COHERENCE GRAPHS

ENRIQUE MIRANDA AND MARCO ZAFFALON

ABSTRACT. We study the consistency of a number of probability distributions, which are allowed to be imprecise. To make the treatment as general as possible, we represent those probabilistic assessments as a collection of *conditional lower previsions*. The problem then becomes proving Walley's (*strong*) *coherence* of the assessments. 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; and we provide a polynomial-time algorithm that finds the subsets efficiently. We show some of the implications of our results by focusing on three models and problems: *Bayesian* and *credal networks*, of which we prove coherence; the *compatibility problem*, for which we provide an optimal graphical decomposition; *probabilistic satisfiability*, of which we show that some intractable instances can instead be solved efficiently by exploiting coherence graphs.

1. INTRODUCTION

We focus on studying the consistency of a number of conditional and unconditional distributions of some variables. In order to make our treatment as general as possible, we are going to represent these probabilistic assessments using the theory of *coherent lower previsions* developed by Walley in [33], which is based on de Finetti's work about *subjective probability* [10, 11]. This allows us to study the case where the above distributions are imprecise, i.e., where each of them is actually a closed convex set of precise distributions, which includes as a particular case that where our assessments are precise probabilities. It also allows us to work with any type of variable, without placing restrictions on the admissible possibility spaces (finite, countable, continuous). The approach by Walley includes also as particular cases most of the other imprecise probability models appearing in the literature.

Studying the consistency problem is important both for theoretical and applied reasons. On the theoretical side, it has been shown by de Finetti that a subjective theory of precise probability (such as the Bayesian theory) can be founded on a single axiom of consistency. Williams [35] and later Walley have shown that this continues to hold when such a theory is generalised to handle imprecision in probability. In these theories proving consistency is therefore a necessary step to exploit all the tools they provide us with, such as Bayes' rule and its generalisations. The application of these tools alone, on the other hand, does not necessarily lead

Key words and phrases. Walley's strong and weak coherence, coherent lower previsions, graphical models, probabilistic logic, satisfiability.

to self-consistent inference, even in the case of precise probabilities, as shown by Walley.

On the applied side, it is a very common requirement that an inference method should not give rise to inconsistencies. This requirement is present, for example, in *probabilistic logic* [26], where one has to check first of all that the available assessments are self-consistent. It is also present in the many other models and methods designed so as to give rise to a joint distribution, which is often regarded as a feature that ensures global consistency. Exactly this argument was used, for example, to support *Bayesian networks* versus rule-based systems already at the time of Pearl's seminal work [27]. But consistency is quite a subtle concept to deal with. The following striking example adapted from [33, Section 7.3.5] shows that the existence of a compatible joint is not always a good way to get rid of inconsistencies.

Example 1. Let X_1, X_2 be two variables taking values in $\{1, 2, 3\}$, and assume that $X_1 = 3$ if and only if $X_2 = 3$, and for the other cases we have the contradictory information $X_1 = X_2$ and $X_1 \neq X_2$. We can model this by the conditional probabilities $P(X_1 = 3|X_2 = 3) = 1 = P(X_2 = 3|X_1 = 3)$, $P(X_1 = 1|X_2 = 1) = 1 = P(X_1 = 2|X_2 = 2)$ and $P(X_2 = 1|X_1 = 2) = 1 = P(X_2 = 2|X_1 = 1)$. Despite the contradiction, it can be checked that the assessments are compatible with the joint mass function determined by $P(X_1 = 3, X_2 = 3) = 1$, in the sense that this joint induces the above conditionals by Bayes' rule when the conditioning events have positive probability.[1] ⧫

The key here is that Bayes' rule cannot be always applied because of the presence of events with zero probability; this technical issue prevents the contradiction from being identified. It follows that in order to check consistency we generally need stronger tools than those based on the existence of a compatible joint distribution. Walley's notion of coherence appears to be one such tool.

In fact, Walley considers two different consistency concepts for conditional lower previsions, called *weak coherence* and (*strong*) *coherence* (these will be introduced in Section 2, along with other material about Walley's theory). What we show in Section 3 is that a number of conditional lower previsions are weakly coherent when they can all be induced by the same joint via Bayes' rule (or its generalisation for the imprecise case) and marginalisations. In other words, we show that weak coherence is the generalisation to imprecise probability of the consistency criterion based on the existence of a compatible joint. Coherence, on the other hand, strengthens weak coherence and it can be shown that the difference between weak and strong coherence is indeed related to conditioning on sets of probability zero (see [21]).

Our goal in this paper is to simplify the verification of the weak or the strong coherence of a number of assessments. To achieve this, we introduce in Section 5 a new graphical representation called *coherence graphs*. We prove in Section 6 that coherence graphs allow us to decompose the task of verifying weak and strong coherence in a number of simpler tasks. Specifically, they help us determine a partition of the set of assessments with the property that coherence (resp., weak coherence) within each of the elements of the partition implies coherence (resp.,

[1]This consistency notion is what we shall call later in this paper *weak coherence*. Note that the contradictory assessments $X_1 = X_2$ and $X_1 \neq X_2$ can also be modelled by other conditional probabilities that do not even satisfy this consistency notion, for instance $P(X_1 = 1|X_2 = 1) = 1 = P(X_1 = 2|X_2 = 1)$.

weak coherence) of all the assessments. We prove moreover that this is the finest partition with this property in the case of weak coherence. Besides, this partition of our set of assessments can be determined with a polynomial-time algorithm, which we present in Section 7.

Then we move to show some of the implications of our results for artificial intelligence by considering three well-known related research fields. In Section 8.1, we consider *Bayesian networks* [27] and their extension to imprecise probability called *credal networks* [8]. By joining coherence graphs with a notion of probabilistic independence, we show for the first time and to a very large extent that Bayesian and credal networks are coherent models. In Section 8.2 we focus on the so-called *compatibility problem* (see [9] and the references therein for a recent overview), i.e., the problem of deciding whether a number of distributions has a compatible joint. In this case we exploit our results about weak coherence to deliver new graphical criteria that enable one to *optimally* decompose such a problem, under a very general formulation. Finally, in Section 8.3 we relate our results to a powerful form of probabilistic satisfiability based on consistency that has been recently proposed in [34]. In particular, we discuss how probabilistic satisfiability can be used to check the consistency of a number of (possibly imprecise) conditional and unconditional mass functions, and we outline that this task can easily become intractable as a consequence of the *NP-hardness* of the problem [5]. Moreover, we show that coherence graphs can decompose such a task in a way that makes it possible to solve some instances of the problem that would otherwise be intractable.

As we said, our results are very general, in the sense that they are applicable for variables taking values on finite or infinite spaces, and that we can also consider precise or imprecise representations. We make nevertheless some assumptions, like the logical independence of the variables studied or the representation of our assessments through a functional defined on a sufficiently large domain. In Section 9, we comment on the extent to which these assumptions can be relaxed. This is an important problem in order to relate our work more tightly with other areas of research. Finally, we conclude the paper in Section 10 with some additional discussion. To make the paper easier to read, we have relegated all the proofs to an Appendix.

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 [33] for an in-depth study of these and other properties, and for an interpretation of the notions we shall introduce below.

Given a possibility space Ω, a *gamble* is a bounded real-valued function on Ω.[2] 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.

We can also consider the supremum acceptable buying prices for a gamble *conditional* on a subset of Ω. Given such a set B and a gamble f on Ω, the lower

[2] Although we only deal in this paper with *bounded* gambles, an extension of the theory to unbounded random variables can be found in [30].

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 variables, which are the focus of our attention in this paper. Consider 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) 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 $\mathcal{X}^n$ for $\mathcal{X}_{\{1,\dots,n\}}$. 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 $\pi_J : \mathcal{X}^n \to \mathcal{X}_J$ 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 $\mathcal{X}_J$. We shall denote by $\mathcal{K}_J$ the set of $\mathcal{X}_J$-measurable gambles.

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., on the set $\mathcal{K}_{O \cup I}$ of the $\mathcal{X}_{O \cup I}$-measurable gambles on $\mathcal{X}^n$. 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 $\mathcal{X}^n$. 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.

This type of conditional previsions is what we are going to consider throughout the paper. We refer to [22, 33] 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. A definition of conditional previsions where we do not necessarily condition on a partition can be found in [35].

The $\mathcal{X}_I$*-support* $S(f)$ of a gamble f in $\mathcal{K}_{O \cup I}$ is given by

$$S(f) := \{\pi_I^{-1}(x) : x \in \mathcal{X}_I, f\mathbb{I}_{\pi_I^{-1}(x)} \neq 0\}, \tag{1}$$

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, $\mathbb{I}_A$ 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 $\mathbb{I}_{\pi_I^{-1}(x)}(f - \underline{P}(f|x))$, and by $G(f|X_I)$ the gamble that takes the value $G(f|\pi_I(y))$ for any $y \in \mathcal{X}^n$.

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 lower previsions. Then we should verify that all the assessments modelled by these conditional lower previsions are consistent with one another. 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*. In this case, where the domain is a linear set of gambles, separate coherence holds if and only if the following conditions are satisfied for any $x \in \mathcal{X}_I$, $f, g \in \mathcal{K}_{O\cup I}$, and $\lambda > 0$:

(SC1) $\underline{P}(f|x) \geq \inf_{y\in\pi_I^{-1}(x)} f(y)$.
(SC2) $\underline{P}(\lambda f|x) = \lambda \underline{P}(f|x)$.
(SC3) $\underline{P}(f+g|x) \geq \underline{P}(f|x) + \underline{P}(g|x)$.

It shall also be interesting for this paper to consider the particular case where $I = \emptyset$, that is, when we have (unconditional) information about the variables X_O. We have then an *(unconditional) lower prevision* $\underline{P}(X_O)$ on the set $\mathcal{K}_O$ of $\mathcal{X}_O$-measurable gambles. Then separate coherence is simply called *coherence*, and it holds if and only if the following three conditions hold for any $f, g \in \mathcal{K}_O$, and $\lambda > 0$:

(C1) $\underline{P}(f) \geq \inf f$.
(C2) $\underline{P}(\lambda f) = \lambda \underline{P}(f)$.
(C3) $\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g)$.

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 one another. Formally, we are going to consider what we shall call *collections* of conditional lower previsions.

Definition 2. Let $\{\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})\}$ be conditional lower previsions with respective domains $\mathcal{K}^1,\dots,\mathcal{K}^m \subseteq \mathcal{L}(\mathcal{X}^n)$, where $\mathcal{K}^j$ is the set of $\mathcal{X}_{O_j\cup I_j}$-measurable gambles,[3] for $j = 1,\dots,m$. Then this is called a *collection* on $\mathcal{X}^n$ when for each $j_1 \neq j_2$ in $\{1,\dots,m\}$, either $O_{j_1} \neq O_{j_2}$ or $I_{j_1} \neq I_{j_2}$.

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. Given a collection $\{\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})\}$ of conditional lower previsions, there are different ways in which we can guarantee their consistency.

Definition 3. Let $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ be separately coherent conditional lower previsions. We say that they *avoid uniform sure loss* if and only if

$$\sup_{x\in\mathcal{X}^n} [\sum_{j=1}^m G_j(f_j|X_{I_j})](x) \geq 0,$$

for every $f_j \in \mathcal{K}^j$, $j = 1,\dots,m$.

[3]We use $\mathcal{K}^j$ instead of $\mathcal{K}_{O_j\cup I_j}$ in order to alleviate the notation when no confusion is possible about the variables involved.

Definition 4. Let $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ be separately coherent conditional lower previsions. We say that they *avoid partial loss* when for every $f_j \in \mathcal{K}^j$, $j = 1,\dots,m$ there is some $B \in \bigcup_{j=1}^m S_j(f_j)$ such that

$$\sup_{x\in B}[\sum_{j=1}^{m} G_j(f_j|X_{I_j})](x) \geq 0, \tag{2}$$

where $S_j(f_j)$ is the $\mathcal{X}_{I_j}$-support of f_j given by Eq. (1).

The notions of avoiding partial or uniform sure loss are minimal consistency requirements that we shall use in Section 8 to connect our work with probabilistic logic; nevertheless, the main focus in this paper will be made on some stronger notions, that we shall call weak and strong coherence:

Definition 5. Let $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ be separately coherent conditional lower previsions. We say that they are *weakly coherent* when for any $f_j \in \mathcal{K}^j$, $j = 1,\dots,m$, $j_0 \in \{1,\dots,m\}$, $f_0 \in \mathcal{K}^{j_0}, z_0 \in \mathcal{X}_{I_{j_0}}$,

$$\sup_{x\in \mathcal{X}^n}[\sum_{j=1}^{m} G_j(f_j|X_{I_j}) - G_{j_0}(f_0|z_0)](x) \geq 0. \tag{3}$$

Under the behavioural interpretation, a number of weakly coherent conditional lower previsions can still present some forms of inconsistency with one another; see [33, Example 7.3.5] for an example and [33, Chapter 7] and [34] for some discussion. Because of this, we consider yet a stronger notion, called (*joint* or *strong*) coherence:[4]

Definition 6. Let $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ be separately coherent conditional lower previsions. We say that they are *coherent* when for every $f_j \in \mathcal{K}^j$, $j = 1,\dots,m$, $j_0 \in \{1,\dots,m\}$, $f_0 \in \mathcal{K}^{j_0}, z_0 \in \mathcal{X}_{I_{j_0}}$, there exists some $B \in \{\pi^{-1}_{I_{j_0}}(z_0)\} \cup \bigcup_{j=1}^m S_j(f_j)$ such that

$$\sup_{x\in B}[\sum_{j=1}^{m} G_j(f_j|X_{I_j}) - G_{j_0}(f_0|z_0)](x) \geq 0, \tag{4}$$

where, again, $S_j(f_j)$ is the $\mathcal{X}_{I_j}$-support of f_j given by Eq. (1).

The coherence of a collection of conditional lower previsions implies their weak coherence; although the converse does not hold in general, it does in the particular case when we only have a conditional and an unconditional lower prevision. Note for instance that the conditional previsions in Example 1 in the Introduction are not coherent: if we consider $f_1 := \mathbb{I}_{\{(1,1),(2,2)\}}$ and $f_2 := \mathbb{I}_{\{(2,1),(1,2)\}}$, then $G(f_1|X_2) + G(f_2|X_1) < 0$ in the union of the supports, which is the set $\{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2)\}$.

In the next section, we prove a number of results that will help to better understand the differences between weak and strong coherence. But before we do that, we introduce a special case that will be of 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 $f, g \in \mathcal{K}_{O\cup I}$. When a separately coherent conditional lower

[4] The distinction between this and the unconditional notion of coherence mentioned above will always be clear from the context.

prevision $\underline{P}(X_O|X_I)$ is linear we shall denote it by $P(X_O|X_I)$; in the unconditional case, we shall use the notation $P(X_O)$. A separately coherent unconditional linear prevision corresponds to the expectation operator (the Dunford integral [4]) with respect to a finitely additive probability.

If we consider 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, and they are weakly coherent if and only if they avoid uniform sure loss [33, Section 7.1, page 347].

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 $x\in\mathcal{X}_I$ (this is a consequence of [33, Sections 6.7.2 and 6.7.4]). We shall denote this set of dominating conditional linear previsions by $\mathcal{M}(\underline{P}(X_O|X_I))$.

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 $x\in\mathcal{X}_I$,

(JC1) $\underline{P}(G(f|X_I))\geq 0$
(JC2) $\underline{P}(G(f|x))=0$.

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 $P(f)=P(P(f|X_I))$.

Before concluding this section, it is important to remark that if a lower prevision $\underline{P}$ is coherent with a conditional lower prevision $\underline{P}(\cdot|\mathcal{B})$, then $\underline{P}$ must satisfy the property of *conglomerability*, which is discussed in some detail in [33, Section 6.8]. This property is one of the points of disagreement between Walley's and de Finetti's [13] approach to conditional previsions.

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, for any $j=1,\dots,m$,*

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

In particular, conditional linear previsions $P_1(X_{O_1}|X_{I_1}),\dots,P_m(X_{O_m}|X_{I_m})$ are weakly coherent if and only if there exists a coherent prevision P on $\mathcal{X}^n$ such that $P(f)=P(P_j(f|X_{I_j}))$ for any f in $\mathcal{K}^j$, $j=1,\dots,m$.

Remark 1. When all the conditional previsions are linear and 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, because of how these conditions deal with the problem of conditioning on sets of probability zero. For instance, the conditional previsions in Example 1 are weakly coherent because they are compatible with the linear prevision determined by the assessment $P(X_1 = 3, X_2 = 3) = 1$. We shall come back to this in Section 8.2. ♦

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. *Let $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ be conditional lower previsions. They 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 two results allow us to understand a bit better the conceptual difference between weak coherence and (strong) coherence: from Corollary 1 and Theorem 1, weak coherence amounts to the existence of a joint that is pairwise coherent with each of the conditional lower previsions; from Proposition 1, coherence means that there is a joint that is coherent with all the conditional lower previsions, *taken together*.

This difference is perhaps easier to grasp in the particular case where we deal with finite spaces and with linear conditional previsions. In that case, weak coherence is equivalent to the existence of a linear prevision (the expectation with respect to a finitely additive probability) that induces each of the conditional previsions by means of Bayes' rule. But this does not guarantee that the conditional previsions are coherent, because the joint mentioned above might not be coherent with all of them *considered as a whole*. That is, the conditional previsions may provoke some behavioural inconsistencies when taken together even if they can all be induced from the same joint. This is due to the fact that when this joint gives zero probability to a set B, then any conditional prevision $P(\cdot|B)$ is coherent with the joint.

4. Collection templates

With this paper we should like to deliver tools to prove coherence that are sufficiently general to be applied in most situations. To do this, we have to assume very little about the conditional lower previsions that are the subject of our study. In particular, we are not going to assume anything about the numbers that make up the lower previsions themselves, other than they produce separately coherent assessments. We do require separate coherence as it is really a minimal requirement of self-consistency for a conditional lower prevision.

Abstracting away from the numbers implies that for each lower prevision we only know, apart from its being separately coherent, what are the variables on both sides of the conditioning bar. This can be regarded as the 'form' of a conditional lower prevision, or its *template*, as we call it in the following.

Definition 7. Let $\underline{P}_{j_1}(X_{O_{j_1}}|X_{I_{j_1}})$ and $\underline{P}_{j_2}(X_{O_{j_2}}|X_{I_{j_2}})$ be two lower previsions on $\mathcal{X}^n$. We say that they *have the same template* if $O_{j_1} = O_{j_2}$ and $I_{j_1} = I_{j_2}$. The class of all the lower previsions on $\mathcal{X}^n$ with the same template is just called a *lower prevision template on* $\mathcal{X}^n$ (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 $\underline{P}_j(X_{O_j}|X_{I_j})$.

We can identify a template with a disjoint pair of indices. A collection template is then determined by a finite number of such pairs. This is formalised in the following definition.

Definition 8. Two collections of lower previsions on $\mathcal{X}^n$ *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 $\{1, \ldots, m\}$ the two respective j-th lower previsions have the same template. The class of all the collections on $\mathcal{X}^n$ with the same template is just called a *collection template on* $\mathcal{X}^n$ (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 $\{\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})\}$.

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 prevision templates in a collection template.

The definitions just introduced allow us to state the task of this paper more precisely: i.e., to characterise what we know about the coherence of a collection of (separately coherent) lower previsions once we know its template alone. To this extent, it is useful to introduce a graphical characterisation of a collection template. This is done in the next section.

5. Coherence graphs

In this section, we introduce a graphical representation of collection templates based on directed graphs. 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 *endpoints*. 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*, respectively). When the origin and destination coincide, we say that the path is a *directed cycle*, or just a cycle, for short. If a cycle does not contain any proper cycle, then it is said to be *elementary*. 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.

It is useful to introduce also the notion of *strong component* of a directed graph. This is a *maximal strongly connected subgraph*, where a strongly connected graph is one for which there is a path for each ordered pair of nodes. A strong component is said to be trivial if it is made by a single node.

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 *neighbours*. 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 9. 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 $< \mathcal{Z}, \mathcal{D}, \mathcal{A} >$ is called a *coherence graph* on $\mathcal{Z}$ if the following properties hold:

(CG1) $\mathcal{Z}$ is non-empty.
(CG2) All neighbours 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.

Fig. 1 displays the coherence graph of the assessments

$$\underline{P}_1(X_1), \underline{P}_2(X_4|X_1), \underline{P}_3(X_6|X_5), \underline{P}_4(X_7|X_2), \underline{P}_5(X_7|X_3), \underline{P}_6(X_8|X_3),$$
$$\underline{P}_7(X_8|X_4), \underline{P}_8(X_9|X_1, X_5), \underline{P}_9(X_{10}|X_{13}), \underline{P}_{10}(X_{11}|X_7), \underline{P}_{11}(X_{12}|X_8),$$
$$\underline{P}_{12}(X_{13}|X_{14}), \underline{P}_{13}(X_{14}|X_6, X_{10}), \underline{P}_{14}(X_{15}, X_{16}|X_9, X_{12}, X_{13}).$$

Here the actual nodes are $X_1, \dots, X_{16}$. 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 create a problem since by CG5 each dummy node is univocally identified by its neighbours); 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).

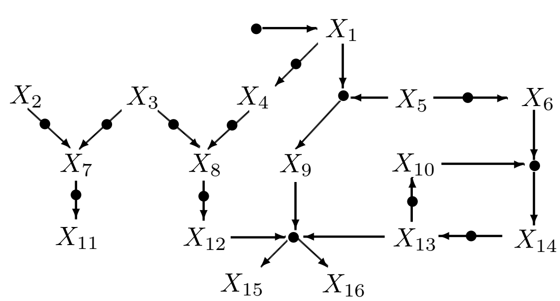

FIGURE 1. The coherence graph for $\underline{P}_1, \dots, \underline{P}_{14}$.

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 $\mathcal{X}^n$. To this extent, it is useful to isolate the notion of a D-structure in a coherence graph.

Definition 10. Given a dummy node D of a coherence graph, we call *D-structure* the subgraph whose nodes are D and its neighbours, and whose arcs are those connecting D to its neighbours.

At this point we consider two functions: the first one, which 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:

(Γ1) Let $\mathcal{Z} := \{X_1, \ldots, X_n\}$ be the set of actual nodes.
(Γ2) Let $\mathcal{D} := \{D_1, \ldots, D_m\}$ be the set of dummy nodes.
(Γ3) Let $\mathcal{A} := \emptyset$.
(Γ4) For all $j \in \{1, \ldots, m\}$, all $i_1 \in I_j$, all $i_2 \in O_j$, add the arcs (X_{i_1}, D_j) and (D_j, X_{i_2}) to $\mathcal{A}$.

The second function, which we denote by Γ^{-1}, maps a coherence graph on $\mathcal{Z} = \{X_1, \ldots, X_n\}$, with dummy nodes $\{D_1, \ldots, D_m\}$, into the collection template

$$\{\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})\}$$

related to the variables $\{X_1, \ldots, X_n\}$. This mapping is determined by the following procedure:

(Γ^{-1}1) Set the collection of lower prevision templates equal to the empty set.
(Γ^{-1}2) For all $j \in \{1, \ldots, m\}$, 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. This makes the two functions to be each other's inverses as it is established in the following proposition, whose elementary proof is omitted:

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

This proposition enables us to graphically display some basic configurations of collection templates that are problematic with respect to the coherence of the collection. One such configuration is created by collection templates whose coherence graph contains an actual node with more than one parent, such as X_8 in Fig. 1. In this case there are two different conditional lower previsions, $\underline{P}_6(X_8|X_3)$ and $\underline{P}_7(X_8|X_4)$, that express knowledge about X_8. In this situation it is not possible to deduce the coherence of the collection only taking its template into account. The reason is that it is always possible to find a specific instance of lower previsions with the given template that is incoherent. This is enough because any claim that we make based only on the template must be valid, by definition, for all the collections of lower previsions with the considered template. For instance, consider $x_1^8 \neq x_2^8$ in $\mathcal{X}_8$, and define, for any gamble f on $\mathcal{X}_8$, $\underline{P}_6(f|x_3) := f(x_1^8)$ for all $x_3 \in \mathcal{X}_3$, and $\underline{P}_7(f|x_4) := f(x_2^8)$ for all $x_4 \in \mathcal{X}_4$. This specific choice corresponds to use (precise) degenerate distributions that put all the probability mass on x_1^8 and x_2^8, respectively, and this irrespective of their conditioning events. It follows that $\underline{P}_6(X_8|X_3)$ implies that $X_8 = x_1^8$ and $\underline{P}_7(X_8|X_4)$ that $X_8 = x_2^8$, a contradiction.[5]

Another problematic configuration arises out of collection templates whose coherence graph contains a cycle; this is the case of the actual nodes X_{10}, X_{13}, X_{14},

[5] In this example, as well as in some of the proofs in the Appendix, we use 0-1 valued probabilities to make things simpler, but we can equivalently use probabilities that are never degenerate as above. In this example, for instance, define $P_6(f|x_3) := 0.1f(x_1^8) + 0.9f(x_2^8)$ for all $x_3 \in \mathcal{X}_3$, and $P_7(f|x_4) := 0.1f(x_2^8) + 0.9f(x_1^8)$ for all $x_4 \in \mathcal{X}_4$. This corresponds to using probability masses on (x_1^8, x_2^8) equal to $(0.1, 0.9)$ and $(0.9, 0.1)$, respectively, irrespective of their conditioning events. To see that these two conditional previsions are not compatible with any joint mass function P on $\mathcal{X}_{\{3,4\}}$, note that given any such P we should deduce on the one hand that $P(\{x_1^8\}) = 0.1$ and on the other that $P(\{x_1^8\}) = 0.9$, a contradiction. Also the following example based on cycles can be re-worked, with some additional complications, using only non-degenerate probabilities.

in Fig. 1. In this case we can create a contradiction by defining the lower previsions involved in the cycle as follows. Consider $x_1^{10} \neq x_2^{10}$ in $\mathcal{X}_{10}$, $x_1^{14} \neq x_2^{14}$ in $\mathcal{X}_{14}$, $x_1^{13} \neq x_2^{13}$ in $\mathcal{X}_{13}$. Let

$$\underline{P}_9(f|x^{13}) := \begin{cases} f(x_1^{10}) & \text{if } x^{13} = x_1^{13} \\ f(x_2^{10}) & \text{otherwise,} \end{cases}$$

$$\underline{P}_{13}(g|x^6, x^{10}) := \begin{cases} g(x_1^{14}) & \text{if } x^{10} = x_1^{10} \\ g(x_2^{14}) & \text{otherwise,} \end{cases}$$

$$\underline{P}_{12}(h|x^{14}) := \begin{cases} h(x_1^{13}) & \text{if } x^{14} = x_2^{14} \\ h(x_2^{13}) & \text{otherwise,} \end{cases}$$

for arbitrary gambles f on $\mathcal{X}_{10}$, g on $\mathcal{X}_{14}$, h on $\mathcal{X}_{13}$. These definitions again correspond to use degenerate (and precise) distributions: for example, $\underline{P}_9(X_{10}|X_{13})$ corresponds to a distribution that puts all the probability mass on x_1^{10} if $X_{13} = x_1^{13}$, and on x_2^{10} otherwise; in other words, $\underline{P}_9(X_{10}|X_{13})$ models the fact that $X_{10} = x_1^{10}$ holds with probability one assuming it is known that $X_{13} = x_1^{13}$, and otherwise, if $X_{13} \neq x_1^{13}$, that $X_{10} = x_2^{10}$, also with probability one. Analogously, $\underline{P}_{13}(X_{14}|X_6, X_{10})$ states that $X_{14} = x_1^{14}$ resp. $X_{14} = x_2^{14}$ hold with probability one provided that $X_{10} = x_1^{10}$ resp. $X_{10} \neq x_1^{10}$; finally, $\underline{P}_{12}(X_{13}|X_{14})$ that $X_{13} = x_1^{13}$ resp. $X_{13} = x_2^{13}$ hold with probability one provided that $X_{14} = x_2^{14}$ resp. $X_{14} \neq x_2^{14}$.

At this point it is easy to see that the cycle gives rise to a contradiction. Assume that $X_{13} = x_1^{13}$. Using the above considerations, this implies with probability one that $X_{10} = x_1^{10}$, then that $X_{14} = x_1^{14}$, and finally that $X_{13} = x_2^{13}$, a contradiction. Repeating the argument starting with $X_{13} \neq x_1^{13}$, we obtain that $X_{13} = x_1^{13}$, another contradiction. In summary, the cycle, together with the specific choice of lower previsions, codes the contradictory statement that X_{13} should be equal to two different values. And since we have been able to find some specific lower previsions with the given template that are incoherent, then we cannot deduce coherence only considering the related coherence graph—just because it contains a cycle.

These considerations motivate the following two definitions, which introduce some graph-based terminology that is more directly relevant to our subsequent results.

Definition 11. We say that an actual node of a coherence graph is a (potential) *source of contradiction* if it has more than one parent or if it belongs to a cycle.

More formally, if an actual node X_ℓ has more than one parent, this means that there are $1 \leq i_1 \neq i_2 \leq m$ such that $\ell \in O_{i_1} \cap O_{i_2}$. On the other hand, the fact that X_ℓ belongs to a cycle implies that it is involved in (at least) an elementary cycle, and so that there are different $j_1, \dots, j_p$ in $\{1, \dots, m\}$ such that

$$O_{j_1} \cap I_{j_2} \neq \emptyset, O_{j_2} \cap I_{j_3} \neq \emptyset, \dots, O_{j_p} \cap I_{j_1} \neq \emptyset, \tag{5}$$

and with ℓ belonging to one of these non-empty intersections.

Definition 12. 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 Fig. 1 is obviously not of type A1, as there are five sources of contradiction: X_7 and X_8, because each of them has two parents; and X_{10}, X_{13} and X_{14}, because they are part of a cycle. An example of an A1 graph is given in Fig. 2. This is a subgraph of that in Fig. 1 where we have eliminated a number of elements creating sources of contradiction. This also shows that A1 graphs can take complicated forms.

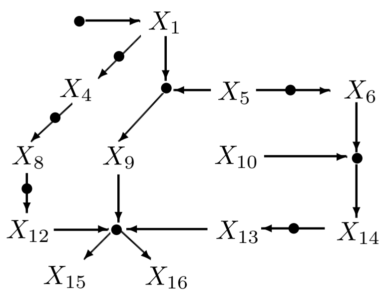

FIGURE 2. An example of an A1 coherence graph.

When there are sources of contradiction in a coherence graph, it is useful to isolate some special subgraphs that are related to them, and which we call blocks.

Definition 13. 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.

The reason why we introduce blocks is to capture the idea of the 'area of influence' of a certain source of contradiction. This is easy to see in the case of an actual node with more than one parent, such as X_8 in Fig. 1. In this case we can again use degenerate distributions in such a way that $\underline{P}_6(X_8|X_3)$ forces X_8 to take a certain value x_1^8; that $\underline{P}_7(X_8|X_4)$ forces X_4 to take a certain value x_1^4 when $X_8 = x_1^8$, and in turn, that $\underline{P}_2(X_4|X_1)$ forces X_1 to take a certain value x_1^1 when $X_4 = x_1^4$. In this case, exploiting the source of contradiction, we can force a specific value on a node such as X_1 that can be far away from the source itself. This is useful in the proofs to show how to use a source of contradiction to actually create a contradiction somewhere: in the example above, for instance, we can define $\underline{P}_1(X_1)$ so as to force X_1 to take a value different from x_1^1, thus originating a contradiction at X_1. The situation is a bit more complicated with cycles, but the effect is eventually the same.

It is useful to study a bit more in detail the situations just sketched. For this, we are going to introduce the notion of *constraining sub-block* for an actual node in a block. This will be a subset of the previsions of the block with the property that we can assign them certain values for which there is a unique value in the actual node compatible with them. In the previous example the constraining sub-block for X_1 would be determined by the previsions $\underline{P}_6(X_8|X_3)$, $\underline{P}_7(X_8|X_4)$ and $\underline{P}_2(X_4|X_1)$.

In a constraining sub-block, as we shall show in Proposition 3, the information flows (possibly opposite to the directions of the arcs) so as to eventually force an actual node to take on a chosen value. This property will be used in Theorem 4 to decompose the task of verifying weak coherence in an optimal way.

Consider first of all a block B_Z originated by an actual node X_s with more than one parent. Then there exist $1 \leq i_1 \neq i_2 \leq m$ such that $s \in O_{i_1} \cap O_{i_2}$.

For any actual node X_ℓ in B_Z, there exists a path leading from X_ℓ to the node X_s originating the block. Hence, there are $i_1,\dots,i_p$ in $\{1,\dots,m\}$ such that $s \in O_{i_1} \cap O_{i_2}, I_{i_2} \cap O_{i_3} \neq \emptyset, \dots, I_{i_{p-1}} \cap O_{i_p} \neq \emptyset$, and $\ell \in I_{i_p}$. Note that we can assume without loss of generality that the nodes $\{i_1,\dots,i_p\}$ are all different: otherwise, we could establish a cycle in the path from X_ℓ to X_s, and by eliminating these cycles we should obtain another (shorter) path going from X_ℓ to X_s where all the indices are different. We shall refer to a set of indices $\{i_1,\dots,i_p\}$ such that $s \in O_{i_1} \cap O_{i_2}, I_{i_2} \cap O_{i_3} \neq \emptyset, \dots, I_{i_{p-1}} \cap O_{i_p} \neq \emptyset, \ell \in I_{i_p}$ as an *the constraining sub-block* for X_ℓ in the block generated by X_s. Note that such a sub-block is not necessarily unique.

Consider now a block B_Z which is generated by a cycle. Then this cycle corresponds to a strong component in the coherence graph, in the sense that for any two nodes X_{s_1} and X_{s_2} in the component there is a path going from X_{s_1} to X_{s_2} and another one going from X_{s_2} to X_{s_1}; hence, X_{s_1} and X_{s_2} belong to a cycle.

Again, we have two possibilities: that a node X_s in the block B_Z belongs to the strong component of the block, or that it is a predecessor of this strong component. In the first case, X_s belongs to an elementary cycle, meaning that there exist $j_1,\dots,j_p \in \{1,\dots,m\}$ satisfying Eq. (5) and $s \in O_{j_1} \cap I_{j_2}$. Consider on the other hand a predecessor X_ℓ of this source of contradiction. Then X_ℓ is a predecessor of the nodes in an elementary cycle, which we denote by X_s, and it is not itself in the cycle. Let $j_1,\dots,j_p$ be the indices of the previsions in the cycle, meaning that they satisfy Eq. (5), with $s \in O_{j_1} \cap I_{j_2}$. Then there exist $k_1,\dots,k_r$ such that $\ell \in O_{k_1} \cap I_{k_2}, O_{k_2} \cap I_{k_3} \neq \emptyset, \dots, O_{k_{r-1}} \cap I_{k_r} \neq \emptyset$, and where k_r is one of the indices in the cycle, for instance $k_r = j_1$.

The indices $j_1,\dots,j_p$ are all different, because the cycle is elementary. Moreover, we can also assume that the indices $\{k_1,\dots,k_{r-1}\}$ are different from $\{j_1,\dots,j_p\}$: otherwise, we can eliminate the indices they have in common and we should still have a path from X_ℓ to the cycle. Finally, we can assume without loss of generality that the indices $\{k_1,\dots,k_{r-1}\}$ are all different, because otherwise we should be able to establish a cycle in them, and by eliminating these cycles we should obtain another (shorter) path from X_ℓ to the cycle. We shall refer to the indices $\{k_2,\dots,k_{r-1},j_1,\dots,j_p\}$ as a *constraining sub-block* for X_ℓ in the block associated to the source of contradiction X_s. Again, such a constraining sub-block may not be unique.

Remark 2. Note that in this constraining sub-block we have two possibilities:

- $O_{k_{r-1}} \cap I_{k_r} \cap O_{j_p} \neq \emptyset$, and then in particular $O_{k_{r-1}} \cap O_{j_p} \neq \emptyset$. Then X_ℓ is a predecessor of a node with more than one parent, and $\{k_2,\dots,k_{r-1},k_r,j_p\}$ would also be a constraining sub-block for X_ℓ (see an example in Fig. 3).
- If $O_{k_{r-1}} \cap I_{k_r} \cap O_{j_p} = \emptyset$, then X_ℓ is a predecessor of a dummy node in the cycle, and only precedes the actual nodes in the cycle through this dummy node (see an example in Fig. 4).

The distinction between these two cases will simplify the proofs of our subsequent results. ♦

As mentioned previously, the reason why we have defined the notion of constraining sub-block is because it can be used to determine the value of the variable X_ℓ. This is stated in the following proposition:

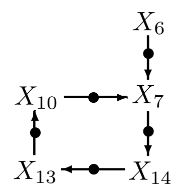

FIGURE 3. Constraining sub-block for X_6 in the block associated to the source of contradiction, here a node with two predecessors.

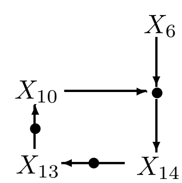

FIGURE 4. Constraining sub-block for X_6 in the block associated to the source of contradiction, in this case a cycle.

Proposition 3. *Let us consider a block B_Z, and any actual node X_ℓ that belongs to B_Z. Consider $x \in \mathcal{X}_\ell$.*

(1) If X_ℓ is a source of contradiction in the block B_Z, let j_1 be an element of $\{1, \ldots, m\}$ such that $\ell \in O_{j_1}$. Then there are weakly coherent previsions $P_1(X_{O_1}|X_{I_1}), \ldots, P_m(X_{O_m}|X_{I_m})$ such that any prevision P which is coherent with $P_{j_1}(X_{O_{j_1}}|X_{I_{j_1}})$ satisfies $P(x) = 1$.

(2) If X_ℓ is not a source of contradiction in B_Z, let $\{j_1, \ldots, j_k\}$ be the indices of the previsions in a constraining sub-block for X_ℓ. Then there are $P_1(X_{O_1}|X_{I_1}), \ldots, P_m(X_{O_m}|X_{I_m})$ weakly coherent such that any prevision P coherent with $P_{j_i}(X_{O_{j_i}}|X_{I_{j_i}})$ for $i = 1, \ldots, k$ satisfies $P(x) = 1$.

An important consequence of this proposition is that if two blocks happen to share an actual node, then it is possible to create a contradiction at that node by forcing a certain value on it from one block, and a different value on it from the other block. This suggests that the blocks that share some actual nodes should be considered as a single structure, rather than as separate ones, in order to avoid contradictions; this is the reason for the following definition.

Definition 14. Call *superblock* of a coherence graph, any union of all the blocks that share at least an actual node.

Fig. 5 displays the three different blocks of the coherence graph under consideration: the block for X_7, the block for X_8, and the one for X_{14} (or, equivalently, for X_{10} or X_{13}). The blocks for X_7 and X_8 have a node in common (X_3), so their union is in a superblock, also shown in the figure. The other superblock in the graph is just $B_{X_{14}}$.

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 (for instance $B_{X_{14}}$ coincides

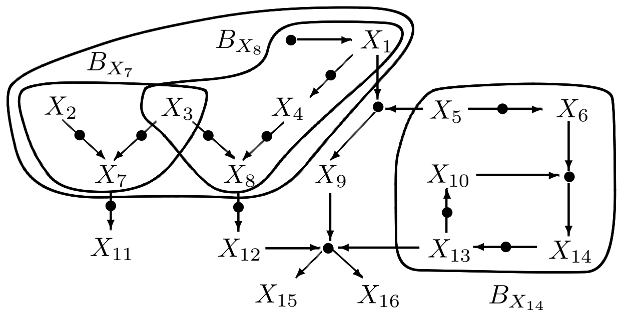

FIGURE 5. The areas delimited by closed lines denote the blocks and superblocks of the coherence graph.

with $B_{X_{10}}$ and $B_{X_{13}}$), that one of them is included in another, or that two of them share only some nodes (as B_{X_7} and B_{X_8} in the same figure).

It is useful to make two observations at this point. The first is that the intuition behind the notion of superblock made up by different blocks is to delimit the joint area of influence of multiple sources of contradictions that belong to different blocks connected by some actual nodes. The second is more formal and concerns the actual nodes of different blocks and their relation with superblocks: in fact, for any block B_Z we can consider the set $A_Z := \cup_{i \in B_Z}(I_i \cup O_i)$ of the actual nodes involved in B_Z; then, if $A_{Z_1} \cap A_{Z_2} \neq \emptyset$, B_{Z_1} and B_{Z_2} belong to the same superblock. It follows from this that if we consider two different superblocks B_1 and B_2, then $(I_i \cup O_i) \cap (I_j \cup O_j) = \emptyset$ for any $i \in B_1, j \in B_2$.

Now we use the notion of superblock in order to build a partition of the dummy nodes. The point here is that, similarly to the case of blocks, sources of contradictions can extend their influence to every actual node in a superblock under an opportune definition of the lower previsions involved. Because of this, it is not possible to prove the coherence of the lower previsions in a superblock by simply inspecting the coherence graph: to prove it, it is necessary to know something more than the bare collection template. It follows that superblocks are a kind of core entities in that we cannot prove the coherence of a collection template without first being able to prove the coherence of the lower previsions in each superblock. Those core entities constitute the elements of the partition defined below, together with the lower previsions that do not belong to any superblock.

Definition 15. 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 $\{1, \ldots, m\}$ 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 Fig. 5, we obtain the following partition of the related collection template:

$$\{\{\underline{P}_1(X_1), \underline{P}_2(X_4|X_1), \underline{P}_4(X_7|X_2), \underline{P}_5(X_7|X_3), \underline{P}_6(X_8|X_3), \underline{P}_7(X_8|X_4)\},$$
$$\{\underline{P}_3(X_6|X_5), \underline{P}_9(X_{10}|X_{13}), \underline{P}_{12}(X_{13}|X_{14}), \underline{P}_{13}(X_{14}|X_6, X_{10})\},$$
$$\{\underline{P}_8(X_9|X_1, X_5)\}, \{\underline{P}_{10}(X_{11}|X_7)\}, \{\underline{P}_{11}(X_{12}|X_8)\}, \{\underline{P}_{14}(X_{15}, X_{16}|X_9, X_{12}, X_{13})\}\}.$$

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, and has therefore m elements.

We conclude this section by remarking that if we consider B_1, B_2 in the minimal partition with $|B_1| > 1, |B_2| = 1$, i.e., such that B_1 is associated to a superblock and B_2 to an A1 component of the coherence graph, we must have $O_i \cap (I_j \cup O_j) = \emptyset$ for any $i \in B_2, j \in B_1$: if $O_i \cap O_j \neq \emptyset$ we have a node with more than one parent, and i should belong to the associated superblock; and if $O_i \cap I_j \neq \emptyset$ then i would be a predecessor of j, which is a predecessor of a source of contradiction, and therefore i should be in the same block as j. Nevertheless, we may have $I_i \cap (I_j \cup O_j) \neq \emptyset$ for some $j \in B_1$. This will be important in the proofs of the results we formulate in the next section.

6. Coherence graphs as tools to prove coherence

This section formally relates the graphical notions introduced in the previous section with the notions of weak and strong coherence, and doing so gives the most important results of this paper.

In particular, 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 $\mathcal{B}$ be coherent.

Theorem 2. *Let $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ be a collection of separately coherent conditional lower previsions, and let $\mathcal{B}$ be their associated minimal partition. If $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in B}$ are coherent for any $B \in \mathcal{B}$, then 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.*

The intuition behind the proof of the theorem, which we include in the Appendix, is the following. We exploit the properties of the coherence graph to create a total order on a set of coherent lower previsions tightly related to our collection template. That order allows us to use the generalisation of the *marginal extension theorem* established in [23] 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 that a similar result holds when we work with weak coherence instead of coherence:

Theorem 3. *Let $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ be a collection of separately coherent conditional lower previsions such that 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 15 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 that respect is that the partition $\mathcal{B}$ is indeed minimal when we are studying the problem for weak coherence:

Theorem 4. *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 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.

However, a similar result to Theorem 4 does not apply for coherence: there are 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.

Example 2. Let us consider the collection template associated to the conditional lower previsions $\{\underline{P}_1(X_1),\underline{P}_2(X_2|X_1),\underline{P}_3(X_2,X_3|X_1)\}$. Its coherence graph is given in Fig. 6, and the minimal partition $\mathcal{B}$ associated to that graph is $\{1,2,3\}$. However,

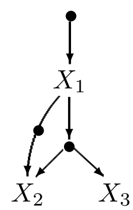

FIGURE 6. Coherence graph of $\{\underline{P}_1(X_1),\underline{P}_2(X_2|X_1),\underline{P}_3(X_2,X_3|X_1)\}$.

we can deduce the coherence of the collection template with a smaller subset. For this, let us show that the coherence of $\underline{P}_2(X_2|X_1),\underline{P}_3(X_2,X_3|X_1)$ implies (actually, it is equivalent to the fact) that for any $\mathcal{X}_1\times\mathcal{X}_2$-measurable gamble f_0 and for any $x_1\in\mathcal{X}_1$,

$$\underline{P}_2(f_0|x_1)=\underline{P}_3(f_0|x_1). \tag{6}$$

To see this, apply Eq. (4) to $\underline{P}_2(X_2|X_1)$ and $\underline{P}_3(X_2,X_3|X_1)$ using $j_0:=2$, $z_0:=x_1\in\mathcal{X}_1$, $f_2:=0$, and $f_3:=f_0\mathbb{I}_{\{x_1\}\times\mathcal{X}_{\{2,3\}}}$:

$$\begin{aligned}0\le&\sup_B[G_2(0|X_1)+G_3(f_0\mathbb{I}_{\{x_1\}\times\mathcal{X}_{\{2,3\}}}|X_1)-G_2(f_0|x_1)]\\=&\sup_{\pi_1^{-1}(\{x_1\})}[G_3(f_0\mathbb{I}_{\{x_1\}\times\mathcal{X}_{\{2,3\}}}|x_1)-G_2(f_0|x_1)]\\=&\sup_{\pi_1^{-1}(\{x_1\})}[f_0-\underline{P}_3(f_0|x_1)-f_0+\underline{P}_2(f_0|x_1)]\\=&-\underline{P}_3(f_0|x_1)+\underline{P}_2(f_0|x_1),\end{aligned}$$

i.e., $\underline{P}_2(f_0|x_1)\ge\underline{P}_3(f_0|x_1)$. The converse inequality, $\underline{P}_2(f_0|x_1)\le\underline{P}_3(f_0|x_1)$, follows by repeating the same argument with $j_0:=3$, $z_0:=x_1$, $f_2:=f_0\mathbb{I}_{\{x_1\}\times\mathcal{X}_2}$ and $f_3:=0$.

At this point, assume that $\underline{P}_2(X_2|X_1),\underline{P}_3(X_2,X_3|X_1)$ are coherent, and hence that Eq. (6) holds. Consider the expression used to prove the coherence of the previsions $\underline{P}_1(X_1)$, $\underline{P}_2(X_2|X_1)$, and $\underline{P}_3(X_2,X_3|X_1)$: i.e.,

$$\begin{cases}G_1(f_1)+G_2(f_2|X_1)+G_3(f_3|X_1)-G_{j_0}(f_0|x_1) & \text{if } j_0\in\{2,3\}\\ G_1(f_1)+G_2(f_2|X_1)+G_3(f_3|X_1)-G_{j_0}(f_0) & \text{if } j_0=1,\end{cases}$$

for any $f_j\in\mathcal{K}^j$, $j=1,2,3$, $j_0\in\{1,2,3\},f_0\in\mathcal{K}^{j_0}$ and $x_1\in\mathcal{X}_1$; verify that G_2 can be replaced by G_3 under (6). As a result, we obtain that the conditional lower previsions $\underline{P}_1(X_1)$, $\underline{P}_2(X_2|X_1)$ and $\underline{P}_3(X_2,X_3|X_1)$ are coherent if and only if $\underline{P}_1(X_1), \underline{P}_3(X_2, X_3|X_1)$ are. But $\underline{P}_1(X_1), \underline{P}_3(X_2, X_3|X_1)$ are always coherent because of the marginal extension theorem [33, Theorem 6.7.2], and so 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. ♦

It remains an open problem at this stage to determine a partition with the property that the coherence within each of its elements 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, the most interesting particular case would be under which conditions this minimal partition is equal to $\{\{1\}, \ldots, \{m\}\}$, that is, when we can deduce the coherence of $\{\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ from their separate coherence. We can deduce from Theorem 2 that when the coherence graph associated to $\{\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ is of type A1, then the separate coherence of these previsions implies their coherence. Using Theorem 4, we can prove that being of type A1 is also necessary for this property.

Proposition 4. *Consider a collection* $\{\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ *of separately coherent conditional lower previsions. The following are equivalent:*

(1) *The separate coherence of* $\{\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ *implies their coherence.*
(2) *The separate coherence of* $\{\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ *implies their weak coherence.*
(3) *The coherence graph of this collection template is of type A1.*

We should like to conclude this section remarking that if the collection template is A1-representable, then we can give the following 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.*

Hence, when the coherence graph is A1, we can interpret our coherent conditional lower previsions as a model for the imprecise knowledge of some precise coherent conditional 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 [33, Section 6.6.10].[6] A sufficient condition for the converse to hold is that the spaces $\mathcal{X}_1, \ldots, \mathcal{X}_n$ are finite [33, Theorem 8.1.10]. 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. An algorithm to find the minimal partition

In order to exploit coherence graphs as tools to check coherence, one should be able to compute the minimal partition of a coherence graph. This is what we set out to do in the following: we give an algorithm to find the minimal partition and then discuss its computational complexity, thus showing the efficiency of the algorithm.

[6]The previsions in the example given in that section can be written in our language as $\underline{P}(X_1, X_2), \underline{P}(X_2|X_1)$. Note that they both belong to the same block, and consequently their coherence graph is not A1.

The rationale behind the algorithm is very simple. We start a visit of the graph from each source of contradiction, going backwards with respect to the direction of the arrows, so as to identify the related block. This is done by a recursive procedure that tags the nodes found in the visit, assigning different tags to blocks originated by different sources of contradiction. The only complication is that some blocks might have non-empty intersection, and as a consequence they should be regarded as a single superblock. When this happens, the tags of the different blocks should be regarded as the same. To this extent, we implement a data structure that acts as a dictionary, i.e., which maps the tags of the blocks in the same superblock into a unique tag, referred to as the *true* tag, which can be regarded as the class of equivalence of those tags. The dictionary is filled by a procedure during the visits of the graph every time a node is found at the intersection of two or more blocks.

Below we describe the algorithm more precisely in a C-like language (as opposed to C, we assume that the first element of an array has index 1 to make the code simpler to read). To make things simpler, we assume that the nodes of the graph have been re-indexed from 1 to $m+n$, where the first m nodes are $D_1, \dots, D_m$ and the following ones are the actual nodes. We also assume that, as a result of previous computations, the following *global* data structures are available:

- two integer numbers, `m` and `n`, corresponding to the number of dummy and actual nodes, respectively.
- An array called `node`, of size $m+n$, whose generic element `node[i]` is a structure that contains the following components related to node `i`:
  - an integer called `node[i].nParents` containing the number of parents of node `i`;
  - an integer array called `node[i].parent`, of size `node[i].nParents`, containing the indexes of the parents of node `i`;
  - an integer called `node[i].tag` initialised to `0` to denote that node `i` is not tagged.
- An integer called `nContradictions` containing the number of actual nodes that are sources of contradictions.
- An integer array called `contradiction`, of size `nContradictions`, containing the indexes of the actual nodes that are sources of contradictions.
- An integer array called `minPartition`, of size m, which implements the dictionary.

Fig. 7 reports the software code to find the minimal partition based on such data structures.

The code has three procedures: `findMinPartition`, which is the main one; its subroutine `findBlock`, with parameters `i` and `currentTag`; and `mergeBlocks`, with parameters `tagFound` and `currentTag`, which is a subroutine of `findBlock`.

The purpose of `findMinPartition` is to assign tags to the dummy nodes and fill the array `minPartition` in such a way that for any $j_1, j_2 \in \{1, \dots, m\}$, j_1 and j_2 belong to the same element of the minimal partition if and only if the tags `minPartition[node[`j_1`].tag]` and `minPartition[node[`j_2`].tag]` coincide.

`findMinPartition` works in two steps. In the first, up to and including line `07`, the procedure enumerates the sources of contradiction and, when one is found that is not tagged, calls `findBlock` to tag it and all its predecessors, thus identifying a block. `findBlock` also takes care, by means of `mergeBlocks`, to merge blocks with non-empty intersection into a superblock by properly filling the array

```
01 void findMinPartition(){ // identify blocks and superblocks
02   int currentTag=0, i, l;
03   for(l=1;l<=nContradictions;++l) // enumerate sources of contradiction
04     if(!node[contradiction[l]].tag){ // contradiction not tagged yet?
05       findBlock(contradiction[l],++currentTag); // give tag to nodes in block
06       minPartition[currentTag]=currentTag; // update dictionary
07     }
08   for(i=1;i<=m;++i) // enumerate dummy nodes
09     if(!node[i].tag){ // that are not tagged, i.e., not in any superblock
10       node[i].tag=++currentTag; // give (newly created) tag
11       minPartition[currentTag]=currentTag; // update dictionary
12     }
13 }
14
15 void findBlock(int i, int currentTag){ // identify block for given actual node
16   int j, k;
17   if(!node[i].tag){ // node not tagged yet?
18     node[i].tag=currentTag; // give tag
19     for(j=1;j<=node[i].nParents;++j) // enumerate parents
20       if(!node[node[i].parent[j]].tag){ // parent not tagged yet?
21         node[node[i].parent[j]].tag=currentTag; // tag it
22         for(k=1;k<=node[node[i].parent[j]].nParents;++k) // enumerate its parents
23           findBlock(node[node[i].parent[j]].parent[k],currentTag); // recursion
24       }else if(minPartition[node[node[i].parent[j]].tag]!=currentTag) // tagged?
25         mergeBlocks(minPartition[node[node[i].parent[j]].tag],currentTag); // merge
26   }else if(minPartition[node[i].tag]!=currentTag) // (already) tagged?
27     mergeBlocks(minPartition[node[i].tag],currentTag); // merge (the two blocks)
28 }
29
30 void mergeBlocks(int tagFound, int currentTag){ // two tags must become one
31   int l;
32   for(l=1;l<currentTag;++l) // enumerate existing (true) tags
33     if(minPartition[l]==tagFound) // any tag ‘l’ previously mapped into ‘tagfound’
34       minPartition[l]=currentTag; // is now mapped into ‘currentTag’
35 }
```

FIGURE 7. The software code in C-like language to find the minimal partition of a coherence graph.

`minPartition`, so that at line `08` of `findMinPartition` all the superblocks have been identified. In the second step, from line `08` onwards, the procedure simply considers the dummy nodes that are not in any superblock and tags them with new and increasing values of `currentTag`, thus identifying the remaining elements of the minimal partition.

The way `findBlock` works is more complicated. Its main purpose is to tag the actual node `i` and all its predecessors by `currentTag` in a recursive way. It may happen that some of the nodes found during the visit in the graph are already tagged (this happens at line 25 if the node found is dummy and at line 29 if it is actual). In this case, if the true tag found is different from the current one, `findBlock` calls `mergeBlocks` to merge the two blocks related to the two different tags. `mergeBlocks` does so by identifying the tags with each other.

Let us recall at this point that to be really in the condition of computing the minimal partition, we should also have methods to fill the global data structures mentioned at the beginning of this section, before running `findMinPartition`. These methods are actually trivial to implement apart from the one that solves the problem of identifying the actual nodes involved in cycles.

A key observation to address such a problem is that a node belongs (at least) to a cycle in a directed graph if and only if it belongs to a *non-trivial strong component* of the graph. This implies that identifying the strong components enables one to find the actual nodes involved in at least a cycle. Fortunately, the task of identifying the strong components is well known and efficiently tackled by Tarjan's algorithm in [29], so that also this part of the problem can be addressed easily.

The complexity of the overall procedures, including the methods to fill the global data structures, is given in the following theorem.

Theorem 6. *The worst-case complexity to compute the minimal partition using the procedures in Fig. 7 is bounded by* $O(m + n + m \cdot n)$.

Note that this bound is derived under the implicit assumption that the input of the problem is a coherence graph, but usually one would rather start with a collection template. This is not problematic because the lower prevision templates in a collection are in one-to-one correspondence with the D-structures of the corresponding coherence graph, as stated in Section 5, so that converting a collection template into a coherence graph takes linear time and thus it does not increase the complexity bound derived above.

8. Applications

Now that the main results of this paper have been presented, we can show how they naturally relate with three important models and tools well known in artificial intelligence. In Section 8.1 we shall use our results to prove for the first time the coherence, to a large extent, of the graphical models called *Bayesian* and *credal networks*. In Section 8.2, we shall establish a tight relationship between weak coherence and the so-called *compatibility problem*: i.e., checking whether a number of assessments admits a compatible joint probabilistic model. In this case, we shall show that coherence graphs allow us to *optimally* decompose a compatibility problems into smaller ones, for a very general version of the problem. Finally, in Section 8.3, we focus on a recently proposed coherence-based version of *probabilistic satisfiability* that enhances and extends similar problems in *probabilistic logic*. In this case we first give new results that detail the connection between these more traditional approaches and Walley's theory. Then we show that coherence graphs can be used also in this case to decompose some important instances of probabilistic satisfiability into smaller ones.

8.1. **Coherence of Bayesian and credal networks.**

Let us focus on proving the coherence of the graphical models called *Bayesian nets* and their extension to imprecise probability called *credal nets*.

Although it may seem surprising at first, these models have not been shown yet to be coherent in a strong sense. What is well known is that these models give rise to a joint coherent lower prevision (in our language, this means that the assessments that make up those models are pairwise coherent with such a joint). For example, a Bayesian net is equivalent to a joint mass (or density) function over the considered variables; a credal net is equivalent to a closed convex set of such joint mass functions (the joint coherent lower prevision is the lower envelope of the expectations computed from those mass functions). This is often implicitly taken as evidence that the models under consideration are self consistent. But, as we have shown in Theorem 1, the existence of such a joint is only equivalent to weak coherence; and weak coherence leaves room to inconsistencies so that Bayesian and credal nets could still express inconsistent beliefs, for all we know.

In Section 8.1.2 we show that this is not the case as the mentioned models satisfy Walley's notion of strong coherence. We do this using coherence graphs and some additional developments introduced in Section 8.1.1 that are needed to connect coherence graphs with the so-called notion of *strong product*, a generalisation of stochastic independence to imprecise probability. The connection is necessary as the strong product underlies the models under consideration. Finally, in Section 8.1.3 we discuss some extensions of the presented results to more general models of credal nets and also to some major statistical applications.

8.1.1. *Relating A1 graphs to the strong product.* In this section we summarise some results that connect A1 graphs with a notion of probabilistic independence called *strong independence* (see [8, 24]) and that have been introduced in [37].

We start with a lemma[7] that shows that A1 graphs naturally entail a notion of order of the corresponding lower previsions: in particular, that it is possible to permute the indexes of the lower previsions in such a way that the only admissible paths between two dummy nodes are those in which the index of the origin precedes that of the destination.[8]

Lemma 1. [37, Lemma 1] *If the coherence graph associated to the collection template $\{\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})\}$ is A1, then we may assume without loss of generality that for any $k=1,\dots,m$, $O_k\cap(\cup_{i=1}^{k-1}I_i)=\emptyset$.*

Note that this result depends only on the properties of the coherence graph, and therefore it is applicable also when dealing with precise conditional previsions.

Now we restrict the attention to the special case of A1 graphs originated by collections of separately coherent conditional lower previsions such that $\{O_1,\dots,O_m\}$ forms a partition of $\{1,\dots,n\}$. This means that for each variable there is exactly one lower prevision expressing beliefs about it. Let us call this type of graphs A1$^+$. From Lemma 1, if a collection template $\{\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})\}$ is $A1^+$-representable, we can assume without loss of generality that $I_1=\emptyset$. Let us define $A_1:=\emptyset, A_j:=\cup_{i=1}^{j-1}(I_i\cup O_i)$ for $j=2,\dots,m+1$, and for $j=1,\dots,m$ let $\underline{P}'_j(X_{O_j}|X_{A_j\cup I_j})$ be given on the set $\mathcal{H}^j$ of $\mathcal{X}_{A_{j+1}}$-measurable gambles by

$$\underline{P}'_j(f|z):=\underline{P}_j(f(z,\cdot)|\pi_{I_j}(z))$$

[7]This lemma will be used not only in this section but also in the proofs of the Appendix for other reasons, through Lemma 6.

[8]This order notion is similar to the graph-theoretic notion of *topological ordering*, but here it is applied only to the dummy nodes.

for any $z \in \mathcal{X}_{A_j \cup I_j}$ and any $f \in \mathcal{H}^j$. Since $\underline{P}_j(X_{O_j}|X_{I_j})$ is separately coherent for $j = 1, \dots, m$, so is $\underline{P}'_j(X_{O_j}|X_{A_j \cup I_j})$. Moreover, thanks to Lemma 1 and the requirement that $\{O_1, \dots, O_m\}$ forms a partition of $\{1, \dots, n\}$, the sets of indices of the conditional variables in the previsions $\underline{P}'_1(X_{O_1}), \dots, \underline{P}'_m(X_{O_m}|X_{A_m \cup I_m})$ form an increasing sequence and hence they satisfy the hypotheses of the generalised marginal extension theorem established in [23, Theorem 4]. As a consequence, $\underline{P}'_1(X_{O_1}), \dots, \underline{P}'_m(X_{O_m}|X_{A_m \cup I_m})$ are also coherent.

A similar reasoning shows that if we take for $j = 1, \dots, m$ a conditional linear prevision $P'_j(X_{O_j}|X_{A_j \cup I_j})$ on the set $\mathcal{H}^j$ that dominates $\underline{P}'_j(X_{O_j}|X_{A_j \cup I_j})$, then $P'_1(X_{O_1}), \dots, P'_m(X_{O_m}|X_{A_m \cup I_m})$ are jointly coherent. Moreover, taking into account that $\{O_1, \dots, O_m\}$ is a partition of $\{1, \dots, n\}$, Theorem 3 in Ref. [23] implies that the *only* prevision P on $\mathcal{X}^n$ which is coherent with the assessments $P'_1(X_{O_1}), \dots, P'_m(X_{O_m}|X_{A_m \cup I_m})$ is

$$P(f) = P'_1(P'_2(\dots(P'_m(f|X_{A_m \cup I_m}))\dots)). \tag{7}$$

In other words, $P'_1(X_{O_1}), \dots, P'_m(X_{O_m}|X_{A_m \cup I_m})$ give rise to a unique joint lower prevision. When $\mathcal{X}_1, \dots, \mathcal{X}_n$ are finite, it can be checked that P is the prevision associated to the probability mass function

$$P(x) = \prod_{j=1}^{m} P'_j(\pi_{O_j}(x)|\pi_{A_j \cup I_j}(x)). \tag{8}$$

At this point we are ready to give the definition of lower envelope model.

Definition 16. Suppose that for each $\lambda \in \Lambda$, the collection of conditional lower previsions $\{P_\lambda(X_{O_1}), \dots, P_\lambda(X_{O_m}|X_{A_m \cup I_m})\}$ is s.t. for all $j = 1, \dots, m$, $P_\lambda(X_{O_j}|X_{A_j \cup I_j})$ dominates $\underline{P}'_j(X_{O_j}|X_{A_j \cup I_j})$ and also $\min_{\lambda \in \Lambda} P_\lambda(X_{O_j}|X_{A_j \cup I_j}) = \underline{P}'_j(X_{O_j}|X_{A_j \cup I_j})$. The coherent lower prevision $\underline{P}$ defined as $\underline{P} := \min_{\lambda \in \Lambda} P_\lambda$, where P_λ is the coherent prevision determined by $P_\lambda(X_{O_1}), \dots, P_\lambda(X_{O_m}|X_{A_m \cup I_m})$ and Eq. (7), is called a *lower envelope model.*

Intuitively, a lower envelope model is a joint lower prevision that is built out of a number of conditional and unconditional assessments. The interest in lower envelope models arises because it is a very common practice to build joint models out of smaller conditional and unconditional ones, and then to use the joint model to draw some conclusions. Lower envelope models abstract this procedure of constructing joint models in the general case of coherent lower previsions. As particular cases of lower envelope models, we can consider the following:

(1) If for each $j = 1, \dots, m$ we consider all the $P_\lambda(X_{O_j}|X_{A_j \cup I_j})$ in the set $\mathcal{M}(\underline{P}'_j(X_{O_j}|X_{A_j \cup I_j}))$, then the lower prevision $\underline{P}$ is the *marginal extension* of $\underline{P}'_1(X_{O_1}), \dots, \underline{P}'_m(X_{O_m}|X_{A_m \cup I_m})$.
(2) If for $j = 1, \dots, m$ we take all the $P_\lambda(X_{O_j}|X_{A_j \cup I_j})$ in the set of extreme points of $\mathcal{M}(\underline{P}'_j(X_{O_j}|X_{A_j \cup I_j}))$, with the additional requirement that $P_\lambda(X_{O_j}|z) = P_\lambda(X_{O_j}|z')$ if $\pi_{I_j}(z) = \pi_{I_j}(z')$, then the lower envelope model $\underline{P}$ is called the *strong product* of $\underline{P}_1(X_{O_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$.

The marginal extension represents the most conservative lower envelope model built out of the assessments defined in (7). The strong product is also the most conservative lower envelope model built out of those assessments with the additional assumption of strong independence (see for instance [37] for additional information).

Theorem 7. [37, Theorem 2] *Consider an* $A1^+$*-representable collection template* $\{\underline{P}_1(X_{O_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})\}$*, and let* $\underline{P}$ *be a lower envelope model associated to it. Then* $\underline{P}$*,* $\underline{P}_1(X_{O_1})$*, ...,* $\underline{P}_m(X_{O_m}|X_{I_m})$ *are coherent.*

Note that this result is concerned with the original assessments, not with those defined in (7). In particular, it implies that it is always coherent to build the strong product out of an $A1^+$-representable collection. This is the main tool that we shall use in the following.

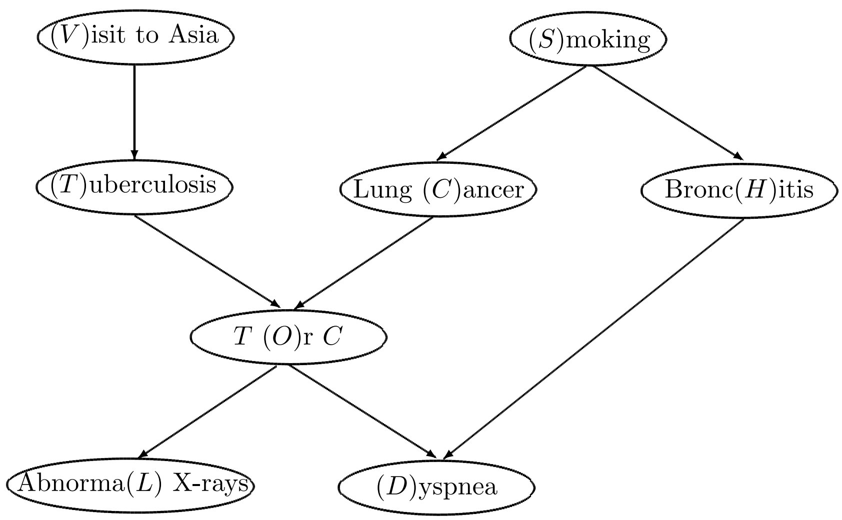

FIGURE 8. The Asia network: a model of an artificial medical problem related to the presence of dyspnea. The nodes correspond to variables (with names in parentheses), which are all 0-1 valued in this example.

8.1.2. *Bayesian and credal networks.* A *Bayesian net* [27] is made of a directed acyclic graph whose nodes are in one-to-one correspondence with the variables $X_1, \ldots, X_n$ (a well-known example of Bayesian net is given in Fig. 8). The arcs of the graph represent stochastic independences among the variables by means of the so-called *Markov condition*: i.e., the fact that a variable is independent of its non-descendant non-parents given its parents. Let X_j be the generic node in the net, and denote by I_j the set of indexes of its parents. Let us assume that the variables take values on finite spaces. In this case each node of the network is associated with a set of conditional mass functions that in our language correspond to the conditional linear previsions $P_j(X_j|X_{I_j})$. Note also that $m = n$ with Bayesian nets. Specifying a Bayesian net is equivalent, through the Markov condition, to specifying a joint mass function over the variables of interest:[9]

$$P(x) = \prod_{j=1}^{n} P_j(\pi_j(x)|\pi_{I_j}(x)), x \in \mathcal{X}^n. \tag{9}$$

[9]Note that in Expression (9) the symbol π refers to the projection operator from Definition 1, and it should not be confused with the symbol used to denote the parents of a node in a Bayesian network.

Credal networks are an extension of Bayesian nets to imprecise probabilities [8]. The extension is achieved, in the case of the so-called *separately specified* credal nets, about which we focus, by allowing the generic node X_j to replace each of its mass functions with the closed convex hull of a finite number of conditional mass functions. In other words, the linear prevision $P_j(X_j|X_{I_j})$ is replaced by a coherent lower prevision $\underline{P}_j(X_j|X_{I_j})$ for all the nodes X_j in the net. As an example, the Asia network can be turned into a credal net by replacing each of its local conditional probabilities with a probability interval, as a closed convex set of mass functions is equivalent to a probability interval in the case of a binary variable.

A credal net is equivalent to a set of Bayesian nets: if one chooses a precise mass function in each local closed convex set of the net, the credal net becomes a Bayesian net that is 'compatible' with the credal net. In the language of coherent lower previsions this means that the graph, together with the choice of a linear prevision $P_j(X_j|X_{I_j}) \geq \underline{P}_j(X_j|X_{I_j})$ at node X_j, for all $j = 1, \dots, n$, is a Bayesian net. Therefore, a credal net can be regarded as the set of Bayesian nets originated by choosing the above dominating linear previsions in all the possible ways. Call $\mathcal{P}$ the set of joint mass functions obtained applying Eq. (9) to each compatible Bayesian net.

Credal nets are rendered consistent with Walley's theory of coherent lower previsions using the notion of *strong extension*, defined as

$$K(\mathcal{P}) := \overline{\mathrm{CH}}(\mathcal{P}),$$

where the symbol $\overline{\mathrm{CH}}$ denotes the operation of taking the closed convex hull. Since $K(\mathcal{P})$ is convex, it has extreme points, i.e., mass functions in $K(\mathcal{P})$ that cannot be expressed as convex combinations of other ones in $K(\mathcal{P})$. Let us denote the set of extreme points by $\mathrm{ext}(K(\mathcal{P}))$. Usually, the definition of credal networks requires that such a subset is finite. It is well known that in this case $\mathrm{ext}(K(\mathcal{P})) \subseteq \mathcal{P}$; this means that the extreme points correspond to a subset of the compatible Bayesian nets. For this reason, a credal net is usually regarded as equivalent to a *finite* set of Bayesian nets even if $\mathcal{P}$ has infinitely many elements. Note also that a Bayesian net is a special case of credal net. In such a case, the strong extension is a singleton containing the joint mass function coded by the Bayesian net through Eq. (9).

The strong extension is nothing else but an equivalent representation of a coherent lower prevision $\underline{P}(X_1, \dots, X_n)$, from which the connection with Walley's theory. To enforce this connection, inferences with credal nets are usually made with respect to the strong extension rather than the initial set $\mathcal{P}$. This should not be controversial because doing inference with credal nets is usually taken to be the computation of lower and upper posterior expectations; and these stay the same irrespective of the fact that one uses $\mathcal{P}$, $K(\mathcal{P})$ or $\mathrm{ext}(K(\mathcal{P}))$.

At this point we are ready to show that credal nets, and consequently Bayesian nets, are coherent models. To this aim, it is sufficient to show that a credal net leads to an A1$^+$ coherence graph (see Fig. 9 for an example) and that the strong extension of a credal net coincides with the strong product of the related coherence graph, as in the next theorem.

Theorem 8. *The local conditional lower previsions $\underline{P}_1(X_1|X_{I_1}), \dots, \underline{P}_n(X_n|X_{I_n})$ of a credal network are A1$^+$-representable. Their strong product coincides with the strong extension of the network.*

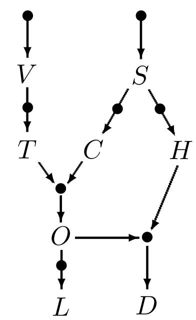

FIGURE 9. The coherence graph obtained from the Asia Bayesian network.

We can then apply Theorem 7 in a straightforward way to obtain the wanted result as an immediate corollary of the previous theorem.

Corollary 2. *Let* $\underline{P}_1(X_1|X_{I_1}),\dots,\underline{P}_n(X_n|X_{I_n})$ *be the local conditional lower previsions of a credal network, and let* $\underline{P}(X_1,\dots,X_n)$ *be its strong extension. Then* $\underline{P}_1(X_1|X_{I_1}),\dots,\underline{P}_n(X_n|X_{I_n})$ *and* $\underline{P}(X_1,\dots,X_n)$ *are coherent.*

8.1.3. *Some concluding remarks.* There are at least two reasons that make Corollary 2 important. One is that credal and Bayesian nets are very general and important models, and it is important to know that they are coherent, both for theoretical and practical reasons. Actually, exploiting tools presented elsewhere [37], it is possible also to give a stronger result: that coherence is preserved even under the so-called *updating* of a credal net, under very general conditions. This means that it is not possible to produce inconsistencies by building and repeatedly using a credal net.

The second reason is the (implicit) generality of the theorem. Remember that there are two basic limitations on the traditional definition of credal nets, on which we have focused so far: the strong extension is assumed to have a finite number of extreme points; and the variables to take finitely many possible values. Both are related to the definition of strong extension and hence it is useful, and even necessary for the second question, to extend such a definition. What seems to be the natural way to extend it, in our view, is just adopting in its place the definition of strong product given in Section 8.1.1. This choice allows us to propose, for the first time, a definition of credal nets for general spaces; moreover, it allows us, through Theorems 7 and 8, to immediately prove the coherence of credal nets in the general case (i.e., any kind of possibility spaces involved, and possibly infinitely many extreme points in the generalised strong extension), which appears to be an important outcome.

More generally speaking, we should like to point out that the approach used to prove the coherence of credal nets can be replayed with opportune changes also in other important contexts. For example, some recent work [37] has again exploited coherence graphs together with the strong product to prove the coherence of some very general statistical models: it has been shown that imprecise-probability based forms of statistical inference, such as some generalised *parametric inference* and *pattern classification*, are coherent. These models generalise Bayesian inference by using sets of *priors* to model prior knowledge (which can also correspond to a condition of *near-ignorance*) and use sets of *likelihood* functions to model very flexibly the possible presence of a process responsible for missing values. Also in this case, coherence graphs are a key tool to prove in a relatively easy way the coherence of the mentioned models, despite their generality and complexity.

A final point seems to be particularly worth of interest: both in the statistical case and with credal nets, the coherence graphs naturally originated are only of type A1. Remember that A1 graphs lead to coherence irrespective of the numbers that make up the related lower previsions (provided that they are separately coherent); in other words, in the case of A1 graphs coherence is a structural component of the collection. With more general graphs this is not necessarily the case: the related collection might no longer be coherent with very small changes in the numbers making up the lower previsions, in a way that might make the check of coherence problematic due to numerical instabilities. Therefore, it is interesting to observe that some of the most commonly used models in artificial intelligence and statistics have naturally been selected with the property that their coherence is relatively insensitive to the mentioned instabilities. On the other hand, this confirms the importance of A1 graphs.

8.2. **Compatibility of marginal and conditional probabilistic assessments.** The results in this paper allow us to provide new insight and solutions to the problem of the compatibility of a collection of marginal and conditional previsions. This problem has received a long-standing interest in the literature, since the seminal works by Boole [6], Hoeffding [16], Fréchet [14] and Vorobev [32]. See [9] and the references therein for recent works in the subject. It is related via Sklar's theorem to the notion of *copulas* [28], which has applications in economics [25].

Consider variables $X_1, \dots, X_n$ taking values in respective sets $\mathcal{X}_1, \dots, \mathcal{X}_n$. Consider $I_j, O_j \subseteq \{1, \dots, n\}$ such that $I_j \cap O_j = \emptyset$, for $j = 1, \dots, m$, and conditional lower previsions $\underline{P}_j(X_{O_j}|X_{I_j})$ for $j = 1, \dots, m$. As we have mentioned before, $\underline{P}_j(X_{O_j}|X_{I_j})$ models the information that the variable $X_{I_j} := (X_i)_{i \in I_j}$ provides about the variable $X_{O_j} := (X_i)_{i \in O_j}$. If in particular $I_j = \emptyset$, then $\underline{P}_j(X_{O_j})$ is simply the marginal information that we have about the variable X_{O_j}.

We formulate the compatibility problem, in a very general way, as studying whether there is an extension $\underline{P}$ on $\mathcal{X}^n$ of the (marginal and conditional) lower previsions in a collection $\{\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$. In the case of marginal lower previsions, this means obviously that $\underline{P}$ is an extension of them to the set of all gambles. The situation for conditional lower previsions is a bit more involved: we shall interpret compatibility of $\underline{P}$ and $\underline{P}(X_O|X_I)$ as coherence of these two lower previsions, in the sense considered throughout this paper (note that, as we have remarked in Section 2, in the particular case of a conditional and an unconditional assessment weak and strong coherence are equivalent).

Note that the study we make for lower previsions is of course also valid for the particular case where the assessments are precise, that is, when we have linear conditional and unconditional previsions $P_j(X_{O_j}|X_{I_j})$, $j = 1, \dots, m$. In such a case, we look for a *precise* joint, that is, for a linear prevision which is compatible with all these assessments. Coherence implies now (and, in the finite case, is equivalent to the fact) that the joint P induces all the conditionals by means of Bayes' rule when conditioning on an event of positive probability (see for instance [2, Chapter 10]). The more traditional formulation of the problem is a special case of the previous formulation obtained by restricting the attention to marginal previsions only, and amounts to study whether there is a compatible joint P on $\mathcal{X}^n$ whose O_j-marginal is $P(X_{O_j})$, for $j = 1, \dots, m$.

Our first, and important, result for the general compatibility problem is just Theorem 1 (see also Remark 1): from this theorem, we know that the lower previsions $\underline{P}_1(X_{O_1}|X_{I_1})$, ..., $\underline{P}_m(X_{O_m}|X_{I_m})$ are compatible if and only if they are weakly coherent. This makes of weak coherence a very general characterisation of compatibility. We are not aware of any other work that has given such a characterisation in the general case that we consider, although a similar result has been given in [20] for the particular situation made of unconditional linear previsions, and of finite sets $\mathcal{X}_1, \dots, \mathcal{X}_n$.

Theorem 1 is a key result also because it allows us to exploit in a direct way the outcomes from this paper, obtained for weak coherence, in the case of compatibility problems. For instance, we can use Proposition 4 to deduce that weak coherence, and hence compatibility, is implied by the separate coherence of $\underline{P}_1(X_{O_1}|X_{I_1})$, ..., $\underline{P}_m(X_{O_m}|X_{I_m})$ alone (i.e., without any other information about the numbers that make up the previsions) if and only if the coherence graph associated to this collection template is $A1$.[10] This means that if a collection of lower previsions is A1-representable, we immediately know that they are compatible; this gives us a graphical criterion that is sufficient for compatibility. The criterion becomes easier, and more obvious, if we focus on the simpler compatibility problem made only of unconditional lower previsions. In this case, the coherence graph cannot possess cycles; hence, it will be A1 if and only if there are no actual nodes with more than one parent, which in turn means that the sets $\{O_1, \dots, O_m\}$ must be pairwise disjoint.

The above graphical criteria may of course not be met in practice; in this case one should study the specific problem at hand in order to be able to check compatibility. Yet, without going into the detail of the numbers that make up the lower previsions, one can still take advantage of coherence graphs to *optimally* decompose a compatibility problem into simpler ones, i.e., those related to the subsets of the graph's minimal partition. Actually, this is the situation in which coherence graphs may prove to be more helpful, by having the potential to greatly reduce the complexity of the original task. Also, in our knowledge, this appears to be quite an original avenue for compatibility problems.

Let us stress also that these results are very general compared to the ones established in the classical works for the compatibility problem, because they can be applied to marginal and conditional *lower* previsions, instead of linear. Hence, they are also valid in situations of ambiguous or scarce information, where the use a precise probability model is not possible (or adequate). Furthermore, in the case of infinite spaces $\mathcal{X}_1, \dots, \mathcal{X}_n$, our work is based on finitely additive probabilities on the class of all subsets of $\mathcal{X}_{I_j \cup O_j}$, for $j = 1, \dots, m$, which allows us to avoid topological and measurability assumptions on the domains.

8.3. **Coherence-based probabilistic satisfiability.** The compatibility problem is close to a different and well-known problem called *probabilistic satisfiability* [15,

[10] Moreover, we have proven that when the coherence graph of the collection template is of type A1, we not only deduce that these previsions are weakly coherent, but also that they are (strongly) coherent. This allows us to give them a behavioural interpretation in terms of betting rates, as in the works by Walley [33] and de Finetti [13], and connects the compatibility problem naturally with decision making.

17]. A major difference is that latter is usually defined only relative to variables with finite support (it is often defined only with respect to events). In the following we shall therefore make this assumption, too.

Let us consider precise probabilities for a moment. In this case, probabilistic satisfiability consists in checking whether a number of precise conditional and unconditional probabilities is consistent with a joint. This is usually done via algorithms based on *linear programming*. Probabilistic satisfiability is tightly related to *probabilistic logic* [26]. In fact, the above check is often a first necessary step to be able to compute, again using linear programming, the probabilistic implications of the initial assessments on new events, which is the goal of probabilistic logic.

Probabilistic satisfiability has been extended to deal also with imprecise probabilistic assessments, and even with lower and upper previsions in some recent work by Walley, Pelessoni and Vicig [34] (from now on we refer to this paper by WPV for short). This work is somewhat atypical as it is based on a coherence-based view of probability rarely employed (another approach in a similar spirit is [5]) in probabilistic logic. And yet the authors of that paper show that relying on coherence is just the key to fix some of the problems of probabilistic logic and to develop truly general and powerful methods. These two characteristics make of WPV a natural candidate to bridge our results and probabilistic satisfiability.

To this aim, in Section 8.3.1 we first consider WPV in some detail. In particular, we consider the specific notions of consistency used there, and give new results that make it easy to move back and forth from those notions and the more traditional ones used in probabilistic logic. Moreover, we extend some of the theory for coherence graphs to show that they can decompose, in Section 8.3.2, some instances of the consistency problem of WPV into smaller ones. This is an important result because probabilistic satisfiability is an *NP-hard* problem [5]; when coherence graphs allow for reducing one such problem into smaller ones, we immediately obtain the possibility to solve bigger problem instances than it was possible before. (For an alternative, and possibly complementary approach, see [3]; in this case the focus is on heuristic considerations to reduce the computational burden.)

8.3.1. *Sufficient conditions for avoiding partial and uniform sure loss.* The WPV work is made of a first part related to the satisfiability problem and a second one for the extension of the assessments to new events or gambles. In the language of that paper, the first problem is one of checking whether the assessments *avoid uniform loss*; the second is the *natural extension* of the assessments.

These notions are taken from the seminal work of Williams [35] about lower previsions. In the setup that we shall consider, they are equivalent to the related ones from Walley [33, Section 7.1]. From this, it follows that avoiding uniform loss is a weaker requirement than (strong) coherence. We recall that in the case of finite spaces of possibilities, which we are considering, avoiding uniform loss is equivalent to the notion of avoiding partial loss, introduced in Section 2. The conclusion follows because coherence is indeed defined as a strengthening of avoiding partial loss.

At first it may look surprising that WPV focuses on a consistency notion that, being weaker than coherence, leaves room to inconsistencies in the original assessments. The reason is related to the second part of that work, i.e., the natural extension. Walley shows that for the natural extension to be well defined, it is necessary that the original assessments avoid partial loss. Remember that the natural extension is a procedure that allows one to compute the (tight) probabilistic implications of the original assessments on any new event or gamble, and in particular even on the original ones themselves. In this case such a procedure automatically corrects the possible inconsistencies left among the original assessments, making them (strongly) coherent. In WPV coherence is, in other words, a byproduct of the natural extension.

WPV discusses at different points the relationship with former approaches, especially those based on probabilistic logic, pointing to some of their problems. Here the key to understand the difference between the approaches is in the different notion of loss used. Unlike WPV, which is based on avoiding uniform loss (i.e., avoiding partial loss), most of the other approaches are based on the weaker notion of *avoiding uniform sure loss* given in Definition 3. To make the connection with the mentioned approaches more explicit, in the following proposition we give an equivalent formulation of avoiding uniform sure loss based on the existence of dominating conditional lower previsions which are weakly coherent.

Proposition 5. *Let* $\{\underline{P}_1(X_{O_1}|X_{I_1}),\ldots,\underline{P}_m(X_{O_m}|X_{I_m})\}$ *be a collection of separately coherent conditional lower previsions. If all the spaces* $\mathcal{X}_1,\ldots,\mathcal{X}_n$ *are finite, then they avoid uniform sure loss if and only if there exist weakly coherent dominating conditional linear previsions* $\{P_1(X_{O_1}|X_{I_1}),\ldots,P_m(X_{O_m}|X_{I_m})\}$.

This proposition can be easily given an intuitive meaning. The existence of dominating linear previsions means that there must be precise probabilities in the feasible set determined by the constraints (on conditional and unconditional probabilities) that define a satisfiability problem; in other words, the feasible set must be non-empty. But this is not enough: the fact that they are weakly coherent means, through Theorem 1 and Remark 1, that there must be a joint mass function from which one can derive, via Bayes' rule and marginalisations, those precise probabilities. Stated differently, if we use a number of conditional lower previsions as a means to express bounds on conditional probabilities, the property of avoiding uniform sure loss is equivalent to the existence of a joint mass function that satisfies all these bounds. This should make it clear that avoiding uniform sure loss is just the implicit condition that the more traditional approaches to probabilistic satisfiability try to test (these questions are discussed at some length in [34, Section 2.4]).

There is a further point of interest. It is possible to show that in a probabilistic satisfiability problem, avoiding uniform loss and avoiding uniform sure loss coincide if the (lower) probabilities of all the involved conditioning events are positive. This is related to the existence of the joint mass function mentioned above: in fact, since the joint mass function must be related to the conditional assessments of the problem via Bayes' rule, it turns out that there is no relation when Bayes' rule cannot be applied, i.e., when the (lower) probability of a conditioning event is zero. It follows that avoiding uniform sure loss does not capture specific inconsistencies in the original assessments that arise on top of those zero probabilities (such as in the example presented in the Introduction). This is, in particular, a source of criticism in WPV of the approaches based on avoiding uniform sure loss (see [34, Section 3.7] for a discussion about this point). It is indeed a considerable feature of the WPV approach that it does not suffer for these kinds of inconsistencies, nor can draw wrong conclusions because of them.[11]

At this point that we have some insight about the WPV approach and its relationship with ours and others, we can move to work more closely on the relationship between coherence graphs and WPV. Our final aim is to be able to use coherence graphs to simplify the check of avoiding uniform loss in WPV. The following result is what we need to reach our goal.

Theorem 9. *Let* $\{\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})\}$ *be a collection of separately coherent conditional lower previsions, and let* $\mathcal{B}$ *be their associated minimal partition, given by Definition 15.*

(1) If $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in B}$ *avoid partial loss for any* $B\in\mathcal{B}$*, then the conditional lower previsions* $\{\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})\}$ *avoid partial loss.*
(2) If $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in B}$ *avoid uniform sure loss for any* $B\in\mathcal{B}$*, then the conditional lower previsions* $\{\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})\}$ *avoid uniform sure loss.*

This theorem states that the properties of coherence graphs that we have investigated in the case of weak and strong coherence hold similarly also in the case of the losses under consideration. The similarity goes even further, as discussed in the next remark.

Remark 3. In the particular case where the coherence graph is A1, it follows from Theorem 9 that we can deduce that the conditional lower previsions avoid partial or uniform sure loss just from their separate coherence; to see that being A1 is also necessary for this property, it suffices to notice that the counterexamples in the proof of Theorem 4 are linear conditional previsions, for which weak coherence is equivalent to avoiding uniform sure loss and coherence is equivalent to avoiding partial loss. Hence, the partition is minimal in the case of A1 coherence graphs.

For general (not necessarily A1) coherence graphs, the partition may not be minimal if we want to deduce the avoiding partial loss condition: if we consider the previsions in Example 2, it is not difficult to show that if $\underline{P}(X_2|X_1)$ and $\underline{P}(X_2,X_3|X_1)$ avoid partial loss, then they also avoid partial loss with $\underline{P}(X_1)$. With respect to avoiding uniform sure loss, using again that the previsions in the counterexamples in the proof of Theorem 4 are linear conditional previsions, it follows that the partition is the minimal one from which we can deduce the avoiding uniform sure loss condition. ♦

8.3.2. *Bridging coherence graphs and probabilistic satisfiability.* We can finally exploit coherence graphs in WPV. We focus in particular on using WPV to the extent of checking whether a collection $\{\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})\}$ of separately coherent lower previsions avoids uniform loss. Recalling that we are considering finite spaces of possibilities, this can be re-phrased as using WPV to check whether a collection of closed and convex, conditional and unconditional, sets of mass functions, avoids uniform loss.

This task can be particularly onerous for WPV. Let us recall that WPV checks that the assessments avoid uniform loss by running a linear program (or more than

[11] For instance, in [34, Section 3.8, Example 10] it is presented a case in which the mentioned inconsistencies lead traditional probabilistic logic to deduce that a given event is certain a posteriori whereas the methods in WPV more correctly deduce that there is complete ignorance about it, i.e., that the posterior probability for such an event lies in the interval $[0,1]$.

one). But probabilistic satisfiability is an NP-hard problem. This means that in practice the size of the linear program, corresponding to [34, Eq. (1)], grows exponentially large with the size of the input. In the setup that we consider, this is a consequence of the number of linear constraints in the program which grows according to the size of the joint possibility space $\mathcal{X}^n$, i.e., as an exponential function of n.

Consider for example the collection of lower previsions that gives rise to the graph in Fig. 1. Say that each of the sixteen variables under consideration takes values on a set with three elements. Then the number of constraints in the program is about 43 millions, quite a prohibitive one.

But we know that we can apply Theorem 9 at this point. Theorem 9 allows us to decompose the above linear problem into two much smaller ones, according to the superblocks in Fig. 5: one superblock of six variables and another of five. That is, in two linear problems with about 700 and 240 constraints, respectively. In this example, coherence graphs make it possible to solve efficiently something that would be intractable otherwise.

This is not an isolated case: every time a coherence graph allows some proper superblocks to be isolated, the size of the largest linear program decreases according to an exponential function of the number of variables contained in the largest superblock: then, similarly to what happens also in the case of Bayesian nets, the computation is no longer exponential in a global feature of the model, such as the number of variables, but in a local one, which has the potential to lead to efficient solutions in a number of real-world problems. There is also another advantage: using smaller linear programs reduces the risks originated by numerical instabilities: both those involved in using collections that are more general than A1, as mentioned at the end of Section 8.1.2, and those more strictly related to the kinds of linear problems needed to check avoiding uniform loss, as reported in [34].

9. Some possible extensions of our results

This work is focused on the problem of the coherence of a collection of closed and convex sets of distributions. The formalism that we have introduced is consequently based on some constant related features. One is that we work with variables. Another one is that the joint space of possibilities $\mathcal{X}^n$ for these variables is the product of the spaces of the individual variables. A final one is that every lower prevision that we consider is defined on the set of all the gambles relative to the involved variables. Although this setup is the more general one for our aims, it is not the more general that one could consider.

In this section, we briefly investigate to what extent our results could be extended to more general setups. This could be useful in particular for problems of probabilistic satisfiability/logic. In fact, in such a field it is not uncommon to focus on problems where the lower previsions are not defined on all gambles; and also on problems where the joint space of possibilities $\mathcal{X}^n$ is not a product space because there are so-called *logical constraints* between the possible values of the variables under consideration that make some of the joint values impossible. This holds for WPV but also for more traditional approaches in artificial intelligence, e.g., [19, 31], as well as other approaches based on coherent probabilities [7].

Actually, the problems of probabilistic satisfiability are usually not even expressed in the language of previsions conditional on variables that we have considered in this paper, and are rather defined using events. Nevertheless, it is not difficult to consider variables that take values in the conditioning events in order to express everything in our language. Let us give an example of this:

Example 3. Let us consider three events A, B and C, and assume that we are given the probabilities $P(A|B), P(C|A \cap B), P(B|C)$. Define the variables X_1, X_2, X_3, where X_1 takes values in $\{B, B^c\}$, X_2 takes values in $\{A \cap B, A^c \cap B, A \cap B^c, A^c \cap B^c\}$ and X_3 takes values in $\{C, C^c\}$. In the language of this paper, the above assessments could be expressed as $P(X_2|X_1), P(X_3|X_2)$ and $P(X_1|X_3)$, where these previsions are defined, respectively, on $\mathbb{I}_A$, $\mathbb{I}_C$ and $\mathbb{I}_B$. In this case, the associated coherence graph is as in Fig. 10. Since we have a cycle in the coherence graph, we can only

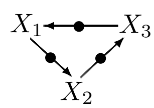

FIGURE 10. Coherence graph of $\{P(X_2|X_1), P(X_3|X_2), P(X_1|X_3)\}$.

know that these assessments are not always coherent. Note however that if we were given the probabilities $P(A|B), P(C|A \cap B)$ and $P(B)$, they would be expressed in terms of the above variables as $P(X_2|X_1), P(X_3|X_2)$ and $P(X_1)$. Their associated coherence graph is A1 and as a consequence they are always coherent. ♦

The following sections detail our investigation on the remaining problems. Section 9.1 deals with the case when the lower previsions $\underline{P}_j(X_{O_j}|X_{I_j})$, $j = 1, \dots, m$ are defined on domains $\mathcal{H}^j$ that are subsets of the sets $\mathcal{K}^j$ of $\mathcal{X}_{O_j \cup I_j}$-measurable gambles. Section 9.2 discusses the problem of logical dependence. We also report on a further possible extension in Section 9.3 that is concerned with the case when we know the coherence of some subsets of the collection of lower previsions, and these subsets do not make up a partition as we have always assumed so far. Finally, we give some concluding remarks in Section 9.4.

9.1. **Considering smaller domains.** One of the assumptions we have made in this paper is that the domain of the conditional lower prevision $\underline{P}_i(X_{O_i}|X_{I_i})$ is the whole set $\mathcal{K}^i$ of $\mathcal{X}_{O_i \cup I_i}$-measurable gambles, i.e., the set of gambles which depend on the value that the variables X_{O_i}, X_{I_i} take. Although when the domain of $\underline{P}_i(X_{O_i}|X_{I_i})$ is a subset $\mathcal{K}'^i$ of $\mathcal{K}^i$ we can always extend it to $\mathcal{K}^i$ using the procedure of natural extension, we think it is also of interest to discuss in some detail this case.

To simplify our reasoning, we are going to assume that $\mathcal{K}'^i$ is a linear subset of $\mathcal{K}^i$. It is not difficult to extend our reasoning to the case where $\mathcal{K}'^i$ is a non-linear set (for instance a finite set), using the work in [22].

The first important thing to remark here is that the sufficient conditions we give in this paper for coherence, weak coherence and avoiding partial and uniform sure loss in Theorems 2, 3 and 9 also hold if the domains are smaller. For this, it suffices to consider that the only assumptions in the domains we make in their proofs are that for any $f \in \mathcal{K}'^i$ and any subset A of $\mathcal{X}_{I_i}$, the gamble $f\mathbb{I}_A$ is also in $\mathcal{K}'^i$; but this can be assumed without loss of generality as a consequence of separate coherence [33, Lemma 6.2.4].

With respect to the necessary condition we establish in Theorem 4 for deducing weak coherence, it does not generalise to conditional lower previsions with arbitrary linear domains: if we think for instance of conditional lower previsions defined on constant gambles only, they are coherent as soon as they are separately coherent (i.e., as soon as they satisfy $\underline{P}_i(\mu|X_{I_i}) = \mu$ for any constant $\mu \in \mathbb{R}$). This holds irrespective of their coherence graph, so this does not need to be A1.

If we want Theorem 4 to hold in this more general situation, we need to impose some additional constraints on $\mathcal{K}'^i$. It can be checked that all the results hold if for all $i = 1, \dots, n$, $\underline{P}_i(X_{O_i}|X_{I_i})$ satisfies the following two conditions:

(SD1) For any subset A of $\mathcal{X}_{O_i \cup I_i}$, its indicator function $\mathbb{I}_A$ belongs to $\mathcal{K}'^i$.

(SD2) For all gambles $f \in \mathcal{K}'^i$ and all subsets A of $\mathcal{X}_{O_i \cup I_i}$, the gamble $f\mathbb{I}_A$ belongs to $\mathcal{K}'^i$.

Note that if $\underline{P}_i(X_{O_i}|X_{I_i})$ is separately coherent, we can assume without loss of generality that its domain $\mathcal{K}'^i$ includes all constant gambles. In that case, condition (SD1) is a consequence of (SD2). Note moreover that (SD2) cannot be seen as a consequence of separate coherence, because $\mathbb{I}_A$ may not only depend on the conditioning variables, but also on the conditional ones. The proof of Theorem 4 implies also that when conditions (SD1) and (SD2) hold, the partition we define is the minimal one which allows to deduce the property of avoiding uniform sure loss.

9.2. **Adding logical dependence considerations.** Another important issue for the applicability of our results is that of logical independence. In this paper, we have assumed that the variables $X_1, \dots, X_n$ are logically independent, meaning that we consider any combination $(x_1, \dots, x_n)$ in $\mathcal{X}^n$ as a possible value for the joint variable X^n. It is not uncommon, however, to consider variables that satisfy some kind of *logical dependence* assumption, which in the end will imply that some elements in $\mathcal{X}^n$ are 'structurally' impossible values for the joint variable $(X_1, \dots, X_n)$ and therefore they should be removed from the joint possibility space.

With respect to this, the first observation we have to make is that the partition of the conditional lower previsions associated to the superblocks is not enough to deduce weak or strong coherence (or avoiding partial or uniform sure loss) when we have in addition some logical dependence considerations; and this can happen even with finite spaces and precise conditional previsions, as we show in the following example:

Example 4. Consider two binary variables X_1, X_2, and take the previsions $P(X_2), P(X_1)$ determined by $P(X_2 = 1) = 1 = P(X_1 = 0)$. These previsions are coherent because their associated coherence graph is A1; however, they are incompatible with the logical dependence assumption $X_1 = X_2$, because any compatible model P will satisfy $P(X_1 = 0, X_2 = 1) = 1$. ♦

One way to model the fact that some of the joint values in $\mathcal{X}^n$ are impossible is to use unconditional lower previsions, by giving upper probability zero to the impossible combinations. If we know for instance that the variables X_1, X_2 can only assume together values in the subset A of $\mathcal{X}_{\{1,2\}}$, we can make the assessment $\underline{P}(A) = 1$. This can be expressed by the unconditional lower prevision $\underline{P}(X_1, X_2)$ on $\mathcal{K}_{\{1,2\}}$ given by $\underline{P}(f) = \inf_{\omega \in A} f(\omega)$. We discuss the use of this method together with coherence graphs below. There is a cautionary note, however: strictly speaking, the fact that some joint values are given zero upper probability is in general only one implication of logical dependence, in the sense that the converse is not necessarily true: an event with upper probability zero needs not be regarded as 'structurally' impossible. In the following we shall refer to the case where some joint values are given upper probability zero as *practical impossibility* to distinguish it from logical dependence. The following discussion leads then to sufficient conditions to check coherence (or loss) notions under practical impossibility. When the focus in on logical dependence instead, and the difference between this and practical impossibility is enforced in applications, our results below should be regarded as necessary conditions. Whether these are also sufficient conditions in general for logical dependence is an open problem that should be considered in future work.

As we said, we can model the practical impossibility of some combinations of values as additional unconditional lower previsions $\underline{P}'_1, \dots, \underline{P}'_k$. Then we should like to verify the coherence or weak coherence of the assessments $\underline{P}_1, \dots, \underline{P}_n, \underline{P}'_1, \dots, \underline{P}'_k$. For this, we can apply the reasoning in this paper and build the associated coherence graph. Its superblocks will produce a partition of the set of previsions with the property that it suffices to verify coherence (resp., weak coherence, avoiding partial loss, avoiding uniform sure loss) within each of the elements of the partition to immediately deduce coherence (resp., weak coherence, avoiding partial loss, avoiding uniform sure loss) of all the assessments.

Note that in order to do this the first thing we must verify is that we are still working with a collection of previsions. If a practical impossibility assumption is expressed in terms of an unconditional $\underline{P}'(X_1, X_2)$ and we already have another assessment $\underline{P}(X_1, X_2)$, then we should check whether also $\underline{P}(X_1, X_2)$ satisfies such an assumption; if it does not, then our assessments are not coherent. If it does, then we do not need to include $\underline{P}'(X_1, X_2)$ in our set of assessments.

Since we are working with more assessments now, the superblocks we have in the coherence graph will be bigger in general, and therefore the associated partition will be coarser. In the worst of cases, if we have a practical impossibility assumption that involves *all* the variables, it will be expressed as an unconditional $\underline{P}'(X_1, \dots, X_n)$, and all the previsions will belong to the same superblock. In that case the coherence graphs will not help to simplify the verification of coherence.

The other extreme case will be that when the practical impossibility assumptions involve only the variables which are already in the same superblock, as in the following example:

Example 5. Assume that we have variables $X_1, \dots, X_5$, and that we make the assessments $\underline{P}_1(X_1|X_2), \underline{P}_2(X_2|X_1), \underline{P}_3(X_5|X_3)$ and $\underline{P}_4(X_5|X_4)$. Their associated coherence graph is given in Fig. 11. If we make now the assumptions $X_1 = X_2$ and

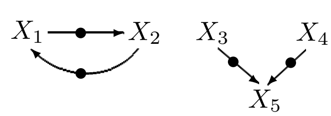

FIGURE 11. Coherence graph of the collection $\{\underline{P}_1(X_1|X_2), \underline{P}_2(X_2|X_1), \underline{P}_3(X_5|X_3), \underline{P}_4(X_5|X_4)\}$.

$X_3 \neq X_4$, they would be included in the graph by means of the unconditional previsions $\underline{P}_5(X_1, X_2), \underline{P}_6(X_3, X_4)$. The new coherence graph would be as in Fig. 12. We see that $\underline{P}_1(X_1|X_2), \underline{P}_2(X_2|X_1), \underline{P}_3(X_5|X_3)$ and $\underline{P}_4(X_5|X_4)$ are coherent and

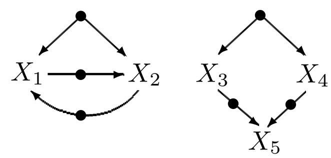

FIGURE 12. The coherence graph in Fig. 11 modified so as to add considerations of practical impossibility.

compatible with the practical impossibility assumptions if $\underline{P}_1(X_1|X_2), \underline{P}_2(X_2|X_1)$ are coherent and compatible with $X_1 = X_2$ on the one hand and $\underline{P}_3(X_5|X_3)$ and $\underline{P}_4(X_5|X_4)$ are coherent and compatible with $X_3 \neq X_4$ on the other. ♦

Hence, in some cases we only need to verify the practical impossibility assumptions within each of the superblocks, and we should be able to deduce coherence. We could expect that in a number of cases there will be a situation intermediate between the two extreme ones just presented: the superblocks will grow by adding practical impossibility considerations and the biggest one will not coincide with the entire graph. In these cases, coherence graphs will be useful as they will still permit to decompose the original problem in smaller ones to some degree.

The situation is a bit simpler if we focus on weak coherence instead of coherence. Assume that we have a number of weakly coherent conditional lower previsions, and that the practical impossibility considerations imply that only the values in $A \subseteq \mathcal{X}^n$ are acceptable for the joint variable $(X_1, \ldots, X_n)$. It can be checked that the unconditional lower prevision $\underline{E}$ defined in the proof of Theorem 1 is the smallest coherent lower prevision which is weakly coherent with $\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})$ [21]. If it also satisfies $\underline{E}(A) = 1$, then we deduce that the assessments are also compatible with the practical impossibility considerations, which is an alternative and easier sufficient condition to check. Using Proposition 5, it should be possible to simplify also the study of the compatibility of the avoiding uniform sure loss condition with practical impossibility considerations.

9.3. **Making non-disjoint assumptions of coherence.** In this paper, we have focused on the problem of deducing the coherence of a number of conditional lower previsions from the coherence of the previsions that belong to some sets. In our formulation, we have always assumed that we are given a partition of the set of conditional lower previsions. We have proven that a tool for verifying coherence or weak coherence can be to compare this partition to the minimal partition that we can derive using the superblocks of the coherence graph.

It would be interesting, however, to consider also the case where the information we are given is not made in terms of a partition, but that we are said that some sets of conditional lower previsions are coherent, and these sets are not disjoint. For instance, we may consider the assessments $\underline{P}(X_1, X_2), \underline{P}(X_2|X_1), \underline{P}(X_1|X_2)$, and we may be told that any two of these assessments are coherent. That is, we know that $\underline{P}(X_1, X_2), \underline{P}(X_2|X_1)$ are coherent, that $\underline{P}(X_1, X_2), \underline{P}(X_1|X_2)$ are also coherent (which implies then that the three assessments are weakly coherent) and also that $\underline{P}(X_2|X_1), \underline{P}(X_1|X_2)$ are coherent, and we would like to know if we can deduce from this the (joint) coherence of $\underline{P}(X_1, X_2), \underline{P}(X_2|X_1), \underline{P}(X_1|X_2)$. Such a situation is considered for instance in [33, Section 7.9.1].

One way of using our results would be to consider partitions which are finer than the information we are given, and to compare these with the minimal partition. This would provide us with a sufficient condition for deducing coherence. Nevertheless, this approach is not always fruitful: in the above example, we should not be able to derive coherence, even though it has been established in [33, Section 7.9.1]. A thorough study of this matter would be one of the main open problems to consider in the future.

9.4. **Concluding remarks.** We summarise here the three main outcomes of the previous sections.

The first is a positive answer for the extension to problems based on smaller domains: coherence graphs can be used as before to decompose also those problems. What we lose in general here is the optimality of the decomposition in the case of weak coherence, but this does not seem to be critical especially if we regard coherence (resp. avoiding partial loss) as the consistency notion on which to focus rather than weak coherence (resp. avoiding uniform sure loss).

The second outcome concerns what we have called statements of practical impossibility, which is a concept related to logical dependence. In this case we can well include this kind of statements in a coherence graph, but we do not know in general how much this will affect the topology of the graph, and hence the minimal partition. It is possible that the partition stays the same (and even that the resulting graph is A1), that it grows, and in the worst case that it coincides with the entire graph. Therefore in some cases coherence graphs will still prove to be useful also under considerations of practical impossibility; but the prospect to fully exploit them together with such considerations is an open problem at this time.

It is also an open problem to extend our results to the case when we know that some subsets of the assessments are coherent and they do not form a partition.

10. Discussion

Coherence can be regarded as the very essence of a theory of personal probability. But working directly with coherence can be particularly onerous. The present paper is an attempt to deal with this difficulty in the case of Walley's important notion of (strong) coherence, and to deliver tools that make easier to check it. 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 minimal 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. The generality of the results is also what has enabled us to apply them to problems as diverse as proving the coherence of Bayesian and credal networks, and decomposing problems of compatibility or probabilistic satisfiability into smaller ones. On the other hand, such a generality has needed proofs that are somewhat long and technically involved. This is also due to the fact that graphs are naturally models of distributed computation and this clashes with the global nature of coherence. This nevertheless, the presented results are easy to exploit in practice, as we have provided a polynomial-time algorithm that computes the minimal partition of a coherence graph very efficiently.

On a more theoretical level, our results 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. They are related to the generalisation of the marginal extension theorem established in [23]: the relationship arises because 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 marginal extension theorem. 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. Moreover, when the associated coherence graph is A1, 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 [33, 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.

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 may be also useful to recall that completely analogous considerations hold when we consider loss notions, such as avoiding partial, or uniform sure, loss. We should also recall that there are differences: for example, 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.

There are some important open problems related to this paper. One would be the possibility to fully exploit coherence graphs under considerations of logical dependence, as we have proposed a partial solution to this problem. It seems to us that to extend our proposal it will be necessary on the one hand to investigate its theoretical properties, and on the other hand to specialise the models used in this paper. This could be done, for example, by strengthening the notion of collection template so as to insert some knowledge about the numbers that make up the lower previsions in the collection; or it could be done by focusing on coherence graphs with special topologies or dealing with variables taking values on finite sets (or even binary variables). These considerations could also simplify problems in probabilistic logic, different from those we have considered in this paper. As another topic for future research in this respect, we suggest the study of the *optimisation compatibility problem* [15, 34], where one looks for the smallest joint which is compatible with a number of assessments. We think that the functional $\underline{E}$ defined in the proof of Theorem 1 should play an important role here.

One possible extension we have not discussed yet would be to consider an infinite set of variables in our assessments; if we still have a finite number of conditional lower previsions, we think it should be possible to use coherence graphs to determine their coherence and their weak coherence, by making compact representations of the variables. The problem is more complicated if we consider an infinite number of variables *and* assessments; in that case, we should first of all generalise the coherence notions in [33] to an infinite number of assessments, and such a generalisation may not be immediate.

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 should also like to thank Renato Pelessoni and Paolo Vicig for instructive discussion about their joint work with Peter Walley cited in the references, Reinhard Diestel and Gregory Gutin for drawing our attention on the strong components as a way to simplify the algorithm to compute the minimal partition, and finally the referees for useful suggestions and comments. We acknowledge financial support by the MCYT project TSI2007-66706-C04-01, by the Swiss NSF grants 200021-113820/1 and 200020-116674/1, and by the Hasler Foundation grant 2233.

Appendix A. Proofs

Proof of Theorem 1. ($\Rightarrow$) This part of the proof is very similar to the one that Walley gives in [33, Theorem 8.1.8] for coherence. For this reason, we only give a brief sketch of this part.

Assume that $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent. Let us define the lower prevision $\underline{E}$ on $\mathcal{L}(\mathcal{X}^n)$ by

$$\underline{E}(f) := \sup\{\alpha : \exists f_j \in \mathcal{K}^j, j = 1,\dots,m, \text{s.t.} \sup_{x\in\mathcal{X}^n}[\sum_{j=1}^m G_j(f_j|X_{I_j}) - (f-\alpha)](x) < 0\}.$$

To see that $\underline{E}$ is well-defined, it suffices to note that $\sup f \geq \underline{E}(f) \geq \inf f$ for any gamble f: given $\alpha > \sup f$, there are no gambles $f_1,\dots,f_m$ satisfying the above equation or we contradict the weak coherence of $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$; and for any $\alpha < \inf f$ we can take $f_1 = \dots = f_m = 0$. It is also easy to see that $\underline{E}$ satisfies conditions (C1)–(C3), and as a consequence it is a coherent lower prevision. Let us show next that $\underline{E},\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent, i.e., that they satisfy Eq. (3). Consider $f \in \mathcal{L}(\mathcal{X}^n)$, and gambles $f_j \in$

$\mathcal{K}^j, j = 1, \dots, m$. Consider $\epsilon > 0$. Then the definition of $\underline{E}$ implies that there are $g_j \in \mathcal{K}^j, j = 1, \dots, m$, such that

$$\sup_{x \in \mathcal{X}^n} [\sum_{j=1}^m G_j(g_j|X_{I_j}) - G(f) - \frac{\epsilon}{2}] < 0, \tag{10}$$

where $G(f) = f - \underline{E}(f)$ (just consider $\alpha = \underline{E}(f) - \frac{\epsilon}{2}$ in the definition of $\underline{E}$). Hence, $G(f) > \sum_{j=1}^m G_j(g_j|X_{I_j}) - \frac{\epsilon}{2}$.

There are two possible cases in Eq. (3): that $j_0 \in \{1, \dots, m\}$ (case (a) below) or that it does not (case (b)).

(a) Consider $f_0 \in \mathcal{K}^{j_0}, z_0 \in \mathcal{X}_{I_{j_0}}$ for some j_0 in $\{1, \dots, m\}$. Then, using Eq. (10), the super-additivity of $\underline{P}_j(X_{O_j}|X_{I_j})$ for $j = 1, \dots, m$ and the weak coherence of $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ we deduce that

$$\begin{aligned} &\sup_{x \in \mathcal{X}^n} [\sum_{j=1}^m G_j(f_j|X_{I_j}) + G(f) - G_{j_0}(f_0|z_0)](x) \\ &\geq \sup_{x \in \mathcal{X}^n} [\sum_{j=1}^m G_j(f_j|X_{I_j}) + \sum_{j=1}^m G_j(g_j|X_{I_j}) - \frac{\epsilon}{2} - G_{j_0}(f_0|z_0)](x) \\ &\geq \sup_{x \in \mathcal{X}^n} [\sum_{j=1}^m G_j(f_j + g_j|X_{I_j}) - \frac{\epsilon}{2} - G_{j_0}(f_0|z_0)](x) \geq -\frac{\epsilon}{2}. \end{aligned}$$

Since this holds for any $\epsilon > 0$, we deduce that $\sup_{x \in \mathcal{X}^n} [\sum_{j=1}^m G_j(f_j|X_{I_j}) + G(f) - G_{j_0}(f_0|z_0)](x) \geq 0$.

(b) Take $f_0 \in \mathcal{L}(\mathcal{X}^n)$. Then, using Eq. (10), we see that

$$\sup_{x \in \mathcal{X}^n} [\sum_{j=1}^m G_j(f_j|X_{I_j}) + G(f) - G(f_0)](x) \geq -\epsilon;$$

otherwise, we should have

$$\begin{aligned} 0 &> \sup_{x \in \mathcal{X}^n} [\sum_{j=1}^m G_j(f_j|X_{I_j}) + G(f) - G(f_0) + \epsilon](x) \\ &\geq \sup_{x \in \mathcal{X}^n} [\sum_{j=1}^m G_j(f_j|X_{I_j}) + \sum_{j=1}^m G_j(g_j|X_{I_j}) - G(f_0) + \frac{\epsilon}{2}](x) \\ &\geq \sup_{x \in \mathcal{X}^n} [\sum_{j=1}^m G_j(f_j + g_j|X_{I_j}) - G(f_0) + \frac{\epsilon}{2}](x), \end{aligned}$$

where the second inequality follows from Eq. (10), and the third from the super-additivity of the conditional lower previsions $\underline{P}_j(X_{O_j}|X_{I_j})$ for $j = 1, \dots, m$. But this means that we can raise the value $\underline{E}(f_0)$ by $\frac{\epsilon}{2}$, which contradicts the definition of $\underline{E}$. Since this holds for any $\epsilon > 0$, we deduce that $\sup_{x \in \mathcal{X}^n} [\sum_{j=1}^m G_j(f_j|X_{I_j}) + G(f) - G(f_0)](x) \geq 0$.

Hence, $\underline{E}, \underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent, and as a consequence, for any $j = 1, \dots, m$, $\underline{E}$ and $\underline{P}_j(X_{O_j}|X_{I_j})$ are weakly coherent. From [33, Section 6.5], we deduce that for any $j = 1, \dots, m$, $f \in \mathcal{K}^j$, and any $x \in \mathcal{X}_{I_j}$, $\underline{E}(G_j(f|X_{I_j})) \geq 0$ and $\underline{E}(G_j(f|x)) = 0$.

($\Leftarrow$) Take $f_j \in \mathcal{K}^j, j = 1,\dots,m, f_0 \in \mathcal{K}^{j_0}, z_0 \in \mathcal{X}_{I_{j_0}}$ for some $j_0 \in \{1,\dots,m\}$. Take $g_j := G_j(f_j|X_{I_j}), j = 1,\dots,m, g_0 := G_{j_0}(f_0|z_0)$. Then we deduce from the assumption that $\underline{P}(g_j) \geq 0$ for $j = 1,\dots,m$, and $\underline{P}(g_0) = 0$, whence $g_j = G_j(f_j|X_{I_j}) \geq G(g_j) := g_j - \underline{P}(g_j)$ for $j = 1,\dots,m$ and $g_0 = G_{j_0}(f_0|z_0) = G(g_0) := g_0 - \underline{P}(g_0)$. As a consequence,

$$\sup_{x\in\mathcal{X}^n}[\sum_{j=1}^m G_j(f_j|X_{I_j}) - G_{j_0}(f_0|z_0)](x) \geq \sup_{x\in\mathcal{X}^n}[\sum_{j=1}^m G(g_j) - G(g_0)](x) \geq 0,$$

where the second inequality follows from the coherence of $\underline{P}$. We deduce that $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent.

For the second statement, let $P_1(X_{O_1}|X_{I_1}),\dots,P_m(X_{O_m}|X_{I_m})$ be weakly coherent conditional linear previsions. Use [33, Section 6.5.5] to deduce that any linear prevision P that dominates the coherent lower prevision $\underline{P}$, that exists because of the first part of the theorem, will satisfy $P(G_j(f|X_{I_j})) = 0$ for any $j = 1,\dots,m$ and any $f \in \mathcal{K}^j$, whence $P(f) = P(P_j(f|X_{I_j}))$. The converse implication follows trivially from the first part of the theorem. □

Proof of Proposition 1. The direct implication is a consequence of Theorem 8.1.8 in [33], taking into account that we can always include in this family an unconditional lower prevision $\underline{P}$ defined in the constant gambles. The converse implication is trivial. □

Proof of Corollary 1. The direct implication has been established in the proof of Theorem 1. The converse implication is trivial. □

Before we prove Proposition 3, we are going to give three lemmas that show how we can constrain the probability of any actual node in a block. The first two lemmas will be applied to blocks associated to a node with more than one parent, while Lemma 4 will be employed when dealing with blocks associated to a cycle. Lemma 2 will also be employed when we show how the notion of weak coherence can be verified through smaller parts in Theorem 4.

Lemma 2. *Let us consider $x_1^i, x_2^i \in \mathcal{X}_i$ for $i = 1,\dots,n$. Define the previsions $P_1(X_{O_1}|X_{I_1}),\dots,P_m(X_{O_m}|X_{I_m})$ with respective domains $\mathcal{K}^1,\dots,\mathcal{K}^m$ by*[12] *$P_j(f) := f((x_1^i)_{i\in O_j})$ if $I_j = \emptyset$, and*

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

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

Proof. First of all, it follows immediately that these previsions satisfy conditions (SC1)–(SC3) in Section 2, and are therefore separately coherent. Moreover, since they are all linear, coherence is equivalent to avoiding partial loss (Eq. (2)). Hence, we must prove that for any f_j in $\mathcal{K}^j$, $j = 1,\dots,m$, not all of which are equal to 0, there exists some B in $\cup_{j=1}^m S_j(f_j)$ such that

$$\sup_{x\in B}\sum_{j=1}^m G_j(f_j|X_{I_j})(x) \geq 0, \tag{11}$$

[12]We use here the one-to-one correspondence between gambles on $\mathcal{X}_{O_j\cup I_j}$ and gambles in $\mathcal{K}^j$.

where, in order to simplify the notation, in the case where $I_j = \emptyset$ we also use $G_j(f_j|X_{I_j})$ to denote $G_j(f_j)$.

If there is some $j \in \{1,\dots,m\}$ such that $I_j = \emptyset$ and $f_j \neq 0$, then $S_j(f_j) = \mathcal{X}^n$ and Eq. (11) becomes

$$\sup_{x\in\mathcal{X}^n} \sum_{j=1}^{m} G_j(f_j|X_{I_j})(x) \geq 0,$$

and this condition holds trivially by considering $x := (x_1^i)_{i=1,\dots,n}$, for which all the terms in the sum are equal to 0.

Let us assume next that $I_j \neq \emptyset$ for $j = 1,\dots,m$, i.e., that we are dealing with conditional linear previsions only. Note that we can assume without loss of generality that $f_j \geq 0$ for all $j = 1,\dots,m$: otherwise, it suffices to consider for each j the gamble $f'_j = f_j - \inf f_j \geq 0$, which satisfies $G_j(f'_j|X_{I_j}) = G_j(f_j|X_{I_j})$.

For any $j = 1,\dots,m$, let us define the sets $A_j := \pi^{-1}_{O_j\cup I_j}(A'_j)$, where

$$A'_j := \{(x_1^i)_{i\in O_j\cup I_j}\} \cup \{(x_2^i)_{i\in O_j} \times (\mathcal{X}_{I_j} \setminus \{(x_1^i)_{i\in I_j}\})\}$$

and the gambles $g_j := f_j\mathbb{I}_{A_j}$, where $\mathbb{I}_{A_j}$ is the indicator function of A_j. Since both f_j and $\mathbb{I}_{A_j}$ are $\mathcal{X}_{O_j\cup I_j}$-measurable, we deduce that g_j belongs to $\mathcal{K}^j$. Moreover, given $y \in \mathcal{X}_{I_j}$,

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

whence $P_j(g_j|X_{I_j}) = P_j(f_j|X_{I_j})$. Since $g_j \leq f_j$ for any $j = 1,\dots,m$ because $f_j \geq 0$, this implies that $G_j(g_j|X_{I_j}) \leq G_j(f_j|X_{I_j})$ for all j. Moreover, given $y \in A_j$, we have that $G_j(g_j|X_{I_j})(y) = 0$, and if $y \notin A_j$, $G_j(g_j|X_{I_j})(y) = -P_j(g_j|\pi_{I_j}(y)) = -P_j(f_j|\pi_{I_j}(y)) \leq 0$, taking into account that the gamble f_j is non-negative by assumption.

Now, if $g_j = 0$ for all $j = 1,\dots,m$, then $G_j(g_j|X_{I_j})(x) = 0$. Since $G_j(f_j|X_{I_j}) \geq G_j(g_j|X_{I_j})$, we deduce that $G_j(f_j|X_{I_j})(x) \geq 0$ for any $x \in \mathcal{X}^n$ and for all $j = 1,\dots,m$, and considering any element in the union of the supports we see that Eq. (11) holds. Assume now that $g_j \neq 0$ for some $j \in \{1,\dots,m\}$. Since $(x_1^i)_{1\leq i\leq n} \in A_j$ for all $j = 1,\dots,m$, $\sum_{j=1}^m G_j(g_j|X_{I_j})(x_1^i)_{1\leq i\leq n} = 0$. Hence, if there is some $B \in \cup_{j=1}^m S_j(g_j)$ such that $(x_1^i)_{1\leq i\leq n} \in B$, then Eq. (11) holds.

If $(x_1^i)_{1\leq i\leq n} \notin B$ for any $B \in \cup_{j=1}^m S_j(g_j)$, let j_1 be the smallest integer such that $g_{j_1} \neq 0$. Consider $B_1 \in S_{j_1}(g_{j_1})$. Then there is some $y_1 \in \mathcal{X}_{I_{j_1}}$ such that $B_1 = \{y_1\} \times \mathcal{X}_{I^c_{j_1}}$.

Let $z_1 := \{y_1\} \times (x_2^i)_{i\in I^c_{j_1}} \in B_1$. That is, the i-th component of z_1 is the i-th component of y_1 if $i \in I_{j_1}$, and is equal to x_2^i otherwise. Note that $y_1 \neq (x_1^i)_{i\in I_{j_1}}$, or we should have that $(x_1^i)_{1\leq i\leq n} \in B_1$, a contradiction. As a consequence, $z_1 \in A_{j_1}$, and therefore $G_{j_1}(g_{j_1}|X_{I_{j_1}})(z_1) = 0$. If $\sum_{j=1}^m G_j(g_j|X_{I_j})(z_1) \geq 0$, we deduce that Eq. (11) holds. If $\sum_{j=1}^m G_j(g_j|X_{I_j})(z_1) < 0$, then there is some $j_2 \in \{1,\dots,m\}$ such that $G_{j_2}(g_{j_2}|X_{I_{j_2}})(z_1) < 0$. Since $G_{j_2}(g_{j_2}|X_{I_{j_2}})(z) = 0$ for any z outside $S_{j_2}(g_{j_2})$, we deduce the existence of some $B_2 \in S_{j_2}(g_{j_2})$ such that $z_1 \in B_2$. For this B_2, there is some $y_2 \in \mathcal{X}_{I_{j_2}}$ such that $B_2 = \{y_2\} \times \mathcal{X}_{I^c_{j_2}}$. Note that $y_2 \neq (x_1^i)_{i\in I_{j_2}}$, using the same reasoning as before. Let $z_2 := \{y_2\} \times (x_2^i)_{i\in I^c_{j_2}}$ denote the element of $\mathcal{X}^n$ whose i-th component is the i-th component of z_1 if $i \in I_{j_2}$, and is x_2^i otherwise. Then $z_2 \in A_{j_2}$, whence $G_{j_2}(g_{j_2}|X_{I_{j_2}})(z_2) = 0$. Moreover, for any $i = 1,\dots,n$, if the

i-th component of z_1 is x_2^i, then so is the i-th component of z_2 by definition; hence, if we define the mapping $h : \mathcal{X}^n \to \{0, \dots, n\}$ such that $h(z)$ is the number of i such that $\pi_i(z)$ is equal to x_2^i, we have that $h(z_2) \geq h(z_1)$. If we had $h(z_2) = h(z_1)$, then it would be $z_2 = z_1$, whence $z_1 \in A_{j_2}$ and $G_{j_2}(g_{j_2}|X_{I_{j_2}})(z_1) = 0$, a contradiction. Hence, it is $h(z_2) > h(z_1)$.

Again, if $\sum_{j=1}^m G_j(g_j|X_{I_j})(z_2) \geq 0$, we deduce that Eq. (11) holds. If this sum is negative, then there is some j_3 such that $G_{j_3}(g_{j_3}|X_{I_{j_3}})(z_2) < 0$. We define then z_3 as the element of $\mathcal{X}^n$ whose i-th component is the i-th component of z_2 if $i \in I_{j_3}$, and is x_2^i otherwise. Then it belongs to the same element of $S_{j_3}(g_{j_3})$ than z_2, and moreover $h(z_3) > h(z_2)$.

If we follow this procedure, we obtain a sequence of elements (z_n) in the union of the supports of g_j such that $h(z_n) > h(z_{n-1})$ for all n. But since $0 \leq h(z) \leq n$ for all $z \in \mathcal{X}^n$, this process must be finite. If we stop with some z for which $h(z) < n$, this means that there is some z in $\cup_{j=1}^m S_j(g_j)$ such that $\sum_{j=1}^m G_j(g_j|X_{I_j})(z) \geq 0$. On the other hand, if we stop with z such that $h(z) = n$, that is, with $z = (x_2^i)_{1 \leq i \leq n}$, this means that $(x_2^i)_{1 \leq i \leq n}$ belongs to $\cup_{j=1}^m S_j(g_j)$ and, since it belongs to A_j for all j, we have $\sum_{j=1}^m G_j(g_j|X_{I_j})(z) = 0$. As a consequence, $\sum_{j=1}^m G_j(f_j|X_{I_j})(z) \geq \sum_{j=1}^m G_j(g_j|X_{I_j})(z) = 0$. This shows that $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ are coherent. □

Lemma 3. *Take $x_1^i \neq x_2^i \in \mathcal{X}_i$ for $i = 1, \dots, n$. Let us define the previsions $P_1(X_{O_1}|X_{I_1})$,…,$P_{m-1}(X_{O_{m-1}}|X_{I_{m-1}})$ with respective domains $\mathcal{K}^1, \dots, \mathcal{K}^{m-1}$ by $P_j(f) := f((x_1^i)_{i \in O_j})$ if $I_j = \emptyset$, and*

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

if $I_j \neq \emptyset$ for any $y \in \mathcal{X}_{I_j}, f \in \mathcal{K}^j$. Let $P_m(X_{O_m}|X_{I_m})$ be given by $P_m(f) := f((x_1^i)_{i \in O_m})$ for any $f \in \mathcal{K}^m$ if $I_m = \emptyset$, and $P_m(f|y) := f((x_1^i)_{i \in O_m}, y)$ for any $y \in \mathcal{X}_{I_m}$ and $f \in \mathcal{K}^m$ if $I_m \neq \emptyset$. Then $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ are weakly coherent.

Proof. It is easy to see that these previsions satisfy conditions (SC1)–(SC3) in Section 2, and are therefore separately coherent. Since moreover they are all linear, they are weakly coherent if and only if for any $f_j \in \mathcal{K}^j$, $j = 1, \dots, m$,

$$\sup_{x \in \mathcal{X}^n} \sum_{j=1}^m G_j(f_j|X_{I_j})(x) \geq 0. \tag{12}$$

Now, since $G_j(f_j|X_{I_j})(y) = 0$ for $y = (x_1^i)_{i=1,\dots,n}$, and for all $j = 1, \dots, m$, we deduce that Eq. (12) holds and therefore $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ are weakly coherent. □

Next, we show that the value of a parent of a dummy node in a cycle which is not in the cycle itself can also be determined by the previsions in the cycle:

Lemma 4. *Consider indices $j_1, \dots, j_p$ satisfying Eq. (5), i.e., determining a cycle, and let $k \in I_{j_1} \setminus O_{j_p}$. Then for any $x \in \mathcal{X}_k$ there are weakly coherent previsions $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ such that any joint prevision P which is coherent with $P_{j_i}(X_{O_{j_i}}|X_{I_{j_i}})$ for $i = 1, \dots, p$ satisfies $P(x) = 1$.*

Proof. Take $\ell_i \in I_{j_i} \cap O_{j_{i-1}}$ for $i = 1, \dots, p$, where for simplicity of notation we make $j_0 := j_p$. For each $i \in \{1, \dots, n\}$, let us consider $x_1^i \neq x_2^i$ in $\mathcal{X}_i$, and such that $x_1^k := x$. Let us define $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ with domains $\mathcal{K}^1, \dots, \mathcal{K}^m$ by

$$P_{j_1}(f|y) := \begin{cases} f((x_2^i)_{i\in O_{j_1}}, y) & \text{if } \pi_{\ell_1}(y) = x_1^{\ell_1}, y \neq (x_1^i)_{i\in I_{j_1}} \\ f((x_1^i)_{i\in O_{j_1}}, y) & \text{otherwise} \end{cases}$$

for any $y \in \mathcal{X}_{I_{j_1}}, f \in \mathcal{K}^{j_1}$,

$$P_{j_s}(f|y) := \begin{cases} f((x_1^i)_{i\in O_{j_s}}, y) & \text{if } \pi_{\ell_s}(y) = x_1^{\ell_s} \\ f((x_2^i)_{i\in O_{j_s}}, y) & \text{otherwise} \end{cases}$$

for any $y \in \mathcal{X}_{I_{j_s}}, f \in \mathcal{K}^{j_s}$, $s = 2, \dots, p$ and let us define, for $j \notin \{j_1, \dots, j_p\}$, $P_j(f) = f((x_1^i)_{i\in O_j})$ for any $f \in \mathcal{K}^j$ if $I_j = \emptyset$, and

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

if $I_j \neq \emptyset$ for any $y \in \mathcal{X}_{I_j}$ and any $f \in \mathcal{K}^j$.

These previsions satisfy conditions (SC1)–(SC3) in Section 2, and are therefore separately coherent. Moreover, they are also weakly coherent. For this, we have to show that for any $f_j \in \mathcal{K}^j$, $j = 1, \dots, m$, Eq. (12) holds. But it is easy to see that $G_j(f_j|X_{I_j})(y) = 0$ for $y = (x_1^i)_{i=1,\dots,n}$, and this for all $j = 1, \dots, m$.

Next, we show that any joint P which is coherent with each of the conditional previsions $P_{j_1}(X_{O_{j_1}}|X_{I_{j_1}}), \dots, P_{j_p}(X_{O_{j_p}}|X_{I_{j_p}})$ must satisfy $P(x_1^k) = 1$. For this, note first of all that for any $i = 1, \dots, p$, $P(x_1^{\ell_i}) + P(x_2^{\ell_i}) = P(P_{j_{i-1}}(x_1^{\ell_i}|X_{I_{j_{i-1}}})) + P(P_{j_{i-1}}(x_2^{\ell_i}|X_{I_{j_{i-1}}})) = 1$, because $P_{j_{i-1}}(x_1^{\ell_i}|X_{I_{j_{i-1}}}) + P_{j_{i-1}}(x_2^{\ell_i}|X_{I_{j_{i-1}}}) = 1$ for any $i = 1, \dots, p$. In particular,

$$1 = P(x_1^{\ell_1}) + P(x_2^{\ell_1}) = P(\{z : \pi_{\ell_1}(z) = x_1^{\ell_1}, \pi_i(z) \in \{x_1^i, x_2^i\}\ \forall i \in I_{j_1}\}) + P(x_2^{\ell_1}),$$

where the second equality holds because $P(\{x_1^i, x_2^i\}) = 1$ for all $i \in I_{j_1}$. We are going to show that the first of these terms is equal to $P((x_1^i)_{i\in I_{j_1}})$ and that the second is equal to 0. In order to prove this, we are going to use that for $i = 2, \dots, p$,

$$\begin{aligned} &P(\{z \in \mathcal{X}^n : \pi_{\ell_{i+1}}(z) \neq \pi_{\ell_i}(z)\}) = P(\{z : \pi_{\ell_i}(z) \in \{x_1^{\ell_i}, x_2^{\ell_i}\}, \pi_{\ell_{i+1}}(z) \neq \pi_{\ell_i}(z)\}) \\ &= P(P_{j_i}(\{z : \pi_{\ell_i}(z) \in \{x_1^{\ell_i}, x_2^{\ell_i}\}, \pi_{\ell_{i+1}}(z) \neq \pi_{\ell_i}(z)\}|X_{I_{j_i}})) = P(0) = 0, \end{aligned}$$

where we are using the notation $\ell_{p+1} := \ell_1$. Hence,

$$P(x_1^{\ell_i}, x_2^{\ell_{i+1}}) = P(x_2^{\ell_i}, x_1^{\ell_{i+1}}) = 0,\ i = 2, \dots, p. \tag{13}$$

The equality $P(\{x_1^{\ell_i}, x_2^{\ell_i}\}) = 1$, valid for $i = 1, \dots, p$, implies that $P(x_1^{\ell_2}) = P(\{z : \pi_{\ell_2}(z) = x_1^{\ell_2}, \pi_{\ell_i}(z) \in \{x_1^{\ell_i}, x_2^{\ell_i}\}, i = 3, \dots, p\})$. Take z such that $\pi_{\ell_2}(z) = x_1^{\ell_2}$, $\pi_{\ell_i}(z) \in \{x_1^{\ell_i}, x_2^{\ell_i}\}, i = 3, \dots, p$ and such that $z \neq (x_1^{\ell_i})_{i=2,\dots,p}$. Then there is some i in $\{2, \dots, p-1\}$ such that $\pi_{\ell_i}(z) = x_1^{\ell_i}, \pi_{\ell_{i+1}}(z) = x_2^{\ell_{i+1}}$, and Eq. (13) implies that $P(z) = 0$. We deduce that $P(x_1^{\ell_2}) = P((x_1^{\ell_i})_{i=2,\dots,p})$. A completely similar argument shows that $P(x_2^{\ell_2}) = P((x_2^{\ell_i})_{i=2,\dots,p})$.

Now, $P(x_2^{\ell_1}, x_2^{\ell_2}) = P(P_{j_1}(x_2^{\ell_1}, x_2^{\ell_2}|X_{I_{j_1}})) = 0$, whence $P(x_2^{\ell_1}) = P(x_2^{\ell_1}, x_1^{\ell_2}) + P(x_2^{\ell_1}, x_2^{\ell_2}) = P(x_2^{\ell_1}, x_1^{\ell_2})$. As a consequence,

$$P(x_2^{\ell_1}) = P(x_2^{\ell_1}, x_1^{\ell_2}) = P(x_2^{\ell_1}, (x_1^{\ell_i})_{i=2,\dots,p}) \leq P(x_2^{\ell_1}, x_1^{\ell_p}) = 0,$$

using Eq. (13), and therefore $P(x_2^{\ell_1}) = 0$. Consider on the other hand $z \in \mathcal{X}_{I_{j_1}\setminus\{\ell_1\}}$ different from $(x_1^i)_{i\in I_{j_1}\setminus\{\ell_1\}}$. Then $P(x_1^{\ell_1}, z) = P(x_1^{\ell_1}, z, x_2^{\ell_2})$, because $P(x_1^{\ell_1}, z, x_1^{\ell_2}) = P(P_{j_1}(x_1^{\ell_1}, z, x_1^{\ell_2}|X_{I_{j_1}})) = 0$. Hence,

$$P(x_1^{\ell_1}, z) = P(x_1^{\ell_1}, z, x_2^{\ell_2}) = P(x_1^{\ell_1}, z, (x_2^{\ell_i})_{i=2,\dots,p}) \leq P(x_1^{\ell_1}, x_2^{\ell_p}) = 0,$$

again using Eq. (13). Hence, $P(\{z : \pi_{\ell_1}(z) = x_1^{\ell_1}, \pi_i(z) \in \{x_1^i, x_2^i\}\ \forall i \in I_{j_1}\}) = P((x_1^i)_{i\in I_{j_1}})$ and therefore $1 = P(x_1^{\ell_1}) + P(x_2^{\ell_1}) = P((x_1^i)_{i\in I_{j_1}})$. In particular, $P(x_1^k) = 1$. □

Proof of Proposition 3. Let X_ℓ be a source of contradiction. Then, because any source of contradiction has parents in the coherence graph, there exists some $i_1 \in \{1,\dots,m\}$ such that $\ell \in O_{i_1}$. Apply Lemma 2 with $x_1^\ell := x_2^\ell := x$. Then we obtain a set of coherent conditional linear previsions, and any joint P which is coherent with $P_{i_1}(X_{O_{i_1}}|X_{I_{i_1}})$ satisfies $P(x) = P(P_{i_1}(x|X_{I_{i_1}})) = P(1) = 1$.

Consider next a node X_ℓ which is a predecessor of a source of contradiction, and $x \in \mathcal{X}_\ell$. We have the following possibilities:

(1) Assume first of all that there is a successor X_s of the node X_ℓ which has more than one parent. Let us consider a constraining sub-block for X_ℓ in this block, i.e., different $i_1,\dots,i_p$ in $\{1,\dots,m\}$ such that $s \in O_{i_1} \cap O_{i_2}$, $I_{i_2} \cap O_{i_3} \neq \emptyset, \dots, I_{i_{p-1}} \cap O_{i_p} \neq \emptyset$, and $\ell \in I_{i_p}$. Apply Lemma 3 with $m := i_1$ and $x_1^\ell := x$; then any joint P which is coherent with $P_{i_1}(X_{O_{i_1}}|X_{I_{i_1}})$ will satisfy $P(x_1^s) = P(P_{i_1}(x_1^s|X_{I_{i_1}})) = 1$. As a consequence, we have that $1 = P(x_1^s) = P(P_{i_2}(x_1^s|X_{I_{i_2}})) = P(x_1^s, (x_1^i)_{i\in I_{i_2}})$, whence $P((x_1^i)_{i\in I_{i_2}}) = 1$, and in particular $P(x_1^t) = 1$ for any $t \in I_{i_2} \cap O_{i_3}$. A similar reasoning allows us to show that $P((x_1^i)_{i\in I_{i_3}}) = 1$, and by following this procedure we also deduce that $P(x_1^\ell) = 1$. Hence, in this case we can define weakly coherent conditional previsions $P_1,\dots,P_m$ such that any prevision P which is coherent with $P_{i_1}(X_{O_{i_1}}|X_{I_{i_1}}),\dots,P_{i_p}(X_{O_{i_p}}|X_{I_{i_p}})$ satisfies $P(x_1^\ell) = 1$.

(2) Consider next the case where none of the successors of X_ℓ has more than one parent. Then X_ℓ must belong to a block originated by a cycle. In particular, X_ℓ is the predecessor of some node in an elementary cycle in the block. If X_ℓ is a parent of a dummy node in the cycle but is not in the cycle itself, the result follows from Lemma 4.

If neither of the previous possibilities holds, then there must be a path in the graph connecting X_ℓ with a dummy node in the elementary cycle. We consider then a constraining sub-block for X_ℓ in the block associated to this cycle. Take then $j_1,\dots,j_p$ in $\{1,\dots,m\}$ such that $O_{j_1} \cap I_{j_2} \neq \emptyset, O_{j_2} \cap I_{j_3} \neq \emptyset, \dots, O_{j_p} \cap I_{j_1} \neq \emptyset$, and $k_1,\dots,k_r$ in $\{1,\dots,n\}$ such that $\{k_1,\dots,k_{r-1}\} \subseteq \{1,\dots,n\} \setminus \{j_1,\dots,j_p\}$, $\ell \in O_{k_1} \cap I_{k_2}$, $O_{k_2} \cap I_{k_3} \neq \emptyset$, $\dots$, $O_{k_{r-1}} \cap I_{k_r} \neq \emptyset$, and $k_r = j_1$. Let us consider $t_i \in O_{k_i} \cap I_{k_{i+1}}$ for $i = 2,\dots,r-1$.

Take $x_1^i \neq x_2^i \in \mathcal{X}_i$ for $i = 1,\dots,n$, and let us consider the previsions $P_1(X_{O_1}|X_{I_1}),\dots,P_m(X_{O_m}|X_{I_m})$ defined in the proof of Lemma 4. Then these previsions are weakly coherent, and any prevision P which is coherent with $P_{j_1}(X_{O_{j_1}}|X_{I_{j_1}}),\dots,P_{j_p}(X_{O_{j_p}}|X_{I_{j_p}})$ satisfies $P((x_1^i)_{i\in I_{j_1}}) = 1$. As a consequence, we have that $P(x_1^{t_{r-1}}) = 1$, because $t_{r-1} \in I_{k_r} = I_{j_1}$, whence $1 = P(x_1^{t_r-1}) = P(P_{k_{r-1}}(x_1^{t_r-1}|X_{I_{k_{r-1}}})) = P(x_1^{t_r-1}, (x_1^i)_{i\in I_{k_{r-1}}})$, whence

$P((x_1^i)_{i\in I_{k_{r-1}}}) = 1$, and in particular $P(x_1^t) = 1$ for any $t \in I_{k_{r-1}} \cap O_{k_{r-2}}$. A similar reasoning allows us to show that $P((x_1^i)_{i\in I_{k_{r-2}}}) = 1$, and by following this procedure we also deduce that $P(x_1^\ell) = 1$ for any P coherent with

$$P_{j_1}(X_{O_{j_1}}|X_{I_{j_1}}), \dots, P_{j_p}(X_{O_{j_p}}|X_{I_{j_p}}), P_{k_1}(X_{O_{k_1}}|X_{I_{k_1}}), \dots, P_{k_{r-1}}(X_{O_{k_{r-1}}}|X_{I_{k_{r-1}}}).$$

This completes the proof. □

In order to simplify the proofs of Theorems 2 and 3, we are going to establish a couple of results (Lemmas 5 and 6) that will be common to these proofs.

Lemma 5. *Let us consider a non-empty subset J of $\{1,\dots,m\}$. Let $\mathcal{B}_1$ be a partition of J, and define, for each $C \in \mathcal{B}_1$, the set $B_C := \{\cup_{j\in C}(I_j \cup O_j)\}$. Then, if the sets $\{B_C : C \in \mathcal{B}_1\}$ are pairwise disjoint,[13] the following statements hold:*

(1) If $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in C}$ are coherent for all $C \in \mathcal{B}_1$, then the lower previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in J}$ are coherent.

(2) If $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in C}$ are weakly coherent for all $C \in \mathcal{B}_1$, then the lower previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in J}$ are weakly coherent.

Proof. (1) Consider $f_j \in \mathcal{K}^j$ for $j \in J$, $f_0 \in \mathcal{K}^{j_0}, z_0 \in \mathcal{X}_{I_{j_0}}$ for some $j_0 \in J$. Assume that $I_j \neq \emptyset$ for any j such that $f_j \neq 0$; otherwise, the result follows from the second statement and the reduction theorem [33, Theorem 7.1.5]. Let C_0 be the element of $\mathcal{B}_1$ that includes j_0. The coherence of $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in C_0}$ implies the existence of some $D_{C_0} \in \cup_{j\in C_0} S_j(f_j) \cup \{z_0\} \times \mathcal{X}_{I^c_{j_0}}$ such that

$$\sup_{x\in D_{C_0}} \left[\sum_{j\in C_0} G_j(f_j|X_{I_j}) - G_{j_0}(f_0|z_0)\right](x) \geq 0;$$

hence, for any $\epsilon > 0$ there is some $x_{C_0} \in D_{C_0}$ such that

$$\left[\sum_{j\in C_0} G_j(f_j|X_{I_j}) - G_{j_0}(f_0|z_0)\right](x_{C_0}) \geq -\epsilon. \tag{14}$$

On the other hand, for any $C \neq C_0$ in $\mathcal{B}_1$, there is some $D_C \in \cup_{j\in C} S_j(f_j)$ such that

$$\sup_{x\in D_C} \sum_{j\in C} G_j(f_j|X_{I_j})(x) \geq 0;$$

hence, given $\epsilon > 0$ there is some $x_C \in D_C$ such that

$$\sum_{j\in C} G_j(f_j|X_{I_j})(x_C) \geq -\epsilon. \tag{15}$$

Let us consider now an element $z \in \mathcal{X}^n$ satisfying $\pi_{B_C}(z) = \pi_{B_C}(x_C)$ for any $C \in \mathcal{B}_1$; such an element exists because the sets $\{B_C : C \in \mathcal{B}_1\}$ are pairwise disjoint. Then we deduce from Eq. (14) and (15) that

$$\left[\sum_{j\in J} G_j(f_j|X_{I_j}) - G_{j_0}(f_0|z_0)\right](z) \geq -|\mathcal{B}_1|\epsilon,$$

[13]This assumption means that given $C_1 \neq C_2 \in \mathcal{B}_1$, the subgraphs associated to B_{C_1} and B_{C_2} are not connected.

and moreover $z \in D_C$ for any $C \in \mathcal{B}_1$ (and in particular for one of them). Since we can do this for any $\epsilon > 0$, we deduce that the conditional lower previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in J}$ are coherent.

(2) Using a similar reasoning and the notations from the previous point, we deduce that for any $\epsilon > 0$ there is some x_{C_0} such that

$$\left[\sum_{j\in C_0} G_j(f_j|X_{I_j}) - G_{j_0}(f_0|z_0)\right](x_{C_0}) \geq -\epsilon.$$

On the other hand, for any $C \neq C_0$ in $\mathcal{B}_1$, there is some x_C such that

$$\sum_{j\in C} G_j(f_j|X_{I_j})(x_C) \geq -\epsilon.$$

Now, since the sets $\{B_C : C \in \mathcal{B}_1\}$ are pairwise disjoint, we deduce from these two equations that for any element z of $\mathcal{X}^n$ such that $\pi_{B_C}(z) = \pi_{B_C}(x_C)$ for all $C \in \mathcal{B}_1$,

$$\left[\sum_{j\in J} G_j(f_j|X_{I_j}) - G_{j_0}(f_0|z_0)\right](z) \geq -|\mathcal{B}_1|\epsilon.$$

Again, since we can do this for any $\epsilon > 0$, we deduce that the conditional lower previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in J}$ are weakly coherent.

□

Lemma 6. *Let* $\{\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})\}$ *be a separately coherent collection template whose graph is A1. Let I be a non-empty subset of* $\{1,\dots,n\}$ *which is disjoint with* $\cup_{j=1}^m O_j$*, and let* $x \in \mathcal{X}_I$*. Then for any* $f_j \in \mathcal{K}^j, j = 1,\dots,m,$ $j_0 \in \{1,\dots,m\}, f_0 \in \mathcal{K}^{j_0}, z_0 \in \mathcal{X}_{I_{j_0}}$,

$$\sup_{y\in\pi_I^{-1}(x)} \left[\sum_{j=1}^m G_j(f_j|X_{O_i}) - G_{j_0}(f_0|z_0)\right](y) \geq 0. \tag{16}$$

Proof. We are going to assume that the gamble $G_{j_0}(f_0|z_0)$ is not identically equal to zero; the result when $G_{j_0}(f_0|z_0) = 0$ follows as a corollary. From Lemma 1, we can assume that $O_k \cap (\cup_{j=1}^{k-1} I_j) = \emptyset$ for any $k = 1,\dots,m$. For any such k, let us define the set $A_k := \cup_{s=1}^k (O_s \cup I_s)$, and the previsions $\underline{Q}_k(X_{O_k}|X_{I\cup I_k\cup A_{k-1}})$ on the class $\mathcal{H}^k$ of $\mathcal{X}_{I\cup A_k}$-measurable gambles by

$$\underline{Q}_k(f|x) := \underline{P}_k(f(x,\cdot)|\pi_{I_k}(x)).$$

Note that this is well-defined because $(I \cup A_{k-1}) \cap O_k = \emptyset$.

For any $k = 1,\dots,m$, $\underline{Q}_k(X_{O_k}|X_{I\cup I_k\cup A_{k-1}})$ satisfies conditions (SC1)–(SC3) and is, therefore, separately coherent. This follows from the separate coherence of $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$. On the other hand, they are conditional on an increasing sequence of variables. Applying the marginal extension theorem for variables in [23], we conclude that $\underline{Q}_1,\dots,\underline{Q}_m$ are coherent.[14]

[14] For this, note that the domain $\mathcal{H}^k$ of $\underline{Q}_k(X_{O_k}|X_{I\cup I_k\cup A_{k-1}})$, which is the set of $\mathcal{X}_{I\cup A_k}$-measurable gambles, is included in the set of $\mathcal{X}_{I\cup I_{k+1}\cup A_k}$-measurable gambles, and that the partitions on the conditioning side are increasingly finer.

Consider $f_j \in \mathcal{K}^j, j = 1,\dots,m,\ j_0 \in \{1,\dots,m\}, f_0 \in \mathcal{K}^{j_0}, z_0 \in \mathcal{X}_{I_{j_0}}$. Let us define the gambles $h_1 := \sum_{j=1}^{j_0-1} G_j(f_j|X_{I_j})$ and $h_2 := \sum_{j=j_0}^{m} G_j(f_j|X_{I_j}) - G_{j_0}(f_0|z_0)$. Let us also define $g_j := f_j \mathbb{I}_{\pi_I^{-1}(x)}$ for $j = 1,\dots,j_0-1$. Then, since both f_j and $\mathbb{I}_{\pi_I^{-1}(x)}$ belong to $\mathcal{H}^j$, it follows that $g_j \in \mathcal{H}^j$ for $j = 1,\dots,j_0-1$. Moreover, any set E in the $\mathcal{X}_{I\cup I_j\cup A_{j-1}}$-support of g_j is included in $\pi_I^{-1}(x)$. Besides, for any $y \in \pi_I^{-1}(x)$, $g_j(y) - \underline{Q}_j(g_j|X_{I\cup I_j\cup A_{j-1}})(y) = f_j(y) - \underline{P}_j(f_j|X_{I_j})(y)$. We deduce from the coherence of $\underline{Q}_1,\dots,\underline{Q}_{j_0-1}$ that there is some $E \in \cup_{j=1}^{j_0-1} S_j(g_j)$ such that

$$\sup_{y\in E}\left[\sum_{j=1}^{j_0-1} g_j - \underline{Q}_j(g_j|X_{I\cup I_j\cup A_{j-1}})\right](y) \geq 0,$$

whence

$$\sup_{y\in\pi_I^{-1}(x)} h_1(y) = \sup_{y\in\pi_I^{-1}(x)}\left[\sum_{j=1}^{j_0-1} f_j - \underline{P}_j(f_j|X_{I_j})\right](y) = $$
$$= \sup_{y\in\pi_I^{-1}(x)}\left[\sum_{j=1}^{j_0-1} g_j - \underline{Q}_j(g_j|X_{I\cup I_j\cup A_{j-1}})\right](y) \geq 0.$$

Consider $\epsilon > 0$, and let $y_1 \in E$ satisfy $h_1(y_1) \geq -\epsilon$. Let $E_1 := \pi^{-1}_{I\cup A_{j_0-1}}(y_1)$. Note that if $j_0 = 1$ then we simply have $h_1 = 0$ (it vanishes from the following equations) and $E_1 := \pi_I^{-1}(x)$. There are two possibilities:

- If $E_1 \cap \pi^{-1}_{I_{j_0}}(z_0) = \emptyset$, then let us define $g_j := f_j\mathbb{I}_{E_1}$ for $j = j_0,\dots,m$. Then $g_j \in \mathcal{H}^j$ for all $j = j_0,\dots,m$, and any element in the support of g_j is included in E_1. We deduce from the coherence of $\underline{Q}_{j_0},\dots,\underline{Q}_m$ that there is some $E_2 \in \cup_{j=j_0}^{m} S_j(g_j)$ such that

$$\sup_{y\in E_2}\left[\sum_{j=j_0}^{m} g_j - \underline{Q}(g_j|X_{I\cup I_j\cup A_{j-1}})\right](y) \geq 0, \tag{17}$$

  whence

$$\sup_{y\in E_1}\left[\sum_{j=1}^{m} f_j - \underline{P}_j(f_j|X_{I_j}) - G_{j_0}(f_0|z_0)\right](y) = \sup_{y\in E_1}\left[\sum_{j=1}^{m} f_j - \underline{P}_j(f_j|X_{I_j})\right](y)$$
$$= \sup_{y\in E_1} h_1(y) + h_2(y) \geq -\epsilon,$$

  taking into account Eq. (17), that the gamble h_1 is identically equal to $h_1(y_1)$ on E_1 and that $\sup_{y\in E_1} h_2(y) \geq \sup_{y\in E_2} h_2(y) \geq 0$. As a consequence,

$$\sup_{y\in\pi_I^{-1}(x)} h_1(y) + h_2(y) \geq \sup_{y\in E_1} h_1(y) + h_2(y) \geq -\epsilon,$$

  and since we can do this for any $\epsilon > 0$, we deduce that Eq. (16) holds.
- If $E_1 \cap \pi^{-1}_{I_{j_0}}(z_0) \neq \emptyset$, we consider z_1 in this intersection, and let $y_2 := \pi_{I\cup I_{j_0}\cup A_{j_0-1}}(z_1)$. Note that $E_2 := \pi^{-1}_{I\cup I_{j_0}\cup A_{j_0-1}}(y_2) \subseteq E_1 \cap \pi^{-1}_{I_{j_0}}(z_0)$. Let us consider the gambles $g_j := f_j\mathbb{I}_{E_2}$ for $j = j_0,\dots,m$. It follows from the

coherence of $\underline{Q}_{j_0}, \ldots, \underline{Q}_m$ that there is $E_3 \in \cup_{j=j_0}^m S_j(g_j) \cup \pi^{-1}_{I \cup I_{j_0} \cup A_{j_0-1}}(y_2)$ such that

$$\sup_{y \in E_3} \left[\sum_{j=j_0}^m g_j - \underline{Q}(g_j | X_{I \cup I_j \cup A_{j-1}}) - G_{j_0}(g_0 | y_1) \right](y) \geq 0; \tag{18}$$

note that it follows from the definition of the gambles $g_j, j = j_0, \ldots, m$ and of y_1 that $E_3 \subseteq E_2$, and as a consequence also $E_3 \subseteq E_1 \subseteq \pi_I^{-1}(x)$. Hence,

$$\begin{aligned}
&\sup_{y \in \pi_I^{-1}(x)} \left[\sum_{j=1}^m f_j - \underline{P}_j(f_j | X_{I_j}) - G_{j_0}(f_0 | z_0) \right](y) \\
&\geq \sup_{y \in E_3} \left[\sum_{j=1}^m f_j - \underline{P}_j(f_j | X_{I_j}) - G_{j_0}(f_0 | z_0) \right](y) \\
&= \sup_{y \in E_3} \left[\sum_{j=1}^m g_j - \underline{Q}_j(g_j | X_{I \cup I_j \cup A_{j-1}}) - G_{j_0}(g_0 | y_1) \right](y) \\
&= \sup_{y \in E_3} h_1(y) + h_2(y) \geq -\epsilon,
\end{aligned}$$

taking into account that the gamble h_1 is identically equal to $h_1(y_1)$ on E_1, which is a superset of E_3, and Eq. (18). Since we can do this for any $\epsilon > 0$, we deduce that Eq. (16) holds.

This completes the proof. □

Corollary 3. *Let $\{\underline{P}_1(X_{O_1} | X_{I_1}), \ldots, \underline{P}_m(X_{O_m} | X_{I_m})\}$ be a separately coherent collection template whose graph is A1. Then, for any $f_j \in \mathcal{K}^j, j = 1, \ldots, m$, $j_0 \in \{1, \ldots, m\}, f_0 \in \mathcal{K}^{j_0}, z_0 \in \mathcal{X}_{I_{j_0}}$,*

$$\sup_{y \in B} \left[\sum_{j=1}^m G_j(f_j | X_{O_j}) - G_{j_0}(f_0 | z_0) \right](y) \geq 0$$

for some $B \in \cup_{j=1}^m S_j(f_j) \cup \pi^{-1}_{I_{j_0}}(z_0)$.

Proof. The result follows by using the same reasoning as in the previous proof, now with $I = \emptyset$. The main difference is that now $\pi_I^{-1}(x) = \mathcal{X}^n$, and we should have $g_j = f_j$ for $j = 1, \ldots, j_0 - 1$. By repeating the same arguments we conclude that

$$\sup_{y \in E_2} \left[\sum_{j=1}^m G_j(f_j | X_{O_j}) - G_{j_0}(f_0 | z_0) \right](y) \geq 0$$

for some E_2 included in some $B \in \cup_{j=1}^m S_j(f_j) \cup \pi^{-1}_{I_{j_0}}(z_0)$. □

Proof of Theorem 2. Let us consider gambles $f_j \in \mathcal{K}^j$ for $j = 1, \ldots, m$, $j_0 \in \{1, \ldots, m\}, f_{j_0} \in \mathcal{K}^{j_0}$ and $z_0 \in \mathcal{X}_{I_{j_0}}$, and let us show that

$$\sup_{x \in E} \left[\sum_{j=1}^m G_j(f_j | X_{O_j}) - G_{j_0}(f_0 | z_0) \right](x) \geq 0 \tag{19}$$

for some $E \in \cup_{j=1}^m S_j(f_j) \cup \pi^{-1}_{I_{j_0}}(z_0)$.

Let $\mathcal{B}_1$ be the subset of $\mathcal{B}$ given by the sets C with $|C| > 1$, i.e., those associated to superblocks. Assume that $\mathcal{B}_1 \neq \emptyset$; if $\mathcal{B}_1 = \emptyset$, then the result follows from Corollary 3. For any $C \in \mathcal{B}_1$ let us consider the set $B_C := \cup_{j\in C}(I_j \cup O_j)$. Then it follows from the definition of the superblocks that the sets $\{B_C : C \in \mathcal{B}_1\}$ are pairwise disjoint subsets of $\{1, \dots, n\}$. Let us define $J := \cup_{C\in\mathcal{B}_1} C$. Define the gambles $h_1 := \sum_{j\in J} G_j(f_j|X_{I_j}) - \mathbb{I}_J(j_0)G_{j_0}(f_0|z_0)$ and $h_2 := \sum_{j\in J^c} G_j(f_j|X_{I_j}) - \mathbb{I}_{J^c}(j_0)G_{j_0}(f_0|z_0)$.

From Lemma 5, the previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in J}$ are coherent. As a consequence, there is some $E \in \cup_{j\in J} S_j(f_j) \cup \mathbb{I}_J(j_0)\pi^{-1}_{I_{j_0}}(z_0)$ such that

$$\sup_{x\in E} h_1(x) \geq 0.$$

Consider $\epsilon > 0$, and let $x_1 \in E$ satisfy $h_1(x_1) \geq -\epsilon$. Let us consider $I := \cup_{C\in\mathcal{B}_1} B_C$ and $y_1 := \pi_I(x_1)$. Note that the gamble h_1 is $\mathcal{X}_I$-measurable, and that $\pi^{-1}_I(y_1)$ is included in E because $I_j \subseteq I$ for any $j \in J$.

The coherence graph of the previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in J^c}$ is A1 because they do not belong to any superblock. Moreover, for any $j \in J^c$, $I \cap O_j = \emptyset$, because of the comments at the end of Section 5. Hence, we can apply Lemma 6 to deduce that

$$\sup_{x\in\pi^{-1}_I(y_1)} h_2(x) = \sup_{x\in\pi^{-1}_I(y_1)} \sum_{j\in J^c} G_j(f_j|X_{I_j}) - \mathbb{I}_{J^c}(j_0)G_{j_0}(f_0|z_0)(x) \geq 0.$$

Since h_1 is identically equal to $h_1(x_1) \geq -\epsilon$ on $\pi^{-1}_I(y_1)$, $\sup_{x\in\pi^{-1}_I(y_1)} h_1(x) + h_2(x) \geq -\epsilon + \sup_{x\in\pi^{-1}_I(y_1)} h_2(x) \geq -\epsilon$, and because $\pi^{-1}_I(y_1)$ is included in E we deduce that $\sup_{x\in E} h_1(x) + h_2(x) \geq -\epsilon$. Since we can make a similar reasoning for any $\epsilon > 0$, we deduce that Eq. (19) holds. □

Proof of Theorem 3. We proceed in the same way as in the proof of Theorem 2. The main difference is that, instead of Eq. (19), we now only need to prove that

$$\sup_{x\in\mathcal{X}^n} [\sum_{j=1}^{m} G_j(f_j|X_{I_j}) - G_{j_0}(f_0|z_0)](x) \geq 0 \tag{20}$$

for any $f_j \in \mathcal{K}^j, j = 1, \dots, m, j_0 \in \{1, \dots, m\}, f_0 \in \mathcal{K}^{j_0}$ and $z_0 \in \mathcal{X}_{I_{j_0}}$.

Using the notations from the previous proof, if $\mathcal{B}_1 = \emptyset$, then the result follows from Corollary 3. Assume then that $\mathcal{B}_1 \neq \emptyset$. We deduce from Lemma 5 that the previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in J}$ are weakly coherent. As a consequence, for any $\epsilon > 0$ there exists some x_1 in $\mathcal{X}^n$ such that $h_1(x_1) \geq -\epsilon$. Let $y_1 := \pi_I(x_1) \in \mathcal{X}_I$. Now, since the coherence graph of the previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in J^c}$ is A1 because they do not belong to any superblock, we can apply Lemma 6 to deduce that

$$\sup_{x\in\pi^{-1}_I(y_1)} h_2(x) = \sup_{x\in\pi^{-1}_I(y_1)} \sum_{j\in J^c} G_j(f_j|X_{I_j}) - \mathbb{I}_{J^c}(j_0)G_{j_0}(f_0|z_0)(x) \geq 0.$$

h_1 is identically equal to $h_1(x_1) \geq -\epsilon$ on $\pi^{-1}_I(y_1)$, whence $\sup_{x\in\pi^{-1}_I(y_1)} h_1(x) + h_2(x) \geq -\epsilon + \sup_{x\in\pi^{-1}_I(y_1)} h_2(x) \geq -\epsilon$. Since we can make a similar reasoning for any $\epsilon > 0$, we deduce that Eq. (20) holds. □

Proof of Theorem 4. The sufficiency part is a consequence of Theorem 3. Let us show the necessity part.

Assume ex absurdo that $\mathcal{B}$ is not finer than $\mathcal{B}'$. Then there exists some $B \in \mathcal{B}$ with non-empty intersection with more than one $B' \in \mathcal{B}'$. It follows that B must be the set of indices of the lower previsions in a superblock of the coherence graph: otherwise, the cardinality of B would be 1. We shall identify the subset B of $\{1, \ldots, m\}$ with the corresponding superblock in the coherence graph. Then B is the union of a finite number of different blocks, $B_{Z_1}, \ldots, B_{Z_k}$, which are originated by different sources of contradiction. We shall denote by J_i the set of indexes of the lower previsions represented in the block B_{Z_i}, for $i = 1, \ldots, k$.

One of the following possibilities must hold:

(a) There exists some $i \in \{1, \ldots, k\}$ such that J_i has non-empty intersections with different B'_1, B'_2 in $\mathcal{B}'$.
(b) Condition (a) does not hold and there are $i_1 \neq i_2$ in $\{1, \ldots, k\}$ such that $J_{i_1} \subseteq B'_1, J_{i_2} \subseteq B'_2$ for different B'_1, B'_2 in $\mathcal{B}'$.

In case (a), we have two possibilities: that the block associated to J_i is related to a node with more than one parent (points (a1) and (a2) below), or that it corresponds to a cycle (points (a3) and (a4) below).

We are going to prove that in any of these cases there is a collection of separately coherent conditional linear previsions $\{P_1(X_{O_1}|X_{I_1}), \ldots, P_m(X_{O_m}|X_{I_m})\}$ that are not weakly coherent and such that for any B' in $\mathcal{B}'$, $\{P_j(X_{O_j}|X_{I_j})\}_{j \in B'}$ are weakly coherent.

(a) Assume that J_i has non-empty intersections with different B'_1, B'_2 in $\mathcal{B}'$. Then one of the following possibilities must hold:
(a1) There exists an actual node X_s in B_{Z_i} with two or more parents, corresponding to lower previsions in different elements of $\mathcal{B}'$ (see Fig. 13).

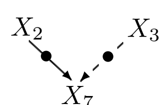

FIGURE 13. An example of the graphical situation considered in point (a1). In this case X_7 has two parents in different sets of $\mathcal{B}'$, and we can create a contradiction on it by inducing two different marginals through the two different parent previsions. We use the dashed line to denote that one of the parent previsions is in an element of $\mathcal{B}'$ and the other is in another element.

Formally, this holds if and only if there are $i_1 \neq i_2 \in J_i$ such that $s \in O_{i_1} \cap O_{i_2}$, and $i_1 \in B'_1, i_2 \in B'_2$ for $B'_1 \neq B'_2$. Consider $y_1^s \neq y_2^s$ in $\mathcal{X}_s$.

Define coherent (and in particular weakly coherent) conditional previsions $P_1(X_{O_1}|X_{I_1}), \ldots, P_m(X_{O_m}|X_{I_m})$ by means of Lemma 2 such that $P_{i_1}(y_1^s|X_{I_{i_1}})$ is identically equal to 1, whence any joint P that is weakly coherent with $P_{i_1}(X_{O_{i_1}}|X_{I_{i_1}})$ satisfies $P(y_1^s) = 1$ (use $x_1^s := x_2^s := y_1^s$). Apply Lemma 2 again with $x_1^s := x_2^s := y_2^s$ to define coherent $P'_1(X_{O_1}|X_{I_1}), \ldots, P'_m(X_{O_m}|X_{I_m})$ such that $P'_{i_2}(y_2^s|X_{I_{i_2}})$ is the constant gamble on 1, whence any joint P weakly coherent with

$P'_{i_2}(X_{O_{i_2}}|X_{I_{i_2}})$ satisfies $P(y_2^s) = 1$. Define now the conditional previsions $Q_1(X_{O_1}|X_{I_1}), \dots, Q_m(X_{O_m}|X_{I_m})$ by

$$Q_j(X_{O_j}|X_{I_j}) := \begin{cases} P_j(X_{O_j}|X_{I_j}) & \text{if } j \in B'_1 \\ P'_j(X_{O_j}|X_{I_j}) & \text{otherwise.} \end{cases}$$

These previsions are separately coherent and moreover for any $B' \in \mathcal{B}'$ the previsions $\{Q_j(X_{O_j}|X_{I_j})\}_{j \in B'}$ are weakly coherent. Given any coherent prevision P on $\mathcal{L}(\mathcal{X}^n)$ such that $P(f) = P(Q_j(f|X_{I_j}))$ for any $f \in \mathcal{K}^j$ and for any $j = 1, \dots, m$, we should have $P(y_1^s) = P(Q_{i_1}(y_1^s|X_{I_{i_1}})) = 1$ on the one hand, and $P(y_2^s) = P(Q_{i_2}(y_2^s|X_{I_{i_2}})) = 1$ on the other; hence, $P(y_1^s) = P(y_2^s) = 1$, a contradiction. Using Theorem 1, we deduce that $Q_1(X_{O_1}|X_{I_1}), \dots, Q_m(X_{O_m}|X_{I_m})$ are not weakly coherent.

(a2) Assume that for any node X_s in the block B_{Z_i} with more than one parent all the previsions corresponding to its parents belong to the same element of $\mathcal{B}'$. Then there must be an actual node X_ℓ in the graph which is a predecessor of a node X_s with more than one parent and such that the previsions in a constraining sub-block for X_ℓ in the block associated to X_s belong to the same element B'_1 of $\mathcal{B}'$, and there is another arc pointing at X_ℓ whose associated prevision belongs to some $B'_2 \neq B'_1$ in $\mathcal{B}'$. See Fig. 14 for an example.

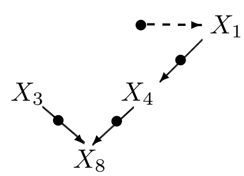

FIGURE 14. An example of the graphical situation considered in point (a2). In this case we can create a contradiction on X_1, (i.e., X_ℓ) by inducing two different marginals, one from its parent and the other one from X_3 through the constraining sub-block connecting X_1 and X_3 via X_8 (i.e., X_s). We use the dashed line to denote that the prevision having X_1 in the conditional side does not belong to the constraining sub-block.

Formally, this holds if and only if there are $i_1, \dots, i_p, i_{p+1}$ in J_i such that $s \in O_{i_1} \cap O_{i_2}, I_{i_2} \cap O_{i_3} \neq \emptyset, \dots, I_{i_{p-1}} \cap O_{i_p} \neq \emptyset, I_{i_p} \cap O_{i_{p+1}} \neq \emptyset, \ell \in I_{i_p} \cap O_{i_{p+1}}$, and $\{i_1, \dots, i_p\} \in B'_1, i_{p+1} \in B'_2$.

Consider $x_1^\ell \neq x_2^\ell$ in $\mathcal{X}_\ell$. Use Proposition 3 to define weakly coherent $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ such that any joint P weakly coherent with $\{P_{i_j}(X_{O_{i_j}}|X_{I_{i_j}})\}_{j=1,\dots,p}$ satisfies $P(x_1^\ell) = 1$. Apply now Lemma 2 to define coherent $P'_1(X_{O_1}|X_{I_1}), \dots, P'_m(X_{O_m}|X_{I_m})$ such that any joint P coherent with $P'_{i_{p+1}}(X_{O_{i_{p+1}}}|X_{I_{i_{p+1}}})$ satisfies $P(x_2^\ell) = 1$.

Define $Q_1(X_{O_1}|X_{I_1}), \dots, Q_m(X_{O_m}|X_{I_m})$ by

$$Q_j(X_{O_j}|X_{I_j}) := \begin{cases} P_j(X_{O_j}|X_{I_j}) & \text{if } j \in B'_1 \\ P'_j(X_{O_j}|X_{I_j}) & \text{otherwise.} \end{cases}$$

Then these previsions are separately coherent and moreover for any $B' \in \mathcal{B}'$ we have that $\{Q_j(X_{O_j}|X_{I_j})\}_{j \in B'}$ are weakly coherent. Now, given any coherent prevision P on $\mathcal{L}(\mathcal{X}^n)$ s.t. $P(f) = P(Q_j(f|X_{I_j}))$ for any $f \in \mathcal{K}^j$ and for any $j = 1, \dots, m$, we should have $P(x_1^\ell) = 1 = P(x_2^\ell)$, a contradiction. Using Theorem 1, we deduce that the conditional previsions $Q_1(X_{O_1}|X_{I_1}), \dots, Q_m(X_{O_m}|X_{I_m})$ are not weakly coherent.

(a3) Both (a1) and (a2) do not hold. This means that the block B_{Z_i}, whose set of indices J_i has non-empty intersections with different B'_1 and B'_2, is not generated by an actual node with more than one parent, but by a cycle. Take any elementary cycle in such a cycle.
Assume first that not all the dummy nodes in the elementary cycle belong to the same element of $\mathcal{B}'$. This implies that there are two adjacent dummy nodes in the elementary cycle that belong to different elements of $\mathcal{B}'$ (see Fig. 15 for an example).

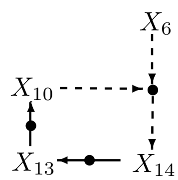

FIGURE 15. An example of the graphical situation considered in point (a3). In this case we can create a contradiction because the parent of X_{14} is in a different element of $\mathcal{B}'$ compared to the previous dummy node in the elementary cycle, as denoted by the dashed lines.

That is, assume the existence of different $j_1, \dots, j_p$ in J_i such that $O_{j_1} \cap I_{j_2} \neq \emptyset, O_{j_2} \cap I_{j_3} \neq \emptyset, \dots, O_{j_p} \cap I_{j_1} \neq \emptyset$ and such that $j_1 \in B'_1, j_2 \in B'_2$. Take $\ell_i \in I_{j_i} \cap O_{j_{i-1}}$ for $i = 1, \dots, p$, where for simplicity of notation we make $j_0 := j_p$.
Consider $x_1^i \neq x_2^i \in \mathcal{X}_i$ for $i = 1, \dots, n$, and apply Lemma 2 to define coherent $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$.
Define then y_1^i, y_2^i in $\mathcal{X}_i$ for $i = 1, \dots, n$ by

$$y_1^i := \begin{cases} x_2^i \text{ if } i = \ell_1 \\ x_1^i \text{ otherwise,} \end{cases} \qquad y_2^i := \begin{cases} x_1^i \text{ if } i = \ell_1 \\ x_2^i \text{ otherwise.} \end{cases}$$

Apply again Lemma 2 with these values to define coherent conditional previsions $P'_1(X_{O_1}|X_{I_1}), \dots, P'_m(X_{O_m}|X_{I_m})$, and define then the conditional previsions $Q_1(X_{O_1}|X_{I_1}), \dots, Q_m(X_{O_m}|X_{I_m})$ by

$$Q_j(X_{O_j}|X_{I_j}) := \begin{cases} P_j(X_{O_j}|X_{I_j}) \text{ if } j \in B'_1 \\ P'_j(X_{O_j}|X_{I_j}) \text{ otherwise.} \end{cases}$$

Let us show that $Q_{j_1}(X_{O_{j_1}}|X_{I_{j_1}}), \dots, Q_{j_p}(X_{O_{j_p}}|X_{I_{j_p}})$ are not weakly coherent: assume ex-absurdo that they are. Then, from Theorem 1, there is a coherent prevision P that is coherent with $Q_{j_1}, \dots, Q_{j_p}$. For any $i = 1, \dots, p$,

$$P(x_1^{\ell_i}) + P(x_2^{\ell_i}) = P(Q_{j_{i-1}}(x_1^{\ell_i}|X_{I_{j_{i-1}}})) + P(Q_{j_{i-1}}(x_2^{\ell_i}|X_{I_{j_{i-1}}})) = 1,$$

because $P_{j_{i-1}}(x_1^{\ell_i}|X_{I_{j_{i-1}}}) + P_{j_{i-1}}(x_2^{\ell_i}|X_{I_{j_{i-1}}}) = P'_{j_{i-1}}(x_1^{\ell_i}|X_{I_{j_{i-1}}}) + P'_{j_{i-1}}(x_2^{\ell_i}|X_{I_{j_{i-1}}}) = 1$ for any $i = 1, \dots, p$, whence $Q_{j_{i-1}}(x_1^{\ell_i}|X_{I_{j_{i-1}}}) + Q_{j_{i-1}}(x_2^{\ell_i}|X_{I_{j_{i-1}}}) = 1$ for any $i = 1, \dots, p$. Hence, for $i = 2, \dots, p$,

$$P(\{z \in \mathcal{X}^n : \pi_{\ell_{i+1}}(z) \neq \pi_{\ell_i}(z)\}) = P(\{z : \pi_{\ell_i}(z) \in \{x_1^{\ell_i}, x_2^{\ell_i}\}, \pi_{\ell_{i+1}}(z) \neq \pi_{\ell_i}(z)\})$$
$$= P(Q_{j_i}(\{z : \pi_{\ell_i}(z) \in \{x_1^{\ell_i}, x_2^{\ell_i}\}, \pi_{\ell_{i+1}}(z) \neq \pi_{\ell_i}(z)\}|X_{I_{j_i}})) = P(0) = 0,$$

where we are using the notation $\ell_{p+1} := \ell_1$. Hence,

$$P(x_1^{\ell_i}, x_2^{\ell_{i+1}}) = P(x_2^{\ell_i}, x_1^{\ell_{i+1}}) = 0 \tag{21}$$

for $i = 2, \dots, p$.

The equality $P(x_1^{\ell_i}) + P(x_2^{\ell_i}) = 1$, valid for $i = 1, \dots, p$, implies that $P(x_1^{\ell_2}) = P(\{z : \pi_{\ell_2}(z) = x_1^{\ell_2}, \pi_{\ell_i}(z) \in \{x_1^{\ell_i}, x_2^{\ell_i}\}, i = 1, \dots, p\})$. Take z such that $\pi_{\ell_2}(z) = x_1^{\ell_2}$, $\pi_{\ell_i}(z) \in \{x_1^{\ell_i}, x_2^{\ell_i}\}, i = 1, \dots, p\}$, and such that $z \neq (x_1^{\ell_i})_{i=1,\dots,p}$. Then there is some i in $\{2, \dots, p\}$ such that $\pi_{\ell_i}(z) = x_1^{\ell_i}, \pi_{\ell_{i+1}}(z) = x_2^{\ell_{i+1}}$, and Eq. (21) implies that $P(z) = 0$. We deduce that $P(x_1^{\ell_2}) = P((x_1^{\ell_i})_{i=1,\dots,p})$. A completely similar argument shows that $P(x_2^{\ell_2}) = P((x_2^{\ell_i})_{i=1,\dots,p})$.

Now, $P(x_1^{\ell_1}, x_1^{\ell_2}) = P(Q_{j_1}(x_1^{\ell_1}, x_1^{\ell_2}|X_{I_{j_1}})) = 0$, and as a consequence $P(x_1^{\ell_2}) = P((x_1^{\ell_i})_{i=1,\dots,p}) = 0$. A completely similar argument shows that $P(x_2^{\ell_1}, x_2^{\ell_2}) = 0$, whence $P(x_2^{\ell_2}) = P((x_2^{\ell_i})_{i=2,\dots,p}) = 0$. But this implies that $1 = P(x_1^{\ell_2}) + P(x_2^{\ell_2}) = 0$, a contradiction. Hence, $Q_{j_1}(X_{O_{j_1}}|X_{I_{j_1}}), \dots, Q_{j_p}(X_{O_{j_p}}|X_{I_{j_p}})$ are not weakly coherent.

(a4) Assume next that all the dummy nodes in the elementary cycle chosen in the previous point belong to the same B'_1, but that there is a predecessor X_ℓ of one of these dummy nodes which belongs to another B'_2 (see Fig. 16 for an example).

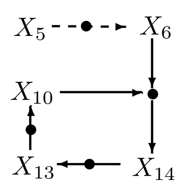

FIGURE 16. An example of the graphical situation considered in point (a4). In this case all the dummy nodes in the elementary cycle belong to the same element of $\mathcal{B}'$, but a dummy predecessor of the elementary cycle belongs to a different element, as denoted by the dashed line. We create then a contradiction on $X_l = X_6$ by inducing a marginal from its parent that is different from the marginal induced through the elementary cycle.

Consider then different $j_1, \dots, j_p$ in $\{1, \dots, m\}$ such that $O_{j_1} \cap I_{j_2} \neq \emptyset, O_{j_2} \cap I_{j_3} \neq \emptyset, \dots, O_{j_p} \cap I_{j_1} \neq \emptyset$, and assume that X_ℓ is a predecessor of the dummy node j_1. Then there are $k_1, \dots, k_r$ in $\{1, \dots, m\}$ such that $\{k_1, \dots, k_{r-1}\} \subseteq \{1, \dots, m\} \setminus \{j_1, \dots, j_p\}$, $\ell \in O_{k_1} \cap I_{k_2}$, $O_{k_2} \cap I_{k_3} \neq \emptyset$, $\dots$, $O_{k_{r-1}} \cap I_{k_r} \neq \emptyset$, and $k_r = j_1$. That is, the indices

$\{k_2,\dots,k_{r-1},j_1,\dots,j_p\}$ determine a constraining sub-block for X_ℓ in the block associated to the elementary cycle.

Assume that $j_1,\dots,j_p,k_2,\dots,k_{r-1}\in B'_1$ and $k_1\in B'_2$ for $B'_2\neq B'_1$. Take $x_1^\ell\neq x_2^\ell$ in $\mathcal{X}_\ell$. Then we can use Proposition 3 to 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 P coherent with $P_j(X_{O_j}|X_{I_j})$ for $j\in\{j_1,\dots,j_p,k_2,\dots,k_{r-1}\}$ satisfies $P(x_1^\ell)=1$. Similarly, we can use Lemma 2 to define coherent conditional lower previsions $P'_1(X_{O_1}|X_{I_1}),\dots,P'_m(X_{O_m}|X_{I_m})$ such that any joint P which is coherent with $P'_{k_1}(X_{O_{k_1}}|X_{I_{k_1}})$ satisfies $P(x_2^\ell)=1$. Let us consider then $Q_1(X_{O_1}|X_{I_1}),\dots,Q_m(X_{O_m}|X_{I_m})$ defined by

$$Q_j(X_{O_j}|X_{I_j}):=\begin{cases}P_j(X_{O_j}|X_{I_j}) & \text{if } j\in B'_1\\ P'_j(X_{O_j}|X_{I_j}) & \text{otherwise ;}\end{cases}$$

then it follows that for any $B\in\mathcal{B}'$ the previsions $\{Q_j(X_{O_j}|X_{I_j})\}_{j\in B'}$ are weakly coherent. However, any coherent prevision P which is coherent with each of them should satisfy $P(x_1^\ell)=1$ on the one hand, and $P(x_2^\ell)$ on the other. Hence, $Q_1(X_{O_1}|X_{I_1}),\dots,Q_m(X_{O_m}|X_{I_m})$ are not weakly coherent.

(b) If (a) does not hold, then for any $i=1,\dots,k$, the set of indices associated to the block B_{Z_i} is included in some $B'\in\mathcal{B}'$. Since B has non-empty intersection with different elements of $\mathcal{B}'$, and the definition of B implies that any block B_{Z_1} in B has another block B_{Z_2} in B such that B_{Z_1} and B_{Z_2} share an actual node, we deduce that there are $1\leq i_1\neq i_2\leq k$ such that $J_{i_1}\subseteq B'_1, J_{i_2}\subseteq B'_2$ and the blocks $B_{Z_{i_1}},B_{Z_{i_2}}$ share some actual node X_ℓ: we can start with a block of B and move to the adjacent blocks until we arrive to one which is not included in the same element of $\mathcal{B}'$. The result then follows from the fact that two adjacent blocks share at least one actual node. See Fig. 17 for an example.

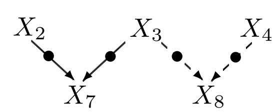

FIGURE 17. An example of the graphical situation considered in point (b). In this case we create a contradiction on $X_l=X_3$, which is a node shared by two adjacent blocks, as denoted by the dashed lines. The contradiction is obtained by inducing a marginal from one block and a different marginal from the other block.

Consider two different values $x_1^\ell\neq x_2^\ell$ in $\mathcal{X}_\ell$. Then applying Proposition 3, there are conditional previsions $P_1(X_{O_1}|X_{I_1}),\dots,P_m(X_{O_m}|X_{I_m})$ that are weakly coherent and such that any joint that is weakly coherent with $\{P_j(X_{O_j}|X_{I_j})\}_{j\in J_{i_1}}$ satisfies $P(x_1^\ell)=1$.

The same proposition guarantees the existence of weakly coherent conditional previsions $P'_1(X_{O_1}|X_{I_1}),\dots,P'_m(X_{O_m}|X_{I_m})$ such that any joint P that is weakly coherent with $\{P'_j(X_{O_j}|X_{I_j})\}_{j\in J_{i_2}}$ satisfies $P(x_2^\ell)=1$.

Define now conditional previsions $Q_1(X_{O_1}|X_{I_1}),\dots,Q_m(X_{O_m}|X_{I_m})$ by

$$Q_j(X_{O_j}|X_{I_j}) := \begin{cases} P_j(X_{O_j}|X_{I_j}) & \text{if } j \in B_1' \\ P_j'(X_{O_j}|X_{I_j}) & \text{otherwise.} \end{cases}$$

Then these previsions are separately coherent and moreover for any $B' \in \mathcal{B}'$ we have that $\{Q_j(X_{O_j}|X_{I_j})\}_{j\in B'}$ are weakly coherent. Now, given any coherent prevision P on $\mathcal{L}(\mathcal{X}^n)$ which is coherent with all these conditional previsions, we should have $P(x_1^\ell) = 1 = P(x_2^\ell)$, a contradiction. Using Theorem 1, we deduce that $Q_1(X_{O_1}|X_{I_1}),\dots,Q_m(X_{O_m}|X_{I_m})$ are not weakly coherent.

□

Proof of Proposition 4. It is trivial that the first statement implies the second. That the second statement implies the third follows from Theorem 4, once we remark that the coherence graph is of type A1 if and only if the minimal partition is $\mathcal{B} = \{1,\dots,m\}$. Finally, Corollary 3 guarantees that the third statement implies the first. □

Proof of Theorem 5. For each $i \in \{1,\dots,m\}$, the separate coherence of the conditional lower prevision $\underline{P}_i(X_{O_i}|X_{I_i})$ implies that it is the lower envelope of a family of separately coherent conditional previsions $\mathcal{M}_i := \{P_i(X_{O_i}|X_{I_i})\}$. Now, if we select a conditional prevision in $\mathcal{M}_i$ for each $i = 1,\dots,m$, they belong to the same collection template as $\{\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})\}$. Proposition 4 implies then that they are jointly coherent, and therefore $\{\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})\}$ is the lower envelope of a family of coherent conditional previsions. □

Proof of Theorem 6. We start by defining a bound on the number of arcs in a coherence graph. Remember that a coherence graph can be regarded as the union of its D-structures. Since a D-structure can be made of n arcs at most, we deduce that the number of arcs in a coherence graph is bounded by $m \cdot n$.

It is also useful to consider the number of *different* blocks in a coherence graph: since a block can be identified with one the sources of contradiction that originates it, and because a source of contradiction is an actual node, the number of different blocks is bounded by the number of actual nodes, namely n.

Now, consider that one effect of the interplay between `findMinPartition` and `findBlock` is a visit of the entire graph. This means that all its arcs are traversed once and hence the visit has worst-case complexity $O(m \cdot n)$, given the previous bound on the number of arcs. This complexity takes into account the fact that `findBlock` stops its visit at a certain node if this is tagged already, thus preventing an arc from being visited more than once.

To find out the overall complexity of the procedures, now we have to consider that the second effect jointly produced by `findMinPartition` and `findBlock`, through `mergeBlocks`, is to fill the array `minPartition`. This array is scanned every time a new, different block is found. Since the size of the array is m and the number of different blocks is bounded by n, as said above, filling the array `minPartition` has worst-case complexity $O(m \cdot n)$.

Overall, we deduce that `findMinPartition`, `findBlock`, and `mergeBlocks`, have a joint worst-case complexity given by $O(m\cdot n)$ (also considering that the remaining calculations done by them have a complexity dominated by such an expression).

Finally, in order to define the complexity to find the minimal partition, we must consider that the three procedures above are based on the global data structures described at the beginning of Section 7. Therefore, we have to calculate also the complexity to create such structures. In this case the computation that dominates the others, with respect to computational complexity, is the identification of the sources of contradiction. Identifying the actual nodes with more than a parent is easy, and takes $O(n)$. Finding out all the nodes that belong at least to a cycle in the graph is more complicated. Yet, as stated at the end of Section 7, this is an immediate byproduct of the identification of the strong components of the graph; and the latter is a task that Tarjan's algorithm in [29] solves in time $O(m+n+m\cdot n)$.

This complexity dominates the one arising from `findMinPartition`, `findBlock`, and `mergeBlocks`, whence we obtain that the complexity to find the minimal partition is just $O(m + n + m \cdot n)$. □

Proof of Theorem 8. We first consider how to convert the graph of a credal net into the related coherence graph: it is sufficient, for each node X_j of the credal net, to insert a dummy node D_j between X_j and its parents X_{I_j}, so as to make D_j parent of X_j and child of X_{I_j}, and to remove the arcs from X_{I_j} to X_j. It is immediate from this conversion procedure to see that the coherence graph resulting from that of a credal net is A1$^+$: the coherence graph is acyclic because so is the graph of a credal net; moreover, each actual node X_j has exactly one parent. Another immediate consequence of the procedure is that permuting the indexes of the conditional lower previsions in the coherence graph as in Lemma 1 can be done simply by making them consistent with the partial order entailed by the graph of the credal net. Therefore, we can assume without loss of generality that the set A_j defined in Section 8.1.1 after Lemma 1 indexes exactly the non-descendant non-parents of X_j, for all $j = 1, \dots, n$.

Now we are ready to show that the strong extension of a credal network and the strong product of the related coherence graph coincide.

Consider a prevision P dominating the strong extension obtained as in the definition of the strong product by applying Eq. (8) to a collection $\{P_1(X_{O_1}|X_{A_1\cup I_1}), \dots, P_n(X_{O_n}|X_{A_n\cup I_n})\}$. Here the generic element $P_j(X_{O_j}|X_{A_j\cup I_j})$ is a synthetic expression for a set of separately coherent conditional previsions $\{P_j(X_{O_j}|z) \in \mathrm{ext}(\mathcal{M}(\underline{P}'_j(X_{O_j}|z))) : z \in \mathcal{X}_{A_j\cup I_j}, P_j(X_{O_j}|z') = P_j(X_{O_j}|z'') \text{ if } \pi_{I_j}(z') = \pi_{I_j}(z'')\}$. For an $x \in \mathcal{X}^n$, it holds that

$$P(x) = \prod_{j=1}^{n} P_j(\pi_{O_j}(x)|\pi_{A_j\cup I_j}(x)) = \prod_{j=1}^{n} P_j(\pi_j(x)|\pi_{I_j}(x)), \tag{22}$$

where the first passage is just Eq. (8) and the second depends on two considerations. The first is that O_j indexes the variable X_j alone when we deal with the coherence graph obtained from a Bayesian net. The second is more important: X_j does not depend stochastically on its non-descendant non-parents (X_{A_j}) given its parents (X_{I_j}). This follows because we know that the chosen element of $\mathcal{M}(\underline{P}'_j(X_{O_j}|z))$ must be the same over all the values that X_{A_j} can take, by definition of $P_j(X_{O_j}|X_{A_j\cup I_j})$; and because, by definition of $\underline{P}'_j(X_{O_j}|X_{A_j\cup I_j})$, an extreme point of $\mathcal{M}(\underline{P}'_j(X_{O_j}|z))$ is an extreme point of $\mathcal{M}(\underline{P}_j(X_{O_j}|\pi_{I_j}(z)))$, and this only depends on X_{I_j}.

It follows that Eqs. (9) and (22) coincide, and hence that the selected element P dominating the strong product corresponds to the joint mass function of a Bayesian net obtained choosing the assessments $P_j(X_j|X_{I_j}) \geq \underline{P}_j(X_j|X_{I_j})$ $j = 1, \ldots, n$, i.e., to a compatible Bayesian net. Therefore P is a linear prevision that dominates the strong extension. Since this holds for all the previsions P in the definition of the strong product, then the strong product dominates the strong extension.

Now consider an extreme point of the strong extension, i.e., equivalently, a joint linear prevision P. It is well known (see [1, Proposition 1] for a proof) that P is obtained by applying Eq. (9) to mass functions corresponding to extreme points of the local conditional credal sets, i.e., equivalently, to a collection $\{P_1(X_1|X_{I_1}), \ldots, P_n(X_n|X_{I_n})\}$, with $P_j(X_j|z) \in \text{ext}(\mathcal{M}(\underline{P}_j(X_j|z)))$ for all $z \in \mathcal{X}_{I_j}$, $j = 1, \ldots, n$. From this, defining the new collection of conditional previsions $\{P_1(X_{O_1}|X_{A_1 \cup I_1}), \ldots, P_n(X_{O_n}|X_{A_n \cup I_n})\}$ as in the definition of the strong product, and applying Eq. (22), we see that there is an element of the strong product that coincides with P. Since this holds for all the extreme points of the strong extension, then the strong extension dominates the strong product. □

Proof of Proposition 5. We start by proving the direct implication. Assume that $\{\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ avoid uniform sure loss. Given $f_j \in \mathcal{K}^j$ for $j = 1, \ldots, m$, this implies that $\sup_{x \in \mathcal{X}^n} \sum_{j=1}^m G_j(f_j|X_{I_j})(x) \geq 0$. Let us define $\mathcal{D} := \{G_j(f_j|X_{I_j}) : f_j \in \mathcal{K}^j \text{ for some } j\}$. Applying [33, Lemma 3.3.2], we deduce the existence of a linear prevision P satisfying $P(G_j(f_j|X_{I_j})) \geq 0$ for all $f_j \in \mathcal{K}^j, j = 1, \ldots, m$, and in particular $P(G_j(f_j|x)) \geq 0$ for all $f_j \in \mathcal{K}^j, x \in \mathcal{X}_{I_j}, j = 1, \ldots, m$.

Let us prove the existence, for every $j = 1, \ldots, m$ and every $x \in \mathcal{X}_j$ of a linear conditional prevision $P_j(X_{O_j}|x)$ dominating $\underline{P}_j(X_{O_j}|x)$ such that $P(\mathbb{I}_x(f - P_j(f|x))) = 0$ for all $f \in \mathcal{K}^j$; from this, taking into account that the spaces are finite, we shall deduce the existence of a conditional linear prevision $P_j(X_{O_j}|X_{I_j})$ that dominates $\underline{P}_j(X_{O_j}|X_{I_j})$ and is coherent with P for all $j = 1, \ldots, m$, and as a consequence of weakly coherent $P_1(X_{O_1}|X_{I_1}), \ldots, P_m(X_{O_m}|X_{I_m})$ dominating our conditional lower previsions.

Consider then $j \in \{1, \ldots, m\}$ and $x \in \mathcal{X}_{I_j}$. There are two possibilities: if $P(x) > 0$, then the conditional linear prevision $P_j(X_{O_j}|x)$ is uniquely determined by P using Bayes' rule. To see that it dominates $\underline{P}_j(X_{O_j}|x)$, assume ex-absurdo the existence of some gamble $f \in \mathcal{K}^j$ for which $\underline{P}_j(f|x) > P_j(f|x)$. Then it follows that $0 \leq P(\mathbb{I}_x(f - \underline{P}_j(f|x))) < P(\mathbb{I}_x(f - P_j(f|x))) = 0$, a contradiction. Finally, it $P(x) = 0$ we simply consider any linear conditional prevision $P_j(X_{O_j}|x)$ that dominates $\underline{P}_j(X_{O_j}|x)$ and it is automatically coherent with P.

The converse implication follows once we realise that (i) weakly coherent conditional previsions in particular avoid uniform sure loss, and (ii) if some conditional lower previsions are dominated by others which avoid uniform sure loss, then they also avoid uniform sure loss. □

The next lemma is needed in the proof of Theorem 9. It is the counterpart, for avoiding partial and uniform sure loss, of Lemma 5.

Lemma 7. *Let us consider a non-empty subset J of $\{1, \ldots, m\}$. Let $\mathcal{B}_1$ be a partition of J, and define, for each $C \in \mathcal{B}_1$, the set $B_C := \{\cup_{j \in C}(I_j \cup O_j)\}$. Then, if the sets $\{B_C : C \in \mathcal{B}_1\}$ are pairwise disjoint, the following statements hold:*

(1) If $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in C}$ avoid partial loss for all $C \in \mathcal{B}_1$, then the lower previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in J}$ avoid partial loss.
(2) If $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in C}$ avoid uniform sure loss for all $C \in \mathcal{B}_1$, then the lower previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in J}$ avoid uniform sure loss.

Proof. (1) Consider $f_j \in \mathcal{K}^j$ for $j \in J$, $f_0 \in \mathcal{K}^{j_0}, z_0 \in \mathcal{X}_{I_{j_0}}$ for some $j_0 \in J$. Assume that $I_j \neq \emptyset$ for any j such that $f_j \neq 0$; otherwise, the result follows from the second statement, using also the reduction theorem [33, Theorem 7.1.5].

For any C in $\mathcal{B}_1$, there is some $D_C \in \cup_{j\in C} S_j(f_j)$ such that

$$\sup_{x\in D_C} \sum_{j\in C} G_j(f_j|X_{I_j})(x) \geq 0;$$

hence, given $\epsilon > 0$ there is some $x_C \in D_C$ such that

$$\sum_{j\in C} G_j(f_j|X_{I_j})(x_C) \geq -\epsilon. \tag{23}$$

Let us consider now an element $z \in \mathcal{X}^n$ satisfying $\pi_{B_C}(z) = \pi_{B_C}(x_C)$ for any $C \in \mathcal{B}_1$; such an element exists because the sets $\{B_C : C \in \mathcal{B}_1\}$ are pairwise disjoint. Then we deduce from Equation (23) that

$$\left[\sum_{j\in J} G_j(f_j|X_{I_j})\right](z) \geq -|\mathcal{B}_1|\epsilon,$$

and moreover $z \in D_C$ for all $C \in \mathcal{B}_1$ (and in particular for one of them). Since we can do this for any $\epsilon > 0$, we deduce that the conditional lower previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in J}$ avoid partial loss.

(2) Using a similar reasoning and the notations from the previous point, we deduce that for any $\epsilon > 0$ and for any C in $\mathcal{B}_1$, there is some x_C such that

$$\sum_{j\in C} G_j(f_j|X_{I_j})(x_C) \geq -\epsilon.$$

Now, since the sets $\{B_C : C \in \mathcal{B}_1\}$ are pairwise disjoint, we deduce from this equation that for any element z of $\mathcal{X}^n$ such that $\pi_{B_C}(z) = \pi_{B_C}(x_C)$ for all $C \in \mathcal{B}_1$,

$$\left[\sum_{j\in J} G_j(f_j|X_{I_j})\right](z) \geq -|\mathcal{B}_1|\epsilon.$$

Again, since we can do this for any $\epsilon > 0$, we deduce that the conditional lower previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j\in J}$ avoid uniform sure loss.

□

Proof of Theorem 9. We shall prove the result for avoiding partial loss; the proof for avoiding uniform sure loss is analogous. Let us consider gambles $f_j \in \mathcal{K}^j$ for $j = 1, \ldots, m$, and let us show that

$$\sup_{x\in E} \left[\sum_{j=1}^{m} G_j(f_j|X_{O_j})\right](x) \geq 0 \tag{24}$$

for some $E \in \cup_{j=1}^{m} S_j(f_j)$.

Let $\mathcal{B}_1$ be the subset of $\mathcal{B}$ given by the sets C with $|C| > 1$, i.e., those associated to superblocks. Assume that $\mathcal{B}_1 \neq \emptyset$; if $\mathcal{B}_1 = \emptyset$, then the result follows from Corollary 3. For any $C \in \mathcal{B}_1$ let us consider the set $B_C := \cup_{j \in C}(I_j \cup O_j)$. Then it follows from the definition of the superblocks that the sets $\{B_C : C \in \mathcal{B}_1\}$ are pairwise disjoint subsets of $\{1, \ldots, n\}$. Let us define $J := \cup_{C \in \mathcal{B}_1} C$. Define the gambles $h_1 := \sum_{j \in J} G_j(f_j|X_{I_j})$ and $h_2 := \sum_{j \in J^c} G_j(f_j|X_{I_j})$.

From Lemma 7, the previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j \in J}$ avoid partial loss. As a consequence, there exists some $E \in \cup_{j \in J} S_j(f_j)$ such that

$$\sup_{x \in E} h_1(x) \geq 0.$$

Consider $\epsilon > 0$, and let $x_1 \in E$ satisfy $h_1(x_1) \geq -\epsilon$. Let us consider $I := \cup_{C \in \mathcal{B}_1} B_C$ and $y_1 := \pi_I(x_1)$. Note that the gamble h_1 is $\mathcal{X}_I$-measurable, and that $\pi_I^{-1}(y_1)$ is included in E because $I_j \subseteq I$ for any $j \in J$.

The coherence graph of the previsions $\{\underline{P}_j(X_{O_j}|X_{I_j})\}_{j \in J^c}$ is A1 because they do not belong to any superblock. Moreover, for any $j \in J^c$, $I \cap O_j = \emptyset$, because of the comments at the end of Section 5. Hence, we can apply Lemma 6 to deduce that

$$\sup_{x \in \pi_I^{-1}(y_1)} h_2(x) = \sup_{x \in \pi_I^{-1}(y_1)} \sum_{j \in J^c} G_j(f_j|X_{I_j})(x) \geq 0.$$

Since h_1 is identically equal to $h_1(x_1) \geq -\epsilon$ on $\pi_I^{-1}(y_1)$, $\sup_{x \in \pi_I^{-1}(y_1)} h_1(x) + h_2(x) \geq -\epsilon + \sup_{x \in \pi_I^{-1}(y_1)} h_2(x) \geq -\epsilon$, and because $\pi_I^{-1}(y_1)$ is included in E we deduce that $\sup_{x \in E} h_1(x) + h_2(x) \geq -\epsilon$. Since we can make a similar reasoning for any $\epsilon > 0$, we deduce that Eq. (24) holds. □

References

[1] A. Antonucci and M. Zaffalon. Decision-theoretic specification of credal networks: a unified language for uncertain modeling with sets of Bayesian networks. *International Journal of Approximate Reasoning*, 49(2):345–361, 2008.

[2] B. Arnold, E. Castillo, and J.M. Sarabia. *Conditional specification of statistical models.* Springer, 1999.

[3] M. Baioletti, A. Capotorti, S. Tulipani, and B. Vantaggi. Elimination of Boolean variables for probabilistic coherence. *Soft Computing*, 4:81–88, 2000.

[4] K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges.* Academic Press, London, 1983.

[5] 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.

[6] G. Boole. *An investigation on the laws of thought, on which are founded the mathematical theories of logic and probabilities.* Walton and Maberley, London, 1854. reprinted by Dover, New York in 1961.

[7] A. Capotorti and B. Vantaggi. Locally strong coherence in inference processes. *Annals of Mathematics and Artificial Intelligence*, 35(1–4):125–149, 2002.

[8] F. G. Cozman. Separation properties of sets of probabilities. In C. Boutilier and M. Goldszmidt, editors, *Uncertainty in Artificial Intelligence (Proceedings of the Sixteenth Conference)*, pages 107–115. Morgan Kaufmann, San Francisco, 2000.

[9] C. Cuadras, J. Fortiana, and J.A. Rodriguez-Lallena. *Distributions with given marginals and statistical modelling.* Kluwer Academic Publishers, 2002.

[10] B. de Finetti. Sul significato soggettivo della probabilità. *Fundamenta Mathematicae*, 17:298–329, 1931.

[11] B. de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937. English translation in [18].

[12] B. de Finetti. *Teoria delle Probabilità*. Einaudi, Turin, 1970.

[13] B. de Finetti. *Theory of Probability*, volume 1. John Wiley & Sons, Chichester, 1974. English Translation of [12].

[14] M. Fréchet. Sur les tableaux de correlation dont les marges sont donns. *Ann. Univ. Lyon, Section A, Series 3*, 14:53–77, 1951.

[15] P. Hansen, B. Jaumard, M. Poggi de Aragão, F. Chauny, and S. Perron. Probabilistic satisfiability with imprecise probabilities. *International Journal of Approximate Reasoning*, 24(2–3):171–189, 2000.

[16] W. Hoeffding. Masstabinvariante korrelationtheorie. *Schriften des Mathematischen Instituts und des Institud für Angewndte Mathematik der Universität Berlin*, 5:181–233, 1940.

[17] B. Jaumard, H. Hansen, and M. Poggi de Aragão. Column generation methods for probabilistic logic. *ORSA Journal on Computing*, 3:135–148, 1991.

[18] H. E. Kyburg Jr. and H. E. Smokler, editors. *Studies in Subjective Probability*. Wiley, New York, 1964. Second edition (with new material) 1980.

[19] T. Lukasiewicz. Probabilistic deduction with conditional constraints over basic events. *Journal of Artificial Intelligence Research*, 10:199–241, 1999.

[20] F. Mauch. On the existence of probability distributions with given marginals. *Theory of Probability and its Applications*, 48(3):541–548, 2004.

[21] E. Miranda. Updating coherent previsions on finite spaces. Technical Report TR 08/04, Rey Juan Carlos University, Technical Reports on Statistics and Decision Sciences, 2008.

[22] E. Miranda and G. de Cooman. Coherence and independence in non-linear spaces. Technical Report TR 05/10, Rey Juan Carlos University, Technical Reports on Statistics and Decision Sciences, 2005.

[23] E. Miranda and G. de Cooman. Marginal extension in the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 46(1):188–225, 2007.

[24] S. Moral and A. Cano. Strong conditional independence for credal sets. *Annals of Mathematics and Artificial Intelligence*, 35(1–4):295–321, 2002.

[25] R. Nelsen. *An introduction to copulas*. Springer, New York, 1999.

[26] N.J. Nilsson. Probabilistic logic. *Artificial intelligence*, 28:71–87, 1986.

[27] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.

[28] A. Sklar. Fonctions de répartition à n-dimensions et leurs marges. *Publications de l'Institute de Statistique de l'Université de Paris*, 8:229–231, 1959.

[29] R. E. Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1:146–160, 1972.

[30] M. C. M. Troffaes and G. de Cooman. Extension of coherent lower previsions to unbounded random variables. In *Intelligent systems for information processing: for information to applications*, pages 277–288. North Holland, Amsterdam, 2003.

[31] L. C. van der Gaag. Computing probability intervals under independency constraints. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *UAI '90: Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, pages 457–466. Elsevier, 1991.

[32] N. N. Vorobev. Consistent families of measures and their extensions. *Theory of Probability and its Applications*, VII(2):147–163, 1962.

[33] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[34] 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.

[35] P. M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, UK, 1975. Reprinted in [36].

[36] P. M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44(3):366–383, 2007.

[37] M. Zaffalon and E. Miranda. Conservative inference rule for uncertain reasoning under incompleteness. 2008. Submitted for publication.

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

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