

Coherence graphs

Enrique Miranda

Rey Juan Carlos University
Dept. of Statistics and Operations Research
C-Tulipán, s/n, 28933 Móstoles,
Spain
enrique.miranda@urjc.es

Marco Zaffalon

IDSIA
Galleria 2, CH-6928 Manno (Lugano),
Switzerland
zaffalon@idsia.ch

Abstract

We consider the task of proving Walley's (*joint* or *strong*) *coherence* of a number of probabilistic assessments, when these assessments are represented as a collection of conditional lower previsions. In order to maintain generality in the analysis, we assume to be given nearly no information about the numbers that make up the lower previsions in the collection. Under this condition, we investigate the extent to which the above global task can be decomposed into simpler and more local ones. This is done by introducing a graphical representation of the conditional lower previsions, that we call the *coherence graph*: we show that the coherence graph allows one to isolate some subsets of the collection whose coherence is sufficient for the coherence of all the assessments. The situation is shown to be completely analogous in the case of Walley's notion of *weak coherence*, for which we prove in addition that the subsets found are optimal, in the sense that they embody the maximal degree to which the task of checking weak coherence can be decomposed. In doing all of this, we obtain a number of related results: we give a new characterisation of weak coherence; we characterise, by means of a special kind of coherence graph, when the local notion of *separate coherence* is sufficient for coherence; and we provide an envelope theorem for collections of lower previsions whose graph is of the latter type.

Keywords. Walley's strong and weak coherence, coherent lower previsions, graphical models, coherence graph.

1 Introduction

Suppose we plan to carry out a statistical analysis about a certain domain modelled by the following

lower previsions:

$$\underline{P}_1(X_1), \underline{P}_2(X_2|X_1), \underline{P}_3(X_3|X_2), \underline{P}_4(X_4|X_3), \\ \underline{P}_5(X_5, X_6|X_1), \underline{P}_6(X_2|X_3), \underline{P}_7(X_7|X_4), \underline{P}_8(X_8|X_5), \\ \underline{P}_9(X_8|X_6), \underline{P}_{10}(X_9, X_{10}|X_6, X_7), \underline{P}_{11}(X_{11}|X_9, X_{10}).$$

Each of them represents a real functional interpreted as a subject's lower prevision (i.e., lower expectation) for every bounded real-valued function of the random variables on the l.h.s. of the bar, conditional on given values of the variables on the r.h.s. of the bar.

In order to carry out the analysis, we should first verify that the assessments represented by the lower previsions are self-consistent, or *coherent*. Indeed coherence is a (minimal) requirement of rationality, and it is the key that enables one to use a number of powerful theoretical tools to do statistical inference.

Yet, checking coherence can be particularly difficult even in the simple setting illustrated above. In fact, this is a common problem. The power of coherence comes with a price: the technical complications that arise when dealing explicitly with it.

This is the case of the coherence notion that is the focus of this paper, i.e., Walley's definition of coherence [4, Section 7.1.4(b)], which we also call *joint* or *strong coherence*, so as to distinguish it from a weaker notion, also developed by Walley, and called *weak coherence*. Weak and strong coherence are reviewed in Section 2 of this paper, along with other introductory material about Walley's theory of coherent lower previsions.

We argue that the mentioned difficulty is strictly related to the fact that coherence, by its very nature, is a *global* notion: as such, it seems to resist being represented and verified in a local fashion. This is enforced by our initial results in Section 3: we show that a number of (conditional) lower previsions, such as $\underline{P}_1, \dots, \underline{P}_{11}$ in the above example, are weakly coherent if and only if there is an extension, i.e., a lower prevision $\underline{P}(X_1, \dots, X_{11})$ in the example, that is *pairwise* coherent with each of them; and they are strongly co-

herent if and only if they are *globally* coherent with such an extension. In other words, strong coherence seems to be much less amenable to local considerations than other, weaker notions of coherence.

Still, locality is an important property: it is the basis for having compact and efficient models of uncertainty, as well as models that are easier to understand, as it is widely acknowledged after the lesson given by *graphical models* in statistics and artificial intelligence.

The question, at this point, is the following: can we preserve both locality and (strong) coherence?

We regard the present paper as a first positive answer to this question; such an answer is made possible by a new graphical model that we propose in Section 5, and that we call *coherence graph*. Coherence graphs are graphical representations of the structural connections of the lower previsions in a given collection. For example, the coherence graph for the lower previsions $\underline{P}_1, \dots, \underline{P}_{11}$ is shown in Figure 1. Its semantics should be obvious once we identify the lower previsions with the black solid circles in the graph.

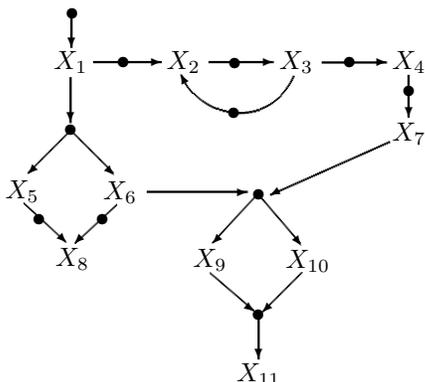


Figure 1: The coherence graph for $\underline{P}_1, \dots, \underline{P}_{11}$.

We talk of structural connections, or of *collection template*, as defined in Section 4, because we do not focus on the numbers that make up the lower previsions; by coherence graphs we rather aim at revealing the *structure* behind the notion of coherence. This structure tells us to what extent the task of checking coherence can be made local. For instance, from the graph in Figure 1 we shall deduce that $\underline{P}_1, \dots, \underline{P}_{11}$ are coherent if so are the lower previsions

$$\underline{P}_1(X_1), \underline{P}_2(X_2|X_1), \underline{P}_3(X_3|X_2), \underline{P}_5(X_5, X_6|X_1), \\ \underline{P}_6(X_2|X_3), \underline{P}_8(X_8|X_5), \underline{P}_9(X_8|X_6).$$

More generally speaking, our main result, stated in Section 6, is that coherence graphs allow us to graphically find out a so-called *minimal* partition of the collection of lower previsions, such that the coherence

of the lower previsions in each set of the partition is sufficient for the coherence of the overall collection.

We show that the situation is completely analogous if we focus on weak coherence: proving weak coherence within each set of the minimal partition is sufficient for the weak coherence of the overall collection. In addition, in the case of weak coherence we can show that the partition found using the coherence graph is indeed minimal, in the sense that it is not possible to use non-coarser¹ alternative partitions to the same extent.

We should mention that these results are fully general with respect to the kind of admissible possibility spaces: they hold irrespective of the fact that we are dealing with finite, countable, or continuous spaces, possibly at the same time. Hence, our results can be used in fields as diverse as expert systems and statistics, just to name a few. Moreover, they are also valid for collections of *linear* previsions, i.e., they do not depend on the precise or imprecise character of the assessments.

We see two major consequences of our results. The first is directly related to proving coherence. We believe that coherence graphs, by making the structure behind coherence explicit, have the potential to give a boost to the theoretical advances in probability, and especially in imprecise probability. Similarly, there seems to be substantial hope for coherence graphs to enhance also the state-of-the-art algorithms for proving coherence. In the case of finite spaces of possibilities, this task is typically addressed by linear programming problems [1, 2, 5] that tend to grow very large as a consequence of the underlying *NP-hardness* of the task itself. But, by exploiting the structure of coherence graphs, it will often be possible to decompose the overall linear programming problem in a number of smaller ones, thus speeding up computation.

The second consequence is more of a principled kind, and is related to a subset of coherence graphs, called of *type A1*, that leads to partitions entirely made up of singletons. This implies that the related collections are immediately known to be coherent, irrespective of their numerical values, provided each of their elements satisfies a local property, called *separate coherence*. We should like to give a special perspective on these collections, by making an analogy with propositional logic. In propositional logic, the formulas that hold irrespective of the values their Boolean variables take, are called *tautologies*, and are regarded as the rules of

¹Proving weak coherence within the sets of a coarser partition would immediately imply weak coherence within the sets of the minimal partition.

logic. We think that collections of lower previsions that have an A1-representation have a special role, and may embody a kind of ‘compositional’ rules that deliver jointly coherent collections by local considerations alone.

2 Coherent lower previsions

Let us give a short introduction to the concepts and results from the behavioural theory of imprecise probabilities that we shall use in the rest of the paper. We refer to [4] for an in-depth study of these and other properties.

Given a possibility space Ω , a *gamble* is a bounded real-valued function on Ω . This function represents a random reward $f(\omega)$, which depends on the a priori unknown value ω of Ω . We shall denote by $\mathcal{L}(\Omega)$ the set of all gambles on Ω . A *lower prevision* \underline{P} is a real functional defined on some set of gambles $\mathcal{K} \subseteq \mathcal{L}(\Omega)$. It is used to represent a subject’s supremum acceptable buying prices for these gambles, in the sense that for any $\epsilon > 0$ and any f in \mathcal{K} the subject is disposed to accept the uncertain reward $f - \underline{P}(f) + \epsilon$.

We can also consider the supremum buying prices for a gamble, *conditional* on a subset of Ω . Given such a set B and a gamble f on Ω , the lower prevision $\underline{P}(f|B)$ represents the subject’s supremum acceptable buying price for the gamble f , updated after coming to know that the unknown value ω belongs to B , and nothing else. If we consider a partition \mathcal{B} of Ω (for instance a set of categories), then we shall represent by $\underline{P}(f|\mathcal{B})$ the gamble on Ω that takes the value $\underline{P}(f|B)$ if and only if $\omega \in B$. The functional $\underline{P}(\cdot|\mathcal{B})$ that maps any gamble f on its domain into the gamble $\underline{P}(f|\mathcal{B})$ is called a *conditional lower prevision*.

Let us now re-formulate the above concepts in terms of random variables, which are the focus of our attention in this paper. Consider random variables X_1, \dots, X_n , taking values in respective sets $\mathcal{X}_1, \dots, \mathcal{X}_n$. For any subset $J \subseteq \{1, \dots, n\}$ we shall denote by X_J the (new) random variable

$$X_J := (X_j)_{j \in J},$$

which takes values in the product space

$$\mathcal{X}_J := \times_{j \in J} \mathcal{X}_j.$$

We shall also use the notation $\mathcal{X}^n := \mathcal{X}_{\{1, \dots, n\}}$. This will be our possibility space in the rest of the paper.

Definition 1. Let J be a subset of $\{1, \dots, n\}$, and let $\pi_J : \mathcal{X}^n \rightarrow \mathcal{X}_J$ be the so-called *projection operator*, i.e., the operator that drops the elements of a vector in \mathcal{X}^n that do not correspond to indexes in J . A gamble f on \mathcal{X}^n is called *\mathcal{X}_J -measurable* when for any $x, y \in \mathcal{X}^n$, $\pi_J(x) = \pi_J(y)$ implies that $f(x) = f(y)$.

There exists a one-to-one correspondence between the gambles on \mathcal{X}^n that are \mathcal{X}_J -measurable and the gambles on \mathcal{X}_J : given an \mathcal{X}_J -measurable gamble f on \mathcal{X}^n , we can define f' on \mathcal{X}_J by $f'(x) := f(x')$, where x' is any element in $\pi_J^{-1}(x)$; conversely, given a gamble g on \mathcal{X}_J , the gamble g' on \mathcal{X}^n given by $g'(x) := g(\pi_J(x))$ is \mathcal{X}_J -measurable.

Consider two disjoint subsets O, I of $\{1, \dots, n\}$. Then, $\underline{P}(X_O|X_I)$ represents a subject’s behavioural dispositions about the gambles that depend on the outcome of the variables $\{X_k, k \in O\}$, after coming to know the outcome of the variables $\{X_k, k \in I\}$. As such, it is defined on the set of gambles that depend on the values of the variables in $O \cup I$ only, i.e., in the set $\mathcal{K}_{O \cup I}$ of the $\mathcal{X}_{O \cup I}$ -measurable gambles on \mathcal{X}^n . Given such a gamble f and $x \in \mathcal{X}_I$, $\underline{P}(f|X_I = x)$ represents his supremum acceptable buying price for the gamble f , if he came to know that the variable X_I took the value x (and nothing else). Under the notation we gave above for lower previsions conditional on events and partitions, this would be $\underline{P}(f|B)$, where $B := \pi_I^{-1}(x)$. When there is no possible confusion about the variables involved in the lower prevision, we shall use the notation $\underline{P}(f|x)$ for $\underline{P}(f|X_I = x)$. The sets $\{\pi_I^{-1}(x) : x \in \mathcal{X}_I\}$ form a partition of Ω . Hence, we can define the gamble $\underline{P}(f|X_I)$, which takes the value $\underline{P}(f|x)$ on $x \in \mathcal{X}_I$. This is a conditional lower prevision.

The \mathcal{X}_I -support $S(f)$ of a gamble f in $\mathcal{K}_{O \cup I}$ is given by $S(f) := \{\pi_I^{-1}(x) : x \in \mathcal{X}_I, f|_{\pi_I^{-1}(x)} \neq 0\}$, i.e., it is the set of conditioning events for which the restriction of f is not identically zero. Here, and in the rest of the paper, $\mathbb{1}_A$ will be used to denote the indicator function of the set A , i.e., the function whose value is 1 in the elements of A and 0 elsewhere. Also, for any gamble f in the domain $\mathcal{K}_{O \cup I}$ of the conditional lower prevision $\underline{P}(X_O|X_I)$, and any $x \in \mathcal{X}_I$, we shall denote by $G(f|x)$ the gamble $\mathbb{1}_{\pi_I^{-1}(x)}(f - \underline{P}(f|x))$, and by $G(f|X_I)$ the gamble that takes the value $G(f|\pi_I(y))$ in all $y \in \mathcal{X}^n$.

These assessments can be made for any disjoint subsets O, I of $\{1, \dots, n\}$, and therefore it is not uncommon to model a subject’s beliefs using a finite number of different conditional previsions. Then, we should verify that all the assessments modelled by these conditional previsions are coherent with each other. The first requirement we make is that for any disjoint $O, I \subseteq \{1, \dots, n\}$, the conditional lower prevision $\underline{P}(X_O|X_I)$ defined on $\mathcal{K}_{O \cup I}$ should be *separately coherent*.² In this case, where the domain is a linear set of gambles, separate coherence holds if and

²We refer to [4] for more general definitions of the following notions in this section in terms of partitions, and for domains that are not necessarily (these) linear sets of gambles.

only if the following conditions are satisfied for any $x \in \mathcal{X}_I$, $f, g \in \mathcal{K}_{O \cup I}$, and $\lambda > 0$:

1. $\underline{P}(f|x) \geq \inf_{y \in \pi_I^{-1}(x)} f(y)$.
2. $\underline{P}(\lambda f|x) = \lambda \underline{P}(f|x)$.
3. $\underline{P}(f + g|x) \geq \underline{P}(f|x) + \underline{P}(g|x)$.

Separate coherence means on the one hand that, if a subject knows that the variable X_I has taken the value x , he cannot raise the (conditional) lower prevision of a gamble by considering the acceptable buying transactions that are implied by other gambles in the domain, and on the other hand that he should bet at any odds on the event that $X_I = x$ after having observed it. In general, separate coherence is not enough to guarantee the consistency of the lower previsions: conditional lower previsions can be conditional on the values of many different variables, and still we should verify that the assessments they provide are consistent not only separately, but also with each other. Formally, we are going to consider what we shall call *collections* of conditional lower previsions.

Definition 2. Consider a set of conditional lower previsions $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ with respective domains $\mathcal{K}^1, \dots, \mathcal{K}^m \subseteq \mathcal{L}(\mathcal{X}^n)$, where \mathcal{K}^i is the set of $\mathcal{X}_{O_i \cup I_i}$ -measurable gambles,³ for $i = 1, \dots, m$. Then, this is called a *collection* on X^n when for each $i \neq j$ in $\{1, \dots, m\}$, either $O_i \neq O_j$ or $I_i \neq I_j$.

This means that we do not have two different conditional lower previsions giving information about the same set of variables X_O , conditional on the same set of variables X_I .

Let $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ be a collection of conditional lower previsions, and let us see the different ways in which we can guarantee their consistency.

Definition 3. $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are *weakly coherent* when for any $f_i \in \mathcal{K}^i$, $i = 1, \dots, m$, $j \in \{1, \dots, m\}$, $f_0 \in \mathcal{K}^j$, $z \in \mathcal{X}_{I_j}$,

$$\sup_{x \in \mathcal{X}^n} \left[\sum_{i=1}^m G_i(f_i|X_{I_i}) - G(f_0|z) \right](x) \geq 0.$$

Although this condition already assures that each of the conditional lower previsions is separately coherent, it does not prevent some inconsistencies from appearing: see [4, Example 7.3.5] for an example. This is the reason why we consider a stronger notion, called (*joint* or *strong*) coherence:

³We use \mathcal{K}^i instead of $\mathcal{K}_{O_i \cup I_i}$ in order to alleviate the notation.

Definition 4. $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are *coherent* when for every $f_i \in \mathcal{K}^i$, $i = 1, \dots, m$, $f_0 \in \mathcal{K}^j$, $z \in \mathcal{X}_{I_j}$, there exists some $B \in \{\pi_J^{-1}(z)\} \cup \bigcup_{i=1}^m S_i(f_i)$ such that

$$\sup_{x \in B} \left[\sum_{i=1}^m G_i(f_i|X_{I_i}) - G(f_0|z) \right](x) \geq 0,$$

where $S_i(f_i)$ is the \mathcal{X}_{I_i} -support of f_i .

In the next section, we prove a number of results that will help to understand better the differences between weak and strong coherence. But before we do that, we introduce a special case that will be of special interest for us: that of *conditional linear previsions*. We say that a conditional lower prevision $\underline{P}(X_O|X_I)$ on the set $\mathcal{K}_{O \cup I}$ is linear if and only if it is separately coherent and moreover $\underline{P}(f + g|x) = \underline{P}(f|x) + \underline{P}(g|x)$ for any $x \in \mathcal{X}_I$ and $f, g \in \mathcal{K}_{O \cup I}$. Conditional linear previsions correspond to the case where a subject's supremum acceptable buying price (lower prevision) coincides with his infimum acceptable selling price (or *upper prevision*) for any gamble on the domain. When a separately coherent conditional lower prevision $\underline{P}(X_O|X_I)$ is linear we shall denote it by $P(X_O|X_I)$.

If we consider a collection of conditional linear previsions $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ with domains $\mathcal{K}^1, \dots, \mathcal{K}^m$, then they are coherent if and only if they *avoid partial loss*: for every $f_i \in \mathcal{K}^i$, $i = 1, \dots, m$, there is some $B \in \bigcup_{i=1}^m S_i(f_i)$ such that

$$\sup_{x \in B} \left[\sum_{i=1}^m G_i(f_i|X_{I_i}) \right](x) \geq 0,$$

where, again, $S_i(f_i) = \{\pi_{I_i}^{-1}(x) : x \in \mathcal{X}_{I_i}, f_i \mathbb{I}_{\pi_{I_i}^{-1}(x)} \neq 0\}$.

One interesting feature of conditional linear previsions allows to easily characterise separate coherence: a conditional lower prevision $\underline{P}(X_O|X_I)$ is separately coherent if and only if it is the lower envelope of a closed (in the *weak-* topology*) convex set of dominating conditional linear previsions, where $P(X_O|X_I)$ is said to *dominate* $\underline{P}(X_O|X_I)$ when for every $\mathcal{X}_{O \cup I}$ -measurable gamble f , $P(f|x) \geq \underline{P}(f|x)$ for every $x \in \mathcal{X}_I$. Note, however, that in general a collection of coherent conditional lower previsions is not necessarily the lower envelope of a collection of coherent (i.e., avoiding partial loss) conditional linear previsions.

Finally, one interesting particular case is that where we are given only an unconditional lower prevision \underline{P} on $\mathcal{L}(\mathcal{X}^n)$ and a conditional lower prevision $\underline{P}(X_O|X_I)$ on $\mathcal{K}_{O \cup I}$. Then, weak and strong coherence are equivalent, and they both hold if and only if, for any $\mathcal{X}_{O \cup I}$ -measurable f and any $x \in \mathcal{X}_I$,

$$(C1) \quad \underline{P}(G(f|X_I)) \geq 0$$

$$(C2) \quad \underline{P}(G(f|x)) = 0.$$

If both P and $P(X_O|X_I)$ are linear previsions, they are coherent if and only if for any $\mathcal{X}_{O \cup I}$ -measurable f it holds that $P(f) = P(P(f|X_I))$.

3 Weak and strong coherence

The following theorem gives a new characterisation of the weak coherence of the conditional lower previsions $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$.

Theorem 1. *$\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent if and only if there is some coherent lower prevision \underline{P} on $\mathcal{L}(\mathcal{X}^n)$ such that*

$$\begin{cases} \underline{P}(G_i(f|X_{I_i})) \geq 0 & \text{for any } f \text{ in } \mathcal{K}^i \\ \underline{P}(G_i(f|x)) = 0 & \text{for any } f \text{ in } \mathcal{K}^i, x \text{ in } \mathcal{X}_{I_i}. \end{cases}$$

Remark 1. When all the conditional previsions are linear, then weak coherence is equivalent to the existence of a *linear* prevision that is coherent with each of the conditionals: we can deduce from Theorem 1 and [4, Section 6.5.5] that any linear prevision P dominating \underline{P} will satisfy $P(G_j(f|X_{I_j})) = 0$ for any $f \in \mathcal{K}^j$, and this implies that P is coherent with $P_j(X_{O_j}|X_{I_j})$.

When moreover all the spaces $\mathcal{X}_1, \dots, \mathcal{X}_n$ are finite, we deduce from Theorem 1 that the weak coherence of the conditional previsions $P_j(X_{O_j}|X_{I_j})$, $j = 1, \dots, m$, is equivalent to the existence of a linear prevision (a finitely additive probability) on \mathcal{X}^n inducing the conditional previsions by means of Bayes rule. This is not enough, however, for the conditional previsions to be coherent. For a counterexample, see [4, Example 7.3.5]. \blacklozenge

From this theorem, we can easily deduce the following two results, that relate (weak or strong) coherence to the existence of an unconditional lower prevision that is (weakly or strongly) coherent with the collection.

Proposition 1. *The conditional lower previsions $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are coherent if and only if there is some coherent unconditional lower prevision \underline{P} on $\mathcal{L}(\mathcal{X}^n)$ such that $\underline{P}, \underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are coherent.*

Corollary 1. *The conditional lower previsions $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent if and only if there is some coherent lower prevision \underline{P} on $\mathcal{L}(\mathcal{X}^n)$ such that $\underline{P}, \underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent.*

These results allow us to understand a bit better the conceptual difference between weak coherence and

(strong) coherence: weak coherence amounts to the existence of a joint that is pairwise coherent with each of the conditional lower previsions; coherence means that there is a joint that is coherent with all the conditional lower previsions, *taken together*.

4 Collection templates

In this paper we are interested in proving coherence properties of lower previsions without assuming to be given information about the numbers that make up the lower previsions themselves, other than they produce separately coherent assessments. For this we need at least to focus on the ‘form’ of the lower previsions, which we call *template*.

Definition 5. Let $\underline{P}_{j'}(X_{O_{j'}}|X_{I_{j'}})$ and $\underline{P}_{j''}(X_{O_{j''}}|X_{I_{j''}})$ be two lower previsions on X^n . We say that they *have the same template* if $O_{j'} = O_{j''}$ and $I_{j'} = I_{j''}$. The class of all the lower previsions on X^n with the same template is just called *lower prevision template on X^n* (of the generic lower previsions in the class). We denote a lower prevision template in the same way as we denote a lower prevision (the distinction should be clear from the context): i.e., by $\underline{P}_j(X_{O_j}|X_{I_j})$.

Definition 6. Similarly, we say that *two collections of lower previsions on X^n have the same template* if they contain the same number m of lower previsions, and if it is possible to order the elements in each collection in such a way that for all j in $\{1, \dots, m\}$ the two respective j -th lower previsions have the same template. The class of all the collections on X^n with the same template is just called *collection template on X^n* (of the generic collection in the class). We denote a collection template in the same way as we denote a collection of lower previsions (again, the distinction should be clear from the context): i.e., by $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$.

The notion of a collection template should be regarded a special kind of assessment about a collection of lower previsions, in the sense that knowing the template of a collection means knowing that the collection belongs to a certain set. An equivalent way to look at a collection template is as a collection of lower prevision templates. For this reason, we shall sometimes refer to the lower previsions templates in a collection template.

5 Coherence graphs

In this section, we introduce a graphical representation of collection templates based on directed graphs. For this, we start by recalling some terminology from graph theory.

A *directed graph* is a structure made up of a set of *nodes* and a set of *directed arcs* between nodes. Two nodes connected by an arc are also called its *end-points*. A sequence of at least two nodes for which each pair of adjacent nodes is an arc in the graph, is called *directed path* between the first and the last node in the sequence (also called *origin* and *destination nodes*, respectively). When the origin and destination nodes coincide, and this is the only case of repeated node in the sequence, we say that the path is a *directed cycle*, or just a *cycle*, for short. Note that a path uniquely identifies a sequence of arcs; for this reason, by an abuse of terminology, we shall sometimes refer to the arcs of a path.

The *predecessors* of a node are all the nodes that have a directed path towards the given node. The predecessors for which there is a directed path made up of a single arc, are called *parents*. The *indegree* of a node is the number of its parents. A node with indegree equal to zero is called a *root*. Similarly, the *successors* of a node are all the nodes that can be reached from the given node following directed paths. The successors for which there is a directed path made up of a single arc, are called *children*. The *outdegree* of a node is the number of its children. A node with outdegree equal to zero is called a *leaf*.

The union of the set of parents and children of a node is called the set of its *neighbors*. The *union of two graphs* is a graph created by taking the union of their nodes and their arcs, respectively.

Now we are ready to define the most important graphical notion used in this paper.

Definition 7. Consider two finite sets $\mathcal{Z} = \{X_1, \dots, X_n\}$ and $\mathcal{D} = \{D_1, \dots, D_m\}$ of so-called *actual* and *dummy nodes*, respectively. Call $\mathcal{N} := \mathcal{Z} \cup \mathcal{D}$ the set of *nodes*, and a given $\mathcal{A} \subseteq \mathcal{N} \times \mathcal{N}$ the set of *arcs*. The triple $\langle \mathcal{Z}, \mathcal{D}, \mathcal{A} \rangle$ is called a *coherence graph* on \mathcal{Z} if the following properties hold:

- (CG1) \mathcal{Z} is non-empty.
- (CG2) All neighbors of dummy nodes are actual nodes, and vice versa.
- (CG3) The set of the parents and that of the children of any dummy node have an empty intersection.
- (CG4) Dummy nodes are not leaves.
- (CG5) Different dummy nodes do not have both the same parents and the same children.

Figure 1 used in the Introduction is just an example of a coherence graph, with actual nodes X_1, \dots, X_{11} . Note that to make graphs easier to see, we represent

dummy nodes in a simplified way: we do not show their labels and rather represent each of them simply as a black solid circle (this does not pose a problem since each dummy node is univocally identified by its neighbors); moreover, when a dummy node has exactly one parent and one child, we do not represent the arrow entering the dummy node (this is not going to cause ambiguity either).

Next, we show that there is a one-to-one relationship between coherence graphs on $\mathcal{Z} = \{X_1, \dots, X_n\}$ and collection templates on X^n . To this extent, it is useful to isolate the notion of a *D-structure* in a coherence graph.

Definition 8. Given a dummy node D of a coherence graph, we call *D-structure* the subgraph whose nodes are D and its neighbors, and whose arcs are those connecting D to its neighbors.

In the graph of Figure 1 there are 11 D-structures, one per dummy node. For example, a D-structure is the subgraph made by the actual nodes X_9, X_{10}, X_{11} , by the dummy node in the middle, and by the arcs that connect them; another D-structure is the subgraph made by X_1, X_2 , the dummy node in between, and the arc(s) connecting them.

At this point we consider two functions: the first one, that we shall denote by Γ , maps a collection template $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$, related to the variables $\{X_1, \dots, X_n\} =: \mathcal{Z}$, into a coherence graph on \mathcal{Z} , with dummy nodes $\{D_1, \dots, D_m\}$. This mapping is determined by the following procedure:

- ($\Gamma 1$) Let $\mathcal{Z} := \{X_1, \dots, X_n\}$ be the set of actual nodes.
- ($\Gamma 2$) Let $\mathcal{D} := \{D_1, \dots, D_m\}$ be the set of dummy nodes.
- ($\Gamma 3$) Let $\mathcal{A} := \emptyset$.
- ($\Gamma 4$) For all $j \in \{1, \dots, m\}$, all $i' \in I_j$, all $i'' \in O_j$, add the arcs $(X_{i'}, D_j)$ and $(D_j, X_{i''})$ to \mathcal{A} .

The second function, that we denote by Γ^{-1} , maps a coherence graph on $\mathcal{Z} = \{X_1, \dots, X_n\}$, with dummy nodes $\{D_1, \dots, D_m\}$, into the collection template $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$, related to the variables $\{X_1, \dots, X_n\}$. This mapping is determined by the following procedure:

- ($\Gamma^{-1} 1$) Set the collection of lower prevision templates equal to the empty set.
- ($\Gamma^{-1} 2$) For all $j \in \{1, \dots, m\}$, add $\underline{P}_j(X_{O_j}|X_{I_j})$ to the collection template, where O_j and I_j are the set of indexes of the children and the parents of D_j , respectively.

The idea behind the two functions is very simple: identifying lower prevision templates in a collection with D -structures in the related coherence graph, and vice versa.

Consider the graph of Figure 1 once again. By applying Function Γ^{-1} we, unsurprisingly, obtain the collection of lower prevision templates used as starting example in the Introduction:

$$\{\underline{P}_1(X_1), \underline{P}_2(X_2|X_1), \underline{P}_3(X_3|X_2), \underline{P}_4(X_4|X_3), \\ \underline{P}_5(X_5, X_6|X_1), \underline{P}_6(X_2|X_3), \underline{P}_7(X_7|X_4), \underline{P}_8(X_8|X_5), \\ \underline{P}_9(X_8|X_6), \underline{P}_{10}(X_9, X_{10}|X_6, X_7), \underline{P}_{11}(X_{11}|X_9, X_{10})\}.$$

It is easy then to see that Function Γ gives back the original graph once it is applied to such a collection template. The reason is that the two functions turn out to be each other's inverses. This is shown by the next theorem, which also allows us to prove the wanted one-to-one relationship between coherence graphs and collection templates.

Theorem 2. *There is one-to-one relationship between coherence graphs and collection templates.*

Next, we introduce some graph-based terminology that is more directly relevant to our subsequent results.

Definition 9. We say that an actual node of a coherence graph is a (potential) *source of contradiction* (or *conflict*) if it has more than one parent or if it belongs to a cycle.

Definition 10. A coherence graph without sources of contradiction is said to be of type $A1$: i.e., acyclic and with maximum indegree for actual nodes equal to one. The corresponding collection template is said to be *representable as a graph of type A1*, or simply *A1-representable*.

The graph in Figure 2 is clearly not of type $A1$, as there are three sources of contradiction: X_8 , given its two parents; X_2 , because it has two parents and also because it is part of a cycle; and X_3 , because it is in such a cycle, too.

Definition 11. Given a source of contradiction Z , call *block for Z*, or B_Z , the subgraph obtained by taking the union of the D -structures of the dummy nodes that are predecessors of Z .

Definition 12. Call *superblock* of a coherence graph, any union of all the blocks that share at least one actual node.

Figure 2 displays the only two different blocks of the coherence graph under consideration: the block for X_8 and that for X_3 (note that the latter coincides with the block for X_2). Those blocks have the node X_1 in common (besides its dummy parent); their union is

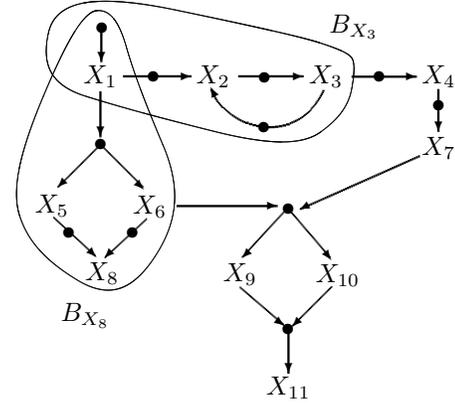


Figure 2: The areas delimited by closed lines contain two blocks of the coherence graph: B_{X_8} and B_{X_3} . Their union is a superblock.

thus a superblock, which is also the only one in the graph.

Observe that there can be many configurations of blocks in a superblock: a superblock can be made up of a single block; if it is made up of more than one block, it may be the case that some blocks coincide (as B_{X_2} and B_{X_3} in Figure 2), that one of them is included in another, or that two of them share only some nodes (as B_{X_3} and B_{X_8} in the same figure).

We use the notion of superblock in order to build a partition of the dummy nodes.

Definition 13. Call *minimal partition of the dummy nodes* in a coherence graph, the partition whose elements are the sets of dummy nodes in each superblock, and the singletons made up of the remaining dummy nodes. The corresponding partition of $\{1, \dots, m\}$ is denoted by \mathcal{B} and is simply called the *minimal partition*.

Note that \mathcal{B} refers also to a partition of the related collection template, given the one-to-one correspondence between dummy nodes and lower prevision templates. With respect to the graph in Figure 2, we obtain the following partition of the related collection template:

$$\{\{\underline{P}_1(X_1), \underline{P}_2(X_2|X_1), \underline{P}_3(X_3|X_2), \underline{P}_5(X_5, X_6|X_1), \\ \underline{P}_6(X_2|X_3), \underline{P}_8(X_8|X_5), \underline{P}_9(X_8|X_6)\}, \\ \{\underline{P}_4(X_4|X_3)\}, \{\underline{P}_7(X_7|X_4)\}, \{\underline{P}_{10}(X_9, X_{10}|X_6, X_7)\}, \\ \{\underline{P}_{11}(X_{11}|X_9, X_{10})\}\}.$$

Moreover, note that for $A1$ -representable collection templates, the minimal partition is entirely made up of singletons, because their coherence graph has no sources of contradiction.

6 Coherence graphs as tools to prove coherence

The following theorem gives us conditions under which the coherence of some subsets of a collection of conditional lower previsions implies the coherence of all the elements in the collection. It shows that it is sufficient that the conditional lower previsions whose indices belong to the same element in \mathcal{B} are coherent.

Theorem 3. *Consider a collection $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ of separately coherent conditional lower previsions with known templates. Then, if for any $B \in \mathcal{B}$, $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j \in B}$ are coherent, then $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ are coherent.*

The intuition behind the proof of the theorem is the following. We exploit the properties of the coherence graph to create a total order on a set of coherent lower previsions strongly related to $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$. That order allows us to use the generalisation of the *Marginal Extension Theorem* (MET, in short) established in [3] to show that the lower previsions in that set are coherent, and from this to derive the coherence of $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$.

It is easy to see a similar result holds when we work with weak coherence instead of coherence:

Theorem 4. *Consider a collection $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ of separately coherent conditional lower previsions with known templates. Then, if for any $B \in \mathcal{B}$, $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j \in B}$ are weakly coherent, then $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ are weakly coherent.*

Next, we investigate in which sense the partition \mathcal{B} given by Definition 13 is minimal. For this, we should like to know if there are other partitions of $\{1, \dots, m\}$ that we can use for the same end, meaning that the coherence of the conditional lower previsions within each of the elements of the partition guarantees the coherence of the collection template.

A first positive result in this regard is that the partition \mathcal{B} is indeed minimal when we are studying the problem for weak coherence:

Proposition 2. *Let \mathcal{B}' be a partition of $\{1, \dots, m\}$, and assume that, for any B' in \mathcal{B}' , $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j \in B'}$ are weakly coherent. Then, this implies the weak coherence of $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ if and only if \mathcal{B} is finer than \mathcal{B}' .*

The sufficiency part in this proposition is actually Theorem 4, which can be proven in a simi-

lar way as Theorem 3. The idea for the necessity part is to show that, when the necessary condition fails, we can create conditional *linear* previsions $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ that are not weakly coherent and yet for all B' in \mathcal{B}' , $\{P_j(X_{O_j}|X_{I_j})\}_{j \in B'}$ are weakly coherent.

A basic step in the construction of such lower previsions is to prove that for any given $j \in \{1, \dots, m\}$ and any $x \in \mathcal{X}_{O_j}$, we can define weakly coherent conditional previsions $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ such that any compatible joint P satisfies $P(\pi_{O_j}^{-1}(x)) = 1$. Even stronger, we can show that any compatible joint with *some* of these conditional previsions satisfies $P(\pi_{O_j}^{-1}(x)) = 1$. This is proven using the following lemmas:⁴

Lemma 1. *For any $i = 1, \dots, n$, let us consider $x_1^i, x_2^i \in \mathcal{X}_i$. Define the conditional previsions $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ with respective domains $\mathcal{K}^1, \dots, \mathcal{K}^m$ by⁵*

$$P_j(f|y) := \begin{cases} f((x_1^i)_{i \in O_j}, y) & \text{if } y = (x_1^i)_{i \in I_j} \\ f((x_2^i)_{i \in O_j}, y) & \text{otherwise,} \end{cases}$$

for any $j = 1, \dots, m$, $y \in \mathcal{X}_{I_j}$ and $f \in \mathcal{K}^j$. Then, $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ are coherent.

Lemma 2. *For any $i = 1, \dots, n$, let us consider $x_1^i, x_2^i \in \mathcal{X}_i$. Define the conditional previsions $P_1(X_{O_1}|X_{I_1}), \dots, P_{m-1}(X_{O_{m-1}}|X_{I_{m-1}})$ with respective domains $\mathcal{K}^1, \dots, \mathcal{K}^{m-1}$ by*

$$P_j(f|y) := \begin{cases} f((x_1^i)_{i \in O_j}, y) & \text{if } y = (x_1^i)_{i \in I_j} \\ f((x_2^i)_{i \in O_j}, y) & \text{otherwise,} \end{cases}$$

for any $y \in \mathcal{X}_{I_j}$, $f \in \mathcal{K}^j$, and define $P_m(X_{O_m}|X_{I_m})$ by

$$P_m(f|y) := f((x_2^i)_{i \in O_m}, y)$$

for any $y \in \mathcal{X}_{I_m}$ and $f \in \mathcal{K}^m$. Then, $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ are weakly coherent.

However, a similar result to Proposition 2 does not apply for coherence, due, among other things, to the fact that the previsions in Lemma 2 are weakly coherent but not coherent. As a consequence, there exist instances of collection templates where the coherence within the elements of a partition which is not coarser than \mathcal{B} guarantees the coherence of all of them. One such case is given in the following example.

⁴Although the previsions in these lemmas correspond to 0-1 valued probabilities, this is not essential for the developments made in the proof of the theorem; it is possible to obtain similar results using probabilities that are not 0-1 valued.

⁵We are using there the one-to-one correspondence between gambles on \mathcal{X}^j and gambles in \mathcal{K}^j .

Example 1. Consider the collection template $\{\underline{P}_1(X_1), \underline{P}_2(X_2|X_1), \underline{P}_3(X_2, X_3|X_1)\}$. Then, the minimal partition \mathcal{B} associated to its coherence graph is $\{1, 2, 3\}$. However, we can deduce the coherence of the collection template using a smaller subset. For this, we must prove first that the coherence of $\underline{P}_2(X_2|X_1), \underline{P}_3(X_2, X_3|X_1)$ holds if and only if for any $\mathcal{X}_1 \times \mathcal{X}_2$ -measurable gamble f and for any $x_1 \in \mathcal{X}_1$,

$$\underline{P}_2(f|x_1) = \underline{P}_3(f|x_1).$$

Using this property, we deduce that, when $\underline{P}_2(X_2|X_1), \underline{P}_3(X_2, X_3|X_1)$ are coherent, then $\{\underline{P}_1(X_1), \underline{P}_2(X_2|X_1), \underline{P}_3(X_2, X_3|X_1)\}$ are coherent if and only if $\underline{P}_1(X_1), \underline{P}_3(X_2, X_3|X_1)$ are. But since $\underline{P}_1(X_1), \underline{P}_3(X_2, X_3|X_1)$ are always coherent because of the marginal extension theorem in [4, Theorem 6.7.2], we deduce that the coherence of $\underline{P}_2(X_2|X_1), \underline{P}_3(X_2, X_3|X_1)$ implies the coherence of the collection template. \blacklozenge

It remains an open problem at this stage to determine a minimal partition with the property that the coherence within each of the elements of the partition guarantees the coherence of the collection template, and that is minimal in the sense that it is finer than any other partition with the same property.

In this respect, we can deduce from Theorem 3 that the separate coherence of the conditional lower previsions $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ implies their joint coherence when their associated coherence graph is of type A1. Using Lemma 1, we can prove that being of type A1 is also necessary for this property.

Proposition 3. *Consider a collection $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ of separately coherent conditional lower previsions with known templates. Then the separate coherence of $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ implies their coherence if and only if their coherence graph is of type A1.*

Note on the other hand that, with respect to weak coherence, we also have a necessary and sufficient condition for the separate coherence to imply the weak coherence, because of Proposition 2:

Corollary 2. *Consider a collection $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ of separately coherent conditional lower previsions with known templates. Then the separate coherence of $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ implies their weak coherence if and only if their coherence graph is of type A1.*

We should like to conclude this section remarking that

if the collection template is A1, then we can give the following Bayesian sensitivity analysis interpretation:

Theorem 5. *Consider a collection of separately coherent conditional lower previsions. If their coherence graph is A1, then these lower previsions are lower envelopes of a family of coherent linear previsions.*

The interest of this result lies in the fact that the lower envelopes of a family of coherent conditional linear previsions are coherent conditional lower previsions, but the converse does not hold in general: there exist instances of coherent conditional lower previsions that are not even dominated by any family of coherent conditional linear previsions. A sufficient condition for the converse to hold is that the spaces $\mathcal{X}_1, \dots, \mathcal{X}_n$ are finite. This theorem shows that, if the coherence graph is A1, then the coherent conditional lower previsions are also lower envelopes of coherent conditional linear previsions, no matter the cardinality of the spaces.

7 Discussion

Coherence can be regarded as the very essence of a theory of personal probability. But working directly with coherence can be particularly onerous.

This paper is an attempt to deal with this difficulty, and to deliver tools that make checking coherence easier. We have been inspired in this by the lesson of graphical models, and have indeed defined a new graphical model called a coherence graph.

Coherence graphs are means to render explicit the structure behind the notion of coherence. We have shown that such a structure induces a partition of the available collection of lower previsions, with the characteristic that the coherence within each set of the partition implies the coherence of the overall collection.

This result is very general: it holds for lower previsions and for any cardinality of the possibility spaces involved. In particular, since it holds for lower previsions, it is also applicable to determine the coherence of a collection of conditional *linear* previsions, and therefore is also useful in the precise context.

More generally speaking, we expect the results in this paper to have substantial theoretical as well as practical consequences, whenever the focus is on the task of proving coherence. They already appear to shed light on specific aspects of coherence, thanks especially to coherence graphs of type A1. These graphs correspond to collections of separately coherent lower previsions that are coherent irrespective of the numerical values that make them up.

Remember that we have shown that there are important conceptual differences between the notions of weak and strong coherence proposed by Walley. Weak coherence is equivalent to the existence of a joint lower prevision that is coherent with each of the assessments. In the particular case of conditional linear previsions and finite spaces, this is equivalent to the existence of a joint mass function inducing each of the conditionals by means of Bayes rule. The introduction of the notion of strong coherence is needed because some conditional lower previsions can have a common joint and still be clearly incoherent with one another. Remarkably, this happens even in the linear and finite case mentioned above.

Taking this into account, we find it noteworthy that, for the problem tackled here, weak and strong coherence exhibit a similar behaviour: if we have a number of assessments and all we know about them is that each of them is separately coherent, we can guarantee that they are weakly coherent exactly under the same conditions for which we can deduce their joint coherence: we just need the graph representing the collection template to be A1. More generally, we have established a partition of the graph for which weak coherence inside implies weak coherence of them all, and we have proven that strong coherence inside this partition also implies the strong coherence of all the assessments. It is worth pointing out that there are also differences: we have shown that the minimal partition obtained using a coherence graph is indeed minimal in the case of weak coherence and not necessarily so for strong coherence.

Another point worth emphasising is the connection, used repeatedly in the proofs of this paper, between the A1 condition and the generalisation of the MET established in [3]: the relationship arises as from the A1 condition we can establish a total order on the conditional lower previsions in our collection template, and such an order is just what allows us to use the generalised MET. In this way, we have also given an easy graphical characterisation of the extent to which the theorem can be applied: to A1-representable collection templates.

Finally, we have proven that if the separate coherence of the lower previsions in a collection template implies their joint coherence (that is, if the associated coherence graph is A1), then the conditional lower previsions in the template are lower envelopes of coherent linear previsions. This does not hold for all collections of coherent conditional lower previsions, as is shown in [4, Section 6.6]. So it is remarkable that our results lead naturally to a Bayesian sensitivity analysis interpretation of the collection of conditional lower previsions.

As a topic for future research, we should like to mention the study of the coherence of collection templates when we have some additional structural assessments, such as considerations of irrelevance or independence.

Acknowledgements

We are grateful to Gert de Cooman for encouraging us to study the problems presented in this paper, and for many helpful comments. We acknowledge financial support by the MCYT projects MTM2004-01269, TSI2004-06801-C04-01, and by the Swiss NSF grants 200020-109295/1, 200021-113820/1.

References

- [1] V. Biazzo, A. Gilio, T. Lukasiewicz, and G. Sanfilippo. Probabilistic logic under coherence: complexity and algorithms. *Annals of Mathematics and Artificial Intelligence*, 45(1–2):35–81, 2005.
- [2] B. Jaumard, H. Hansen, and Poggi de Aragao. Column generation methods for probabilistic logic. *ORSA Journal on Computing*, 3:135–148, 1991.
- [3] E. Miranda and G. de Cooman. Marginal extension in the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 2006. Accepted for publication.
- [4] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
- [5] P. Walley, R. Pelessoni, and P. Vicig. Direct algorithms for checking consistency and making inferences for conditional probability assessments. *Journal of Statistical Planning and Inference*, 126:119–151, 2004.