sum_{x \in \mathscr{X}} p(x) f(x)$. Linear previsions correspond to the Bayesian notion of probability as intended in Section 1. A CLP $\underline{P}$ on $\mathscr{L}(\mathscr{X})$ such that $\underline{P}(f) = \min_{x \in \mathscr{X}} f(x)$ can be easily identified as the most conservative (i.e., less informative) CLP and is therefore called *vacuous*. As both

[2] Although Walley's theory has been developed for bounded gambles only, an extension to the unbounded case can be found in [8].

[3] A real-valued function on a domain is called the *indicator function* of a given subset of this domain if it takes the value one inside the subset and zero otherwise.

linear and vacuous previsions are coherent, we can construct new coherent lower previsions by convex combination of the two [4, Chapter 2]. If P is a linear prevision, for each $0 \leq \varepsilon \leq 1$, $\underline{P}(f) := \varepsilon P(f) + (1-\varepsilon)\min_{x\in\mathscr{X}} f(x)$ defines a new CLP which is called *linear-vacuous mixture*. Walley proved that a CLP can be equivalently specified by a convex set of linear previsions, and hence by a convex set of probability distributions [4]. In this respect, the *linear-vacuous mixture* model can be interpreted as the family of all convex combinations of a known nominal distribution (i.e., the distribution associated to $P(f)$) with any arbitrary distribution. This family can be used to address problems in which we take into account that our model $P(f)$ can be inexact and, thus, we perturb (contaminate) it to reflect this modelling uncertainty.

Now consider also a second variable A with values in $\mathscr{A}$. Given a CLP $\underline{P}$ on $\mathscr{L}(\mathscr{X}\times\mathscr{A})$, we can easily compute its *marginal* prevision on $\mathscr{A}$ for each $f \in \mathscr{L}(\mathscr{A})$ by noting that f can be equivalently regarded as a gamble in $\mathscr{L}(\mathscr{X}\times\mathscr{A})$ which is constant with respect to X, and set

$$\underline{P}^{A}(f) := \underline{P}(f), \tag{4}$$

where the superscript A emphasizes the fact that the marginal prevision is defined on $\mathscr{L}(\mathscr{A})$.

For each $h \in \mathscr{L}(\mathscr{X}\times\mathscr{A})$ and $a \in \mathscr{A}$, a subject's *conditional lower prevision* $\underline{P}^{X|A}(h|A=a)$, denoted also as $\underline{P}^{X|A}(h|a)$, is the highest real number r for which the subject would buy the gamble h for any price strictly lower than r, if he knew in addition that the variable A assumes the value a. We denote by $\underline{P}^{X|A}(h|A)$ the gamble on A that assumes the value $\underline{P}^{X|A}(h|A=a)$ for each $a \in \mathscr{A}$. Overall, $\underline{P}^{X|A}(h|A)$ is a gamble on $\mathscr{A}$ for each $h \in \mathscr{L}(\mathscr{X}\times\mathscr{A})$ and $\underline{P}^{X|A}(\cdot|A)$ is a map between $\mathscr{L}(\mathscr{X}\times\mathscr{A})$ and $\mathscr{L}(\mathscr{A})$.

A conditional lower prevision $\underline{P}^{X|A}(\cdot|A)$ is said to be *separately coherent* if $\underline{P}^{X|A}(\cdot|a)$ is a CLP on $\mathscr{L}(\mathscr{X}\times\mathscr{A})$ and $\underline{P}^{X|A}(I_{\mathscr{X}\times\{a\}}|a) = 1$, for each $a \in \mathscr{A}$. The last condition means that if the subject knew that $A = a$, he would be disposed to bet at all non-trivial odds on the event that $A = a$.

If, besides the separately coherent conditional lower prevision $\underline{P}^{X|A}(\cdot|A)$ on $\mathscr{L}(\mathscr{X}\times\mathscr{A})$, the subject has also specified an unconditional CLP $\underline{P}$ on $\mathscr{L}(\mathscr{X}\times\mathscr{A})$, then $\underline{P}$ and $\underline{P}^{X|A}(\cdot|A)$ should in addition satisfy the criterion of *joint coherence*, that requires

$$\underline{P}\left(I_{\mathscr{X}\times\{a\}}\left[h-\underline{P}^{X|A}(h|a)\right]\right) = 0, \tag{5}$$

for each $a \in \mathscr{A}$ and $h \in \mathscr{L}(\mathscr{X}\times\mathscr{A})$. It can be proved [4, Chapter 6] that, if $\underline{P}(I_{\mathscr{X}\times\{a\}}) > 0$, $\underline{P}^{X|A}(h|a)$ is the only solution of (5). Thus, given a joint CLP on $\mathscr{L}(\mathscr{X} \times \mathscr{A})$, a (separately coherent) conditional lower prevision can be obtained from (5). For this reason, this equation is also called *generalized Bayes rule* (GBR). The solution of (5) is not unique if $\underline{P}(I_{\mathscr{X}\times\{a\}}) = 0$. Nevertheless, if $\overline{P}(I_{\mathscr{X}\times\{a\}}) > 0$, a unique conditional prevision $\underline{P}^{X|A}(\cdot|a)$ can be computed by the following *regular extension* [4, Appendix J]:

$$\underline{P}^{X|A}(h|a) = \max\{\mu : \underline{P}\left(I_{\mathscr{X}\times\{a\}}[h-\mu]\right) \geq 0\}. \tag{6}$$

Finally, if also $\overline{P}(I_{\mathscr{X}\times\{a\}}) = 0$, the only coherent extension of $\underline{P}^{X|A}(\cdot|a)$ is the vacuous one.

On the other side, given a (separately coherent) conditional lower prevision $\underline{P}^{X|A}(\cdot|A)$ and a coherent marginal prevision $\underline{P}^{A}$ on $\mathscr{A}$, a joint CLP on $\mathscr{L}(\mathscr{X} \times \mathscr{A})$ can be obtained by *marginal extension*:

$$\underline{P}(h) = \underline{P}^{A}\left(\underline{P}^{X|A}(h|A)\right). \tag{7}$$

The marginal extension $\underline{P}$ in (7) can be proved to be jointly coherent with $\underline{P}^{X|A}$ as in (5), and its marginal on A is still $\underline{P}^{A}$ [4, Chapter 6].

The standard notion of conditional independence considered in the Bayesian theory requires a more general formulation in the framework of CLPs. Given a joint CLP $\underline{P}$ on $\mathscr{L}(\mathscr{X} \times \mathscr{A}_i \times \mathscr{A}_j)$, we say that, according to $\underline{P}$, A_j is *epistemically irrelevant* to A_i given X, if:

$$\underline{P}^{A_i|X,A_j}(h|x,a_j) = \underline{P}^{A_i|X}(h|x), \tag{8}$$

for each $h \in \mathscr{L}(\mathscr{A}_i)$, $x \in \mathscr{X}$ and $a_j \in \mathscr{A}_j$, where both $\underline{P}^{A_i|X,A_j}$ and $\underline{P}^{A_i|X}$ are obtained from $\underline{P}$ through GBR. If A_j is epistemically irrelevant to A_i given X, and A_i is epistemically irrelevant to A_j given X, then A_i and A_j are said to be epistemically independent (given X).

Let us adopt, for sake of compactness, the notation $A^n := (A_1, \ldots, A_n)$ and $\mathscr{A}^n := \times_{j=1}^{n} \mathscr{A}_j$.

Given a collection of separately coherent conditional lower previsions $\underline{P}_j^{A_j|X}$ on $\mathscr{L}(\mathscr{A}_j)$, for each $j = 1, \ldots, n$, the most conservative separately coherent conditional lower prevision $\underline{P}^{A^n|X}$ which is coherent with each $\underline{P}_j^{A_j|X}$, is defined as follows:

$$\underline{P}(g|x) = \sup_{\substack{g_j \in \mathscr{L}(\mathscr{A}_j) \\ j=1,\ldots,n}} \inf_{\substack{a_j \in \mathscr{A}_j \\ j=1,\ldots,n}} \left\{ g(a_1, \ldots, a_n) - \sum_{j=1}^{n} \left[g_j(a_j) - \underline{P}_j(g_j|x) \right] \right\}. \tag{9}$$

This is the *natural extension* [4]. Under the further assumption that, for each $i, j = 1, \ldots, n$ with $i \neq j$, A_i and A_j are epistemically independent given X, a more informative CLP might be obtained. This is the *independent natural extension* [9][4], which is defined as follows:[5]

$$\underline{P}(g|x) = \sup_{\substack{g_j \in \mathscr{L}(\mathscr{A}_j) \\ j=1,\ldots,n}} \inf_{\substack{a_j \in \mathscr{A}_j \\ j=1,\ldots,n}} \Big\{ g(a_1, \ldots, a_n) - \sum_{j=1}^{n} \Big[g_j(a_1, \ldots, a_n) - \underline{P}_j(g_j(a_1, \ldots, a_{j-1}, \cdot, a_{j+1} \ldots, a_n)|x) \Big] \Big\}. \quad (10)$$

The notion of joint coherence between a separately coherent conditional lower prevision and a joint CLP in (5) reflects the fact that our assessments should be consistent not only separately, but also with each other. GBR provides the definition of joint coherence for the simplest collection of (conditional) lower previsions, i.e., an unconditional and a conditional CLP. In the case of a larger collection of separately coherent conditional lower previsions, joint coherence can be characterized by the following theorem.

Theorem 1. *The separately coherent conditional lower previsions* $\underline{P}_j^{A_j|X}$, *with* $j = 1, \ldots, n$, *are jointly coherent if there is a CLP* $\underline{P}$ *on* $\mathscr{L}(\mathscr{X} \times \mathscr{A}^n)$ *such that: (i) its marginal* $\underline{P}^X$ *assigns positive probability to the elements of* $\mathscr{X}$; *(ii) its marginals* $\underline{P}^{A_j,X}$ *are jointly coherent with* $\underline{P}_j^{X|A_j}$, *for each* $j = 1, \ldots, n$, *in the sense of (5).*

The proof of this theorem can be found in [11, Theorem 5]. A more general formulation based on the concept of regular extension can be found in [12, Theorem 3].

3. Aggregating Coherent Lower Previsions

The theoretical results reviewed in Section 2 can be employed for a generalization to imprecise probabilities of the aggregation rule presented in Section 1. Accordingly, we suppose that the j-th source of information, for each $j = 1, \ldots, n$, makes assessments about the value that X assumes in $\mathscr{X}$ conditionally on its internal states $\tilde{a}_j \in \mathscr{A}_j$. Such assessments are expressed through separately coherent

[4]This paper includes a survey of different aggregation rules for CLPs. Our approach differs as we aggregate knowledge referred to a same domain.

[5]A more general formula for non-linear spaces can be found in [10].

conditional lower previsions $\underline{P}_j^{X|A_j}$. Furthermore, also extra assessments about the internal states of the sources are available and again expressed in terms of CLPs $\underline{P}_j^{A_j}$ on $\mathscr{L}(\mathscr{A}_j)$ for $j = 1,\dots,n$. The IFC should therefore gather this information and aggregate it with its prior about X, which is expressed as a CLP $\underline{P}_0^X$ on $\mathscr{L}(\mathscr{X})$.

Our goal is to compute the IFC's joint CLP $\underline{P}_0$ on $\mathscr{L}(\mathscr{X} \times \mathscr{A}^n)$ from which the beliefs about X conditional on the actual internal states of the sources $(\tilde{a}_1,\dots,\tilde{a}_n)$ could be computed. By analogy with the derivation in Section 1, this task is achieved by the following sequential steps:

(a) As outlined in (7), a CLP $\underline{P}_j$ on $\mathscr{L}(\mathscr{X} \times \mathscr{A}_j)$ can be derived from $\underline{P}_j^{X|A_j}$ and $\underline{P}_j^{A_j}$ by marginal extension

$$\underline{P}_j(f_j) := \underline{P}_j^{A_j}\left(\underline{P}_j^{X|A_j}(f_j|A_j)\right), \tag{11}$$

for each $f_j \in \mathscr{L}(\mathscr{X} \times \mathscr{A}_j)$ and $j = 1,\dots,n$.

(b) GBR as in (5) is used to compute, given $\underline{P}_j$, the conditional CLP $\underline{P}_j^{A_j|X}$ on $\mathscr{L}(\mathscr{X} \times \mathscr{A}_j)$.[6] Accordingly, by computing the solution μ of the equation

$$\underline{P}_j\left(I_{\{\tilde{x}\}} \cdot [f_j - \mu]\right) = 0, \tag{12}$$

we have $\underline{P}_j^{A_j|X}(f_j|\tilde{x}) := \mu$, for each $f_j \in \mathscr{L}(\mathscr{A}_j)$, $\tilde{x} \in \mathscr{X}$, and $j = 1,\dots,n$.

The so-obtained separately coherent conditional lower previsions associated to the sources are assumed to induce a *model revision* into the corresponding beliefs of the IFC, i.e.,

$$\underline{P}_0^{A_j|X}(f_j|x) := \underline{P}_j^{A_j|X}(f_j|x), \tag{13}$$

for each $f_j \in \mathscr{L}(\mathscr{A}_j)$ and $x \in \mathscr{X}$.

(c) If the sources are epistemically irrelevant each other given $X = x$, a conditional CLP $\underline{P}_0^{A^n|X}$ can be obtained from $\underline{P}_0^{A^n|X}$ by means of independent natural

[6] We noted that GBR requires $\underline{P}_j^X(I_{\{\tilde{x}\}}) > 0$. If only $\overline{P}_j^X(I_{\{\tilde{x}\}}) > 0$ holds, regular extension (6) should be employed instead. An example of the calculations required in this latter case is in Section 6. Finally, if also $\overline{P}_j^X(I_{\{\tilde{x}\}}) = 0$, the only coherent extension of $\underline{P}_j^{A_j|X}$ is the vacuous one.

extension (10). If the irrelevance assumption cannot be met, the natural extension (9) can be used instead. Due to its generality, i.e., no assumption about the independence among the sources is made, natural extension may produce very conservative (i.e., uninformative) results. When the irrelevance among the sources cannot be guaranteed, in order to obtain more informative results, the conditional CLP $\underline{P}_0^{A^n|X}$ can be defined in the following way:

$$\underline{P}_0^{A^n|X} := (1-\gamma)\underline{P}_{INE}^{A^n|X} + \gamma\underline{P}_{NE}^{A^n|X}, \tag{14}$$

with $0 \leq \gamma \leq 1$. This is the contamination of the CLP $\underline{P}_{INE}^{A^n|X}$, obtained by independent natural extension, with the CLP $\underline{P}_{NE}^{A^n|X}$, obtained by natural extension.[7] The CLP (14) can be used to address cases in which we take into account that the independence assumption can be wrong and, thus, we perturb $\underline{P}_{INE}^{A^n|X}$ with $\underline{P}_{NE}^{A^n|X}$.[8] This approach can be easily proved to preserve coherence (see Section 4), while the same cannot be guaranteed for contamination of $\underline{P}_{INE}^{A^n|X}$ with arbitrary CLPs (e.g., vacuous ones).

(d) Then, the joint CLP $\underline{P}_0$ on $\mathscr{L}(\mathscr{X} \times \mathscr{A}^n)$ is derived by marginal extension (7):

$$\underline{P}_0(g) := \underline{P}_0^X \left(\underline{P}_0^{A^n|X}(g|X) \right), \tag{15}$$

for each $g \in \mathscr{L}(\mathscr{X} \times \mathscr{A}^n)$.

(e) Finally, assuming that $\underline{P}_0^{A^n}(\tilde{a}_1, \ldots, \tilde{a}_n) > 0$, where $(\tilde{a}_1, \ldots, \tilde{a}_n) \in \mathscr{A}^n$ are the observed internal states of the sources, we again apply GBR,

$$\underline{P}_0 \left(I_{\{\tilde{a}_1, \ldots, \tilde{a}_n\}} \cdot [g - \mu] \right) = 0, \tag{16}$$

to compute the separately coherent conditional lower prevision $\underline{P}_0^{X|A^n}(\cdot|A^n)$ on $\mathscr{L}(\mathscr{X})$.[9]

[7] Note that the contamination is between CLPs over the same domain.

[8] The design parameter γ is used to weight the possibility that the assumption of epistemic independence could be inexact. The choice of the value of γ is based on a trade-off between robustness of the model (large γ) and precision (informativeness) of the inferences (small γ).

[9] Note that, also in this case, if we only have that $\overline{P}_0^{A^n}(\tilde{a}_1, \ldots, \tilde{a}_n) > 0$, the regular extension (6) can be used instead. Again, if also $\overline{P}_0^{A^n}(\tilde{a}_1, \ldots, \tilde{a}_n) = 0$, the only coherent extension of $\underline{P}_0^{X|A^n}(\cdot|A^n)$ is the vacuous one.

The above derivation has been achieved by analogy with that in Section 1, but in the more general framework of CLPs. This allows for a more robust modelling of the information reported by the sources, under weaker assumptions about their independence (see (9) or (14)). Notice also that, if the sources directly provide the CLPs $\underline{P}_j^{A_j|X}$, we could still apply our procedure by considering only the steps from (c) to (e). In the case of epistemic independence among the sources, the posterior CLP coincides in this case with that returned by a *naive credal classifier* (e.g., compare Table 2 with the results in [13]). This holds in spite of a different notion of independence (strong independence is assumed in [13]). The same results can be also obtained by means of the algorithm in [14].

Notice that, steps from (c) to (e) hold also if X is a continuous variable, this producing in general a non-vacuous conditional $\underline{P}_0^{X|A^n}(\cdot|A^n)$. Conversely, step (b) for X continuous would produce a vacuous conditional, being in general $\underline{P}_j^X(I_{\{\tilde{x}\}}) = \overline{P}_j^X(I_{\{\tilde{x}\}}) = 0$. Therefore, if the sources directly provide the CLPs $\underline{P}_j^{A_j|X}$, we can apply our aggregation framework also for continuous X. An example of this kind of application is in Section 7, while in the rest of the paper we always assume both X and A_j, $j = 1, \ldots, n$, discrete.

The coherence between the joint CLP obtained at the step (d) and the initial assessments will be investigated in the next section.

4. Checking Coherence

The subjects involved in the derivation formalized in the previous section (i.e., the sources and the IFC) should be regarded as autonomous and distinct individuals. Nevertheless, we have assumed that the uncertain information associated to a subject can induce in another subject a *model revision*, i.e., the second agent can replace his own CLPs (even in the conditional case) with those of the first agent. More specifically, in our architecture, we allow for an *asymmetrical* model revision, as we assume that each source revises the IFC's beliefs as in (1) or in (13), while the contrary cannot take place because of the way the sources and the IFC share the information (no feedback allowed). In this section we discuss the coherence between the different beliefs specified in our model. According to the previous argument, this will be done separately for each subject, by considering also the beliefs induced by other subjects via model revision.

Let us start from the coherence of the IFC's beliefs. In order to do that, we first consider the derivation in the precise case as in Section 1. As outlined in Figure 1, the mass functions to be considered are the conditionals $p_0(A_j|x)$, for each $j = 1,\ldots,n$, which are obtained through model revision from the sources, and the marginal $p_0(X)$. The consistency between these assessments when considered jointly follows from the existence of a joint probability mass function, which is clearly the one in (2), from which these mass functions can be obtained. Concerning the IFC, we should also verify that this joint probability mass function preserves the assumption of independence between the sources given X. This holds since, after marginalization and Bayes rule, the joint probability mass function p_0 in (2) is such that $p_0(a_i|x,a_j) = p_0(a_i|x)$ for each $i, j = 1,\ldots,n$, $a_i \in \mathscr{A}_i$, $a_j \in \mathscr{A}_j$ and $x \in \mathscr{X}$. Analogous results, in the more general framework of imprecise probability, can be obtained by considering the joint CLP $\underline{P}_0$ in (15), which is the basis to prove the following result.

Theorem 2. *The separately coherent conditional lower previsions $\underline{P}_0^{A_j|X}$ in (13) and $\underline{P}_0^X$ are jointly coherent.*

Proof. The joint coherence of the assessments $\underline{P}_0^X$ and $\underline{P}_j^{A_j|X}(\cdot|x)$, considered for each $j = 1,\ldots,n$, can be proved by considering the joint CLP $\underline{P}_0^{X,A^n}$ in (15). As a consequence of marginal extension, $\underline{P}_0^{X,A^n}$ is jointly coherent with both $\underline{P}_0^X$ and $\underline{P}_0^{A^n|X}(\cdot|x)$. For each $f_j \in \mathscr{L}(\mathscr{A}_j)$ and $x \in \mathscr{X}$, Equation (14) states that:[10]

$$\underline{P}_0^{A^n|X}(f_j|x) = (1-\gamma)\underline{P}_{INE}^{A^n|X}(f_j|x) + \gamma\underline{P}_{NE}^{A^n|X}(f_j|x). \tag{17}$$

But $\underline{P}_{INE}^{A^n|X}(f_j|x) = \underline{P}_0^{A_j|X}(f_j|x)$ because of definition of independent natural extension and similarly $\underline{P}_{NE}^{A^n|X}(f_j|x) = \underline{P}_0^{A_j|X}(f_j|x)$ because of definition of natural extension. Thus, for each $f_j \in \mathscr{L}(\mathscr{A}_j)$ and $x \in \mathscr{X}$, $\underline{P}_0^{A^n|X}(f_j|x) = \underline{P}_0^{A_j|X}(f_j|x)$, i.e, $\underline{P}_0^{A^n|X}$ and $\underline{P}_0^{A_j|X}$ are jointly coherent for each $i = 1,\ldots,n$.

Finally, if $\underline{P}_j^{A^n}(I_{\{a_1,\ldots,a_n\}}) > 0$ $\underline{P}^{X|A^n}(\cdot|a_1,\ldots,a_n)$ can be obtained by GBR and its coherence follows from Theorem 1. On the other side, if $\underline{P}_j^{A^n}(I_{\{a_1,\ldots,a_n\}}) = 0$, but $\overline{P}_j^{A^n}(I_{\{a_1,\ldots,a_n\}}) > 0$, $\underline{P}_j^{X|A^n}(\cdot|a_1,\ldots,a_n)$ can be obtained by regular extension, and its coherence follows from [12, Theorem 3], where it is proved that $\underline{P}_0^X$ and $\underline{P}^{A^n|X}$ are jointly coherent with their so-called *strong product*. This result applies to our case because the joint $\underline{P}_0^{A^n,X}$, as defined in (15) can be easily verified to be

[10]We proceed as in Equation (4) in order to regard a gamble over a single variable as a gamble in a joint domain.

the *strong product* of $\underline{P}_0$ and $\underline{P}^{A^n|X}$. This proof holds also if X is a continuous variable, see [15, Appendix]. $\square$

On the other side, checking the coherence of the beliefs associated to a particular source is trivial, as $\underline{P}_j^{X|A_j}$ and $\underline{P}_j^{A_j}$ are jointly coherent because of (5), for each $j = 1, \ldots, n$. We have argued that the IFC's beliefs are not required to be coherent with those of the sources, as they refer to separate subjects. Nevertheless, let us consider what can be said about the consistency between different subjects in the Bayesian (i.e., precise) formulation. By exploiting the independence between the sources, (2) rewrites as:

$$p_0(x, a_1, \ldots, a_n) = \prod_{j=1}^{n} \frac{p_0(x|a_j)p_0(a_j)}{p_0(x)} p_0(x) \ . \tag{18}$$

By comparing (18) with (2), it can be noticed that the joint coherence between the IFC's beliefs and those of the j-th source cannot be guaranteed in general. In fact, we can always impose $p_0(x|a_j) := p_j(x|a_j)$ and $p_0(a_j) := p_j(a_j)$, but, at least in general, it is not possible to have at the same time $p_0(X) = p_j(X)$, for each $j = 1, \ldots, n$. In fact, since each source and the IFC are considered as autonomous subjects and the information flows from the sources to the IFC, we cannot require that the sources agree on their marginals over $\mathscr{X}$, i.e., $p_i(X) = p_j(X)$ for each $i, j = 1, \ldots, n$. Thus, the IFC can define a single global probabilistic model over all the variables that reproduces all the inputs from the sources only if the IFC and all the sources share the same prior over X.

5. Mathematical Derivation for Linear-Vacuous Mixtures under Epistemic Independence

Let us detail the derivation described in Section 3 for the special case of epistemic independence among the sources given X when the marginal associated to the IFC and the separately coherent conditional lower previsions specified for the sources are linear-vacuous mixtures, while the marginals over $\mathscr{A}_j$ are linear.[11]

[11] The last assumption will be relaxed at the end of this section.

This corresponds to the following settings:

$$\begin{aligned}
\underline{P}_0^X(h) &:= \varepsilon_0 \sum_{x\in\mathscr{X}} p_0(x)h(x) + (1-\varepsilon_0)\min_{x\in\mathscr{X}} h(x),\\
\underline{P}_j^{X|A_j}(f_j|a_j) &:= \varepsilon_j^{a_j} \sum_{x\in\mathscr{X}} p_j(x|a_j) f_j(x,a_j) \\
&+ (1-\varepsilon_j^{a_j})\min_{x\in\mathscr{X}} f_j(x,a_j), \quad \forall a_j \in \mathscr{A}_j
\end{aligned} \tag{19}$$

$$P_j^{A_j}(g_j) := \sum_{a_j\in\mathscr{A}_j} p_j(a_j) g_j(a_j), \tag{20}$$

where $p_j(X|a_j)$, $p_j(A_j)$ and $p_0(X)$ are probability mass functions, $f_j \in \mathscr{L}(\mathscr{X}\times\mathscr{A}_j)$, $g_j \in \mathscr{L}(\mathscr{A}_j)$, and $h \in \mathscr{L}(\mathscr{X})$, for all $j = 1,\ldots,n$. The derivation is as follows.

(a) In this particular case, (11) rewrites as

$$\begin{aligned}
\underline{P}_j(f_j) &= \sum_{a_j\in\mathscr{A}_j} p_j(a_j)\cdot\Big(\varepsilon_j^{a_j}\sum_{x\in\mathscr{X}} p_j(x|a_j)\cdot f_j(x,a_j)\\
&+ (1-\varepsilon_j^{a_j})\min_{x\in\mathscr{X}} f_j(x,a_j)\Big),
\end{aligned} \tag{21}$$

for each $f_j \in \mathscr{L}(\mathscr{X}\times\mathscr{A}_j)$ and $j = 1,\ldots,n$.

(b) Thus, for each $\tilde{x}\in\mathscr{X}$, (12) becomes:

$$\begin{aligned}
&\sum_{a_j\in\mathscr{A}_j} p_j(a_j)\cdot\Big(\varepsilon_j^{a_j}[f_j(\tilde{x},a_j)-\mu]p_j(\tilde{x}|a_j)\\
&+(1-\varepsilon_j^{a_j})\min\{0, f_j(\tilde{x},a_j)-\mu\}\Big) = 0.
\end{aligned} \tag{22}$$

Define the subset $\mathscr{A}_j^*(\mu)$ of $\mathscr{A}_j$ as follows:

$$\mathscr{A}_j^*(\mu) := \{a_j \in \mathscr{A}_j :\ f_j(\tilde{x},a_j)-\mu < 0\}, \tag{23}$$

where $f_j, \tilde{x}$ are omitted from the arguments of $\mathscr{A}_j^*$ for sake of simpler notation. Equation (22) rewrites as:

$$\begin{aligned}
&\sum_{a_j\in\mathscr{A}_j} p_j(a_j)[\varepsilon_j^{a_j} p_j(\tilde{x}|a_j) + (1-\varepsilon_j^{a_j}) I_{\mathscr{A}_j^*(\mu)}(a_j)] f_j(\tilde{x},a_j)\\
&-\mu \sum_{a_j\in\mathscr{A}_j} p_j(a_j)[\varepsilon_j^{a_j} p_j(\tilde{x}|a_j) + (1-\varepsilon_j^{a_j}) I_{\mathscr{A}_j^*(\mu)}(a_j)] = 0.
\end{aligned} \tag{24}$$

The solution of (22) is non-trivial because $\mathscr{A}_j^*$ is a function of μ. Yet, we can compute $\mathscr{A}_j^*(\mu)$ for the particular value $\tilde{\mu}$ of μ that solves (22), without explicitly solving this equation. Accordingly, we set $\mathscr{A}_j^* := \mathscr{A}_j^*(\tilde{\mu})$, and the solution $\underline{P}_j^{A_j|X}(f_j|\tilde{x})$ of (24) is:[12]

$$\frac{\sum_{a_j\in\mathscr{A}_j} p_j(a_j)[\varepsilon_j^{a_j} p_j(\tilde{x}|a_j) + (1-\varepsilon_j^{a_j}) I_{\mathscr{A}_j^*}(a_j)] f_j(\tilde{x},a_j)}{\sum_{a_j\in\mathscr{A}_j} p_j(a_j)[\varepsilon_j^{a_j} p_j(\tilde{x}|a_j) + (1-\varepsilon_j^{a_j}) I_{\mathscr{A}_j^*}(a_j)]}. \tag{25}$$

(c) The (separately coherent) conditional lower previsions associated to the sources and defined as in (25) induce the following *model revision* into the IFC's beliefs,

$$\underline{P}_0^{A_j|X}(f_j|x) := \underline{P}_j^{A_j|X}(f_j|x), \tag{26}$$

for each $f_j \in \mathscr{L}(\mathscr{A}_j)$, $j = 1,\dots,n$ and $x \in \mathscr{X}$. As a consequence of the epistemic irrelevance assumption for the sources given X, the independent natural extension to $\mathscr{A}^n$ should be considered. According to (10), we have:

$$\underline{P}_0^{A^n|X}(g|\tilde{x}) = \sup_{g_j\in\mathscr{L}(\mathscr{X}\times\mathscr{A}_j)\, j=1,\dots,n} \quad \inf_{a_j\in\mathscr{A}_j\, j=1,\dots,n} \Big\{ g(\tilde{x},a_1,\dots,a_n) \\ - \sum_{j=1}^n \Big[g_j(\tilde{x},a_1,\dots,a_n) - \underline{P}_0^{A_j|X}(g_j(\tilde{x},a_1,\dots,a_{j-1},\cdot,a_{j+1}\dots,a_n)|\tilde{x}) \Big] \Big\}, \tag{27}$$

for each $\tilde{x} \in \mathscr{X}$. Notice that the gamble $g_j(\tilde{x},a_1,\dots,a_{j-1},\cdot,a_{j+1}\dots,a_n)$ is in $\mathscr{L}(\mathscr{X}\times\mathscr{A}_j)$. Let us consider, in (27), only gambles $g \in \mathscr{L}(\mathscr{X}\times\mathscr{A}^n)$ such that, for $X = \tilde{x}$ and each $(a_1,\dots,a_n) \in \mathscr{A}^n$, factorize as follows:

$$g(\tilde{x},a_1,\dots,a_n) = \prod_{j=1}^n g'_j(\tilde{x},a_j), \tag{28}$$

with $g'_j \in \mathscr{L}(\mathscr{X}\times\mathscr{A}_j)$ for each $j = 1,\dots,n$. Assume also that the gamble $g'_j(\tilde{x},\cdot) \in \mathscr{L}(\mathscr{A}_j)$ has a constant sign in $\mathscr{A}_j$, and denote its sign by $\sigma_j = \sigma_j(\tilde{x})$[13]. Under these assumptions, if we intend, for fixed $\tilde{x}$, g as a gamble on

[12] This is possible if $\underline{P}_j(I_{\{\tilde{x}\}\times\mathscr{A}_j}) = \sum_{a_j\in\mathscr{A}_j} p_j(a_j)\varepsilon_j^{a_j} p_j(\tilde{x}|a_j) > 0$.

[13] Set $\sigma_j = +1$ if $g'_j(\tilde{x},\cdot) > 0$, $\sigma_j = -1$ if $g'_j(\tilde{x},\cdot) < 0$ and $\sigma_j = 0$ otherwise.

$\mathscr{A}^n$, we have that g has constant sign and (27) reduces to:

$$\underline{P}_0^{A^n|X}(g|\tilde{x}) = \begin{cases} \prod_{j=1}^n \underline{P}_0^{A_j|X}(g'_j|\tilde{x}) & \text{if } g \geq 0 \\ - \prod_{j=1}^n \overline{P}_0^{A_j|X}(\sigma_j g'_j|\tilde{x}) & \text{if } g < 0 \end{cases}, \tag{29}$$

where g'_j is the above defined g_j, for each $j = 1, \ldots, n$. The proof is in [10, Section 5]. The gambles we consider in the following factorize as in (28), and we can therefore use (29) instead of (27).

(d) By marginal extension (15), the following joint CLP can be calculated:

$$\begin{aligned} \underline{P}_0(h) = \underline{P}_0^X\left(\underline{P}_0^{A^n|X}(h|x)\right) = \varepsilon_0 \sum_{x \in \mathscr{X}} \underline{P}_0^{A^n|X}(h|x) p_0(x) \\ +(1-\varepsilon_0) \min_{x \in \mathscr{X}} \underline{P}_0^{A^n|X}(h|x). \end{aligned} \tag{30}$$

(e) Thus, by GBR, given $\{\tilde{a}_1, \ldots, \tilde{a}_n\} \in \mathscr{A}^n$, the conditional CLP $\underline{P}_0^{X|A^n}(g|\tilde{a}_1, \ldots, \tilde{a}_n)$ is the solution of:

$$\underline{P}_0(I_{\{\tilde{a}_1\} \times \cdots \times \{\tilde{a}_n\}}(g - \mu)) = 0, \tag{31}$$

where we assume $\underline{P}_0(I_{\{\tilde{a}_1\} \times \cdots \times \{\tilde{a}_n\}}) > 0$. Note also that the only values of the gamble g that should be considered for the solution of (31) are those such that $A^n \neq \tilde{a}^n$, because otherwise the argument of $\underline{P}_0$ is zero. Furthermore, for fixed x, $g(x, \tilde{a}_1, \ldots, \tilde{a}_n) - \mu$ is constant. Thus, the gamble factorizes as in (28), with $g'_i(\tilde{x}, a_i) = I_{\{\tilde{a}_i\}}$, for each $i < n$ and $g'_i(\tilde{x}, a_n) = I_{\{\tilde{a}_n\}}(g(\cdot) - \mu)$. Therefore, notice that $\sigma_i = 1$, for each $i < n$ and $\sigma_n = \text{sgn}(g(\cdot) - \mu)$. Thus, (29) holds and we can write:[14]

$$\begin{aligned} \underline{P}^{A^n|X}(h|x) &= \underline{P}_1^{A_1|X}(I_{\{\tilde{a}_1\}}|x) \cdots \underline{P}_n^{A_n|X}(I_{\{\tilde{a}_n\}}|x) \\ &\quad [g(x, \tilde{a}_1, \ldots, \tilde{a}_n) - \mu] I_{\{g(x, \tilde{a}_1, \ldots, \tilde{a}_n) - \mu \geq 0\}} \\ &+ \overline{P}_1^{A_1|X}(I_{\{\tilde{a}_1\}}|x) \cdots \overline{P}_n^{A_n|X}(I_{\{\tilde{a}_n\}}|x) \\ &\quad [g(x, \tilde{a}_1, \ldots, \tilde{a}_n) - \mu] I_{\{g(x, \tilde{a}_1, \ldots, \tilde{a}_n) - \mu < 0\}}. \end{aligned} \tag{32}$$

According to (32), (31) can be written as in Table 1, where from (25) it can be derived that:

$$\underline{P}_j^{A_j|X}(I_{\{\tilde{a}_j\}}|\tilde{x}) = \frac{p_j(\tilde{a}_j)\varepsilon_j^{\tilde{a}_j} p_j(x|\tilde{a}_j)}{\sum_{a_j \in \mathscr{A}_j} p_j(a_j)[\varepsilon_j^{a_j} p_j(x|a_j) + (1-\varepsilon_j^{a_j}) I_{\mathscr{A}_j \setminus \{\tilde{a}_j\}}(a_j)]}. \tag{33}$$

[14]Note that the indicator functions in (32) refer to sets that are implicitly defined through inequalities over gambles. This kind of specification will be employed also in the followings.

Table 1: The unique solution μ of GBR corresponding to the conditional CLP $\underline{P}_0^{X|A^n}(g|\tilde{a}_1,\ldots,\tilde{a}_n)$

$$\begin{aligned} 0 &= \varepsilon_0 \sum_{x\in\mathscr{X}} \Big\{ \Big[\underline{P}_0^{A_1|X}(I_{\{\tilde{a}_1\}}|x)\cdots\underline{P}_0^{A_n|X}(I_{\{\tilde{a}_n\}}|x)I_{\{g(x,\tilde{a}_1,\ldots,\tilde{a}_n)-\mu\geq 0\}} \\ &+ \overline{P}_0^{A_1|X}(I_{\{\tilde{a}_1\}}|x)\cdots\overline{P}_0^{A_n|X}(I_{\{\tilde{a}_n\}}|x)I_{\{g(x,\tilde{a}_1,\ldots,\tilde{a}_n)-\mu<0\}}\Big](g(x,\tilde{a}_1,\ldots,\tilde{a}_n)-\mu)p_0(x)\Big\} \\ &+ (1-\varepsilon_0)\min_{x\in\mathscr{X}} \Big\{ \Big[\underline{P}_0^{A_1|X}(I_{\{\tilde{a}_1\}}|x)\cdots\underline{P}_0^{A_n|X}(I_{\{\tilde{a}_n\}}|x)I_{\{g(x,\tilde{a}_1,\ldots,\tilde{a}_n)-\mu\geq 0\}} \\ &+ \overline{P}_0^{A_1|X}(I_{\{\tilde{a}_1\}}|x)\cdots\overline{P}_0^{A_n|X}(I_{\{\tilde{a}_n\}}|x)I_{\{g(x,\tilde{a}_1,\ldots,\tilde{a}_n)-\mu<0\}}\Big](g(x,\tilde{a}_1,\ldots,\tilde{a}_n)-\mu)\Big\} \end{aligned}$$

It can be easily verified that $\mathscr{A}_j^* = \mathscr{A}_j \setminus \{\tilde{a}_j\}$ in this case. Again from (25):

$$\underline{P}_j^{A_j|X}\left(I_{\{\mathscr{A}_j\setminus\tilde{a}_j\}}|x\right) = \frac{\sum_{a_j\in\mathscr{A}_j, a_j\neq\tilde{a}_j} p_j(a_i)\varepsilon_j^{a_j} p_j(x|a_j)}{\sum_{a_j\in\mathscr{A}_j} p_j(a_j)[\varepsilon_j^{a_j} p_j(x|a_j) + (1-\varepsilon_j^{a_j})I_{\{\tilde{a}_j\}}(a_j)]}, \tag{34}$$

where, in this case, $\mathscr{A}_j^* = \{\tilde{a}_j\}$. According to the conjugacy relation reviewed in Section 2, the corresponding upper probability is one minus the lower probability in (34), and hence:

$$\overline{P}_j^{A_j|X}\left(I_{\{\tilde{a}_j\}}|x\right) = \frac{p_j(\tilde{a}_j)[\varepsilon_j^{\tilde{a}_j} p_j(x|\tilde{a}_j) + (1-\varepsilon_j^{\tilde{a}_j})]}{\sum_{a_j\in\mathscr{A}_j} p_j(a_i)[\varepsilon_j^{a_j} p_j(x|a_j) + (1-\varepsilon_j^{a_j})I_{\{\tilde{a}_j\}}(a_j)]}. \tag{35}$$

Finally, by solving the equation in Table 1 with respect to μ, the conditional CLPs $\underline{P}_0^{X|A^n}(g|\tilde{a}_1,\ldots,\tilde{a}_n)$ can be calculated for each $\{\tilde{a}_1,\ldots,\tilde{a}_n\} \in \mathscr{A}^n$.

The assumption of linearity for the prior beliefs over the sources can be relaxed to the case where the previsions $\underline{P}_j^{A_j}$ are CLPs generated by the lower envelope of a finite set of linear previsions [4, Chapter 3]. In this case, we solve the equation in Table 1 for each element of this set, and the minimum over these values is the solution in the general case. The following results can be easily verified to follow from our derivation.

1. If $\underline{P}_0^X$ is vacuous (i.e., $\varepsilon_0 = 0$), then also $\underline{P}_0^{X|A^n}$ is vacuous. This is consistent with the results in [16].

2. If $\underline{P}_j^{X|A_j}$ is vacuous (i.e., $\varepsilon_j^{\tilde{a}_j} = 0$) for each $j = 1, \ldots, n$, then $\underline{P}_j(I_{\{\tilde{x}\}\times\mathscr{A}_j}) = 0$ and, (22) cannot be solved by (25). In this case, from (22) it is straightforward to verify that $\underline{P}_j^{A_j|X}(f_j|\tilde{x})$ is vacuous (if $p_j(a_i) > 0$ for each i), that $\underline{P}^{A^n|X}(g|\tilde{x})$ is also vacuous and that $\underline{P}_0^{X|A^n}(g|\tilde{a}_1, \ldots, \tilde{a}_n)$ is equal to $\underline{P}_0^X(g)$.

3. In order to derive an analytical derivation for $\underline{P}_0^{X|A^n}(g|\tilde{a}_1, \ldots, \tilde{a}_n)$, in (24) and (31) we have assumed non-zero lower probability for the conditioning events. However, if this is not the case, but the upper probability of the conditioning events is positive, we can apply regular extension to compute a non-vacuous conditional CLP $\underline{P}_0^{X|A^n}(g|\tilde{a}_1, \ldots, \tilde{a}_n)$. However, for this case, a general analytical expression of $\underline{P}_0^{X|A^n}(g|\tilde{a}_1, \ldots, \tilde{a}_n)$ cannot be derived, but the posterior CLP can be computed numerically.

4. In (3), it is shown that, since the posterior $p_0(x|a_1, \ldots, a_n)$ does not depend on $p(a_j)$, the only pieces of information to be shared between sources and IFC are $p_j(x)$ and $p_j(x|a_j)$. In the imprecise case, additional information must be shared between sources and IFC. In fact, from Table 1 and from (33) and (35), it can be seen that $\underline{P}_0^{X|A^n}(g|\tilde{a}_1, \ldots, \tilde{a}_n)$ depends on the sources' priors $\underline{P}_j^X$, on the conditional $\underline{P}_j^{X|A_j}$ and on $(1-\varepsilon_j^{\tilde{a}_j})p(\tilde{a}_j)$. In fact, that the denominator in (33) is just equal to $\overline{P}_j^X(I_{\{x\}}) - (1-\varepsilon_j^{\tilde{a}_j})p(\tilde{a}_j) = \underline{P}_j^X(I_{\mathscr{X}\setminus\{x\}}) - (1-\varepsilon_j^{\tilde{a}_j})p(\tilde{a}_j)$, while the denominator in (35) is $\underline{P}_j^X(I_{\{x\}}) + (1-\varepsilon_j^{\tilde{a}_j})p(\tilde{a}_j)$. Conversely, the dependency on $p(\tilde{a}_j)$ in the numerators of (33) and (35) is dropped in Table 1, since the sum and the minimum are over x and, thus $p(\tilde{a}_j)$ can be simplified. Summarizing, the pieces of information to be shared between sources and IFC are: the conditional $\underline{P}_j^{X|A_j}$, the marginal CLP $\underline{P}_j^X$, which corresponds to the prior CLP of the sources and the quantity $(1-\varepsilon_j^{\tilde{a}_j})p(\tilde{a}_j)$, which is equal to the probability that the j-th source is in the state $p(\tilde{a}_j)$ multiplied by the *degree of imprecision* $\overline{P}_j^{X|A_j}(I_{\{x\}}) - \underline{P}_j^{X|A_j}(I_{\{x\}}) = 1 - \varepsilon_j^{\tilde{a}_j}$.

6. Zadeh's Paradox

The problem of aggregating beliefs over the same variable and under the assumption of independence has been already considered in other uncertainty theories. In the case of Dempster-Shafer (DS) theory [1], Dempster's combination rule allows for the following aggregation of two mass functions m_1 and m_2:[15]

$$m_{12}(X) \propto \sum_{X_1,X_2:X_1\cap X_2=X} m_1(X_1)\cdot m_2(X_2). \tag{36}$$

Yet, in the 1980s, DS theory suffered a serious blow when Zadeh proposed his "paradox", an example for which the Dempster's rule of combination gives an apparently counter-intuitive result [6].

Zadeh's example is as follows. Two doctors examine a patient and agree that he suffers from either meningitis (x_1), concussion (x_2) or brain tumor (x_3). Thus, $\mathscr{X} = \{x_1, x_2, x_3\}$ is the state space (frame of discernment) of the variable of interest X. The doctors agree in considering a tumor quite unlikely, but they disagree in deciding the likely cause, thus providing the following diagnosis:

$$\begin{array}{lll} \text{Doctor 1} \rightarrow & m_1(x_1) = 0.99, & m_1(x_3) = 0.01, \\ \text{Doctor 2} \rightarrow & m_2(x_2) = 0.99, & m_2(x_3) = 0.01, \end{array} \tag{37}$$

while the basic belief masses of the other elements of the power set of $\mathscr{X}$ are null. From (36) one gets

$$m_{12}(x_1) = 0, \quad m_{12}(x_2) = 0, \quad m_{12}(x_3) = 1. \tag{38}$$

Hence, from direct application of the DS theory, it turns out that the patient certainly suffers from brain tumor. This result arises from the fact that the two doctors agree that the patient most likely does not suffer from tumor but are in almost full contradiction for the other causes of the disease. Since doctors' diagnoses are modelled by precise probability mass functions, also Bayesian approaches like the one in Section 1 might be applied to Zadeh's example; yet the same result is obtained.

Haenni has shown that the controversy of Zadeh's example can be overcome by assuming that the doctors are not fully reliable [17]. To take this into account, one has to build a model that includes two more variables, modelling the reliabilities of the doctors. Let $A_1 = a_1$ correspond to the statement "Doctor 1 is reliable", and $A_1 = \neg a_1$ to "Doctor 1 is unreliable"; then $p_1(a_1)$ can be interpreted as the probability that the first source is reliable, and $p_1(\neg a_1) = 1 - p_1(a_1)$ that it is not reliable, and similarly for Doctor 2. By following this idea, our aggregation framework can be applied to Zadeh's example. The doctors' diagnoses (37)

[15] We point to [1] for details about DS theory.

Table 2: Upper and lower conditional probability for the Zadeh's example for $i, j, k = 1, 2, 3$ and $i \neq j \neq k$

$$\underline{P}_0^{X|A_1,A_2}(I_{\{x_i\}}|\tilde{a}_1,\tilde{a}_2) = \frac{\underline{P}_1^{A_1|X}(I_{\{\tilde{a}_1\}}|x_i)\underline{P}_2^{A_2|X}(I_{\{\tilde{a}_2\}}|x_i)}{\underline{P}_1^{A_1|X}(I_{\{\tilde{a}_1\}}|x_i)\underline{P}_2^{A_2|X}(I_{\{\tilde{a}_2\}}|x_i) + \overline{P}_1^{A_1|X}(I_{\{\tilde{a}_1\}}|x_j)\overline{P}_2^{A_2|X}(I_{\{\tilde{a}_2\}}|x_j) + \overline{P}_1^{A_1|X}(I_{\{\tilde{a}_1\}}|x_k)\overline{P}_2^{A_2|X}(I_{\{\tilde{a}_2\}}|x_k)}$$

$$\overline{P}_0^{X|A_1,A_2}(I_{\{x_i\}}|\tilde{a}_1,\tilde{a}_2) = \frac{\overline{P}_a^{A_1|X}(I_{\{\tilde{a}_1\}}|x_i)\overline{P}_2^{A_2|X}(I_{\{\tilde{a}_2\}}|x_i)}{\overline{P}_1^{A_1|X}(I_{\{\tilde{a}_1\}}|x_i)\overline{P}_2^{A_2|X}(I_{\{\tilde{a}_2\}}|x_i) + \underline{P}_1^{A_1|X}(I_{\{\tilde{a}_1\}}|x_j)\underline{P}_2^{A_2|X}(I_{\{\tilde{a}_2\}}|x_j) + \underline{P}_1^{A_1|X}(I_{\{\tilde{a}_1\}}|x_k)\underline{P}_2^{A_2|X}(I_{\{\tilde{a}_2\}}|x_k)}$$

can be formalized as in (19) by setting $\varepsilon_1^{a_1} = 1$, $p_1(x_1|a_1) = 0.99$, $p_1(x_2|a_1) = 0$, $p_1(x_3|a_1) = 0.01$ and $\varepsilon_1^{\neg a_1} = 0$ for Doctor 1, and similarly but with $p_2(x_1|a_2) = 0$ and $p_2(x_2|a_2) = 0.99$ for Doctor 2. Notice that, by setting $\varepsilon_1^{\neg a_1} = \varepsilon_2^{\neg a_2} = 0$, it has been assumed that $\underline{P}_1^{X|\neg a_1}$ and $\underline{P}_1^{X|\neg a_2}$ are vacuous, i.e., if the doctors are unreliable they do not provide any useful information. Furthermore, let us assume $p_1(a_1) = p_2(a_2) = \delta$ with $\delta \in (0,1)$ and $\varepsilon_0 = 1$, $p_0(x_1) = p_0(x_2) = p_0(x_3) = 1/3$. Under the assumption of independence, the goal is to evaluate the posterior belief $\underline{P}_0^{X|A_1,A_2}(I_{\{\tilde{x}\}}|\tilde{a}_1,\tilde{a}_2)$, which represents the lower probability of the diagnosis $\tilde{x} \in \mathscr{X}$ conditional on the fact that the sources are in a particular state $(\tilde{a}_1,\tilde{a}_2)$. In this case, we can compute the lower probability $\underline{P}_0^{X|A_1,A_2}(I_{\{\tilde{x}\}}|\tilde{a}_1,\tilde{a}_2)$ by simply putting $g(x,\tilde{a}_1,\tilde{a}_2) = I_{\{\tilde{x}\}}$ in the equation in Table 1. The resulting conditional probabilities are shown in Table 2. For Doctor 1, the CLPs $\underline{P}_1^{A_j|X}$ for $X = x_1$ or $X = x_3$ can be derived from (34) and (35). Conversely, for $X = x_2$, since $\underline{P}_1(I_{\{x_2\}\times\mathscr{A}_1}) = 0$, the GBR cannot be applied to get $\underline{P}_1^{A_j|x_2}$ and, thus, (34) and (35) are not valid anymore. However, since

$$\begin{aligned}\overline{P}_1(I_{\{x_2\}\times\mathscr{A}_1}) &= \sum_{\tilde{a}_j\in\mathscr{A}_1} p_1(\tilde{a}_j)\cdot\Big(\varepsilon_j^{\tilde{a}_j}\sum_{x\in\mathscr{X}} p_1(x|\tilde{a}_j)I_{\{x_2\}\times\mathscr{A}_1}(x,\tilde{a}_j) \\ &+ (1-\varepsilon_j^{\tilde{a}_j})\max_{x\in\mathscr{X}} I_{\{x_2\}\times\mathscr{A}_1}(x,\tilde{a}_j)\Big) = p_1(\neg a_1) > 0,\end{aligned}$$

the regular extension (6) can be used to derive

$$\underline{P}_1^{A_j|x_2}(g|x_2) = \max_{\mu} \underline{P}\left(I_{\{x_2\}\times\mathscr{A}_1}[g-\mu]\right) \geq 0,$$

where the gambles we are interested in are only $I_{\{a_1\}}$ and $I_{\{\neg a_1\}}$. From (24),

$\underline{P}_1^{A_j|x_2}(g|x_2)$ can be calculated by finding the maximum value of μ for which

$$\begin{aligned}&\sum_{\tilde{a}_j\in\mathscr{A}_1} p_j(\tilde{a}_j)[\varepsilon_j^{\tilde{a}_j} p_j(x_2|\tilde{a}_j)+(1-\varepsilon_j^{\tilde{a}_j})I_{\mathscr{A}_1^*(\mu)}(\tilde{a}_j)]g(\tilde{a}_j)\\ &-\mu\sum_{\tilde{a}_j\in\mathscr{A}_1} p_j(\tilde{a}_j)[\varepsilon_j^{\tilde{a}_j} p_j(x_2|\tilde{a}_j)+(1-\varepsilon_j^{\tilde{a}_j})I_{\mathscr{A}_1^*(\mu)}(\tilde{a}_j)]\geq 0\end{aligned} \tag{39}$$

The values of μ which satisfy (39) in the cases $g=I_{\{a_1\}}$ and $g=I_{\{\neg a_1\}}$ are $\mu=0$ and, respectively, $\mu=1$. Hence, it follows that $\underline{P}_1^{A_j|x_2}(I_{\{a_1\}}|x_2)=\overline{P}_1^{A_j|x_2}(I_{\{a_1\}}|x_2)=0$ and $\underline{P}_1^{A_j|x_2}(I_{\{\neg a_1\}}|x_2)=\overline{P}_1^{A_j|x_2}(I_{\{\neg a_1\}}|x_2)=1$. A similar derivation can be clearly achieved for Doctor 2. Posterior lower and upper probabilities calculated for the reliability value $\delta=0.8$ are shown in Table 3. The values of the conditionals which depend on δ are highlighted in bold-face. It can be noticed that, in the case the sources are in the states $\tilde{a}_1=a_1$ and $\tilde{a}_2=a_2$, i.e., both sources are reliable, one gets the following precise conditional probability $\underline{P}_0^{X|A_1,A_2}(I_{\{x_1\}}|a_1,a_2)=\overline{P}_0^{X|A_1,A_2}(I_{\{x_1\}}|a_1,a_2)=0$, $\underline{P}_0^{X|A_1,A_2}(I_{\{x_2\}}|a_1,a_2)=\overline{P}_0^{X|A_1,A_2}(I_{\{x_2\}}|a_1,a_2)=0$, and $\underline{P}_0^{X|A_1,A_2}(I_{\{x_3\}}|a_1,a_2)=\overline{P}_0^{X|A_1,A_2}(I_{\{x_3\}}|a_1,a_2)=1$. This result holds for each value of δ and shows that, when both the sources are reliable, the answer provided in (38) by both DS and Bayesian theory is coherent with the initial assessments. In fact, since Doctor 1 says implicitly that x_2 is wrong (with almost absolute certainty), and Doctor 2 says that x_1 is wrong, it follows then that x_3 must be the true diagnosis when both doctors are reliable.

According to Table 3 it can also be noticed that when both doctors are unreliable the conditionals are vacuous for all the diseases. Conversely, if only one doctor is reliable, e.g., Doctor 1 in Table 3, the disease that he believes wrong has precisely zero probability. For $\delta>0.9$, it can be verified that $\underline{P}_0^{X|A_1,A_2}(I_{\{x_1\}}|a_1,\neg a_2)>\overline{P}_0^{X|A_1,A_2}(I_{\{x_3\}}|a_1,\neg a_2)$ and, thus, the lower probability of x_1 dominates the upper probability of the other element. In this case, the IFC can decide, without a doubt, that the patient suffers from disease x_1.

In general, in this kind of problems, the sources of information do not provide their reliability status (i.e., $\{\tilde{a}_1,\tilde{a}_2\}$). However, since the doctors' diagnoses are almost in full contradiction, the IFC can infer that at least one of the doctors must be unreliable and, thus, apply the aggregation rule by computing the following lower conditional probability $\underline{P}_0^{X|A_1,A_2}(\cdot|\mathscr{A}^2\setminus\{a_1,a_2\})$. In practice, the conditioning event is the complementary event of $\{a_1,a_2\}$, which means that at least one doctor is unreliable.

Table 3: Posterior lower and upper probabilities in the case $\delta = 0.8$

	$\underline{P}_0^{X\|A_1,A_2}(\cdot\|a_1,a_2)$	$\overline{P}_0^{X\|A_1,A_2}(\cdot\|a_1,a_2)$	$\underline{P}_0^{X\|A_1,A_2}(\cdot\|a_1,\neg a_2)$	$\overline{P}_0^{X\|A_1,A_2}(\cdot\|a_1,\neg a_2)$	$\underline{P}_0^{X\|A_1,A_2}(\cdot\|\neg a_1,\neg a_2)$	$\overline{P}_0^{X\|A_1,A_2}(\cdot\|\neg a_1,\neg a_2)$
x_1	0	0	**0.45**	1	0	1
x_2	0	0	0	0	0	1
x_3	1	1	0	**0.54**	0	1

Table 4: Upper and lower conditional probabilities conditioned on $I_{\mathscr{A}^2\setminus\{a_1,a_2\}}$ for $i = 1,2$

$$\underline{P}_0^{X|A_1,A_2}(I_{\{x_i\}}|I_{\mathscr{A}^2\setminus\{a_1,a_2\}}) = \frac{1}{3-\underline{P}^{A_1|X}(I_{\{a_1\}}|x_3)\underline{P}^{A_2|X}(I_{\{a_2\}}|x_3)}, \quad \overline{P}_0^{X|A_1,A_2}(I_{\{x_i\}}|I_{\mathscr{A}^2\setminus\{a_1,a_2\}}) = \frac{1}{2}$$

$$\underline{P}_0^{X|A_1,A_2}(I_{\{x_3\}}|I_{\mathscr{A}^2\setminus\{a_1,a_2\}}) = 0, \quad \overline{P}_0^{X|A_1,A_2}(I_{\{x_3\}}|I_{\mathscr{A}^2\setminus\{a_1,a_2\}}) = \frac{1-\underline{P}^{A_1|X}(I_{\{a_1\}}|x_3)\underline{P}^{A_2|X}(I_{\{a_2\}}|x_3)}{3-\underline{P}^{A_1|X}(I_{\{a_1\}}|x_3)\underline{P}^{A_2|X}(I_{\{a_2\}}|x_3)}$$

Since $I_{\mathscr{A}^2\setminus\{a_1,a_2\}}$ do not factorize as in (28), we cannot apply (32) to compute $\underline{P}^{A^2|X}(\cdot|x)$. However, since $\underline{P}^{A^2|X}(\cdot|x)$ is a CLP, we can exploit the following property: $\underline{P}^{A^2|X}(I_{\mathscr{A}^2\setminus\{a_1,a_2\}}|x) = 1-\overline{P}^{A^2|X}(I_{\{a_1,a_2\}}|x) = 1-\overline{P}^{A_1|X}(I_{\{a_1\}}|x)\overline{P}^{A_2|X}(I_{\{a_2\}}|x)$ and $\overline{P}^{A^2|X}(I_{\mathscr{A}^2\setminus\{a_1,a_2\}}|x) = 1-\underline{P}^{A^2|X}(I_{\{a_1,a_2\}}|x) = 1-\underline{P}^{A_1|X}(I_{\{a_1\}}|x)\underline{P}^{A_2|X}(I_{\{a_2\}}|x)$.

Since $\overline{P}^{A_1|X}(I_{\{a_1\}}|x_i)\overline{P}^{A_2|X}(I_{\{a_2\}}|x_i) = 0$ and $\underline{P}^{A_1|X}(I_{\{a_1\}}|x_i)\underline{P}^{A_2|X}(I_{\{a_2\}}|x_i) = 0$ for $i = 1,2$, and $\overline{P}^{A_1|X}(I_{\{a_1\}}|x_3)\overline{P}^{A_2|X}(I_{\{a_2\}}|x_3) = 1$, the lower and upper probabilities are those in Table 4. We can therefore note that:

$$\underline{P}_0^{X|A_1,A_2}(I_{\{x_1\}}|I_{\mathscr{A}^2\setminus\{a_1,a_2\}}) = \underline{P}_0^{X|A_1,A_2}(I_{\{x_2\}}|I_{\mathscr{A}^2\setminus\{a_1,a_2\}}) \geq \overline{P}_0^{X|A_1,A_2}(I_{\{x_3\}}|I_{\mathscr{A}^2\setminus\{a_1,a_2\}}),$$

this means that the IFC can infer that the patient suffers from x_1 or x_2 but not from x_3. It can be noticed that when the reliability δ approaches one, the lower and upper probabilities converge to the following precise probability mass function:

$$p_0^{X|A_1,A_2}(I_{\{x_1\}}|I_{\mathscr{A}^2\setminus\{a_1,a_2\}}) = p_0^{X|A_1,A_2}(I_{\{x_2\}}|I_{\mathscr{A}^2\setminus\{a_1,a_2\}}) = 1/2.$$

Summarizing, the results of this section generalize those in [17, 18] to CLPs by showing that: (i) if both the doctors are reliable the result obtained by the Bayes' and Dempster's rule in (38) is correct and coherent with the initial assessments;

(ii) if we assume that at least one of the doctors is unreliable, we obtain that the patient must suffer from either x_1 or x_2.

7. Application to Sensor Networks: the Case of Unknown Dependence

In many applications, multiple distributed sensors are used to collect measurements about entities of interest [19]. Measurements from different sensors have to be fused before useful information could be extracted. In centralized processing, all measurements from all sensors are sent to a common fusion center taking care of the aggregation. Such an architecture seems to be optimal, but it has practical disadvantages like high bandwidth to collect all the measurements into a single site, high computation load, and low survivability due to a single point of failure [19]. A distributed processing architecture consists instead of multiple processing agents, each responsible for collecting and processing measurements from some local sensors. The agents communicate indeed their local estimates to other agents in order to share information. The advantages are a reduced communication bandwidth, distribution of processing load, and improved survivability.

An important issue in distributed estimation is how to handle the dependence in the estimates to be fused. This dependence may be due to common information from previous communication (e.g., some measurements may be used several times) or hidden variables affecting the measurements. Several approaches (e.g., [19]) have been considered to perform distributed estimation, each handling dependence differently.

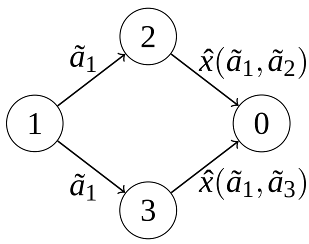

Figure 2: A sensor network.

In order to better explain the problem let us consider a practical example, which is illustrated in Figure 2. Node 1 transmits its measurement $\tilde{a}_1$ to both node 2 and node 3. These nodes fuse their measurements, respectively $\tilde{a}_2$ and $\tilde{a}_3$, with the received information $\tilde{a}_1$, and then they transmit their estimates, respectively $\hat{x}(\tilde{a}_1,\tilde{a}_2)$ and $\hat{x}(\tilde{a}_1,\tilde{a}_3)$, to node 0. Finally, node 0 fuses both estimates in order to obtain an aggregated estimate. As both the estimates $\hat{x}(\tilde{a}_1,\tilde{a}_2)$ and $\hat{x}(\tilde{a}_1,\tilde{a}_3)$ contain the same information from node 1, they are dependent. On the other side, we assume that node 0 does not know the network architecture and, thus, it cannot know this correlation. For this reason, node 0 can be too confident about the quality of the estimate (and underestimate the corresponding variance).

Bayesian case

We consider the problem in the Bayesian (i.e., precise) framework. The variables X, A_1, A_2 and A_3 are all assumed to be real, while for the corresponding probability distributions we have $p_j(a_j|x) = \mathcal{N}(a_j; x, \sigma_m^2)$, $p_j(x) = \mathcal{N}(x; \hat{x}_j, \sigma_j^2)$, and $p_0(x) = \mathcal{N}(x; \hat{x}_0, \sigma_0^2)$, where $\mathcal{N}(\cdot; \mu, \sigma^2)$ is a Gaussian distribution with mean μ and variance σ^2.

By assuming A_1, A_2 and A_3 conditionally independent given X, we can use the aggregation scheme of Figure 1 to fuse node 1 and node 2, with node 2 playing the role of the IFC (and similarly proceed for node 3). Since node 1 sends the measurement to node 2 (node 3), we can use the scheme of Figure 1 starting from step (b). Then, we have $p_j(x|\tilde{a}_1,\tilde{a}_j) \propto p_1(\tilde{a}_1|x)p_j(\tilde{a}_j|x)p_j(x)$, with $j = 2, 3$. Because of the Gaussian assumptions, it follows that $p_j(x|\tilde{a}_1,\tilde{a}_j) = \mathcal{N}(x;\ \hat{x}(\tilde{a}_1,\tilde{a}_j),\ \sigma^2(\tilde{a}_1,\tilde{a}_j))$ with

$$
\begin{aligned}
\hat{x}(\tilde{a}_1,\tilde{a}_j) &= \tfrac{1}{3}(\hat{x}_j + \tilde{a}_1 + \tilde{a}_j), \\
\sigma^2(\tilde{a}_1,\tilde{a}_j) &= \tfrac{1}{3},
\end{aligned}
\tag{40}
$$

where, just for sake of simplicity, $\sigma_j = \sigma_m = 1$. Now we can use the aggregation scheme of Figure 1 to combine the estimates of nodes 2 and 3 at the level of node 0, which now plays the role of the IFC, and conclude:

$$
\begin{aligned}
\hat{x}_{fin} &= \tfrac{1}{5}(\hat{x}_0 + \tilde{a}_1 + \tilde{a}_1 + \tilde{a}_2 + \tilde{a}_3), \\
\sigma^2_{fin} &= \tfrac{1}{5}.
\end{aligned}
\tag{41}
$$

Since in the scheme of Figure 1 we assumed that sources share posteriors and priors with node 0, node 0 does not know the value of the single measurements but it can only derive the value of their sum, i.e., $\tilde{a}_{1j} = \tilde{a}_1 + \tilde{a}_j$ with $j = 2, 3$. Hence, it is not able to detect that the same measurement $\tilde{a}_1$ is used two times in (41). Conversely, if the sources would directly share their measurements with node 0, it would compute the true fused estimate:

$$
\begin{aligned}
\hat{x}_{true} &= \tfrac{1}{4}(\hat{x}_0 + \tilde{a}_1 + \tilde{a}_2 + \tilde{a}_3), \\
\sigma^2_{true} &= \tfrac{1}{4}.
\end{aligned}
\tag{42}
$$

Summarizing, if we assume independence, a consequence of the unknown correlation at the level of node 0 is that the measurement $\tilde{a}_1$ is used two times and, thus, that $\sigma^2_{fin} \leq \sigma^2_{true}$.[16]

Imprecise case

Here we show how it is possible to gain robustness by solving the fusion problem in the framework of CLPs. As discussed in Section 3 we consider the case in which the CLP in is a convex combination of the two CLPs that we obtain by independent natural extension and, respectively, natural extension. These two CLPs correspond to the extreme cases: (i) variables are assumed to be epistemic independent; (ii) no assumption is made about the possible independence among variables. An alternative approach, still based on set of distributions, to gain robustness for this kind of problem is presented in [20].

In our case, the CLP to be combined are the linear previsions $P^{A_{12}|X}$ and $P^{A_{13}|X}$, where $A_{1j} = A_1 + A_j$. Our goal is to compute their conditional natural extension and their independent natural extension to the space $\mathscr{A}_{12} \times \mathscr{A}_{13}$ for each $x \in \mathscr{X}$. Actually, since our final goal is to apply GBR, we are only interested to compute their extensions to the gamble $I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}$, i.e. $\underline{P}^{A_{12},A_{13}|X}_{INE}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x)$ and $\underline{P}^{A_{12},A_{13}|X}_{NE}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x)$.

Unfortunately, in the continuous case probabilities that random variables assume a particular value are zero, as we have $\underline{P}^{A_{12},A_{13},X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}) = \overline{P}^{A_{12},A_{13},X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}) = 0$. Strictly speaking this would imply that neither GBR nor regular extension can be used to determine the final conditional CLP. This problem might be overcome by replacing the measurement a_j with the set $B(a_j,\delta_j) \subset \mathscr{A}_j$, where $\{B(a_j,\delta_j)\}_{\delta_j \in \mathbb{R}}$ are nested neighbourhoods of a_j with positive probability and converging to $\{a_j\}$ as their radius $\delta_j > 0$ decreases to zero [4, Section 6.10]. Coping with these interval-valued measurements makes also sense in practice because of the finite precision of the instruments. Having these ideas in mind, we may assume a_j to be in fact a representation of $B(a_j,\delta_j)$, and hence $\underline{P}^{A_{12},A_{13},X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}) > 0$. This allows us to apply GBR and then solve (16). For δ_j sufficiently small, $p_j(a_j|x)$ can be approximated with

[16]Notice that, sharing measurements is optimal under the estimation point of view but it requires more information to be transmitted among the sources, i.e., three quantities $\tilde{a}_1, \tilde{a}_2, \tilde{a}_3$ should be transmitted instead of two $\hat{x}(\tilde{a}_1, \tilde{a}_2)$ and $\hat{x}(\tilde{a}_1, \tilde{a}_3)$. In large networks, reducing the information to be transmitted is a main requirement. This is one of the reasons why we have chosen the aggregation scheme of Figure 1 in which sources share posteriors (i.e., estimates) instead of measurements.

$\rho(\delta_j)\mathscr{N}(a_j;x,\sigma_m^2)$, where $\rho(\delta_j) > 0$ is the Lebesgue measure of $B(a_j,\delta_j)$, which has been assumed independent of a_j. See [15] for further details about this discretization.[17] Under this hypothesis we can therefore evaluate $P^{A_{12}|X}$ and $P^{A_{13}|X}$, which are clearly linear previsions. For linear prevision the epistemic irrelevance implies standard stochastic independence, and for the independent natural extension we can therefore write:

$$
\begin{aligned}
P_{INE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) &= P^{A_{12}|X}(I_{\{\tilde{a}_{12}\}}|x)P^{A_{13}|X}(I_{\{\tilde{a}_{13}\}}|x) \\
&= \rho(\delta_2)\mathscr{N}(\tilde{a}_{12}/2;x,\sigma_m^2/2)\rho(\delta_3)\mathscr{N}(\tilde{a}_{13}/2;x,\sigma_m^2/2) \\
&= \frac{\rho(\delta_2)\mathscr{N}(x;\hat{x}(\tilde{a}_1,\tilde{a}_2),\sigma^2(\tilde{a}_1,\tilde{a}_2))p(\tilde{a}_{12})}{\mathscr{N}(x;\hat{x}_2,\sigma_2^2)} \\
&\quad \frac{\rho(\delta_3)\mathscr{N}(x;\hat{x}(\tilde{a}_1,\tilde{a}_3),\sigma^2(\tilde{a}_1,\tilde{a}_3))p(\tilde{a}_{13})}{\mathscr{N}(x;\hat{x}_3,\sigma_3^2)}.
\end{aligned}
\tag{43}
$$

Conversely, for the natural extension, we have shown in Appendix A that:

$$
\begin{aligned}
\underline{P}_{NE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) &= \max\left(0, P^{A_{12}|X}(I_{\{\tilde{a}_{12}\}}|x) + P^{A_{13}|X}(I_{\{\tilde{a}_{13}\}}|x) - 1\right) \\
&= \max\left(0, \rho(\delta_2)\mathscr{N}(\tilde{a}_{12}/2;x,\sigma_m^2/2)\right. \\
&+ \left.\rho(\delta_3)\mathscr{N}(\tilde{a}_{13}/2;x,\sigma_m^2/2) - 1\right) \\
&= \max\left(0, \rho(\delta_2)\frac{\mathscr{N}(x;\hat{x}(\tilde{a}_1,\tilde{a}_2),\sigma^2(\tilde{a}_1,\tilde{a}_2))p(\tilde{a}_{12})}{\mathscr{N}(x;\hat{x}_2,\sigma_2^2)}\right. \\
&+ \left.\rho(\delta_3)\frac{\mathscr{N}(x;\hat{x}(\tilde{a}_1,\tilde{a}_3),\sigma^2(\tilde{a}_1,\tilde{a}_3))p(\tilde{a}_{13})}{\mathscr{N}(x;\hat{x}_3,\sigma_3^2)} - 1\right)
\end{aligned}
\tag{44}
$$

[17] The fact that the variable X is continuous is not a problem for the subsequent derivations, as discussed at the end of Section 3. In fact, by using the results in [15], Theorem 1 continues to hold also for continuous variables. Conversely, for X continuous the results in [15] cannot be used to prove coherence at the sources level (see Section 3) for general CLPs. However, in this example, since we are dealing with linear previsions at the sources level, we can use the results of standard probability to assess coherence also at this level.

$$
\begin{aligned}
\overline{P}_{NE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) &= \min\Big(P^{A_{12}|X}(I_{\{\tilde{a}_{12}\}}|x), P^{A_{13}|X}(I_{\{\tilde{a}_{13}\}}|x)\Big) \\
&= \min\Bigg(\rho(\delta_2)\frac{\mathscr{N}(x;\hat{x}(\tilde{a}_1,\tilde{a}_2),\sigma^2(\tilde{a}_1,\tilde{a}_2))p(\tilde{a}_{12})}{\mathscr{N}(x;\hat{x}_2,\sigma_2^2)}, \\
&\quad \rho(\delta_3)\frac{\mathscr{N}(x;\hat{x}(\tilde{a}_1,\tilde{a}_3),\sigma^2(\tilde{a}_1,\tilde{a}_3))p(\tilde{a}_{13})}{\mathscr{N}(x;\hat{x}_3,\sigma_3^2)}\Bigg)
\end{aligned} \tag{45}
$$

Let us report some remarks.

- The lower and upper previsions in (44) and (45) coincide with the lower and upper Fréchet bounds [21].
- Since $\rho(\delta_j)$ are small numbers, we might assume $\rho(\delta_j) < \sigma_m$, in (44) the second term in the maximization always negative. This implies that $\underline{P}_{NE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) = 0$.

Thus, following the procedure indicated in (14), our final CLP is:

$$
\begin{aligned}
\underline{P}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) &= \gamma P_{INE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) + (1-\gamma)\underline{P}_{NE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) \\
&= \gamma P_{INE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x)
\end{aligned} \tag{46}
$$

and

$$
\overline{P}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) = \gamma P_{INE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) + (1-\gamma)\overline{P}_{NE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x). \tag{47}
$$

Thus, by using the results in (30)–(32), since in this case the prior CLP of node 0 is a linear prevision, i.e., $\varepsilon_0 = 1$, we obtain that $\underline{P}_0(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}(g-\mu)) = 0$ reduces to:

$$
\begin{aligned}
0 &= \int_{x\in\mathscr{X}} (g(x)-\mu)\mathscr{N}(x;\hat{x}_0,\sigma_0^2)\Big[I_{\{g(x)-\mu\geq 0\}}\underline{P}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) \\
&+ I_{\{g(x)-\mu<0\}}\overline{P}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x)\Big]\, dx \\
&= \int_{x\in\mathscr{X}} (g(x)-\mu)\mathscr{N}(x;\hat{x}_0,\sigma_0^2)\Bigg[\gamma\mathscr{N}(\tilde{a}_{12};x,2\sigma_m^2)\mathscr{N}(\tilde{a}_{13};x,2\sigma_m^2) \\
&+ (1-\gamma)I_{\{g(x)-\mu<0\}}\min\big(\mathscr{N}(\tilde{a}_{12};x,2\sigma_m^2),\mathscr{N}(\tilde{a}_{13};x,2\sigma_m^2)\big)\Bigg]\, dx,
\end{aligned} \tag{48}
$$

where we have exploited the fact that

$$
\gamma[I_{\{g(x)-\mu\geq 0\}} + I_{\{g(x)-\mu<0\}}]P_{INE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) = \gamma P_{INE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x),
$$

γ	$\lvert\chi\rvert$	%
1	1.2	49
0.9999	2.51	85
0.999	2.56	87
0.9	2.71	89
0.3	3.47	97
0.1	3.89	98

Table 5: Simulation results.

and assumed $\rho(\delta_2) = \rho(\delta_3)$. By solving (48) w.r.t μ, at the end we obtain $\mu = \underline{P}_0(g|\tilde{a}_{12}, \tilde{a}_{13})$.

Simulations

Monte Carlo (MC) simulations have been performed in order to show the robustness of the proposed solution for different values of γ. As a performance metric, we have evaluated how many times in the MC simulations the true value of X was included in the following "robust" 99% credibility region, i.e, the minimum volume region χ such that $\underline{P}(I_{\{x\in\chi\}}|\tilde{a}_{12}, \tilde{a}_{13}) \geq 0.99$.

The simulations results for 1000 MC runs, performed w.r.t the realisations of the measurements and the variable X, are shown in Table 5 for different values of γ and for $\hat{x}_0 = 0$, $\sigma_0 = \sigma_m = 1$. In particular, Table 5 reports the average volume of the 99% credibility region, i.e., $|\mathscr{X}|$, and the percentage of cases in which the true value of X is included in the credibility region for different values of γ. Notice that, for $\gamma = 1$, $\underline{P}_0(g|\tilde{a}_{12}, \tilde{a}_{13})$ reduces to (41), which is taken as performance reference. From Table 5, it can be seen that for $\gamma < 1$ the proposed solution guarantees robustness without increasing too much the volume of the credibility region.

8. Conclusions and Outlooks

A general aggregation framework for coherent lower previsions defined on a common domain has been proposed. This is achieved by a simultaneous *model revision* of beliefs associated to different sources of information. The coherence of the aggregated beliefs is also discussed. Furthermore, in the particular case of linear-vacuous mixtures and epistemic independence among the sources, a closed formula for the aggregated beliefs has been derived. In this context, as an example of application of this approach, Zadeh's paradox is treated and an explanation based on our aggregation framework is achieved. The proposed approach, in its general form, has also been applied to a problem of estimation fusion under unknown correlation for sensor networks.

As a future work, we aim to study the possibility of a recursive application of our approach. In particular, for sensor networks, we plan to apply the algorithm in [15] to develop a recursive rule for estimation fusion in the case of unknown correlation. We also want to investigate the relationships between our approach in the case of a single source and Jeffrey's updating. Finally, we intend to apply our rule to practical problems of information fusion in signal and data processing and communications.

Appendix A. Derivation of Equation (44)

In order to compute $\underline{P}_{NE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x)$ we can make a coarsening of the possibility spaces $\mathscr{A}_{12}$ and $\mathscr{A}_{13}$, which are reduced to spaces of cardinality two: $\mathscr{A}'_{1j} = \{\tilde{a}_{1j}, \neg\tilde{a}_{1j}\}$ for $j = 2,3$, with $\neg\tilde{a}_{1j}$ complement of $\tilde{a}_{1j}$ with respect to $\mathscr{A}_j$. Hence, we induce $\underline{P}_{NE}^{A'_{12},A'_{13}|X}$ from $\underline{P}_{NE}^{A_{12},A_{13}|X}$ by natural extension. The coherence of $\underline{P}_{NE}^{A_{12},A_{13}|X}$ implies that $\underline{P}_{NE}^{A_{12},A_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) = \underline{P}_{NE}^{A'_{12},A'_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x)$. Let $\mathscr{M}$ be the set of dominating linear previsions of $\underline{P}_{NE}^{A'_{12},A'_{13}|X}$. Then, by definition of CLP [4, Section 2.8], it follows that

$$\underline{P}_{NE}^{A'_{12},A'_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) = \min_{P\in\mathscr{M}} P^{A'_{12},A'_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x)$$

where

$$\begin{aligned} P^{A'_{12},A'_{13}|X}(I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}|x) &= \sum_{i\in\{\tilde{a}_{12},\neg\tilde{a}_{12}\}} \sum_{j\in\{\tilde{a}_{13},\neg\tilde{a}_{13}\}} I_{\{\tilde{a}_{12},\tilde{a}_{13}\}}(i,j)p(i,j) \\ &= p(\tilde{a}_{12},\tilde{a}_{13}) \end{aligned}$$

and $p(\cdot)$ is the probability mass function associated to $P(\cdot)$. In [22], it is shown that natural extension can be equivalently formulated as an optimization problem (primal form) which involves distributions. By exploiting this result, it can be verified that the set of linear previsions $\mathscr{M}$ that we obtain by applying natural extension (9) to $P^{A_{12}|X}$ and $P^{A_{13}|X}$ is equal to

$$\mathscr{M} = \begin{cases} \sum_{i \in \{\tilde{a}_{12}, \neg\tilde{a}_{12}\}} \sum_{j \in \{\tilde{a}_{13}, \neg\tilde{a}_{13}\}} I_{\{\tilde{a}_{12}\}}(i) p(i,j) = P^{A_{12}|X}(I_{\{\tilde{a}_{12}\}}|x) \\ \sum_{i \in \{\tilde{a}_{12}, \neg\tilde{a}_{12}\}} \sum_{j \in \{\tilde{a}_{13}, \neg\tilde{a}_{13}\}} I_{\{\tilde{a}_{13}\}}(j) p(i,j) = P^{A_{13}|X}(I_{\{\tilde{a}_{13}\}}|x) \\ \sum_{i \in \{\tilde{a}_{12}, \neg\tilde{a}_{12}\}} \sum_{j \in \{\tilde{a}_{13}, \neg\tilde{a}_{13}\}} p(i,j) = 1 \\ p(i,j) \geq 0 \text{ for } i \in \{\tilde{a}_{12}, \neg\tilde{a}_{12}\}, j \in \{\tilde{a}_{13}, \neg\tilde{a}_{13}\} \end{cases}$$

Thus, $\mathscr{M}$ is just the set of linear previsions in $\mathscr{A}'_{12} \times \mathscr{A}'_{13}$ whose marginal distributions in $\mathscr{A}'_{12}$ and $\mathscr{A}'_{13}$ are $P^{A'_{12}|X} = P^{A_{12}|X}$ and, respectively, $P^{A'_{13}|X} = P^{A_{13}|X}$. This is the minimum set of constraints which must be satisfied by $P_{NE}^{A'_{12},A'_{13}|X}$ to be coherent with $P^{A_{12}|X}$ and $P^{A_{13}|X}$. In fact, when no independence assumptions are made, the only requirement for coherence is that the marginal on A_{12} and A_{13} must be equal to $P^{A_{12}|X}$ and, respectively, $P^{A_{13}|X}$. From the above definition of $\mathscr{M}$, after some algebraic manipulation it can be derived:

$$p(\tilde{a}_{12}, \tilde{a}_{13}) = P^{A_{12}|X}(I_{\{\tilde{a}_{12}\}}|x) + P^{A_{13}|X}(I_{\{\tilde{a}_{13}\}}|x) - 1 + p(\neg\tilde{a}_{12}, \neg\tilde{a}_{13})$$

Hence, being $p(\neg\tilde{a}_{12}, \neg\tilde{a}_{13}) \geq 0$, if $P^{A_{12}|X}(I_{\{\tilde{a}_{12}\}}|x) + P^{A_{13}|X}(I_{\{\tilde{a}_{13}\}}|x) - 1 > 0$ then $P^{A_{12}|X}(I_{\{\tilde{a}_{12}\}}|x) + P^{A_{13}|X}(I_{\{\tilde{a}_{13}\}}|x) - 1$ is the minimum value of $p(\tilde{a}_{12}, \tilde{a}_{13})$ which satisfies the constraints in $\mathscr{M}$. Conversely, if $P^{A_{12}|X}(I_{\{\tilde{a}_{12}\}}|x) + P^{A_{13}|X}(I_{\{\tilde{a}_{13}\}}|x) - 1 \leq 0$, the minimum value of $p(\tilde{a}_{12}, \tilde{a}_{13})$ is clearly 0.

Acknowledgments

This research has been partially supported by the Swiss NSF grants n. 200020-121785/1.

[1] G. Shafer, A mathematical theory of evidence, Princeton University Press, 1976.

[2] D. Dubois, H. Prade, Possibility theory, Plenum Press, New York, 1988.

[3] P. Walley, Measures of uncertainty in expert systems, Artificial Intelligence 83 (1) (1996) 1–58.

[4] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, New York, 1991.

[5] Y. Bar-Shalom, Multitarget-Multisensor Tracking: Advanced Applications., Artech-House, Norwood, MA, 1990.

[6] L. A. Zadeh, On the validity of Dempster rule of combination, in: Memo M 79/24, University of California, Berkeley, 1979, pp. 3–28.

[7] E. Miranda, A survey of the theory of coherent lower previsions, Int. J. Approx. Reasoning 48 (2) (2008) 628–658.

[8] M. Troffaes, G. de Cooman, Extension of coherent lower previsions to unbounded random variables, in: Ninth International Conference IPMU 2002, Anney, France, 2002, pp. 735–742.

[9] G. de Cooman, M. Troffaes, Coherent lower previsions in systems modelling: products and aggregation rules, Reliability Engineering and System Safety 85 (2004) 113–134.

[10] E. Miranda, G. de Cooman, Coherence and independence in non-linear spaces, Technical reports of statistics and decision sciences, Rey Juan Carlos University (2005).

[11] E. Miranda, Updating coherent previsions on finite spaces, Fuzzy Sets and Systems 160 (9) (2009) 1286–1307.

[12] M. Zaffalon, E. Miranda, Conservative inference rule for uncertain reasoning under incompleteness, Journal of Artificial Intelligence Research 34 (2009) 757–821.

[13] M. Zaffalon, The naive credal classifier, Journal of Statistical Planning and Inference 105 (1) (2002) 5–21.

[14] G. de Cooman, F. Hermans, A. Antonucci, M. Zaffalon, Epistemic irrelevance in credal networks: the case of imprecise Markov trees, Int. J. Approx. Reasoning.Accepted for publication.

[15] A. Benavoli, M. Zaffalon, E. Miranda, Reliable hidden Markov model filtering through coherent lower previsions, in: Proc. 12th Int. Conf. Information Fusion, Seattle (USA), 2009, pp. 1743–1750.

[16] A. Piatti, M. Zaffalon, F. Trojani, M. Hutter, Limits of learning about a categorical latent variable under prior near-ignorance, International Journal of Approximate Reasoning 50 (2009) 597–611.

[17] R. Haenni, Shedding new light on Zadeh's criticism of Dempster's rule of combination, in: Proc. 8th Int. Conf. Information Fusion, contribution No. C8-1, Philadelphia, USA, 2005.

[18] S. Arnborg, Robust Bayesianism: Imprecise and paradoxical reasoning, in: Proceedings of the Seventh International Conference on Information Fusion, Vol. I, International Society of Information Fusion, 2004, pp. 407–414.

[19] C. Chong, S. Mori, Graphical models for nonlinear distributed estimation, in: the Seventh International Conference on Information Fusion, 2004, pp. 206–213.

[20] K. Vesa, Bayesian estimation with uncertain parameters of probability density functions, in: Proc. 12th Int. Conf. Information Fusion, Seattle (USA), 2009, pp. 587–594.

[21] H. Joe, Multivariate models and dependence concepts: Monographs on Statistics and Applied Probability, 73, Chapmann & Hall, London, 1997.

[22] L. Utkin, I. Kozine, Different faces of the natural extension, in: Imprecise Probabilities and Their Applications. Proc. of the 2nd Int. Symposium ISIPTA, Vol. 1, 2001, pp. 316–323.