Aggregating Imprecise Probabilistic Knowledge

Alessio Benavoli
IDSIA
Lugano, Switzerland
alessio@idsia.ch

Alessandro Antonucci
IDSIA
Lugano, Switzerland
alessandro@idsia.ch

Abstract

The problem of aggregating two or more sources of information containing knowledge about a same domain is considered. We propose an aggregation rule for the case where the available information is modeled 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. Finally, an alternative explanation of Zadeh's paradox is provided.

Keywords. Information fusion, coherent lower previsions, independent natural extension, generalized Bayes 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, that 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 Choquet capacities [3], belief functions [7], possibility measures [6], and fuzzy measures [15]. As shown in [14], all these models represent uncertainty through sets instead of single probability measures, and can be all regarded as special cases of Walley's *coherent lower previsions* [13]. This theory, which is usually referred to as *imprecise probability*, provides a very general model of uncertain knowledge, for which also some rationality criteria, that can be used to identify conflicts among the different sources and determine whether the model is self-consistent, are provided. All these features seem to be particularly suited for the aggregation of different sources of information, that might 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 method of aggregation for uncertain information coming from different sources. In order to describe this aggregation task, let us us first formalize 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 modeled 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 modeled 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, modeled 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:

$$p_0(a_j|x) := p_j(a_j|x) = \frac{p_j(x|a_j)p_j(a_j)}{\sum_{a_j\in\mathscr{A}_j} p_j(x|a_j)p_j(a_j)}, \qquad (1)$$

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 the variables in $(A_1,\ldots,A_n)$ given X, we can aggregate those beliefs into the following joint:

$$p_0(x,a_1,\ldots,a_n) = \prod_{j=1}^{n} \frac{p_j(x|a_j)p_j(a_j)}{p_j(x)} p_0(x) \ , \qquad (2)$$

[1] Variables are denoted in this paper by uppercase letters; the corresponding calligraphic and lowercase letters denote respectively their sets of possible values and the generic values of these sets.

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 is:

$$p_0(x|\tilde{a}_1,\ldots,\tilde{a}_n) \propto \prod_{j=1}^{n} \frac{p_j(x|\tilde{a}_j)}{p_j(x)} p_0(x) \ , \qquad (3)$$

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 also that the prior over the internal states $p_j(A_j)$ has been dropped from (3) because of normalization.

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, that 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).

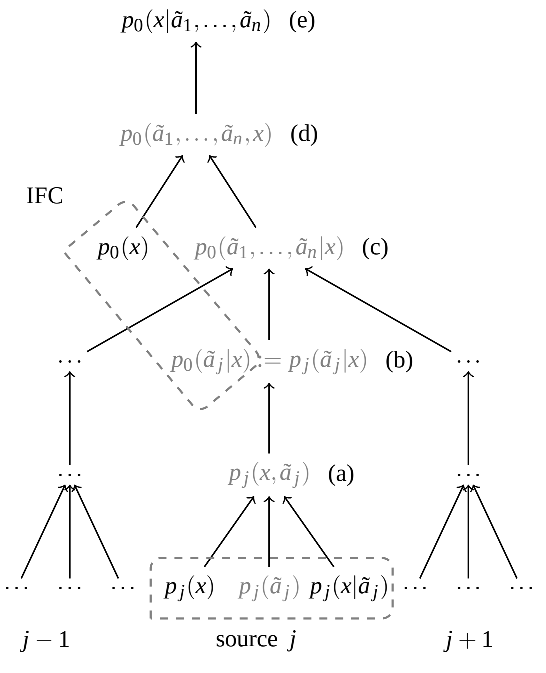

Figure 1: Aggregation of the sources of information in the Bayesian framework. The black-highlighted text describes the information used by the IFC to compute the final posterior density (still in black). The gray-highlighted text denotes the intermediate steps needed to aggregate the information. The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

In this 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 [2, Chapter 8]).

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 a combination rule for the general case of coherent lower previsions. The consistency between the obtained results and the original assessment is discussed in Section 4. The rule is indeed specialized in Section 5 for a special class of coherent lower previsions, called *linear-vacuous mixtures*. Finally, in Section 6, we show how this rule can be applied in practice for a possible explanation of Zadeh's paradox [16]. Conclusions and outlooks for future developments are in Section 7.

2 Coherent Lower Previsions

The *imprecise probability* theory [13] 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*.

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 test 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 modeled by the willingness to accept certain gambles and to reject others. Bayesian theory assumes that subjects are always able to provide a 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 a 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 $x \in \mathscr{X}'$, 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 [13, 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, $P(f) = \sum_{x\in\mathscr{X}} p(x) f(x)$. 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 linear and vacuous previsions are coherent, we can construct new coherent lower previsions by convex combination of the two [13, 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 a convex set of probability distributions [13].

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 [13, 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). GBR cannot be applied if $\underline{P}(I_{\mathscr{X}\times\{a\}}) = 0$. Nevertheless, if $\overline{P}(I_{\mathscr{X}\times\{a\}}) > 0$, a conditional prevision $\underline{P}^{X|A}(\cdot|a)$ can be computed by the following *regular extension*

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

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 $\underline{P}^A$ [13, 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 $\underline{P}$

[2] Although Walley's theory has been developed for bounded gambles only, an extension to the unbounded case can be found in [12].

[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.

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 $\underline{P}_j^{A_j|X}$, under the assumption that, for each $i,j=1,\ldots,n$ with $i\neq j$, A_i and A_j are epistemically independent given X, is defined as follows:[4]

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

This is the *independent natural extension* [5][5]. 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. For this case, 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).*

A more general formulation of Theorem 1 and its proof can be found in [9].

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 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 $j=1,\ldots,n$. 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 $(\tilde{a}_1,\ldots,\tilde{a}_n)$ 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

$$\underline{P}_j(f_j):=\underline{P}_j^{A_j}\left(\underline{P}_j^{X|A_j}(f_j|A_j)\right), \tag{10}$$

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

(b) GBR 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

$$\underline{P}_j\left(I_{\{\tilde{x}\}}\cdot[f_j-\mu]\right)=0, \tag{11}$$

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 $j=1,\ldots,n$.

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.,

$$P_0^{A_j|X}(f_j|x):=P_j^{A_j|X}(f_j|x), \tag{12}$$

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

(c) A conditional CLP $\underline{P}_0^{A^n|X}$ is obtained from $\underline{P}_0^{A_j|X}$ by independent natural extension (9):

$$\underline{P}_0^{A^n|X}(g|x)=\sup_{\substack{g_j\in\mathscr{L}(\mathscr{A}_j)\\ j=1,\ldots,n}}\ \inf_{\substack{a_j\in\mathscr{A}_j\\ j=1,\ldots,n}}\Big\{g(a_1,\ldots,a_n)-\sum_{j=1}^n\Big[g_j(a_1,\ldots,a_n)-\underline{P}_0^{A_j|X}(g_j(a_1,\ldots,a_{j-1},\cdot,a_{j+1}\ldots,a_n)|x)\Big]\Big\}. \tag{13}$$

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

$$\underline{P}_0(g):=\underline{P}_0^X\left(\underline{P}_0^{A^n|X}(g|X)\right), \tag{14}$$

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

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

[5] This paper includes a survey of different aggregation rules for CLPs. Yet, our approach differs in aggregating knowledge referred to the same domain.

[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.

(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,

$$\underline{P}_0\left(I_{\{\tilde{a}_1,\ldots,\tilde{a}_n\}}\cdot[g-\mu]\right)=0, \tag{15}$$

to compute the separately coherent conditional lower prevision $\underline{P}_0^{X|A^n}(\cdot|A^n)$ on $\mathscr{L}(\mathscr{X})$.[7]

The above derivation has been achieved by complete analogy with that in Section 1, but in the more general framework of CLPs. Notice 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 this case, the posterior probabilities coincide with those returned by a *naive credal classifier* (e.g., compare the equation in Table 2 with the results in [17]). This holds in spite of the different notion of independence assumed in [17], and can be verified by means of the algorithm in [4].

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 (12), while the contrary cannot take place because of the way the sources and the IFC share the information. 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 (14), which is the basis to prove the following result.

Theorem 2. *The separately coherent conditional lower previsions $\underline{P}_0^{A_j|X}$ in (12) 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 $\underline{P}^{X,A^n}$ in (14). As a consequence of marginal extension, $\underline{P}^{X,A^n}$ is jointly coherent with both $\underline{P}_0^X$ and $\underline{P}^{A^n|X}(\cdot|x)$. Furthermore, as a consequence of independent natural extension, $\underline{P}^{A^n|X}(\cdot|x)$ is jointly coherent with all the $\underline{P}_j^{A_j|X}(\cdot|x)$ for $j=1,\ldots,n$, because of the epistemic independence between the variables in $(A_1,\ldots,A_n)$ given X. Finally, assuming $\underline{P}_j^{A^n}(I_{\{a_1,\ldots,a_n\}})>0$ because of GBR, the coherence of $\underline{P}^{X|A^n}(\cdot|a_1,\ldots,a_n)$ follows from Theorem 1. □

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 independencies between the sources, (2) rewrites as:

$$p_0(x,a_1,\ldots,a_n) \quad = \quad \prod_{j=1}^{n}\frac{p_0(x|a_j)p_0(a_j)}{p_0(x)}p_0(x) \ . \tag{16}$$

By comparing (16) 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 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 have the same prior over X.

[7] 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.

5 Mathematical Derivation for Linear-Vacuous Mixtures

Let us detail the derivation described in Section 3 in the special case where 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.[8] This corresponds to the following settings:

$$\underline{P}_0^X(h) := \varepsilon_0 \sum_{x\in\mathscr{X}} p_0(x)h(x) + (1-\varepsilon_0)\min_{x\in\mathscr{X}} h(x),$$

$$\underline{P}_j^{X|A_j}(f_j|a_j) := \varepsilon_j^{a_j} \sum_{x\in\mathscr{X}} p_j(x|a_j) f_j(x,a_j) + (1-\varepsilon_j^{a_j})\min_{x\in\mathscr{X}} f_j(x,a_j), \quad \forall a_j \in \mathscr{A}_j \tag{17}$$

$$P_j^{A_j}(g_j) := \sum_{a_j\in\mathscr{A}_j} p_j(a_j) g_j(a_j), \tag{18}$$

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, (10) rewrites as

$$\underline{P}_j(f_j) = \sum_{a_j\in\mathscr{A}_j} p_j(a_j)\cdot\Big(\varepsilon_j^{a_j}\sum_{x\in\mathscr{X}} p_j(x|a_j)\cdot f_j(x,a_j) + (1-\varepsilon_j^{a_j})\min_{x\in\mathscr{X}} f_j(x,a_j)\Big), \tag{19}$$

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

(b) Thus, for each $\tilde{x}\in\mathscr{X}$, (11) becomes:

$$\sum_{a_j\in\mathscr{A}_j} p_j(a_j)\cdot\Big(\varepsilon_j^{a_j}[f_j(\tilde{x},a_j)-\mu]p_j(\tilde{x}|a_j) + (1-\varepsilon_j^{a_j})\min\{0, f_j(\tilde{x},a_j)-\mu\}\Big) = 0. \tag{20}$$

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

$$\mathscr{A}_j^*(\mu) := \{a_j \in \mathscr{A}_j :\ f_j(\tilde{x},a_j)-\mu < 0\}, \tag{21}$$

where $f_j, \tilde{x}$ are omitted from the arguments of $\mathscr{A}_j^*$ for sake of simpler notation. Equation (20) rewrites as:

$$\sum_{a_j\in\mathscr{A}_j} p_j(a_j)[\varepsilon_j^{a_j} p_j(\tilde{x}|a_j) + (1-\varepsilon_j^{a_j}) I_{\mathscr{A}_j^*(\mu)}(a_j)] f_j(\tilde{x},a_j) - \mu \sum_{a_j\in\mathscr{A}_j} p_j(a_j)[\varepsilon_j^{a_j} p_j(\tilde{x}|a_j) + (1-\varepsilon_j^{a_j}) I_{\mathscr{A}_j^*(\mu)}(a_j)] = 0. \tag{22}$$

The solution of (20) 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 (20), 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 (22) is:[9]

$$\frac{\sum_{a_j\in\mathscr{A}_j} p_j(a_j)[\varepsilon_j^{a_j} p_j(\tilde{x}|a_j) + (1-\varepsilon_j^{a_j}) I_{\mathscr{A}_j^*}(a_j)] f_j(\tilde{x},a_j)}{\sum_{a_j\in\mathscr{A}_j} p_j(a_j)[\varepsilon_j^{a_j} p_j(\tilde{x}|a_j) + (1-\varepsilon_j^{a_j}) I_{\mathscr{A}_j^*}(a_j)]}. \tag{23}$$

(c) The (separately coherent) conditional lower previsions associated to the sources and defined as in (23) induce the following *model revision* into the IFC's beliefs,

$$P_0^{A_j|X}(f_j|x) := P_j^{A_j|X}(f_j|x), \tag{24}$$

for each $f_j \in \mathscr{L}(\mathscr{A}_j)$, $j = 1,\ldots,n$ and $x \in \mathscr{X}$. Their independent natural extension to $\mathscr{A}^n$ can be therefore considered:

$$\underline{P}_0^{A^n|X}(g|\tilde{x}) = \sup_{\substack{g_j\in\mathscr{L}(\mathscr{X}\times\mathscr{A}_j)\\ j=1,\ldots,n}} \inf_{\substack{a_j\in\mathscr{A}_j\\ j=1,\ldots,n}} \Big\{ g(\tilde{x},a_1,\ldots,a_n) - \sum_{j=1}^n \Big[g_j(\tilde{x},a_1,\ldots,a_n) - \underline{P}_0^{A_j|X}(g_j(\tilde{x},a_1,\ldots,a_{j-1},\cdot,a_{j+1}\ldots,a_n)|\tilde{x}) \Big] \Big\}, \tag{25}$$

for each $\tilde{x}\in\mathscr{X}$. Notice that the gamble $g_j(\tilde{x},a_1,\ldots,a_{j-1},\cdot,a_{j+1}\ldots,a_n)$ is in $\mathscr{L}(\mathscr{X}\times\mathscr{A}_j)$. Let us consider, in (25), only gambles $g \in \mathscr{L}(\mathscr{X}\times\mathscr{A}^n)$ such that, for $X = \tilde{x}$ and each $(a_1,\ldots,a_n)\in\mathscr{A}^n$, factorize as follows:

$$g(\tilde{x},a_1,\ldots,a_n) = \prod_{j=1}^n g'_j(\tilde{x},a_j), \tag{26}$$

with $g'_j \in \mathscr{L}(\mathscr{X}\times\mathscr{A}_j)$ for each $j = 1,\ldots,n$. 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 $\sigma_j = \sigma_j(\tilde{x})$[10]. Under these assumptions, if we intend, for fixed $\tilde{x}$, g as a gamble on $\mathscr{A}^n$, we have that g has constant sign and (25) reduces to:

$$\underline{P}_0^{A^n|X}(g|\tilde{x}) = \begin{cases} \prod_{j=1}^n \underline{P}_0^{A_j|X}(g'_j|\tilde{x}) & \text{if } g \geq 0 \\ -\prod_{j=1}^n \overline{P}_0^{A_j|X}(\sigma_j g'_j|\tilde{x}) & \text{if } g < 0 \end{cases} \tag{27}$$

where g'_j is the g_j defined in (25), for each $j = 1,\ldots,n$. The proof is in [10]. The gambles we consider in the following factorize as in (26), and we can therefore use (27) instead of (25).

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

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

[10] 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.

(d) By marginal extension (14), the following joint CLP can be calculated:

$$\underline{P}_0(h) = \underline{P}_0^X\left(\underline{P}_0^{A^n|X}(h|x)\right) = \varepsilon_0 \sum_{x\in\mathscr{X}} \underline{P}_0^{A^n|X}(h|x)p_0(x) + (1-\varepsilon_0)\min_{x\in\mathscr{X}} \underline{P}_0^{A^n|X}(h|x). \quad (28)$$

(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:

$$\underline{P}_0(I_{\{\tilde{a}_1\}\times\cdots\times\{\tilde{a}_n\}}(g-\mu)) = 0, \quad (29)$$

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 (29) 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 (26), with $g_i'(\tilde{x},a_i) = I_{\{\tilde{a}_i\}}\ \forall i<n$ and $g_i'(\tilde{x},a_n) = I_{\{\tilde{a}_n\}}(g(\cdot)-\mu)$. Therefore, notice that $\sigma_i = 1\ \forall i<n$ and $\sigma_n = \operatorname{sgn}(g(\cdot)-\mu)$. Thus, (27) holds and we can write:[11]

$$\begin{aligned}\underline{P}^{A^n|X}(h|x) = {} & \underline{P}_1^{A_1|X}(I_{\{\tilde{a}_1\}}|x)\cdots\underline{P}_n^{A_n|X}(I_{\{\tilde{a}_n\}}|x) \\ & [g(x,\tilde{a}_1,\ldots,\tilde{a}_n)-\mu]I_{\{g(x,\tilde{a}_1,\ldots,\tilde{a}_n)-\mu\geq 0\}} \\ & + \overline{P}_1^{A_1|X}(I_{\{\tilde{a}_1\}}|x)\cdots\overline{P}_n^{A_n|X}(I_{\{\tilde{a}_n\}}|x) \\ & [g(x,\tilde{a}_1,\ldots,\tilde{a}_n)-\mu]I_{\{g(x,\tilde{a}_1,\ldots,\tilde{a}_n)-\mu<0\}}\end{aligned} \quad (30)$$

According to (30), (29) can be written as in Table 1, where from (23) it can be derived that:

$$\underline{P}_j^{A_j|X}(I_{\{\tilde{a}_j\}}|\tilde{x}) = \frac{p_j(\tilde{a}_j)\varepsilon_j^{\tilde{a}_j}p_j(x|\tilde{a}_j)}{\sum_{a_j\in\mathscr{A}_j} p_j(a_j)[\varepsilon_j^{a_j}p_j(x|a_j) + (1-\varepsilon_j^{a_j})I_{\mathscr{A}_j\setminus\{\tilde{a}_j\}}(a_j)]}. \quad (31)$$

It can be easily verified that $\mathscr{A}_j^* = \mathscr{A}_j\setminus\{\tilde{a}_j\}$ in this case. Again from (23) it follows that:

$$\underline{P}_j^{A_j|X}\left(I_{\{\mathscr{A}_j\setminus\tilde{a}_j\}}|x\right) = \frac{\sum_{a_j\in\mathscr{A}_j, a_j\neq\tilde{a}_j} p_j(a_i)\varepsilon_j^{a_j}p_j(x|a_j)}{\sum_{a_j\in\mathscr{A}_j} p_j(a_j)[\varepsilon_j^{a_j}p_j(x|a_j) + (1-\varepsilon_j^{a_j})I_{\{\tilde{a}_j\}}(a_j)]}, \quad (32)$$

where, in this case, $\mathscr{A}_j^* = \{\tilde{a}_j\}$. According to the duality relation reviewed in Section 2, the corresponding upper probability is one minus the lower probability in (32), and hence:

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

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 $\{\tilde{a}_1,\ldots,\tilde{a}_n\} \in \mathscr{A}^n$.

[11] Note that the indicator functions in (30) refer to sets that are implicitly defined through inequalities over gambles. This kind of specification will be employed also in the followings.

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 [13, 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 [11].
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, (20) cannot be solved by (23). In this case, from (20) 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 (3), it is shown that, since the posterior probability distribution $p_0(x|a_1,\ldots,a_n)$ 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 (31) and (33), it can be seen that $\underline{P}_0^{X|A^n}(g|\tilde{a}_1,\ldots,\tilde{a}_n)$ depends on the sources' priors $\underline{P}_j^X$ and on $(1-\varepsilon_j^{\tilde{a}_j})p(\tilde{a}_j)$. Notice, in fact, that the denominator in (31) 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 (33) 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 (31) and (33) is dropped in Table 1, since the sum and the minimum are over x and, thus, the $p(\tilde{a}_j)$ can be simplified. Summarizing, the pieces of information to be shared between sources and IFC are: the marginal CLP $\underline{P}_j^X$, which corresponds to the prior CLP of the sources; 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 uncertainty* $\overline{P}_j^{X|A_j}(I_{\{x\}}) - \underline{P}_j^{X|A_j}(I_{\{x\}}) = 1-\varepsilon_j^{\tilde{a}_j}$.

6 Zadeh's Paradox

The problem of aggregating beliefs over the same variable has been already considered in other uncertainty theories. In the case of Dempster-Shafer (DS) theory [7], Dempster's combination rule allows for the following aggregation of two belief functions m_1 and m_2:[12]

$$m_{12}(X) \propto \sum_{X_1,X_2:X_1\cap X_2=X} m_1(X_1)\cdot m_2(X_2). \quad (34)$$

[12] We point to [7] for details about DS theory.

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

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

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 gave an apparently counter-intuitive result [16].

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

$$\begin{array}{lll} \text{Doctor 1} \rightarrow & m_1(x_1)=0.99, & m_1(x_3)=0.01, \\ \text{Doctor 2} \rightarrow & m_2(x_2)=0.99, & m_2(x_3)=0.01, \end{array} \quad (35)$$

while the basic belief masses of the other elements of the power set of $\mathscr{X}$ are null. By (34) one gets

$$m_{12}(x_1)=0,\quad m_{12}(x_2)=0,\quad m_{12}(x_3)=1. \quad (36)$$

Hence, from direct application of the DS theory, it turns out that the patient suffers from brain tumor with certainty. 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 modeled 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 [8]. To take this into account, one has to build a model that includes two more variables, modeling 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", $p_1(a_1)$ can be therefore interpreted as the probability that the first source is reliable, $p_1(\neg a_1)=1-p_1(a_1)$ that is unreliable, and similarly for Doctor 2. By following this idea, our aggregation rule can be applied to Zadeh's example. The doctors' diagnoses (35) can be formalized as in (17) 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., when the doctors are unreliable they do not provide any useful information. Furthermore, we assume that $p_1(a_1)=p_2(a_2)=\delta$ with $\delta\in(0,1)$ and $\varepsilon_0=1$, $p_0(x_1)=p_0(x_2)=p_0(x_3)=1/3$. The goal is the 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 final conditional 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 by applying equations (32)-(33). 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, (32)-(33) are not valid anymore. However, since

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

the regular extension (6) can be used to derive

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

where the gambles we are interested in are only $I_{\{a_1\}}$ and $I_{\{\neg a_1\}}$. From (22), $\underline{P}_1^{A_j|x_2}(g|x_2)$ can be calculated by finding the maximum value of μ for which

$$\begin{aligned} &\sum_{\tilde{a}_j\in\mathscr{A}_1} p_j(\tilde{a}_j)[\varepsilon_j^{\tilde{a}_j}p_j(x_2|\tilde{a}_j)+(1-\varepsilon_j^{\tilde{a}_j})I_{\mathscr{A}_1^*(\mu)}(\tilde{a}_j)]g(\tilde{a}_j) \\ &-\mu\sum_{\tilde{a}_j\in\mathscr{A}_1} p_j(\tilde{a}_j)[\varepsilon_j^{\tilde{a}_j}p_j(x_2|\tilde{a}_j)+(1-\varepsilon_j^{\tilde{a}_j})I_{\mathscr{A}_1^*(\mu)}(\tilde{a}_j)]\ge 0. \end{aligned} \quad (37)$$

The values of μ which satisfy (37) in the cases $g=I_{\{a_1\}}$ and $g=I_{\{\neg a_1\}}$ are $\mu=0$ and, respectively, $\mu=1$. Hence, it follows that $\underline{P}_1^{A_j|x_2}(I_{\{a_1\}}|x_2)=\overline{P}_1^{A_j|x_2}(I_{\{a_1\}}|x_2)=$

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

$$\underline{P}_0^{X|A_1,A_2}(I_{\{x_i\}}|\tilde a_1,\tilde a_2)=\frac{\underline{P}_1^{A_1|X}(I_{\{\tilde a_1\}}|x_i)\underline{P}_2^{A_2|X}(I_{\{\tilde a_2\}}|x_i)}{\underline{P}_1^{A_1|X}(I_{\{\tilde a_1\}}|x_i)\underline{P}_2^{A_2|X}(I_{\{\tilde a_2\}}|x_i)+\overline{P}_1^{A_1|X}(I_{\{\tilde a_1\}}|x_j)\overline{P}_2^{A_2|X}(I_{\{\tilde a_2\}}|x_j)+\overline{P}_1^{A_1|X}(I_{\{\tilde a_1\}}|x_k)\overline{P}_2^{A_2|X}(I_{\{\tilde a_2\}}|x_k)}$$
$$\overline{P}_0^{X|A_1,A_2}(I_{\{x_i\}}|\tilde a_1,\tilde a_2)=\frac{\overline{P}_a^{A_1|X}(I_{\{\tilde a_1\}}|x_i)\overline{P}_2^{A_2|X}(I_{\{\tilde a_2\}}|x_i)}{\overline{P}_1^{A_1|X}(I_{\{\tilde a_1\}}|x_i)\overline{P}_2^{A_2|X}(I_{\{\tilde a_2\}}|x_i)+\underline{P}_1^{A_1|X}(I_{\{\tilde a_1\}}|x_j)\underline{P}_2^{A_2|X}(I_{\{\tilde a_2\}}|x_j)+\underline{P}_1^{A_1|X}(I_{\{\tilde a_1\}}|x_k)\underline{P}_2^{A_2|X}(I_{\{\tilde a_2\}}|x_k)}$$

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. The 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 (36) 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, in the case 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 doubts, that the patient suffers from the disease x_1.

In general, in this kind of reliability problems, the sources of information do not provide their reliability status $\{\tilde a_1,\tilde a_2\}$ and, thus, the IFC cannot know it. 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 $\underline{P}_0^{X|A_1,A_2}(\cdot|\mathscr{A}^2\backslash\{a_1,a_2\})$. In practice, the conditioning event is the complementary event of $\{a_1,a_2\}$, which means that at least one doctor is unreliable.

Since $I_{\mathscr{A}^2\backslash\{a_1,a_2\}}$ do not factorize as in (26), we cannot apply (30) 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: $\underline{P}^{A^2|X}(I_{\mathscr{A}^2\backslash\{a_1,a_2\}}|x)=1-\overline{P}^{A^2|X}(I_{\{a_1,a_2\}}|x)=1-\overline{P}^{A_1|X}(I_{\{a_1\}}|x)\overline{P}^{A_2|X}(I_{\{a_2\}}|x)$ and $\overline{P}^{A^2|X}(I_{\mathscr{A}^2\backslash\{a_1,a_2\}}|x)=1-\underline{P}^{A^2|X}(I_{\{a_1,a_2\}}|x)=1-\underline{P}^{A_1|X}(I_{\{a_1\}}|x)\underline{P}^{A_2|X}(I_{\{a_2\}}|x)$.

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. Because of $\underline{P}_0^{X|A_1,A_2}(I_{\{x_1\}}|I_{\mathscr{A}^2\backslash\{a_1,a_2\}})=\underline{P}_0^{X|A_1,A_2}(I_{\{x_2\}}|I_{\mathscr{A}^2\backslash\{a_1,a_2\}})\geq\overline{P}_0^{X|A_1,A_2}(I_{\{x_3\}}|I_{\mathscr{A}^2\backslash\{a_1,a_2\}})$, 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: $p_0^{X|A_1,A_2}(I_{\{x_1\}}|I_{\mathscr{A}^2\backslash\{a_1,a_2\}})=p_0^{X|A_1,A_2}(I_{\{x_2\}}|I_{\mathscr{A}^2\backslash\{a_1,a_2\}})=1/2$.

Summarizing, the results of this section generalize those in [8, 1] to CLPs by showing that: (i) if both the doctors are reliable the result obtained by the Bayes' and Dempster's rule in (36) 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 Conclusions and Outlooks

A general aggregation rule for coherent lower previsions defined on the same 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, a closed formula for the aggregated beliefs has been derived. As an example of applications of this approach, Zadeh's paradox is treated and an alternative explanation is concluded.

As a future work, we aim to generalize our formula for linear-vacuous mixtures to the more general case of 2-monotone capacities. That would be the basis for a recursive application of our approach. Furthermore, although the size of the possibility space of the variable of interest has been assumed finite, it seems possible to extend our results to the infinite case. Yet, further investigations

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 $i = 1,2$

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

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

about the coherence of the corresponding model should be considered. 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.

Acknowledgments

This research has been partially supported by the Swiss NSF grants n. 200020-121785/1. The authors are grateful to Marco Zaffalon and Enrique Miranda for their useful suggestions.

References

[1] S. Arnborg. Robust Bayesianism: Imprecise and paradoxical reasoning. In *Proceedings of the Seventh International Conference on Information Fusion*, volume I, pages 407–414. International Society of Information Fusion, 2004.

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

[3] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, pages 131–295, 1953–1954.

[4] G. de Cooman, F. Hermans, A. Antonucci, and M. Zaffalon. Epistemic irrelevance in credal networks: the case of Markov trees. (under preparation).

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

[6] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, New York, 1988.

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

[8] 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.

[9] E Miranda. Updating coherent previsions on finite spaces. Technical reports of statistics and decision sciences, Rey Juan Carlos University, 2008.

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

[11] A. Piatti, M. Zaffalon, F. Trojani, and M. Hutter. Limits of learning about a categorical latent variable under prior near-ignorance. *International Journal of Approximate Reasoning*. (accepted for pubblication).

[12] M. Troffaes and G. de Cooman. Extension of coherent lower previsions to unbounded random variables. In *Ninth International Conference IPMU 2002*, pages 735–742, Anney, France, 2002.

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

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

[15] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

[16] L. A. Zadeh. On the validity of Dempster rule of combination. In *Memo M 79/24*, pages 3–28. Univ. of California, Berkeley, 1979.

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