International Journal of Approximate Reasoning 56 (2015) 178–207

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Credal networks under epistemic irrelevance: The sets of desirable gambles approach

Jasper De Bock *, Gert de Cooman

Ghent University, SYSTeMS Research Group, Technologiepark – Zwijnaarde 914, 9052 Zwijnaarde, Belgium

ARTICLE INFO

Article history:
Received 25 October 2013
Received in revised form 24 June 2014
Accepted 2 July 2014
Available online 9 July 2014

Keywords:
Credal networks
Epistemic irrelevance
AD-separation
Irrelevant natural extension
Graphoid properties
Sets of desirable gambles

ABSTRACT

We present a new approach to credal networks, which are graphical models that generalise Bayesian networks to deal with imprecise probabilities. Instead of applying the commonly used notion of strong independence, we replace it by the weaker, asymmetrical notion of epistemic irrelevance. We show how assessments of epistemic irrelevance allow us to construct a global model out of given local uncertainty models, leading to an intuitive expression for the so-called irrelevant natural extension of a credal network. In contrast with Cozman [4], who introduced this notion in terms of credal sets, our main results are presented using the language of sets of desirable gambles. This has allowed us to derive some remarkable properties of the irrelevant natural extension, including marginalisation properties and a tight connection with the notion of independent natural extension. Our perhaps most important result is that the irrelevant natural extension satisfies a collection of epistemic irrelevancies that is induced by AD-separation, an asymmetrical adaptation of d-separation. Both AD-separation and the induced collection of irrelevancies are shown to satisfy all graphoid properties except symmetry.

1. Introduction

A *Bayesian network* [26] is a probabilistic graphical model [20] that is popular in fields such as statistics, machine learning and artificial intelligence. It identifies the nodes of a Directed Acyclic Graph (DAG) with random variables and interprets the graphical structure of the DAG as an assessment of the independencies amongst these variables; nodes that are not connected represent variables that are conditionally independent of each other. By exploiting these independencies, a global uncertainty model can be constructed easily out of local ones, allowing for a compact representation of the model. Efficient algorithms have been developed for performing inferences in such Bayesian networks, leading to their successful application in a multitude of real-life problems [20,28].

Despite their success, Bayesian networks have an important limitation: the construction of a Bayesian network requires the exact specification of a conditional probability distribution for all variables in the network. In case of limited data or disagreeing and/or partial expert opinions, this requirement is clearly unrealistic and renders the resulting model arbitrary; see Ref. [37, Section 1.1.4] for numerous other arguments against this 'precision requirement'. In order to avoid those problems, one can use the theory of *credal networks*, which, simply put, are Bayesian networks that allow for imprecisely specified local models. Initially, these were taken to be *credal sets* [21] (closed and convex sets of probability distributions), which explains

* Corresponding author.
E-mail addresses: jasper.debock@ugent.be (J. De Bock), gert.decooman@ugent.be (G. de Cooman).
URLs: http://users.ugent.be/~jdbock (J. De Bock), http://users.ugent.be/~gdcooma (G. de Cooman).

http://dx.doi.org/10.1016/j.ijar.2014.07.002

the terminology. However, as the theory progressed, other imprecise-probabilistic models such as lower previsions [24,37] and sets of desirable gambles entered the field as well [9–11,14,23].

In the present paper, we use the very general theory of *sets of desirable gambles* [3,17,38]. The key idea of this theory is that a subject's beliefs about the unknown outcome of an experiment can be modelled by means of the bets on this outcome—referred to as gambles—that he is willing to accept. Although sets of desirable gambles are not as well known as other (imprecise) probability models, they have definite advantages. To begin with, they are more expressive than most—if not all—other imprecise-probabilistic models, including the theories of credal sets and coherent lower previsions [32,37,38,42]: every set of desirable gambles has an associated lower prevision and credal set, but—in general—the original set of desirable gambles cannot be recovered from these derived models. A particularly interesting consequence of this added expressiveness is that conditioning on events with (lower or upper) probability zero becomes non-problematic [3,16,30]. Secondly, sets of desirable gambles are strongly connected to classical propositional logic [13,16,41], thereby providing a unified language for both logic and probability. Thirdly, they have the advantage of being operational, meaning that there is a practical way of constructing a model that represents a subject's beliefs [31,37].[1] And finally, our experience tells us that it is usually easier to construct proofs in the geometrically flavoured language of coherent sets of desirable gambles than in other, perhaps more familiar frameworks.

Three main kinds of credal networks can be distinguished, the difference between them being the notion of independence they adopt: strong independence, epistemic independence or epistemic irrelevance; see Cozman's pioneering work [4] for an overview. In a precise-probabilistic context, all these approaches coincide and reduce to a Bayesian network. Credal networks under strong independence are by far the most popular ones; see Refs. [2,6,27] for some nice overviews, containing numerous references to both theoretical results, algorithms and applications. In contrast, credal networks under epistemic independence have received almost no attention [4,12], a situation which is likely to persist due to their computational intractability. The current paper deals with the remaining option: credal networks under epistemic irrelevance. Let us start by stating some of their advantages.

(i) If the topology of the network is a *tree*, there is a *polynomial-time updating algorithm* that can compute posterior beliefs about a single target variable conditional on the observation of others [14]. We believe this to be promising, especially since the same inference problem is NP-hard for credal trees under strong independence [22]. Other promising algorithmic developments have also been made for the special case of imprecise hidden Markov models under epistemic irrelevance [1,9].
(ii) Epistemic irrelevance has a very *intuitive definition*: Y is irrelevant to X if knowing the value of Y does not change your beliefs about X; this is especially compelling if local models are elicited from expert knowledge. More technically, and in contrast with strong independence, it can also be given a direct *behavioural interpretation* [37].
(iii) Unlike the other two notions of independence, epistemic irrelevance is an *asymmetrical* concept. Since the graphical structure that underlies a credal network—a Directed Acyclic Graph—is asymmetrical in nature as well, they are particularly well suited to be combined with one another, both from a mathematical and a philosophical point of view.
(iv) Credal networks under epistemic irrelevance are based on assessments that are *less committal*. Therefore, they provide conservative outer approximations for the other two approaches.

Despite these advantages, credal networks under epistemic irrelevance have received relatively little attention so far; to our knowledge, Refs. [1,4,9–12,14,22,23] are the main contributions to the field. One of the main persisting problems is that—except for networks that are sufficiently small or have a tree topology—no efficient, exact or even approximate inference algorithm is known. We believe that this is to a great extent due to a profound lack of known theoretical properties. In the present paper, we start to remedy this situation by providing a firm theoretical foundation for credal networks under epistemic irrelevance.

We begin in Section 2 by providing a short introduction to the theory of sets of desirable gambles. We then go on to introduce and discuss important concepts such as directed acyclic graphs and epistemic irrelevance in Section 3, and use these in Section 4 to show how assessments of epistemic irrelevance can be combined with given local sets of desirable gambles to construct a joint model for a credal network under epistemic irrelevance. We call this the *irrelevant natural extension of the credal network* and prove that it is the most conservative coherent model that extends the local models and expresses all conditional irrelevancies encoded in the network. In the remainder of the paper, we develop some remarkable properties of this irrelevant natural extension. Section 5 presents what we consider to be our main technical achievement: a very general factorisation result and a closely related marginalisation property. In Section 6, we develop a tight connection with the *independent natural extension* [15,16] and show that it corresponds to a special case of the irrelevant natural extension. Our perhaps most important result is presented in Section 7: the irrelevant natural extension satisfies separation properties similar to the ones that are induced by d-separation in Bayesian networks. We introduce an asymmetrical version of d-separation, called *AD-separation*, and show that it implies epistemic irrelevance. Furthermore, since AD-separation is shown to satisfy all asymmetric graphoid properties (all graphoid properties except symmetry), the induced set of epistemic irrelevancies does so as well. We conclude the paper in Section 8, comment on how to translate our results to the framework

[1] This can be done, for example, by offering the subject certain gambles and asking him whether or not he strictly prefers them to the status quo.

of coherent lower previsions, discuss some algorithmic applications and present future avenues of research. In order to make our main argumentation as readable as possible, all technical proofs are collected in Appendix A.

We should note that some of our results have already been published in an earlier conference version of this paper [10]. The current version gives a more detailed exposition of these results, provides them with proofs—which were omitted in the conference version—, and extends them; notable examples of additional results are the connections with independent natural extension, presented in Section 6.

2. Sets of desirable gambles

Consider a variable X taking values in some non-empty finite set $\mathcal{X}$. Beliefs about the possible values this variable may assume can be modelled in various ways: probability mass functions, credal sets and coherent lower previsions are only a few of the many options. We choose to adopt a different approach: sets of desirable gambles. We will model a subject's beliefs regarding the value of a variable X by means of his behaviour: which gambles (or bets) on the unknown value of X does our subject strictly prefer to the status quo (the zero gamble). We give a brief survey of the basics of sets of desirable gambles; see Refs. [3,17,30–32,37,38,42] for more details, further discussion and connections with other imprecise-probabilistic models.

2.1. Desirable gambles

A gamble f is a real-valued map on $\mathcal{X}$ that is interpreted as an uncertain reward. If the value of the variable X turns out to be x, the (possibly negative) reward is $f(x)$. A non-zero gamble is called *desirable* to a subject if he strictly prefers to zero the transaction in which (i) the actual value x of the variable is determined, and (ii) he receives the reward $f(x)$. The zero gamble is therefore not considered to be desirable.

We model a subject's beliefs regarding the possible values $\mathcal{X}$ that a variable X can assume by means of a set $\mathcal{D}$ of desirable gambles—some subset of the set $\mathcal{G}(\mathcal{X})$ of all gambles on $\mathcal{X}$. For any two gambles f and g in $\mathcal{G}(\mathcal{X})$, we say that $f \geq g$ if $f(x) \geq g(x)$ for all x in $\mathcal{X}$ and $f > g$ if both $f \geq g$ and $f \neq g$. We use $\mathcal{G}(\mathcal{X})_{>0}$ to denote the set of all gambles $f \in \mathcal{G}(\mathcal{X})$ for which $f > 0$ and $\mathcal{G}(\mathcal{X})_{\leq 0}$ to denote the set of all gambles $f \in \mathcal{G}(\mathcal{X})$ for which $f \leq 0$. As a special kind of gambles we consider *indicators* $\mathbb{I}_A$ of events $A \subseteq \mathcal{X}$. $\mathbb{I}_A$ is equal to 1 if the event A occurs—the variable X assumes a value in A—and zero otherwise.

2.2. Coherence

In order to represent a rational subject's beliefs about the values a variable can assume, a set $\mathcal{D} \subseteq \mathcal{G}(\mathcal{X})$ of desirable gambles should satisfy some rationality requirements. If these requirements are met, we call the set $\mathcal{D}$ *coherent*. We require that for all $f, f_1, f_2 \in \mathcal{G}(\mathcal{X})$ and all real $\lambda > 0$:

D1. if $f \leq 0$ then $f \notin \mathcal{D}$; [avoiding null gain]
D2. if $f > 0$ then $f \in \mathcal{D}$; [desiring partial gain]
D3. if $f \in \mathcal{D}$ then $\lambda f \in \mathcal{D}$; [scaling]
D4. if $f_1, f_2 \in \mathcal{D}$ then $f_1 + f_2 \in \mathcal{D}$. [combination]

Requirements D3 and D4 turn $\mathcal{D}$ into a convex cone: $\operatorname{posi}(\mathcal{D}) = \mathcal{D}$, where we use the positive hull operator 'posi' that generates the set of finite strictly positive linear combinations of elements of its argument set:

$$\operatorname{posi}(\mathcal{D}) := \left\{ \sum_{k=1}^{n} \lambda_k f_k \colon f_k \in \mathcal{D}, \lambda_k \in \mathbb{R}_0^+, n \in \mathbb{N}_0 \right\}.$$

Here $\mathbb{R}_0^+$ is the set of all (strictly) positive real numbers, and $\mathbb{N}_0$ the set of all natural numbers (zero not included).

2.3. Natural extension

In practice, a set of desirable gambles is often elicited by presenting an expert a number of gambles and asking him whether or not he finds them desirable, resulting in an *assessment* of desirable gambles $\mathcal{A} \subseteq \mathcal{G}(\mathcal{X})$. However, such an assessment is not guaranteed to be coherent. Hence, the question arises whether $\mathcal{A}$ can be extended to—included in—a coherent set $\mathcal{D}$. It turns out that this is easily done; by applying D2–D4, we can use $\mathcal{A}$ to infer the desirability of other gambles. The largest set of desirable gambles that can be constructed in this way is

$$\mathcal{E}(\mathcal{A}) := \operatorname{posi}\big(\mathcal{A} \cup \mathcal{G}(\mathcal{X})_{>0}\big).$$

Since $\mathcal{E}(\mathcal{A})$ trivially satisfies D2–D4, we see that $\mathcal{E}(\mathcal{A})$ is coherent if and only if it avoids null gain [D1]. Furthermore, if $\mathcal{E}(\mathcal{A})$ is coherent, then it is the smallest—most conservative—coherent set of desirable gambles that contains $\mathcal{A}$ and it is then also equal to the intersection of all the coherent supersets of $\mathcal{A}$ [17]; in that case, we call $\mathcal{E}(\mathcal{A})$ the *natural extension* of $\mathcal{A}$.

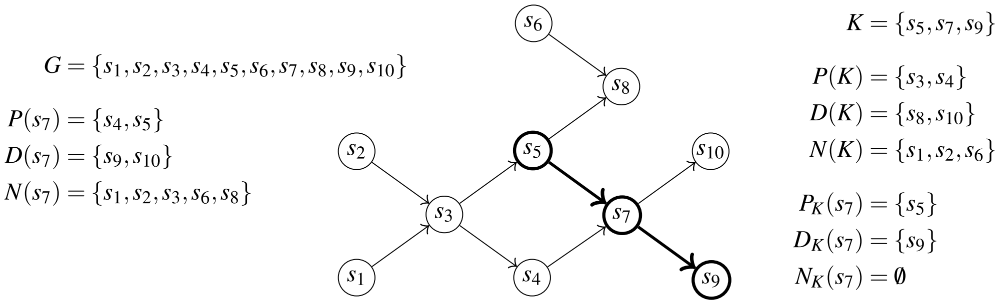

Fig. 1. Example of a directed acyclic graph (DAG). The sub-DAG that is associated with K is depicted in boldface.

3. Credal networks under epistemic irrelevance

In order to develop our results, we need to introduce some elementary, but nevertheless essential concepts. We start with important terminology related to Directed Acyclic Graphs (DAGs). Next, we show how we can use sets of desirable gambles to model our beliefs about the variables that are associated with such a DAG and how to express epistemic irrelevance in this language. Finally, we introduce local uncertainty models and explain our interpretation of the DAG that is associated with a credal network.

3.1. Directed acyclic graphs

A directed acyclic graph (DAG) is a graphical model that is well known for its use in Bayesian networks. It consists of a finite set of nodes (vertices), joined into a network by a set of directed edges, each edge connecting one node with another. Since this directed graph is assumed to be acyclic, it is not possible to follow a sequence of edges from node to node and end up at the same node one started out from.

We will call G the set of nodes s associated with a given DAG. For two nodes s and t, if there is a directed edge from s to t, we denote this as $s \to t$ and say that s is a *parent* of t and t is a *child* of s. For any node s, its set of parents is denoted by $P(s)$ and its set of children by $C(s)$. If a node s has no parents, that is, $P(s) = \emptyset$, then we call s a *root node*. If $C(s) = \emptyset$, then we call s a *leaf*, or *terminal node*.

Two nodes s and t are said to have a *path* between them if one can start from s, follow the edges of the DAG regardless of their direction and end up in t. In other words: one can find a sequence of nodes $s = s_1, \dots, s_n = t$, $n \geq 1$, such that for all $i \in \{1, \dots, n-1\}$ either $s_i \to s_{i+1}$ or $s_i \leftarrow s_{i+1}$. If this sequence is such that $s_i \to s_{i+1}$ for all $i \in \{1, \dots, n-1\}$ (all edges in the path point away from s), we say that there is a *directed path* from s to t and write $s \sqsubseteq t$. In that case we also say that s *precedes* t. If $s \sqsubseteq t$ and $s \neq t$, we say that s *strictly precedes* t and write $s \sqsubset t$. For any node s, we denote its set of *descendants* by $D(s) := \{t \in G : s \sqsubset t\}$ and its set of *non-parent non-descendants* by $N(s) := G \setminus (P(s) \cup \{s\} \cup D(s))$. We also use the shorthand notation $PN(s) := P(s) \cup N(s) = G \setminus (\{s\} \cup D(s))$ to refer to the so-called *non-descendants* of s.

Example 1. Consider the DAG in Fig. 1. For the node $s_7 \in G$, we find that $P(s_7) = \{s_4, s_5\}$, $D(s_7) = \{s_9, s_{10}\}$ and $N(s_7) = \{s_1, s_2, s_3, s_6, s_8\}$.

We extend these notions to subsets of G in the following way. For any $K \subseteq G$, $P(K) := (\bigcup_{s \in K} P(s)) \setminus K$ is its set of parents and $D(K) := (\bigcup_{s \in K} D(s)) \setminus K$ is its set of descendants. The non-parent non-descendants of K are given by[2] $N(K) := G \setminus (P(K) \cup K \cup D(K)) = \bigcap_{s \in K} N(s)$ and we also define $PN(K) := P(K) \cup N(K)$. In general, this last set cannot be referred to as the non-descendants of K since $P(K)$ and $D(K)$ are not necessarily disjoint. We call those subsets of G for which they are disjoint *closed*: a set $K \subseteq G$ is closed if for all $s, t \in K$ and any $k \in G$ such that $s \sqsubseteq k \sqsubseteq t$, it holds that $k \in K$. For closed $K \subseteq G$, we find that[3] $P(K) \cap D(K) = \emptyset$ and therefore $PN(K) = G \setminus (K \cup D(K))$. This means that, for closed K, $PN(K)$ can rightfully be referred to as the non-descendants of K.

Example 2. Let us illustrate these notions by means of the DAG in Fig. 1. For $K = \{s_5, s_7, s_9\}$, which is a closed subset of G, we find that $P(K) = \{s_3, s_4\}$, $D(K) = \{s_8, s_{10}\}$ and $N(K) = \{s_1, s_2, s_6\}$, implying that $P(K) \cap D(K) = \emptyset$.

With any subset K of G, we can associate a so-called *sub-DAG* of the DAG that is associated with G. The nodes of this sub-DAG are the elements of K and the directed edges of this sub-DAG are those edges in the original DAG that connect elements in K. For a sub-DAG that is associated with some subset K of G, we will use similar definitions as those

[2] A proof for the last equality can be found in Lemma 20(vii) in Appendix A.

[3] See Lemma 20(i) in Appendix A for a proof.

for the original DAG, adding the subset K as an index. As an example: for all $k \in K$, we denote by $P_K(k)$ the parents of k in the sub-DAG that is associated with the nodes in K. For all $K \subseteq G$ and $k \in K$, we have[4] $P_K(k) = P(k) \cap K$ and $P(k) \setminus P_K(k) = P(k) \cap P(K)$.

Example 3. Consider again the DAG in Fig. 1, where the sub-DAG that is associated with $K = \{s_5, s_7, s_9\} \subset G$ is depicted in boldface. We find that $P_K(s_7) = \{s_5\}$, $D_K(s_7) = \{s_9\}$ and $N_K(s_7) = \emptyset$.

3.2. Variables and gambles on them

With each node s of the network, we associate a variable X_s assuming values in some non-empty finite set $\mathcal{X}_s$. We denote by $\mathcal{G}(\mathcal{X}_s)$ the set of all gambles on $\mathcal{X}_s$. We extend this notation to more complicated situations as follows. If S is any subset of G, then we denote by X_S the tuple of variables whose components are the X_s for all $s \in S$. This new joint variable assumes values in the finite set $\mathcal{X}_S := \times_{s\in S} \mathcal{X}_s$ and the corresponding set of gambles is denoted by $\mathcal{G}(\mathcal{X}_S)$. When $S = \emptyset$, we let $\mathcal{X}_\emptyset$ be a singleton. The corresponding variable $X_\emptyset$ can then only assume this single value, so there is no uncertainty about it. $\mathcal{G}(\mathcal{X}_\emptyset)$ can then be identified with the set $\mathbb{R}$ of real numbers. Generic elements of $\mathcal{X}_s$ are denoted by x_s or z_s and similarly for x_S and z_S in $\mathcal{X}_S$. Also, if we mention a tuple z_S, then for any $t \in S$, the corresponding element in the tuple will be denoted by z_t.

We will use the simplifying device of identifying a gamble f_S on $\mathcal{X}_S$ with its *cylindrical extension* to $\mathcal{X}_U$, where $S \subseteq U \subseteq G$: the gamble f_U on $\mathcal{X}_U$ defined by $f_U(x_U) := f_S(x_S)$ for all $x_U \in \mathcal{X}_U$. For instance, if $\mathcal{K} \subseteq \mathcal{G}(\mathcal{X}_G)$, this allows us to consider $\mathcal{K} \cap \mathcal{G}(\mathcal{X}_S)$ as the set of those gambles in $\mathcal{K}$ that depend only on the variable X_S.

3.3. Modelling our beliefs about the network

Throughout, we consider sets of desirable gambles as models for a subject's beliefs about the values that certain variables in the network may assume. An important contribution of this paper, further on in Section 4, will be to show how to construct a joint model for our network, being a coherent set $\mathcal{D}_G$ of desirable gambles on $\mathcal{X}_G$.

From such a joint model, one can derive both conditional and marginal models [16,17]. Let us start by explaining how to condition the global model $\mathcal{D}_G$. Consider a non-empty set $A_I \subseteq \mathcal{X}_I$, with $I \subseteq G$, and assume that we want to update the model $\mathcal{D}_G$ with the information that $X_I \in A_I$. This leads to the following updated set of desirable gambles:

$$\mathcal{D}_G \rfloor A_I := \left\{ f \in \mathcal{G}(\mathcal{X}_{G\setminus I}) : \mathbb{I}_{A_I} f \in \mathcal{D}_G \right\},$$

which represents our subject's beliefs about the value of the variable $X_{G\setminus I}$, conditional on the observation that X_I assumes a value in A_I. This definition is very intuitive, since $\mathbb{I}_{A_I} f$ is the unique gamble that is called off (is equal to zero) if $X_I \notin A_I$ and equal to f if $X_I \in A_I$. Since $\mathbb{I}_{\{x_\emptyset\}} = 1$, the special case of conditioning on the certain variable $X_\emptyset$ yields no problems: it amounts to not conditioning at all. The connection with the precise-probabilistic version of conditioning is discussed in Ref. [3].

Marginalisation too is very intuitive in the language of sets of desirable gambles. Suppose we want to derive a marginal model for our subject's beliefs about the variable X_O, where O is some subset of G. This can be done by using the set of desirable gambles that belong to $\mathcal{D}_G$ but only depend on the variable X_O:

$$\operatorname{marg}_O(\mathcal{D}_G) := \left\{ f \in \mathcal{G}(\mathcal{X}_O) : f \in \mathcal{D}_G \right\} = \mathcal{D}_G \cap \mathcal{G}(\mathcal{X}_O).$$

Now let I and O be *disjoint* subsets of G and let A_I be any non-empty subset of $\mathcal{X}_I$. By sequentially applying the process of conditioning and marginalisation we can obtain conditional marginal models for our subject's beliefs about the value of the variable X_O, conditional on the observation that X_I assumes a value in A_I:

$$\operatorname{marg}_O(\mathcal{D}_G \rfloor A_I) = \left\{ f \in \mathcal{G}(\mathcal{X}_O) : \mathbb{I}_{A_I} f \in \mathcal{D}_G \right\}. \tag{1}$$

Conditioning and marginalisation are special cases of Eq. (1); they can be obtained by letting $O = G \setminus I$ or $I = \emptyset$. If A_I is a singleton $\{x_I\}$, with $x_I \in \mathcal{X}_I$, we will use the shorthand notation $\operatorname{marg}_O(\mathcal{D}_G \rfloor x_I) := \operatorname{marg}_O(\mathcal{D}_G \rfloor \{x_I\})$.

Since coherence is trivially preserved under both conditioning and marginalisation, we find that if the joint model $\mathcal{D}_G$ is coherent, all the derived models will also be coherent. For additional properties of these marginalisation and conditioning operators, we refer to Ref. [16].

3.4. Epistemic irrelevance

At this point, we have the necessary tools to introduce one of the most important concepts for this paper, that of epistemic irrelevance. We describe the case of conditional irrelevance, as the unconditional version of epistemic irrelevance can easily be recovered as a special case.

[4] See Lemma 19 in Appendix A for a proof.

Consider three disjoint subsets C, I, and O of G. When a subject judges X_I to be *epistemically irrelevant to X_O conditional on X_C*, denoted as $\mathrm{IR}(I,O|C)$, he assumes that if he knew the value of X_C, then learning in addition which value X_I assumes in $\mathcal{X}_I$ would not affect his beliefs about X_O. More formally put, he assumes for all $x_C \in \mathcal{X}_C$ and $x_I \in \mathcal{X}_I$ that:

$$\mathrm{marg}_O(\mathcal{D}_G \rfloor x_{C\cup I}) = \mathrm{marg}_O(\mathcal{D}_G \rfloor x_C).$$

Alternatively, a subject can make the even stronger statement that he judges X_I to be epistemically *subset-irrelevant* to X_O conditional on X_C, denoted as $\mathrm{SIR}(I,O|C)$. In that case, he assumes that if he knew the value of X_C, then receiving the additional information that X_I is an element of any non-empty subset A_I of $\mathcal{X}_I$ would not affect his beliefs about X_O. In other words, he assumes for all $x_C \in \mathcal{X}_C$ and all non-empty $A_I \subseteq \mathcal{X}_I$ that:

$$\mathrm{marg}_O\big(\mathcal{D}_G \rfloor \{x_C\} \times A_I\big) = \mathrm{marg}_O(\mathcal{D}_G \rfloor x_C).$$

Making a subset-irrelevance statement $\mathrm{SIR}(I,O|C)$ implies the corresponding irrelevance statement $\mathrm{IR}(I,O|C)$. Even stronger, it implies for all $I' \subseteq I$ that $\mathrm{IR}(I',O|C)$. The converse does not hold in general; see Example 4. However, as we will show further on, credal networks under epistemic irrelevance are a useful exception: although we define the joint model by imposing irrelevance, it will also satisfy subset-irrelevance.

For the unconditional irrelevance case it suffices, in the discussion above, to let $C = \emptyset$. This makes sure the variable X_C has only one possible value, so conditioning on that variable amounts to not conditioning at all.

Example 4. Consider two variables X_{s_1} and X_{s_2} that take values in $\mathcal{X}_{s_1} := \{a,b,c\}$ and $\mathcal{X}_{s_2} := \{0,1\}$, respectively. Hence, $G = \{s_1,s_2\}$. Furthermore, let $I := \{s_1\}$, $O := \{s_2\}$, $C := \emptyset$, $A_I := \{a,b\} \subseteq \mathcal{X}_I$ and let $g \in \mathcal{G}(\mathcal{X}_O)$ be the gamble that is defined by $g(1) = -g(0) := 1$. Consider now the set of desirable gambles

$$\mathcal{D}_G := \mathcal{E}\big(\{\mathbb{I}_{A_I} g\}\big) = \big\{f \in \mathcal{G}(\mathcal{X}_G)\colon f \geq \mathbb{I}_{A_I} g \text{ or } f > 0\big\}.$$

Then, for all $x_I \in \mathcal{X}_I$, $\mathrm{marg}_O(\mathcal{D}_G \rfloor x_I) = \mathcal{G}(\mathcal{X}_O)_{>0} = \mathrm{marg}_O(\mathcal{D}_G)$, implying that X_I is epistemically irrelevant to X_O. However, X_I is not subset-irrelevant to X_O because $\mathrm{marg}_O(\mathcal{D}_G \rfloor A_I) = \mathcal{E}(\{g\}) = \{f \in \mathcal{G}(\mathcal{X}_O)\colon f \geq g \text{ or } f > 0\}$.

We consider subset-irrelevance to be the more natural of the two concepts, as it requires *all* information about X_I to be irrelevant, which is what—in our opinion—irrelevance should mean. For example, in Example 4, although X_{s_1} is epistemically irrelevant to X_{s_2}, learning that $X_{s_1} \neq c$ does affect our belief model for X_{s_2}; this would be impossible if X_{s_1} were epistemically subset-irrelevant to X_{s_2}.

Irrelevance and subset-irrelevance can also be extended to cases where I, O and C are not disjoint, but $I \setminus C$ and $O \setminus C$ are. We then call X_I epistemically (subset-)irrelevant to X_O conditional on X_C provided that $X_{I\setminus C}$ is epistemically (subset-)irrelevant to $X_{O\setminus C}$ conditional on X_C. Although these cases are admittedly artificial, they will help us state and prove some of the graphoid properties further on.

3.5. Local uncertainty models

We now add *local uncertainty models* to each of the nodes s in our network. These local models are assumed to be given beforehand and will be used further on in Section 4 as basic building blocks for constructing a joint model for a given network.

If s is not a root node of the network, i.e., has a non-empty set of parents $P(s)$, then we have a conditional local model for every instantiation of its parents: for each $x_{P(s)} \in \mathcal{X}_{P(s)}$, we have a coherent set $\mathcal{D}_{s\rfloor x_{P(s)}}$ of desirable gambles on $\mathcal{X}_s$. It represents our subject's beliefs about the variable X_s conditional on its parents $X_{P(s)}$ assuming the value $x_{P(s)}$.

If s is a root node, i.e., has no parents, then our subject's local beliefs about the variable X_s are represented by an unconditional local model. It should be a coherent set of desirable gambles and will be denoted by $\mathcal{D}_s$. As was explained in Section 3.3, we can also use the common generic notation $\mathcal{D}_{s\rfloor x_{P(s)}}$ in this unconditional case, since for a root node s, its set of parents $P(s)$ is equal to the empty set $\emptyset$.

3.6. The interpretation of the graphical model

In classical Bayesian networks, the graphical structure is taken to represent the following assessments: for any node s, conditional on its parent variables, the associated variable is independent of its non-parent non-descendant variables [20, Section 3.2.2]. When generalising this interpretation to credal networks, the classical notion of independence gets replaced by a more general, imprecise-probabilistic notion of independence. In this paper, we choose to use epistemic irrelevance, as introduced in Section 3.4; see the Introduction for discussion, motivation, and relevant references. It is useful to know that in the special case of precise uncertainty models, epistemic irrelevance is equivalent to the classical notion of independence, making the interpretation of the graphical structure of the network equivalent to the one in Bayesian networks.

Let us state our interpretation more formally. We assume that the graphical structure of the network embodies the following conditional irrelevance assessments, turning the network into a *credal network under epistemic irrelevance*. Consider any node s in the network, its set of parents $P(s)$ and its set of non-parent non-descendants $N(s)$. Then *conditional on* $X_{P(s)}$, $X_{N(s)}$ *is assumed to be epistemically irrelevant to* X_s:

$$\mathrm{IR}\big(N(s), \{s\}\big|P(s)\big).$$

For a coherent set of desirable gambles $\mathcal{D}_G$ that describes our subject's global beliefs about all the variables in the network, this has the following consequences. For every $s \in G$ and all $x_{PN(s)} \in \mathcal{X}_{PN(s)}$, $\mathcal{D}_G$ must satisfy:

$$\operatorname{marg}_s(\mathcal{D}_G \rfloor x_{PN(s)}) = \operatorname{marg}_s(\mathcal{D}_G \rfloor x_{P(s)}). \tag{2}$$

4. Constructing a joint model

We now show how to construct a joint model for the variables in the network, and argue that it is the most conservative coherent model that extends the local models and expresses all conditional irrelevancies encoded in the network. But before we do so, we need to ask ourselves the following question: suppose we have a global set of desirable gambles $\mathcal{D}_G$, how do we express that such a model is compatible with the assessments encoded in the network?

4.1. Defining properties of the joint model

We will require our joint model to satisfy the following four properties. First of all, our global model should extend the local assessments, in the sense that the local models that are derived from the global one by marginalisation should include—be at least as informative as—the given local models:

G1. The joint model $\mathcal{D}_G$ marginalises to supersets of the local uncertainty models:

$$\mathcal{D}_{s\rfloor x_{P(s)}} \subseteq \operatorname{marg}_s(\mathcal{D}_G \rfloor x_{P(s)}) \quad \text{for all } s \in G \text{ and } x_{P(s)} \in \mathcal{X}_{P(s)}.$$

The second requirement is that our model should reflect all epistemic irrelevancies encoded in the graphical structure of the network:

G2. $\mathcal{D}_G$ satisfies all equalities that are imposed by Eq. (2).

The third requirement is that our model should be coherent:

G3. $\mathcal{D}_G$ is coherent (satisfies requirements D1–D4).

Since requirements G1–G3 do not determine a unique global model, we impose a final requirement to ensure that all inferences we make on the basis of our global models are as conservative as possible, and are therefore based on no other considerations than what is encoded in the network:

G4. $\mathcal{D}_G$ is the smallest set of desirable gambles on $\mathcal{X}_G$ satisfying requirements G1–G3: it is a subset of any other set that satisfies them.

We will now show how to construct the unique global model $\mathcal{D}_G$ that satisfies all four requirements G1–G4.

4.2. An intuitive expression for the joint model

Let us start by looking at a single given marginal model $\mathcal{D}_{s\rfloor z_{P(s)}}$ and investigate some of its implications for the joint model $\mathcal{D}_G$. Consider any node s and fix values $z_{P(s)}$ and $z_{N(s)}$ for its parents and non-parent non-descendants. Due to requirements G1 and G2, any gamble $f \in \mathcal{D}_{s\rfloor z_{P(s)}}$ should also be an element of $\operatorname{marg}_s(\mathcal{D}_G \rfloor z_{PN(s)})$, which by definition means that $\mathbb{I}_{\{z_{PN(s)}\}} f \in \mathcal{D}_G$. Inspired by this observation, we introduce the following set of gambles on $\mathcal{X}_G$:

$$\mathcal{A}_G^{\mathrm{irr}} := \{\mathbb{I}_{\{z_{PN(s)}\}} f : s \in G,\ z_{PN(s)} \in \mathcal{X}_{PN(s)},\ f \in \mathcal{D}_{s\rfloor z_{P(s)}}\}. \tag{3}$$

It should now be clear that $\mathcal{A}_G^{\mathrm{irr}}$ must be a subset of our joint model $\mathcal{D}_G$.

Proposition 1. $\mathcal{A}_G^{\mathrm{irr}}$ *is a subset of any joint model* $\mathcal{D}_G$ *that satisfies requirements* G1 *and* G2.

Since our eventual joint model should also be coherent (satisfy requirement G3), and thus in particular should be a convex cone, we can derive the following corollary.

Corollary 2. $\operatorname{posi}(\mathcal{A}_G^{\mathrm{irr}})$ *is a subset of any joint model* $\mathcal{D}_G$ *that satisfies requirements* G1–G3.

We now suggest the following expression for the joint model describing our subject's beliefs about the variables in the network:

$$\mathcal{D}_G^{\mathrm{irr}} := \operatorname{posi}\big(\mathcal{A}_G^{\mathrm{irr}}\big). \tag{4}$$

We will refer to $\mathcal{D}_G^{\mathrm{irr}}$ as the *irrelevant natural extension* of the local models $\mathcal{D}_{s\rfloor x_{P(s)}}$. Since we know from Corollary 2 that it is guaranteed to be a subset of the joint model we are looking for, it is rather natural to propose it as a candidate for the joint model itself. In the next section, we set out to prove that $\mathcal{D}_G^{\mathrm{irr}}$ is indeed the unique joint model satisfying all four requirements G1–G4.

4.3. Justifying our expression for the joint model

We start by proving a number of useful properties of $\mathcal{D}_G^{\mathrm{irr}}$.

Proposition 3. *A gamble $f \in \mathcal{G}(\mathcal{X}_G)$ is an element of $\mathcal{D}_G^{\mathrm{irr}}$ if and only if it can be written as:*

$$f = \sum_{s\in G} \sum_{z_{PN(s)}\in\mathcal{X}_{PN(s)}} \mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}}, \tag{5}$$

where $f_{s,z_{PN(s)}} \in \mathcal{D}_{s\rfloor z_{P(s)}} \cup \{0\}$ for every $s \in G$ and all $z_{PN(s)} \in \mathcal{X}_{PN(s)}$, and at least one of them is non-zero.

Proposition 4. *$\mathcal{G}(\mathcal{X}_G)_{>0}$ is a subset of $\mathcal{D}_G^{\mathrm{irr}}$ and, consequently, $\mathcal{D}_G^{\mathrm{irr}} = \mathcal{E}(\mathcal{A}_G^{\mathrm{irr}})$.*

These two propositions serve as a first step towards the following coherence result, which states that our joint model $\mathcal{D}_G^{\mathrm{irr}}$ satisfies requirement G3.

Proposition 5. *$\mathcal{D}_G^{\mathrm{irr}}$ satisfies requirement* G3*: it is a coherent set of desirable gambles.*

The crucial step in our proof for this result is to consider a specific Bayesian network that has the same topology as our credal network and to use the corresponding joint probability mass function to construct a separating hyperplane argument. In this way, we are using existing coherence results for Bayesian networks to prove their counterparts for credal networks.

Next, we turn to an important factorisation result that is essential in order to prove that our joint model extends the local models and expresses all conditional irrelevancies encoded in the network, and therefore satisfies G1 and G2.

Proposition 6. *Fix arbitrary $s \in G$, $x_{P(s)} \in \mathcal{X}_{P(s)}$ and $g \in \mathcal{G}(\mathcal{X}_{N(s)})_{>0}$. Then for every $f \in \mathcal{G}(\mathcal{X}_s)$:*

$$g\mathbb{I}_{\{x_{P(s)}\}} f \in \mathcal{D}_G^{\mathrm{irr}} \quad\Leftrightarrow\quad f \in \mathcal{D}_{s\rfloor x_{P(s)}}.$$

Corollary 7. *$\mathcal{D}_G^{\mathrm{irr}}$ satisfies requirements* G1 *and* G2*. Even stronger: it holds for every $s \in G$ and all $x_{PN(s)} \in \mathcal{X}_{PN(s)}$ that*

$$\operatorname{marg}_s\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor x_{PN(s)}\big) = \operatorname{marg}_s\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor x_{P(s)}\big) = \mathcal{D}_{s\rfloor x_{P(s)}}.$$

Notice that, although G1 only requires $\operatorname{marg}_s(\mathcal{D}_G\rfloor x_{P(s)})$ to be a superset of $\mathcal{D}_{s\rfloor x_{P(s)}}$, the irrelevant natural extension $\mathcal{D}_G^{\mathrm{irr}}$ also satisfies a modified, stronger version of G1: $\operatorname{marg}_s(\mathcal{D}_G^{\mathrm{irr}}\rfloor x_{P(s)}) = \mathcal{D}_{s\rfloor x_{P(s)}}$.

We now have all tools necessary to formulate a crucial result. It is the first important contribution of this paper and provides a justification for the joint model $\mathcal{D}_G^{\mathrm{irr}}$ that was proposed in Eq. (4).

Theorem 8. *The irrelevant natural extension $\mathcal{D}_G^{\mathrm{irr}}$ is the unique set of desirable gambles on $\mathcal{X}_G$ that satisfies all four requirements* G1–G4*.*

It is already apparent from Proposition 6 that the properties of the irrelevant natural extension $\mathcal{D}_G^{\mathrm{irr}}$ are not limited to G1–G4. As a first example, Proposition 6 implies that for any node s, conditional on its parent variables $X_{P(s)}$, the non-parent non-descendant variables $X_{N(s)}$ are not only epistemically irrelevant, but also subset-irrelevant to X_s.

Corollary 9. *For all nodes $s \in G$, the subset-irrelevance statement* $\mathrm{SIR}(N(s), \{s\}|P(s))$ *is satisfied: for any $x_{P(s)} \in \mathcal{X}_{P(s)}$ and non-empty $A_{N(s)} \subseteq \mathcal{X}_{N(s)}$, it holds that*

$$\operatorname{marg}_s\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor \{x_{P(s)}\}\times A_{N(s)}\big) = \operatorname{marg}_s\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor x_{P(s)}\big).$$

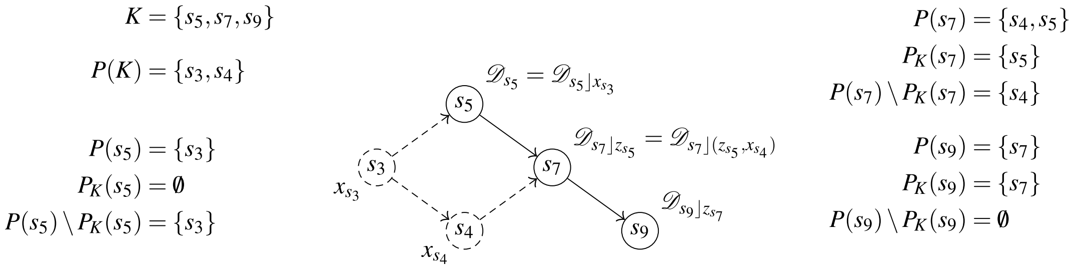

Fig. 2. Local models for a sub-DAG.

In the remainder of this paper, we establish a number of even stronger properties of $\mathcal{D}_G^{\mathrm{irr}}$. However, before we do so, let us take a small step back to take a look at the larger picture.

If we were to drop the irrelevance assessments, that is, if we were only to impose requirements G1, G3 and G4, then the unique model to satisfy these three requirements would be the smallest coherent set of desirable gambles that extends the local models or, in other words, the natural extension of

$$\mathcal{A}_G := \{\mathbb{I}_{\{z_{P(s)}\}} f : s \in G,\ z_{P(s)} \in \mathcal{X}_{P(s)},\ f \in \mathcal{D}_{s\rfloor z_{P(s)}}\}.$$

By including G2, we are applying a more general form of natural extension that combines an assessment of gambles with structural assessments. In our case, these structural assessments consist of a very specific set of epistemic irrelevancies—G2—and the resulting 'generalised' natural extension of $\mathcal{A}_G$, which we call the irrelevant natural extension, turns out to be the 'traditional' natural extension of the extended assessment $\mathcal{A}_G^{\mathrm{irr}}$, which is obtained by combining the original assessment $\mathcal{A}_G$ with the irrelevancies that are imposed by G2.

In principle, this generalised form of natural extension can be applied to other structural assessments as well. Cozman [4, p. 208] discusses the possibility of considering credal networks with arbitrary assessments of epistemic irrelevance; the special case of credal networks under epistemic independence is discussed in Ref. [12]. However, unlike in our special case, it is not always easy—and sometimes even impossible—to obtain a closed-form expression for the resulting joint model of these credal networks. Finally, one can also combine natural extension with structural judgements other than epistemic irrelevance. For example: Ref. [17] considers the structural judgement of exchangeability. That being said, for the remainder of this paper, we focus on the special case of the irrelevant natural extension $\mathcal{D}_G^{\mathrm{irr}}$.

5. Additional marginalisation properties

As explained in Section 3.1, a subset K of G can be associated with a so-called sub-DAG of the original DAG. Similarly to what we have done for the original DAG, we can use Eq. (4) to construct a joint model for this sub-DAG. All we need to do is to provide, for every $s \in K$ and $z_{P_K(s)} \in \mathcal{X}_{P_K(s)}$, a local model $\mathcal{D}_{s\rfloor z_{P_K(s)}}$.

One particular way of providing these local models is to derive them from the ones of the original DAG. The starting point to do so is fixing a value $x_{P(K)} \in \mathcal{X}_{P(K)}$ for the parent variables of K. This provides us, for every $s \in K$, with a value $x_{P(s)\setminus P_K(s)} \in \mathcal{X}_{P(s)\setminus P_K(s)}$ because $P(s) \setminus P_K(s) \subseteq P(K)$. For every $s \in K$ and $z_{P_K(s)} \in \mathcal{X}_{P_K(s)}$, we can then identify the local model $\mathcal{D}_{s\rfloor z_{P_K(s)}}$ of the sub-DAG with the local model $\mathcal{D}_{s\rfloor z_{P(s)}}$ of the original DAG, where $z_{P(s)\setminus P_K(s)} = x_{P(s)\setminus P_K(s)}$. In other words, for every $s \in K$ and $z_{P_K(s)} \in \mathcal{X}_{P_K(s)}$

$$\mathcal{D}_{s\rfloor z_{P_K(s)}} = \mathcal{D}_{s\rfloor (z_{P_K(s)}, x_{P(s)\setminus P_K(s)})}. \tag{6}$$

For every $K \subseteq G$ and all $x_{P(K)} \in \mathcal{X}_{P(K)}$, the resulting joint model for the sub-DAG that is associated with K is given by

$$\mathcal{D}_{K\rfloor x_{P(K)}}^{\mathrm{irr}} := \mathrm{posi}\big(\mathcal{A}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}\big), \tag{7}$$

where

$$\mathcal{A}_{K\rfloor x_{P(K)}}^{\mathrm{irr}} := \{\mathbb{I}_{\{z_{PN_K(s)}\}} f : s \in K,\ z_{PN_K(s)} \in \mathcal{X}_{PN_K(s)},\ f \in \mathcal{D}_{s\rfloor (z_{P_K(s)}, x_{P(s)\setminus P_K(s)})}\}.$$

Example 5. Consider again the DAG in Fig. 1 and the sub-DAG induced by the closed subset $K = \{s_5, s_7, s_9\} \subset G$. In order to use Eq. (7) to construct a joint model for this sub-DAG, we need to fix a value $x_{P(K)} \in \mathcal{X}_{P(K)}$ for the variables associated with the parents of K. Equivalently, since $P(K) = \{s_3, s_4\}$, we need to fix values $x_{s_3} \in \mathcal{X}_{s_3}$ and $x_{s_4} \in \mathcal{X}_{s_4}$. We can now construct local models by means of Eq. (6). For the node s_5, we obtain an unconditional local model $\mathcal{D}_{s_5} = \mathcal{D}_{s_5\rfloor x_{s_3}}$. For the node s_7, this yields, for every $z_{s_5} \in \mathcal{X}_{s_5}$, a conditional local model $\mathcal{D}_{s_7\rfloor z_{s_5}} = \mathcal{D}_{s_7\rfloor (z_{s_5}, x_{s_4})}$. Finally, for the node s_9, we obtain, for every $z_{s_7} \in \mathcal{X}_{s_7}$, a conditional local model $\mathcal{D}_{s_9\rfloor z_{s_7}}$. Fig. 2 provides a graphical representation of the construction of these local models.

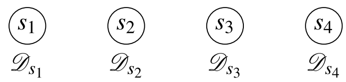

Fig. 3. Example of a credal network consisting of disconnected nodes.

A question that now naturally arises is whether these joint models for sub-DAGs, as given by Eq. (7), can be related to the original joint model $\mathcal{D}_G^{\mathrm{irr}}$. It turns out that, for subsets K of G that are closed, this is indeed the case.

Theorem 10. *If K is a closed subset of G, then for any $x_{P(K)} \in \mathcal{X}_{P(K)}$, $g \in \mathcal{G}(\mathcal{X}_{N(K)})_{>0}$ and $f \in \mathcal{G}(\mathcal{X}_K)$:*

$$g\mathbb{I}_{\{x_{P(K)}\}} f \in \mathcal{D}_G^{\mathrm{irr}} \quad \Leftrightarrow \quad f \in \mathcal{D}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}.$$

Our proof for the reverse implication is complex and elaborate, but its core is a simple separating hyperplane argument. Similar to what we have done in the proof of Proposition 6, we construct a joint probability mass function to perform the separation. However, in contrast with the proof of Proposition 6, a factorising probability mass function is no longer sufficient. This makes constructing the joint probability mass function that is used in Theorem 10 both complex and elaborate. We consider Theorem 10 to be the main technical achievement of this paper. It is a significant generalisation of Proposition 6 [with $K = \{s\}$] and has a number of important consequences. As a first example, it implies the following generalisations of Corollaries 7 and 9.

Corollary 11. *For all closed $K \subseteq G$, $x_{P(K)} \in \mathcal{X}_{P(K)}$ and non-empty $A_{N(K)} \subseteq \mathcal{X}_{N(K)}$, we have that*

$$\mathrm{marg}_K\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor\{x_{P(K)}\} \times A_{N(K)}\big) = \mathcal{D}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}.$$

Corollary 12. *For all closed sets $K \subseteq G$, we have that the subset-irrelevance statement $\mathrm{SIR}(N(K), K|P(K))$ is satisfied: for any $x_{P(K)} \in \mathcal{X}_{P(K)}$ and non-empty $A_{N(K)} \subseteq \mathcal{X}_{N(K)}$, it holds that*

$$\mathrm{marg}_K\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor\{x_{P(K)}\} \times A_{N(K)}\big) = \mathrm{marg}_K\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor x_{P(K)}\big).$$

6. Connections with independent natural extension

Let us now consider the special case of a credal network for which the underlying DAG has no edges or, equivalently, consists of disconnected nodes only; see Fig. 3 for an example with four nodes. In that case, these nodes clearly have neither parents nor descendants. Consequently, for every $s \in G$, the local model $\mathcal{D}_s$ is unconditional and the non-descendants are given by $PN(s) = G \setminus \{s\}$. It is therefore easy to see that in this particular case, Eq. (4) reduces to

$$\mathcal{D}_G^{\mathrm{irr}} = \mathrm{posi}\big(\{\mathbb{I}_{\{z_{G\setminus\{s\}}\}} f : s \in G,\ z_{G\setminus\{s\}} \in \mathcal{X}_{G\setminus\{s\}},\ f \in \mathcal{D}_s\}\big). \tag{8}$$

Due to Corollary 12, we know that $\mathcal{D}_G^{\mathrm{irr}}$, as given by Eq. (8), is *independent* [16, Definition 5], meaning that

$$\mathrm{marg}_O\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor x_I\big) = \mathrm{marg}_O\big(\mathcal{D}_G^{\mathrm{irr}}\big) \quad \text{for all disjoint subsets } I \text{ and } O \text{ of } G, \text{ and all } x_I \in \mathcal{X}_I.$$

Since $\mathcal{D}_G^{\mathrm{irr}}$ also marginalises to its local models (Corollary 7), we find that $\mathcal{D}_G^{\mathrm{irr}}$, as given by Eq. (8), is an *independent product* of the local models $\mathcal{D}_s$, $s \in G$, that is also coherent (Proposition 5).[5] Moreover, due to Corollary 2 and Eq. (4), it is a subset of any other coherent independent product of $\mathcal{D}_s$, $s \in G$. This makes $\mathcal{D}_G^{\mathrm{irr}}$, as given by Eq. (8), the unique smallest coherent set of desirable gambles on $\mathcal{X}_G$ that is an independent product of $\mathcal{D}_s$, $s \in G$. This is called the *independent natural extension* of the local models $\mathcal{D}_s$, $s \in G$, and is denoted as $\bigotimes_{s\in G}\mathcal{D}_s$ [16, Section 7]. We can therefore conclude that, for credal networks for which the underlying DAG consists of disconnected nodes only,

$$\mathcal{D}_G^{\mathrm{irr}} = \bigotimes_{s\in G}\mathcal{D}_s = \mathrm{posi}\big(\{\mathbb{I}_{\{z_{G\setminus\{s\}}\}} f : s \in G,\ z_{G\setminus\{s\}} \in \mathcal{X}_{G\setminus\{s\}},\ f \in \mathcal{D}_s\}\big), \tag{9}$$

a result that was already mentioned in Ref. [16, Section 10]. Consequently, quite a few of the results we obtain in the present paper can be regarded as generalisations of those in Ref. [16].[6] Our next two results show that the connection between our irrelevant natural extension of a network and the independent natural extension, as defined in Ref. [16], goes much further than Eq. (9).

[5] $\mathcal{D}_G$ is an independent product of the local models $\mathcal{D}_s$, $s \in G$, if it marginalises to these local models and is furthermore independent [16, p. 618].

[6] Ref. [16, Proposition 15] is a special case of Proposition 5, Ref. [16, Propositions 17 and 18] are both special cases of Theorem 11 and by combining Theorem 8 with Corollary 12, we can generalise Ref. [16, Theorem 19]. Finally, the associativity result in Ref. [16, Theorem 20] can be regarded as a special case of Proposition 13. Indeed, it suffices to (i) apply Proposition 13 to a DAG consisting of two separate, disconnected sub-DAGs, each of which consists of disconnected nodes only, and (ii) to subsequently apply Eq. (9) to $\mathcal{D}_G^{\mathrm{irr}}$, $\mathcal{D}_{G_1}^{\mathrm{irr}}$ and $\mathcal{D}_{G_2}^{\mathrm{irr}}$.

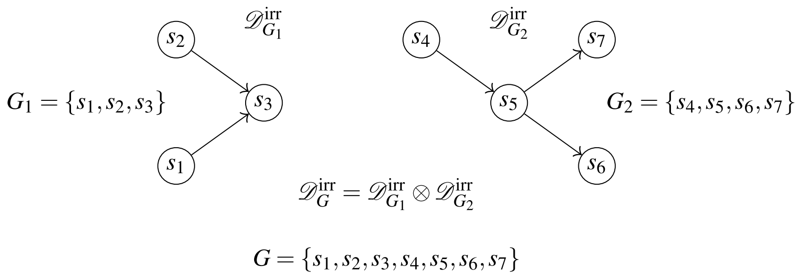

Fig. 4. Illustration of Proposition 13.

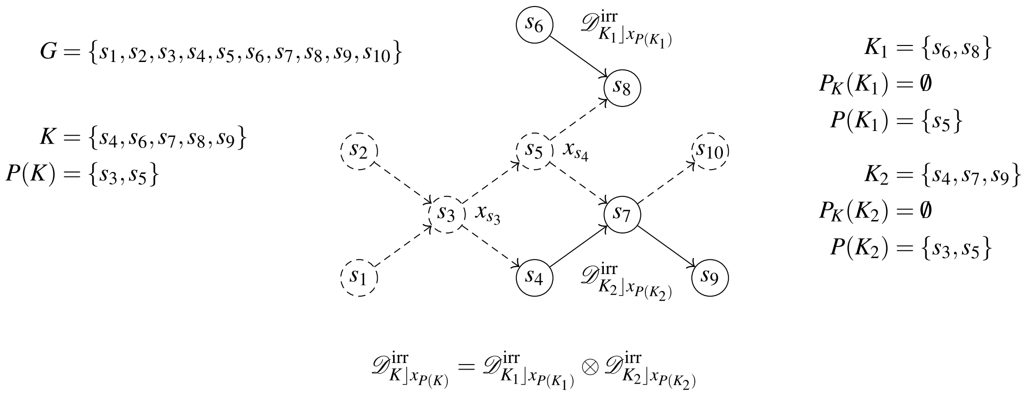

Fig. 5. Illustration of Theorem 14.

Proposition 13. *Consider any partition $G_1, \ldots, G_n$ of G such that $P(G_i) = \emptyset$ for all $i \in \{1, \ldots, n\}$. Or equivalently, let the DAG of the complete network consist of n separate, disconnected sub-DAGs, each of which has G_i as its set of nodes, with $i \in \{1, \ldots, n\}$. Then*

$$\mathcal{D}^{\text{irr}}_G = \bigotimes_{i=1}^{n} \mathcal{D}^{\text{irr}}_{G_i} = \operatorname{posi}\left(\left\{\mathbb{I}_{\{z_{G \setminus G_i}\}} f : i \in \{1, \ldots, n\},\ z_{G \setminus G_i} \in \mathcal{X}_{G \setminus G_i},\ f \in \mathcal{D}^{\text{irr}}_{G_i}\right\}\right),$$

where $\mathcal{D}^{\text{irr}}_G$ is the irrelevant natural extension of the complete network, as given by Eq. (4), and, for all $i \in \{1, \ldots, n\}$, $\mathcal{D}^{\text{irr}}_{G_i}$ is the irrelevant natural extension of the network that has the sub-DAG associated with G_i as its graphical structure, as given by Eq. (7).[7]

Fig. 4 illustrates this result with a simple example. It should be clear that Eq. (9) is a special case of Proposition 13.[8] Theorem 14 generalises Proposition 13 even further.

Theorem 14. *Consider a closed subset K of G and a partition $K_1, \ldots, K_n$ of K such that $P_K(K_i) = \emptyset$ for all $i \in \{1, \ldots, n\}$. Or equivalently, let the sub-DAG that corresponds to the set K consist of n separate, disconnected sub-DAGs, each of which has K_i as its set of nodes, with $i \in \{1, \ldots, n\}$. Then $P(K_i) \subseteq P(K)$ for all $i \in \{1, \ldots, n\}$ and, for all $x_{P(K)} \in \mathcal{X}_{P(K)}$,*

$$\mathcal{D}^{\text{irr}}_{K \rfloor x_{P(K)}} = \bigotimes_{i=1}^{n} \mathcal{D}^{\text{irr}}_{K_i \rfloor x_{P(K_i)}} = \operatorname{posi}\left(\left\{\mathbb{I}_{\{z_{K \setminus K_i}\}} f : i \in \{1, \ldots, n\},\ z_{K \setminus K_i} \in \mathcal{X}_{K \setminus K_i},\ f \in \mathcal{D}^{\text{irr}}_{K_i \rfloor x_{P(K_i)}}\right\}\right),$$

where $\mathcal{D}^{\text{irr}}_{K \rfloor x_{P(K)}}$ and $\mathcal{D}^{\text{irr}}_{K_i \rfloor x_{P(K_i)}}$, with $i \in \{1, \ldots, n\}$, are given by Eq. (9).

Fig. 5 illustrates this result with an example.

7. AD-separation and its consequences for the irrelevant natural extension

In Bayesian networks, there exists a simple criterion, called d-separation, that is capable of detecting independencies in the joint model, based only on the graphical structure of the underlying DAG [26]. Due to their close connection with Bayesian networks, credal networks under strong independence inherit this property almost trivially [4]: every d-separation

[7] We would like to thank one of the referees of a previous conference version of this paper [10] for suggesting this result.

[8] Choose $n = |G|$ and let G_i be singletons, each of which contains a different $s \in G$.

in the DAG corresponds to a (strong) independence in the joint model. For credal networks under epistemic independence, no such result exists. We do know that for general credal networks under epistemic independence, d-separation does not imply epistemic independence [5]. However, it is considered an important open problem [4,5,12,14] whether or not this holds for the most conservative joint model, also referred to as the independent natural extension[9] of a credal network [5]; see Ref. [12] for some partial but promising results for Markov chains. For the irrelevant natural extension of a credal network, which is the subject of the current paper, d-separation does not imply epistemic irrelevance; see Ref. [14, Section 7] for a counterexample. However, as we will show, it is possible to derive very similar results by employing an asymmetrical version of d-separation, which we call AD-separation. As we will see, AD-separation satisfies all graphoid properties except symmetry. Furthermore, and perhaps most importantly: we show that in the irrelevant natural extension of a credal network, AD-separation implies epistemic irrelevance, thereby establishing an asymmetric version of the classical d-separation result. We should note that our results are inspired by the work of Moral [23], who developed similar results in a much more restricted context; we comment on the restrictions he imposes further on in Section 7.2.

7.1. AD-separation

In probabilistic graphical networks that are defined by means of a symmetrical independence concept, the notion of d-separation is a very powerful tool [26]. However, for asymmetrical independence concepts such as epistemic irrelevance, there seems to be no convincing reason for using a symmetrical separation criterion such as d-separation. If learning Y is irrelevant to X, must it follow that learning X is irrelevant to Y? We agree with Dawid [8] that such a requirement is not obvious. Hence, we prefer to consider a modified version of d-separation that does not require symmetry. Moral [23] speaks of *asymmetrical D-separation*[10] (AD-separation) and Vantaggi [33–36] has introduced the very similar *L-separation* criterion. Here, we do not use one of these existing concepts, but choose to introduce a slightly modified version, which we will call *AD-separation* (asymmetrical d-separation) as well.[11] We prefer our version because our definition is weaker than—in the sense that it is implied by—Moral's AD-separation, slightly more general[12] than Vantaggi's L-separation and yet, it has stronger properties than both of these other concepts.

Consider any path $s_1, \dots, s_n$ in G, with $n \geq 1$. We say that this path is *blocked* by a set of nodes $C \subseteq G$ whenever at least one of the following four conditions holds:

B1. $s_1 \in C$;
B2. there is some $1 < i < n$ such that $s_i \to s_{i+1}$ and $s_i \in C$;
B3. there is some $1 < i < n$ such that $s_{i-1} \to s_i \leftarrow s_{i+1}$, $s_i \notin C$ and $D(s_i) \cap C = \emptyset$;
B4. $s_n \in C$.

In Moral's version of AD-separation, the notion of a blocked path is very similar. The only difference is condition B1, which he strengthens by requiring that $s_1 \to s_2$. Clearly, our condition is implied by Moral's. Vantaggi uses the same notion of blocked path as we do,[13] but leaves out conditions B1 and B4. They are redundant in her case, because she does not need to consider cases in which s_1 or s_n are elements of C.[14]

Example 6. Fig. 6 illustrates how each of the blocking conditions B1–B4 can block a path. The examples for B1 and B4 are straightforward. Note that in the example of B2, the crucial point is the arrow between s_3 and s_5. If that arrow were reversed, the path would no longer be blocked. In the example of B3, it is essential that s_5, s_6 and s_7 are not elements of C. If any of them were, the path would not be blocked. Notice also that the path in the example for B1 is not blocked according to Moral's version of AD-separation, the reason being that the arrow between s_3 and s_2 is pointing in the wrong direction.

Now consider (not necessarily disjoint) subsets I, O and C of G. We say that O is *AD-separated* from I by C, denoted as $\mathrm{AD}(I, O|C)$, if every path $i = s_1, \dots, s_n = o$, $n \geq 1$, from a node $i \in I$ to a node $o \in O$, is blocked by C; see Fig. 7 for an example of AD-separated sets.

[9] Not to be confused with the independent natural extension of a number of separate, unconditional models, which was the subject of our Section 6. However, there is no conflict in terminology since the latter corresponds to a special case.

[10] Judging by the references he provides [25,26], Moral actually seems to mean asymmetrical d-separation rather than asymmetrical D-separation; D-separation is an enhanced version of d-separation that allows for deterministic nodes [18]. However, since d-separation is a special case of D-separation, the term asymmetrical D-separation (AD-separation) does not produce a conflict in terminology and we choose to adopt it as well.

[11] We prefer Moral's terminology over the one by Vantaggi because (i) we think that the term asymmetrical d-separation really captures the meaning of the concept and (ii) the L in L-separation refers to the logical constraints that can be imposed in Vantaggi's framework, which do not seem relevant for our current purposes. It is however important to keep in mind that our notion of AD-separation is different than the one by Moral. We consider the resulting conflict in terminology to be minor, especially since our version of AD-separation is implied by Moral's version.

[12] At least as far as the sets on which it can defined is concerned: L-separation is only defined for disjoint sets. We should however mention that, if one restricts himself to disjoint sets, Vantaggi's L-separation criterion is more general than ours because it also includes the possibility to include logical constraints, which our notion of AD-separation does not.

[13] At first sight, it might seem as if she does not; loosely speaking, the confusion arises because she applies her definition to the reversed path.

[14] Because L-separation is defined for disjoint sets only; see the definition of AD-separation further on.

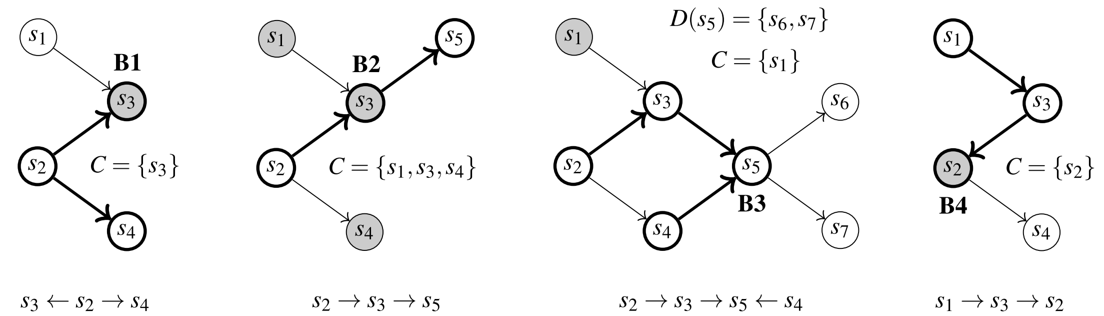

Fig. 6. Illustration of paths that are blocked by conditions B1–B4.

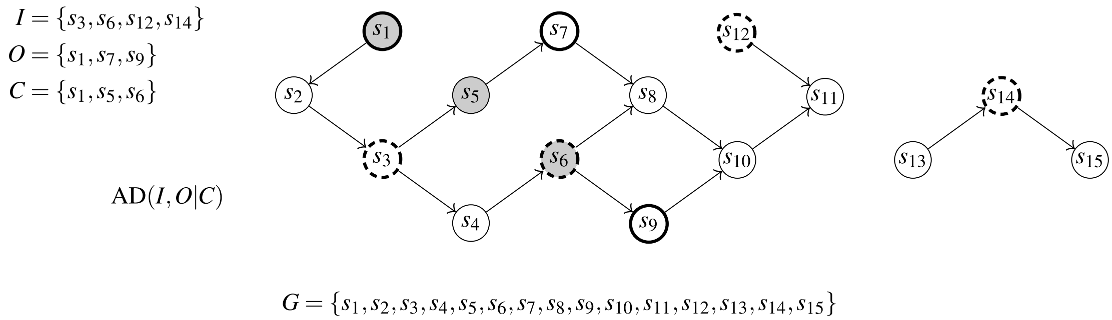

Fig. 7. Illustration of AD-separation.

Moral and Vantaggi define their separation criteria in much the same way. The only difference with Moral's version of AD-separation is his notion of a blocked path, as explained earlier. Clearly, AD-separation in Moral's sense implies AD-separation in our sense. The difference with Vantaggi's criterium is that L-separation is defined for disjoint sets only. Notice that if we restrict ourselves to disjoint sets, AD-separation (both our version and the one by Moral) is identical to L-separation.[15]

It turns out that our version of AD-separation satisfies all graphoid properties except symmetry.

Theorem 15. *For any subsets I, O, S and C of G, the following properties hold*[16]*:*

Direct redundancy: $\mathrm{AD}(I, O|I)$
Reverse redundancy: $\mathrm{AD}(I, O|O)$
Direct decomposition: $\mathrm{AD}(I, O \cup S|C) \Rightarrow \mathrm{AD}(I, O|C)$
Reverse decomposition: $\mathrm{AD}(I \cup S, O|C) \Rightarrow \mathrm{AD}(I, O|C)$
Direct weak union: $\mathrm{AD}(I, O \cup S|C) \Rightarrow \mathrm{AD}(I, O|C \cup S)$
Reverse weak union: $\mathrm{AD}(I \cup S, O|C) \Rightarrow \mathrm{AD}(I, O|C \cup S)$
Direct contraction: $\mathrm{AD}(I, O|C)$ & $\mathrm{AD}(I, S|C \cup O) \Rightarrow \mathrm{AD}(I, O \cup S|C)$
Reverse contraction: $\mathrm{AD}(I, O|C)$ & $\mathrm{AD}(S, O|C \cup I) \Rightarrow \mathrm{AD}(I \cup S, O|C)$
Direct intersection: *if $O \cap S = \emptyset$, then* $\mathrm{AD}(I, O|C \cup S)$ & $\mathrm{AD}(I, S|C \cup O) \Rightarrow \mathrm{AD}(I, O \cup S|C)$
Reverse intersection: *if $I \cap S = \emptyset$, then* $\mathrm{AD}(I, O|C \cup S)$ & $\mathrm{AD}(S, O|C \cup I) \Rightarrow \mathrm{AD}(I \cup S, O|C)$

This result (and our proof for it) is very similar to, and heavily inspired by, the work of Vantaggi [34, Theorem 7.1].[17] The main difference is that Vantaggi does not include the two redundancy properties, since L-separation is defined only for

[15] Making abstraction of the logical component of L-separation.

[16] We follow Refs. [7,23] in naming these properties. Vantaggi [34] uses a different terminology: for example, her notion of reverse decomposition refers to a property denoted as $(I \cup S, O|C)^l_G \Rightarrow (I, O|C)^l_G$, which seems similar to our notion of reverse decomposition, but actually, corresponds to what we call direct decomposition, since, loosely speaking, Vantaggi reverses the order in which I and O occur in the notation. Care should therefore be taken in comparing results.

[17] We provide a direct proof for Theorem 15. However, as suggested to us by Barbara Vantaggi, our result can probably be derived as a corollary of Ref. [34, Theorem 7.1] as well. A possible way of doing so could be to first prove (direct and reverse) redundancy and decomposition (which is trivial) and to use these properties to try and infer (direct and reverse) weak union, contraction and intersection from their 'disjoint' versions (which are proven in Ref. [34, Theorem 7.1]).

disjoint subsets I, O and C of G. Moral's version of AD-separation does not require I, O and C to be disjoint, but it does not satisfy direct redundancy, and proofs for a number of other properties are not given [23, Theorem 4].

7.2. Separation properties of the irrelevant natural extension

The reason why we have introduced AD-separation, is because it can be used to state the following very general factorisation result, the proof of which relies heavily on Theorem 10.

Theorem 16. *Consider any $I, O, C \subseteq G$ such that $\mathrm{AD}(I, O|C)$. Then for all $x_C \in \mathcal{X}_C$, $g \in \mathcal{G}(\mathcal{X}_I)_{>0}$ and $f \in \mathcal{G}(\mathcal{X}_O)$*

$$g\mathbb{I}_{\{x_C\}} f \in \mathcal{D}_G^{\mathrm{irr}} \quad \Leftrightarrow \quad \mathbb{I}_{\{x_C\}} f \in \mathcal{D}_G^{\mathrm{irr}}.$$

By combining this with Theorem 15, we can prove a result that is very similar to the classical d-separation result: AD-separation implies epistemic irrelevance in the irrelevant natural extension of a credal network.

Corollary 17. *For any $I, O, C \subseteq G$ such that $\mathrm{AD}(I, O|C)$ we have that $\mathrm{SIR}(I, O|C)$ (and thus also $\mathrm{IR}(I, O|C)$): for all $x_C \in \mathcal{X}_C$ and non-empty $A_{I\setminus C} \subseteq \mathcal{X}_{I\setminus C}$ it holds that*

$$\mathrm{marg}_{O\setminus C}\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor\{x_C\} \times A_{I\setminus C}\big) = \mathrm{marg}_{O\setminus C}\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor x_C\big).$$

This family of subset-irrelevance statements satisfies all graphoid properties except symmetry: it satisfies redundancy, decomposition, weak union, contraction and intersection, both in their direct and reverse form.

We leave it to the reader to show that Theorem 16 is a generalisation of Theorem 10 and that Corollary 17 generalises the first part of Corollary 12. In other words: for any closed subset K of G, it holds that $\mathrm{AD}(N(K), K|P(K))$.

What is particularly nice about Corollary 17 is that it allows us to detect epistemic irrelevancies in the joint model in a purely graphical way, without resorting to numerical computations; all we have to do is to check for AD-separation. Note however that AD-separation is only a sufficient condition for epistemic irrelevance. An important—and so far open—question is therefore whether the epistemic irrelevancies that are detected by AD-separation are the only ones that can be detected based on the graphical structure of the network. In other words, to put it more technically: is AD-separation complete with respect to epistemic irrelevance? We conjecture that it is, but provide no proof. Another important question is whether or not AD-separation can be checked efficiently. We suspect that this is indeed the case and that polynomial time solutions can be obtained by suitably adapting existing algorithms for d-separation [19,20]. However, this too, we leave as a possible topic for future research.

7.3. A crucial difference with earlier work by Moral

Readers who are familiar with the work in Ref. [23] might have noticed the similarity between Ref. [23, Theorem 5] and the first part of Corollary 17. The main difference between our approach and Moral's approach [23], besides the fact that we use a slightly different separation criterion, is that he enforces a more stringent version of epistemic irrelevance than we do. He calls X_I epistemically irrelevant to X_O if and only if the model $\mathcal{D}_{I\cup O}$ for the variable $X_{I\cup O}$ is the unique smallest set that marginalises to the marginal models $\mathcal{D}_I$ and $\mathcal{D}_O$ and for which X_I is irrelevant to X_O in our sense. He refers to our concept of irrelevance as 'weak' epistemic irrelevance. Consequently, *Moral's results in Ref. [23] are not applicable to all directed acyclic networks*. As a simple example: his concept of irrelevance does not allow for two variables to be mutually irrelevant, except in some degenerate uninformative cases. Therefore, his results cannot be applied to a network consisting of two unconnected nodes. More generally speaking, it seems to us his results can only be applied to networks in which every pair of nodes can be connected by means of a directed path.

7.4. Further comments and some clarification

As far as the second part of Corollary 17 is concerned, some clarification is perhaps in order. We do not claim that epistemic irrelevance satisfies the graphoid axioms that are stated in Theorem 15. As was proven in Ref. [7], epistemic irrelevance can violate direct contraction and both direct and reverse intersection. In fact, we believe that this negative result might even be one of the main reasons why a result such as Corollary 17 has thus far not appeared in any literature.

Indeed, in Bayesian networks, proving the counterpart to Corollary 17—with AD-separation replaced by d-separation and epistemic irrelevance replaced by stochastic independence—is usually done by using the fact that stochastic independence satisfies the graphoid axioms [26]. By applying these axioms to the independence assessments that are used to define a Bayesian network, one can infer new independencies, namely those that correspond to d-separations in the DAG of that network.

If one tries to mimic this approach in our context, then since epistemic irrelevance can fail some of the graphoid axioms, one might suspect that Corollary 17 cannot be proven. However, it is not necessary to use the axioms: our proof for

Theorem 16—of which the first part of Corollary 17 is a straightforward consequence—uses only Theorem 10 and a number of properties of AD-separation. At no point does it invoke graphoid properties of epistemic irrelevance. The second part of Corollary 17 is then but a mere consequence of the first part and Theorem 15. It states that the family of irrelevance statements that are proven to hold in the first part, are closed under the graphoid properties in Theorem 15.

So in order to conclude this section: epistemic irrelevance can fail a number of graphoid axioms, which implies that the irrelevance statements that are proven in Corollary 17 do not necessarily hold for every joint model $\mathcal{D}_G$ that satisfies requirements G1–G3. However for the unique one that also satisfies G4, being the irrelevant natural extension $\mathcal{D}_G^{\mathrm{irr}}$ of the network, this family of irrelevance statements does hold, the reason being that for this specific model, one can provide a direct proof that does not invoke any graphoid axioms of epistemic irrelevance.

8. Summary and conclusions

This paper has developed the notion of a credal network under epistemic irrelevance within the framework of sets of desirable gambles. By combining local sets of desirable gambles with assessments of epistemic irrelevance, and by doing so in the most conservative way possible, we have constructed an intuitive expression for the irrelevant natural extension of a credal network. We then went on to establish a number of theoretical properties of this irrelevant natural extension. It satisfies a powerful factorisation property, marginalises in an intuitive way and has tight connections with the independent natural extension. Furthermore, the irrelevant natural extension satisfies a result that is very similar to the classical d-separation result in Bayesian networks. We have introduced the notion of AD-separation, an asymmetrical adaptation of d-separation, have shown that it satisfies all graphoid properties except symmetry and, most importantly, that it implies epistemic irrelevance.

As far as future work is concerned, the most immediate task seems to be translating our results to the framework of coherent lower previsions. Although the expressiveness of sets of desirable gambles renders them extremely useful from a theoretical point of view, and as such allowed us to develop the results in this paper, representing them in a computer in order to manipulate them using algorithms quickly becomes overly complicated; see for example Refs. [3,29]. Since part of the expressiveness of sets of desirable gambles is not relevant as far as probabilistic inference is concerned [42, Example 10], it seems preferable to use less expressive frameworks for the development of algorithms. Inspired by the recent linear time algorithm for credal trees [14], coherent lower previsions seem to provide the right balance between expressiveness and algorithmic power.

We are confident that it is indeed possible to translate our results to the framework of coherent lower previsions. In fact, we have concrete ideas on how to do so; see Refs. [40,42] for some essential theoretical results, establishing an elegant connection between the theories of sets of desirable gambles and coherent lower previsions in a very general setup. For the particular case of credal networks under epistemic irrelevance, some preliminary results in terms of coherent lower previsions (and credal sets) can already be found in Refs. [10,11]. Using these techniques, it is possible to express our more fundamental results—such as those concerned with marginalisation, independent natural extension and AD-separation—directly in terms of coherent lower previsions. We intend to publish this in the near future.

Another important avenue for future research would be to try and establish similar results for credal networks under epistemic independence [4,12]. Do these networks satisfy marginalisation properties such as the one presented in Section 5? Do they exhibit the connection with the independent natural extension that was discussed in Section 6? And perhaps the most important open question related to credal networks under epistemic independence: do they satisfy the same separation properties as Bayesian networks? See the introduction of Section 7 for some comments on this last question, including relevant references.

Our ultimate goal, which served as a motivation to develop the present results, is to develop an efficient inference algorithm for credal networks under epistemic irrelevance whose underlying graphical structure is not necessarily a tree, extending the developments in Ref. [14]. We would like to point out that some of the properties in this paper already provide a direct tool to do so. Similar to what is usually done in Bayesian networks, our marginalisation properties and the separation properties that are induced by AD-separation can be used to reduce the size of the network in which the inference problem at hand needs to be solved; see Ref. [12, Section 5] as well. If after such a preprocessing step, the graphical structure of the network is reduced to a tree, the algorithm in Ref. [14] can be applied. On top of these preprocessing steps, we think that further algorithmic developments could also benefit greatly from our results in Section 6 on the connection with the independent natural extension, especially since the strong factorisation [15] property of the independent natural extension has been an essential tool in the development of the aforementioned algorithm for credal trees under epistemic irrelevance [14].

Acknowledgements

Jasper De Bock is a Ph.D. Fellow of the Fund for Scientific Research – Flanders (FWO) and wishes to acknowledge the financial support of the FWO. Gert de Cooman was supported by FWO-project 3G012512. The authors also wish to thank the anonymous referees of this paper and a previous conference version [10] for their constructive feedback.

Appendix A. Proofs of our results

This appendix provides proofs for the results in this paper. It also contains some additional results that are needed in the proofs, a short introduction to maximal sets of desirable gambles (included just before the proof of Proposition 6) and two definitions that are used in some of the proofs related to AD-separation (included just before the proof of Proposition 23).

Proof of Proposition 1. Consider any $s \in G$ and $z_{PN(s)} \in \mathcal{X}_{PN(s)}$. As a consequence of requirements G1 and G2, we see that $\mathrm{marg}_s(\mathcal{D}_G \rfloor z_{PN(s)})$ should be a superset of the given local model $\mathcal{D}_{s \rfloor z_{P(s)}}$. If we now choose $f \in \mathcal{D}_{s \rfloor z_{P(s)}}$ and apply Eq. (1), it follows immediately that $\mathbb{I}_{\{z_{PN(s)}\}} f$ is an element of $\mathcal{D}_G$. □

Proof of Corollary 2. We know from Proposition 1 that $\mathcal{A}_G^{\mathrm{irr}}$ is a subset of any joint model that satisfies requirements G1 and G2: $\mathcal{A}_G^{\mathrm{irr}} \subseteq \mathcal{D}_G$. Applying the posi operator to both sides, we obtain that $\mathrm{posi}(\mathcal{A}_G^{\mathrm{irr}}) \subseteq \mathrm{posi}(\mathcal{D}_G)$. If in addition to satisfying requirements G1 and G2, $\mathcal{D}_G$ is also coherent (satisfies requirement G3), and thus in particular is a convex cone (satisfies properties D3 and D4), then $\mathrm{posi}(\mathcal{D}_G) = \mathcal{D}_G$ and we get that $\mathrm{posi}(\mathcal{A}_G^{\mathrm{irr}}) \subseteq \mathcal{D}_G$. □

Proof of Proposition 3. Since $\mathcal{D}_G^{\mathrm{irr}} := \mathrm{posi}(\mathcal{A}_G^{\mathrm{irr}})$, the if part of this proof is trivial. For the only if part, fix any $f \in \mathcal{D}_G^{\mathrm{irr}}$. We know by definition that

$$f = \sum_{s \in G} \sum_{z_{PN(s)} \in \mathcal{X}_{PN(s)}} \sum_{i \in I(s, z_{PN(s)})} \lambda_{s, z_{PN(s)}, i} \mathbb{I}_{\{z_{PN(s)}\}} f_{s, z_{PN(s)}, i},$$

where, for all $s \in G$ and $z_{PN(s)} \in \mathcal{X}_{PN(s)}$, $I(s, z_{PN(s)})$ is a (possibly empty) finite index set (but with at least one of them non-empty) and for all $s \in G$, $z_{PN(s)} \in \mathcal{X}_{PN(s)}$ and $i \in I(s, z_{PN(s)})$, $\lambda_{s, z_{PN(s)}, i}$ is a strictly positive real number and $f_{s, z_{PN(s)}, i}$ is an element of $\mathcal{D}_{s \rfloor z_{P(s)}}$.

We now construct, for all $s \in G$ and $z_{PN(s)} \in \mathcal{X}_{PN(s)}$, a gamble $f_{s, z_{PN(s)}} \in \mathcal{G}(\mathcal{X}_s)$. If $I(s, z_{PN(s)}) = \emptyset$, we let $f_{s, z_{PN(s)}} = 0$. If $I(s, z_{PN(s)}) \neq \emptyset$, we let

$$f_{s, z_{PN(s)}} = \sum_{i \in I(s, z_{PN(s)})} \lambda_{s, z_{PN(s)}, i} f_{s, z_{PN(s)}, i},$$

which is an element of $\mathcal{D}_{s \rfloor z_{P(s)}}$ (and thus different from zero) due to the fact that $f_{s, z_{PN(s)}, i} \in \mathcal{D}_{s \rfloor z_{P(s)}}$ for all $i \in I(s, z_{PN(s)})$ and because $\mathcal{D}_{s \rfloor z_{P(s)}}$ is assumed to be coherent. It should now be clear that

$$f = \sum_{s \in G} \sum_{z_{PN(s)} \in \mathcal{X}_{PN(s)}} \mathbb{I}_{\{z_{PN(s)}\}} f_{s, z_{PN(s)}},$$

in which the gambles $f_{s, z_{PN(s)}}$ are elements of $\mathcal{D}_{s \rfloor z_{P(s)}} \cup \{0\}$ and at least one of them is non-zero. □

Proof of Proposition 4. The crucial step in the proof consists in showing that for any $z_G \in \mathcal{X}_G$, the indicator $\mathbb{I}_{\{z_G\}}$ is an element of $\mathcal{A}_G^{\mathrm{irr}}$. To prove this, pick an arbitrary leaf $s \in G$. This is possible because a DAG with a finite amount of nodes always has at least one leaf. Since s is a leaf, it has no descendants and we therefore have that $G = \{s\} \cup PN(s)$. Due to the coherence of the local models, and in particular property D2, we know that the indicator $\mathbb{I}_{\{z_s\}}$ is an element of $\mathcal{D}_{s \rfloor z_{P(s)}}$. We can therefore apply Eq. (3) to see that $\mathbb{I}_{\{z_s\}} \mathbb{I}_{\{z_{PN(s)}\}} = \mathbb{I}_{\{z_{\{s\} \cup PN(s)}\}} = \mathbb{I}_{\{z_G\}}$ is an element of $\mathcal{A}_G^{\mathrm{irr}}$. Since every $f \in \mathcal{G}(\mathcal{X}_G)_{>0}$ is a finite strictly positive linear combination of indicators $\mathbb{I}_{\{z_G\}}$ that were constructed above, it follows that $\mathcal{G}(\mathcal{X}_G)_{>0}$ is indeed a subset of $\mathrm{posi}(\mathcal{A}_G^{\mathrm{irr}})$.

To prove the second part, notice that any gamble in $\mathcal{E}(\mathcal{A}_G^{\mathrm{irr}})$ is a finite, strictly positive linear combination of gambles in $\mathcal{A}_G^{\mathrm{irr}}$ and gambles in $\mathcal{G}(\mathcal{X}_G)_{>0}$. However, since we have just shown that gambles in $\mathcal{G}(\mathcal{X}_G)_{>0}$ are themselves finite strictly positive linear combinations of specific indicators in $\mathcal{A}_G^{\mathrm{irr}}$, this implies that $\mathcal{E}(\mathcal{A}_G^{\mathrm{irr}}) \subseteq \mathrm{posi}(\mathcal{A}_G^{\mathrm{irr}})$. The converse inclusion is trivial and we thus find that $\mathcal{E}(\mathcal{A}_G^{\mathrm{irr}}) = \mathrm{posi}(\mathcal{A}_G^{\mathrm{irr}}) =: \mathcal{D}_G^{\mathrm{irr}}$. □

Our proof for Proposition 5 and Theorem 10 uses the following convenient version of the separating hyperplane theorem. It is proved in Ref. [16, Lemma 2] and repeated here to make the paper more self-contained.

Lemma 18. *Consider any finite subset $\mathcal{A}$ of $\mathcal{G}(\mathcal{X})$. Then $0 \notin \mathcal{E}(\mathcal{A})$ if and only if there is some probability mass function p such that $\sum_{x \in \mathcal{X}} p(x) f(x) > 0$ for all $f \in \mathcal{A}$ and $p(x) > 0$ for all $x \in \mathcal{X}$.*

Proof of Proposition 5. Since, by Proposition 4, $\mathcal{D}_G^{\mathrm{irr}} = \mathcal{E}(\mathcal{A})$, we know that $\mathcal{D}_G^{\mathrm{irr}}$ is coherent if and only if it satisfies D1, which states that any gamble $f \in \mathcal{G}(\mathcal{X}_G)$ for which $f \leq 0$ cannot be an element of $\mathcal{D}_G^{\mathrm{irr}}$. So consider any $f \in \mathcal{D}_G^{\mathrm{irr}}$ and assume *ex absurdo* that $f \leq 0$. We will show that this leads to a contradiction.

Since f is an element of $\mathcal{D}_G^{\text{irr}}$, we know from Proposition 3 that

$$f = \sum_{s \in G} \sum_{z_{PN(s)} \in \mathcal{X}_{PN(s)}} \mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}},$$

where every $f_{s,z_{PN(s)}}$ is an element of $\mathcal{D}_{s\rfloor z_{P(s)}} \cup \{0\}$ and at least one of them is non-zero. We now construct, for every $s \in G$ and $z_{P(s)} \in \mathcal{X}_{P(s)}$, a finite subset of the local model $\mathcal{D}_{s\rfloor z_{P(s)}}$:

$$\mathcal{A}^f_{s\rfloor z_{P(s)}} := \{f_{s,x_{PN(s)}} \colon x_{PN(s)} \in \mathcal{X}_{PN(s)},\ x_{P(s)} = z_{P(s)} \text{ and } f_{s,x_{PN(s)}} \neq 0\}.$$

As a consequence of the coherence of $\mathcal{D}_{s\rfloor z_{P(s)}}$, we have that $0 \notin \mathcal{D}_{s\rfloor z_{P(s)}} = \operatorname{posi}(\mathcal{D}_{s\rfloor z_{P(s)}})$. This in turn implies that $0 \notin \operatorname{posi}(\mathcal{A}^f_{s\rfloor z_{P(s)}} \cup \mathcal{G}(\mathcal{X}_s)_{>0}) =: \mathcal{E}(\mathcal{A}^f_{s\rfloor z_{P(s)}})$, because both $\mathcal{A}^f_{s\rfloor z_{P(s)}}$ and $\mathcal{G}(\mathcal{X}_s)_{>0}$ are subsets of $\mathcal{D}_{s\rfloor z_{P(s)}}$, and we can therefore apply Lemma 18. This yields for every $s \in G$ and $z_{P(s)} \in \mathcal{X}_{P(s)}$ a mass function $p_s(\cdot|z_{P(s)})$ on $\mathcal{X}_s$ with expectation operator $E_s(\cdot|z_{P(s)})$ on $\mathcal{G}(\mathcal{X}_s)$ such that $p_s(z_s|z_{P(s)}) > 0$ for all $z_s \in \mathcal{X}_s$ and $E_s(g|z_{P(s)}) > 0$ for each $g \in \mathcal{A}^f_{s\rfloor z_{P(s)}}$.

The trick is now to create a Bayesian network that has the conditional mass functions $p_s(\cdot|z_{P(s)})$ as its local models and has the same graphical structure as our credal network under epistemic irrelevance. If we let E_G be the joint expectation operator for this Bayesian net, we find that

$$\begin{aligned} E_G(f) = \sum_{s \in G} \sum_{z_{PN(s)} \in \mathcal{X}_{PN(s)}} E_G(\mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}}) &= \sum_{s \in G} \sum_{z_{PN(s)} \in \mathcal{X}_{PN(s)}} p_G(z_{PN(s)}) E_s(f_{s,z_{PN(s)}}|z_{PN(s)}) \\ &= \sum_{s \in G} \sum_{z_{PN(s)} \in \mathcal{X}_{PN(s)}} p_G(z_{PN(s)}) E_s(f_{s,z_{PN(s)}}|z_{P(s)}), \end{aligned}$$

where p_G is the global mass function of the Bayesian net and where we have applied Bayes' rule and the conditional independencies encoded in the graph. Since all the local probabilities $p_s(\cdot|z_{P(s)})$ are strictly positive, this is also true for the global ones and we find that $p_G(z_{PN(s)}) > 0$. For the conditional expectations $E_s(f_{s,z_{PN(s)}}|z_{P(s)})$ there are two possibilities. Either $f_{s,z_{PN(s)}} = 0$, in which case $E_s(f_{s,z_{PN(s)}}|z_{P(s)}) = 0$, or $f_{s,z_{PN(s)}} \in \mathcal{A}^f_{s\rfloor z_{P(s)}}$, in which case $E_s(f_{s,z_{PN(s)}}|z_{P(s)}) > 0$. However, since at least one of the gambles $f_{s,z_{PN(s)}}$ in Eq. (5) has to be non-zero, it is not possible that $E_s(f_{s,z_{PN(s)}}|z_{P(s)}) = 0$ for all gambles $f_{s,z_{PN(s)}}$ and we can therefore conclude that $E_G(f) > 0$. If we now apply our assumption that $f \leq 0$, we also obtain that $E_G(f) \leq 0$, a contradiction. $\square$

Since Theorem 10 generalises Proposition 6 without building upon it, it is not necessary to provide Proposition 6 with a separate proof. However, we feel that the complexity of the proof for Theorem 10 obscures the ease with which Proposition 6 can be proved. We therefore choose to provide Proposition 6 with a proof of its own. As it makes use of so-called maximal sets of desirable gambles, a concept that has not been introduced in the main text of this paper, we provide a short introduction here.

A coherent set $\mathcal{D}$ of desirable gambles on $\mathcal{X}$ is called *maximal* if it is not included in any other coherent set of desirable gambles on $\mathcal{X}$—in other words, if adding any gamble f to $\mathcal{D}$ makes sure we can no longer extend the resulting $\mathcal{D} \cup \{f\}$ to a coherent set. We generically denote maximal sets of desirable gambles as $\mathcal{M}$ instead of using the general notation $\mathcal{D}$.

Maximal sets of desirable gambles have a number of useful properties. For example, a coherent set $\mathcal{D}$ of desirable gambles on $\mathcal{X}$ is always the intersection of all the maximal coherent sets $\mathcal{M}$ of desirable gambles on $\mathcal{X}$ that include it; see Ref. [17]. In other words, $f \in \mathcal{D}$ if and only if $f \in \mathcal{M}$ for every $\mathcal{M} \supseteq \mathcal{D}$. As a consequence, we have the following separation property: if a gamble $f \in \mathcal{G}(\mathcal{X})$ is not an element of $\mathcal{D}$, there is at least one maximal set $\mathcal{M} \supseteq \mathcal{D}$ for which $f \notin \mathcal{M}$. Another useful property is that maximal sets of desirable gambles resolve points: for any maximal set $\mathcal{M}$ and non-zero gamble f in $\mathcal{G}(\mathcal{X})$, either f or $-f$ is an element of $\mathcal{M}$; see Ref. [3].

Proof of Proposition 6. Fix $s \in G$, $g \in \mathcal{G}(\mathcal{X}_{N(s)})_{>0}$, $f \in \mathcal{G}(\mathcal{X}_s)$ and $x_{P(s)} \in \mathcal{X}_{P(s)}$.

We begin by proving the converse implication: $f \in \mathcal{D}_{s\rfloor x_{P(s)}} \Rightarrow g\mathbb{I}_{\{x_{P(s)}\}} f \in \mathcal{D}_G^{\text{irr}}$. As explained both in Section 4.2 and the proof of Proposition 1, it holds for any $x_{N(s)} \in \mathcal{X}_{N(s)}$ that $\mathbb{I}_{\{x_{PN(s)}\}} f$ is an element of $\mathcal{D}_G^{\text{irr}}$. If we then realise that $g = \sum_{x_{N(s)} \in \mathcal{X}_{N(s)}} g(x_{N(s)}) \mathbb{I}_{\{x_{N(s)}\}}$, we get that

$$g\mathbb{I}_{\{x_{P(s)}\}} f = \sum_{x_{N(s)} \in \mathcal{X}_{N(s)}} g(x_{N(s)}) \mathbb{I}_{\{x_{PN(s)}\}} f$$

is a finite strictly positive linear combination of elements of $\mathcal{D}_G^{\text{irr}}$ and thus also an element of $\mathcal{D}_G^{\text{irr}}$, due to its coherence.

To prove the direct implication, we assume *ex absurdo* that $f \notin \mathcal{D}_{s\rfloor x_{P(s)}}$ and show that it implies $g\mathbb{I}_{\{x_{P(s)}\}} f \notin \mathcal{D}_G^{\text{irr}}$, a contradiction. The case $f = 0$ is trivial because $g\mathbb{I}_{\{x_{P(s)}\}} f$ is then equal to zero, which cannot be an element of $\mathcal{D}_G^{\text{irr}}$ due to its coherence; see Proposition 5. If $f \neq 0$, we start by applying some of the properties of maximal coherent sets of desirable gambles that were introduced in the text preceding this proof. Due to the first property, we can infer from $f \notin \mathcal{D}_{s\rfloor x_{P(s)}}$ that there is at least one maximal set of desirable gambles $\mathcal{M}^*_{s\rfloor x_{P(s)}} \supseteq \mathcal{D}_{s\rfloor x_{P(s)}}$ for which $f \notin \mathcal{M}^*_{s\rfloor x_{P(s)}}$. Due to the second property and the fact that $f \neq 0$, this in turn implies that $-f \in \mathcal{M}^*_{s\rfloor x_{P(s)}}$. We will now denote by $\mathcal{D}^{\mathrm{irr}*}_G$ the set that is obtained by Eq. (4) if we replace the local model $\mathcal{D}_{s\rfloor x_{P(s)}}$ by the specific maximal superset $\mathcal{M}^*_{s\rfloor x_{P(s)}}$. It should be clear that $\mathcal{D}^{\mathrm{irr}*}_G \supseteq \mathcal{D}^{\mathrm{irr}}_G$. Next, since $-f \in \mathcal{M}^*_{s\rfloor x_{P(s)}}$, it follows from a similar argument as the one used in the proof of the converse implication that was given above, that $g\mathbb{I}_{\{x_{P(s)}\}}(-f) \in \mathcal{D}^{\mathrm{irr}*}_G$. Hence, due to the coherence of $\mathcal{D}^{\mathrm{irr}*}_G$, $g\mathbb{I}_{\{x_{P(s)}\}} f \notin \mathcal{D}^{\mathrm{irr}*}_G$ and therefore, since $\mathcal{D}^{\mathrm{irr}*}_G \supseteq \mathcal{D}^{\mathrm{irr}}_G$, we find that $g\mathbb{I}_{\{x_{P(s)}\}} f \notin \mathcal{D}^{\mathrm{irr}}_G$. □

Proof of Corollary 7. Fix $s \in G$, $f \in \mathcal{G}(\mathcal{X}_s)$ and $x_{PN(s)} \in \mathcal{X}_{PN(s)}$. Since $\mathbb{I}_{\{x_{N(s)}\}}$ and 1 are both elements of $\mathcal{G}(\mathcal{X}_{N(s)})_{>0}$, we derive from Eq. (1) and Proposition 6 that

$$f \in \mathrm{marg}_s\big(\mathcal{D}^{\mathrm{irr}}_G \rfloor x_{PN(s)}\big) \Leftrightarrow \mathbb{I}_{\{x_{PN(s)}\}} f \in \mathcal{D}^{\mathrm{irr}}_G \Leftrightarrow f \in \mathcal{D}_{s\rfloor x_{P(s)}} \Leftrightarrow \mathbb{I}_{\{x_{P(s)}\}} f \in \mathcal{D}^{\mathrm{irr}}_G \Leftrightarrow f \in \mathrm{marg}_s\big(\mathcal{D}^{\mathrm{irr}}_G \rfloor x_{P(s)}\big). \qquad \square$$

Proof of Theorem 8. We know from Proposition 5 and Corollary 7 that $\mathcal{D}^{\mathrm{irr}}_G := \mathrm{posi}(\mathcal{A}^{\mathrm{irr}}_G)$ satisfies requirements G1–G3. Because of Corollary 2, it is also the smallest set of desirable gambles on $\mathcal{X}_G$ that does so and therefore, it is the unique set of desirable gambles on $\mathcal{X}_G$ that satisfies G1–G4. □

Proof of Corollary 9. Fix arbitrary $s \in G$, $f \in \mathcal{G}(\mathcal{X}_s)$, $x_{P(s)} \in \mathcal{X}_{P(s)}$ and any non-empty $A_{N(s)} \subseteq \mathcal{X}_{N(s)}$. Since $\mathbb{I}_{A_{N(s)}}$ and 1 are both elements of $\mathcal{G}(\mathcal{X}_{N(s)})_{>0}$, we derive from Eq. (1) and Proposition 6 that

$$\begin{aligned} f \in \mathrm{marg}_s\big(\mathcal{D}^{\mathrm{irr}}_G \rfloor \{x_{P(s)}\} \times A_{N(s)}\big) &\Leftrightarrow \mathbb{I}_{A_{N(s)}}\mathbb{I}_{\{x_{P(s)}\}} f \in \mathcal{D}^{\mathrm{irr}}_G \Leftrightarrow f \in \mathcal{D}_{s\rfloor x_{P(s)}} \\ &\Leftrightarrow \mathbb{I}_{\{x_{P(s)}\}} f \in \mathcal{D}^{\mathrm{irr}}_G \Leftrightarrow f \in \mathrm{marg}_s\big(\mathcal{D}^{\mathrm{irr}}_G \rfloor x_{P(s)}\big). \qquad \square \end{aligned}$$

Lemma 19. *Consider any $K \subseteq G$ and any $k \in K$. Then $P_K(k) = P(k) \cap K = P(k) \setminus P(K)$ and $P(k) \setminus P_K(k) = P(k) \cap P(K)$.*

Proof. We start by proving that $P_K(k) = P(k) \cap K$. An element $q \in P_K(k)$ is by definition a parent of k according to the sub-DAG that corresponds to K, therefore q is also a parent of k in the original DAG: $q \in P(k)$. Since q is an element of the sub-DAG, we have $q \in K$ and therefore $q \in P(k) \cap K$. Conversely, if $q \in P(k) \cap K$, then q is clearly a parent of k in the sub-DAG that corresponds to K and therefore $q \in P_K(k)$.

Next we show that $P(k) \cap K = P(k) \setminus P(K)$. By definition of $P(K)$ and since $k \in K$, we know that $q \in P(k)$ implies $q \in P(K) \cup K$ and we therefore have that $P(k) \subseteq P(K) \cup K$. Since $P(K)$ and K are disjoint by definition, we infer that $P(k) \cap K = P(k) \setminus P(K)$.

The final property is a direct consequence of the previous equality:

$$P(k) \setminus P_K(k) = P(k) \setminus \big(P(k) \setminus P(K)\big) = P(k) \cap P(K). \qquad \square$$

Lemma 20. *Consider any closed*[18] *$K \subseteq G$, $k \in K$, $s \in G$ and $t \in PN(s)$. Then the following statements hold:*

(i) *$P(K) \cap D(K) = \emptyset$;*
(ii) *$P(K)$, $N(K)$, K and $D(K)$ constitute a partition*[19] *of G;*
(iii) *$PN(K)$ and $D(K)$ are closed subsets of G;*
(iv) *$P(t) \subseteq PN(s)$;*
(v) *$P(PN(K)) = \emptyset$;*
(vi) *$PN(D(K)) = PN(K) \cup K$;*
(vii) *$N(K) = \bigcap_{q \in K} N(q)$;*
(viii) *$N(K) \subseteq N(k)$;*
(ix) *$P(K) \subseteq PN(k)$;*
(x) *$PN(K) \subseteq PN(k)$;*
(xi) *$P(K) \setminus P(k) \subseteq N(k)$.*

Proof. (i). Assume *ex absurdo* that $q \in P(K)$ and $q \in D(K)$. Then $q \in D(K)$ implies the existence of some $r_1 \in K$ such that $r_1 \sqsubseteq q$ and $q \in P(K)$ implies the existence of some $r_2 \in K$ such that $q \sqsubseteq r_2$. We find that $r_1 \sqsubseteq q \sqsubseteq r_2$, with $r_1, r_2 \in K$. Since K is closed, this implies that $q \in K$, contradicting both $q \in P(K)$ and $q \in D(K)$.

[18] Statements (vii) and (viii) are true for general—not necessarily closed—sets $K \subseteq G$ as well.
[19] We use the term 'partition' in a somewhat looser sense than is usual, as we do not exclude that some of its elements may be empty.

(ii). Direct consequence of (i) and the definition of $P(K)$, $D(K)$ and $N(K)$.

(iii). To prove that $PN(K)$ is closed, consider $q_1, q_2 \in PN(K)$ and $r \in G$ such that $q_1 \sqsubseteq r \sqsubseteq q_2$ and assume *ex absurdo* that $r \notin PN(K)$, implying, due to (ii), that $r \in K \cup D(K)$. This in turn implies that there is some $u \in K$ such that $u \sqsubseteq r$ and therefore $u \sqsubseteq q_2$, which implies that $q_2 \in K \cup D(K)$, contradicting $q_2 \in PN(K)$ due to (ii).

To prove that $D(K)$ is closed, consider $q_1, q_2 \in D(K)$ and $r \in G$ such that $q_1 \sqsubseteq r \sqsubseteq q_2$ and assume *ex absurdo* that $r \notin D(K)$. $q_1 \in D(K)$ implies that there is some $u \in K$ such that $u \sqsubseteq q_1$ and therefore $u \sqsubseteq r$, implying that $r \in K \cup D(K)$ and, since $r \notin D(K)$, that $r \in K$. We thus find that $u \sqsubseteq q_1 \sqsubseteq r$, with $u, r \in K$. Because K is closed, this tells us that $q_1 \in K$, contradicting $q_1 \in D(K)$.

(iv). Consider any $q \in P(t)$. By the definition of $PN(s)$, it suffices to show that $q \notin \{s\} \cup D(s)$. Assume *ex absurdo* that $s \sqsubseteq q$, then we derive from $q \sqsubset t$ (since $q \in P(t)$) that $s \sqsubset t$, meaning that $t \in D(s)$, contradicting $t \in PN(s)$.

(v). Assume *ex absurdo* that $q \in P(PN(K))$, so there is some $r \in PN(K)$ such that $q \in P(r)$. By definition of $P(PN(K))$, this implies that $q \notin PN(K)$, which in turn implies, due to (ii), that $q \in K \cup D(K)$. By definition of $D(K)$, this implies that there is some $u \in K$ such that $u \sqsubseteq q$. Since $q \sqsubset r$ (because $q \in P(r)$), we find that $u \sqsubset r$, implying that $r \in K \cup D(K)$. Due to (ii), this contradicts $r \in PN(K)$.

(vi). First notice that it suffices to show that $D(D(K)) = \emptyset$. Indeed, this implies $PN(D(K)) = G \setminus D(K) = PN(K) \cup K$ by applying (ii) once for the closed $D(K)$ and once for the closed K. So assume *ex absurdo* that $q \in D(D(K))$, implying that there is some $r \in D(K)$ such that $r \sqsubset q$. Since $r \in D(K)$ in turn implies that there is some $u \in K$ such that $u \sqsubset r$, we find that $u \sqsubset q$, implying that $q \in K \cup D(K)$. But $q \notin D(K)$ because we know that $q \in D(D(K))$, and therefore $q \in K$. Since $u, q \in K$ and $u \sqsubset r \sqsubset q$, we derive from K being closed that $r \in K$, contradicting $r \in D(K)$.

(vii). This follows at once from:

$$N(K) := G \setminus \big(P(K) \cup K \cup D(K)\big) = G \setminus \left(\bigcup_{q\in K} P(q) \cup K \cup \bigcup_{q\in K} D(q)\right)$$
$$= G \setminus \left(\bigcup_{q\in K} \big(P(q) \cup \{q\} \cup D(q)\big)\right) = \bigcap_{q\in K} \big(G \setminus \big(P(q) \cup \{q\} \cup D(q)\big)\big) = \bigcap_{q\in K} N(q).$$

(viii). Direct consequence of (vii).

(ix). Choose $q \in P(K)$ and assume, *ex absurdo*, that $q \notin PN(k)$. This implies that $q \in \{k\} \cup D(k)$, or equivalently $k \sqsubseteq q$, and therefore that $q \in K \cup D(K)$, contradicting $q \in P(K)$ because of (ii).

(x). Direct consequence of (viii) and (ix).

(xi). Choose $q \in P(K) \setminus P(k)$ and assume *ex absurdo* that $q \notin N(k)$, implying that $q \in P(k) \cup \{k\} \cup D(k)$ or, since $q \notin P(k)$, that $q \in \{k\} \cup D(k)$ and therefore $k \sqsubseteq q$. This in turn implies that $q \in K \cup D(K)$, contradicting $q \in P(K)$ because of (ii). □

Lemma 21. *Consider any closed $K \subseteq G$, and any $k \in K$, $s \in G \setminus K$ and $t \in PN(s) \cap K$. Then the following statements hold:*

(i) $D_K(k) = D(k) \cap K$;
(ii) $N_K(k) = N(k) \cap K$;
(iii) $PN_K(k) = PN(k) \cap K$;
(iv) $N(K)$, $P(K) \setminus P(k)$ *and* $N_K(k)$ *are disjoint subsets of* $N(k)$;
(v) $P_K(t) \subseteq PN(s) \cap K$;
(vi) $P(t) \setminus P_K(t) \subseteq PN(s) \cap P(K)$.

Proof. (i). An element $q \in D_K(k)$ is by definition a descendant of k according to the sub-DAG that corresponds to K, therefore q is also a descendant of k in the original DAG: $q \in D(k)$. Since q is an element of the sub-DAG, we have $q \in K$ and therefore $q \in D(k) \cap K$.

Conversely, if $q \in D(k) \cap K$, then $q \in D(k)$ implies the existence of a directed sequence of nodes $k = r_1, \dots, r_n = q$, $n \geq 1$, in G. Since $k \in K$ and $q \in K$, we can derive from K being closed that for all $i \in \{1, \dots, n\}$, $r_i \in K$, implying that in the sub-DAG that corresponds to K, q is also a descendant of k: $q \in D_K(k)$.

(ii). This follows at once from the definitions of $N_K(k)$ and $N(k)$:

$$N_K(k) = K \setminus \big(P_K(k) \cup \{k\} \cup D_K(k)\big) = K \setminus \big(\big(P(k) \cap K\big) \cup \{k\} \cup \big(D(k) \cap K\big)\big) = K \setminus \big(P(k) \cup \{k\} \cup D(k)\big) = N(k),$$

where the second equality is due to Lemma 19 and (i).

(iii). This follows at once from the definitions of $PN_K(k)$ and $PN(k)$:

$$PN_K(k) = P_K(k) \cup N_K(k) = \big(P(k) \cap K\big) \cup \big(N(k) \cap K\big) = \big(P(k) \cup N(k)\big) \cap K =: PN(k) \cap K,$$

where the second equality is due to Lemma 19 and (ii).

(iv). $N(K)$, $P(K)$ and K are disjoint subsets of G because of Lemma 20(ii). Since $P(K) \setminus P(k) \subseteq P(K)$ and $N_K(k) \subseteq K$, this implies that $N(K)$, $P(K) \setminus P(k)$ and $N_K(k)$ are disjoint as well. It only remains to show that $N(K)$, $P(K) \setminus P(k)$ and $N_K(k)$

are subsets of $N(k)$. For $N(K)$, this is due to Lemma 20(viii). For $P(K) \setminus P(k)$, this is due to Lemma 20(ix) and for $N_K(k)$, this follows from (ii).

(v). By definition, $P_K(t) \subseteq K$. To show that $P_K(t) \subseteq PN(s)$, use Lemma 19 to find that $P_K(t) \subseteq P(t)$, and use Lemma 20(iv) to infer that $P(t) \subseteq PN(s)$.

(vi). Due to Lemma 19, $P(t) \setminus P_K(t) \subseteq P(t) \cap P(K)$. Due to Lemma 20(iv), we have $P(t) \subseteq PN(s)$ and therefore $P(t) \cap P(K) \subseteq PN(s) \cap P(K)$. Combining them yields $P(t) \setminus P_K(t) \subseteq PN(s) \cap P(K)$. □

Lemma 22. *Consider any closed $K \subseteq G$ and $s \in PN(K)$ and let $P_1(K)$ and $P_2(K)$ be an arbitrary partition of $P(K)$.*[20] *Let $K_2 := K \cap D(P_2(K))$ and $K_1 := K \setminus K_2 = K \setminus D(P_2(K))$ and choose any $k_1 \in K_1$ and $k_2 \in K_2$. The following statements hold:*

(i) *K_2 is a closed subset of G;*
(ii) *$P(K_1) \subseteq P_1(K)$;*
(iii) *$P_2(K) \subseteq P(K_2)$;*
(iv) *$K_1 \subseteq PN(k_2)$;*
(v) *$P(k_1) \cap K = P(k_1) \cap K_1$;*
(vi) *$P(k_1) \cap P(K) = P(k_1) \cap P(K_1)$;*
(vii) *$PN(s) \cap K_1 = K \setminus D((P(K) \setminus PN(s)) \cup (PN(s) \cap P_2(K)))$.*

Proof. (i). Consider $q_1, q_2 \in K \cap D(P_2(K))$ and $r \in G$ such that $q_1 \sqsubseteq r \sqsubseteq q_2$. Since K is closed, we have that $r \in K$, and we are left to show that $r \in D(P_2(K))$. That $q_1 \in D(P_2(K))$ implies the existence of some $u \in P_2(K)$ such that $u \sqsubseteq q_1$ and therefore $u \sqsubseteq r$, implying that $r \in P_2(K) \cup D(P_2(K))$. Since $r \in K$, we know that $r \notin P(K)$ and therefore $r \notin P_2(K)$. We infer that indeed $r \in D(P_2(K))$.

(ii). Consider any $q \in P(K_1)$, implying the existence of some $r \in K_1$ such that $q \in P(r)$ and $q \notin K_1$. We are first going to show that $q \notin P_2(K) \cup K_2$. Assume *ex absurdo* that $q \in P_2(K) \cup K_2$, implying that $q \in P_2(K) \cup D(P_2(K))$, which means that there is some $u \in P_2(K)$ for which $u \sqsubseteq q$ and, since $q \in P(r)$, that $u \sqsubseteq r$. From this we infer that $r \in P_2(K) \cup D(P_2(K))$ and therefore that $r \in D(P_2(K))$, since $r \in K_1 \subseteq K$ implies that $r \notin P(K)$, which in turn implies that $r \notin P_2(K)$. We have thus found that $r \in K \cap D(P_2(K)) = K_2$, contradicting $r \in K_1$. Hence indeed $q \notin P_2(K) \cup K_2$, implying $q \notin P_2(K)$ and $q \notin K_2$. Since also $q \notin K_1$, we find that $q \notin K$, which implies that $q \in P(K)$, since $q \in P(r)$ with $r \in K_1 \subseteq K$. Since $P_1(K)$ and $P_2(K)$ form a partition of $P(K)$ and $q \notin P_2(K)$, we conclude that indeed $q \in P_1(K)$.

(iii). Consider any $q \in P_2(K) \subseteq P(K)$, implying the existence of some $r \in K$ such that $q \in P(r)$. From this we infer that $q \sqsubseteq r$ and therefore $r \in P_2(K) \cup D(P_2(K))$. Since $r \in K$, we see that $r \notin P(K)$ and therefore $r \notin P_2(K)$, whence $r \in D(P_2(K))$. Together with $r \in K$, this implies that $r \in K_2$. Since $q \in P(K)$ implies $q \notin K$ and therefore $q \notin K_2$, we can infer from $q \in P(r)$ that $q \in P(K_2)$.

(iv). Consider any $q \in K_1$. Assume *ex absurdo* that $q \notin PN(k_2)$, implying that $q \in \{k_2\} \cup D(k_2)$ and therefore that $k_2 \sqsubseteq q$. Since $k_2 \in K_2$, we infer that $k_2 \in D(P_2(K))$, implying the existence of some $r \in P_2(K)$ such that $r \sqsubseteq k_2$ and therefore $r \sqsubseteq q$, which in turn implies that $q \in P_2(K) \cup D(P_2(K))$. Since $q \in K_1 \subseteq K$, we have that $q \notin P(K)$ and therefore that $q \notin P_2(K)$. Hence $q \in D(P_2(K))$ and therefore also $q \in K_2$, since $q \in K$. This contradicts $q \in K_1$, since K_1 and K_2 form a partition of K.

(v). Since it trivially holds that $P(k_1) \cap K \supseteq P(k_1) \cap K_1$, we only need to prove that $P(k_1) \cap K \subseteq P(k_1) \cap K_1$. So consider any $q \in P(k_1) \cap K$. By definition of $P(K_1)$, we derive from $q \in P(k_1)$ that either $q \in P(K_1)$ or $q \in K_1$. Assume *ex absurdo* that $q \in P(K_1)$, then due to (ii), $q \in P_1(K)$. Since $q \in K$ implies $q \notin P(K)$ and therefore $q \notin P_1(K)$, we have a contradiction. We have thus found that $q \in K_1$ and, since $q \in P(k_1)$, that $q \in P(k_1) \cap K_1$.

(vi). Since due to (ii), $P(K_1) \subseteq P_1(K)$, we find that $P(K_1) \subseteq P(K)$ and therefore $P(k_1) \cap P(K) \supseteq P(k_1) \cap P(K_1)$. To prove that $P(k_1) \cap P(K) \subseteq P(k_1) \cap P(K_1)$, consider any $q \in P(k_1) \cap P(K)$. By definition of $P(K_1)$, we derive from $q \in P(k_1)$ that either $q \in P(K_1)$ or $q \in K_1$. Since $q \in P(K)$, we have that $q \notin K$ and therefore also that $q \notin K_1$. Hence $q \in P(K_1)$ and therefore, since $q \in P(k_1)$, also $q \in P(k_1) \cap P(K_1)$.

(vii). First, notice that by subtracting both sides of the expression from K, we obtain the equivalent statement

$$K_2 \cup \big(K_1 \setminus PN(s)\big) = K \cap D\big(P(K) \setminus PN(s)\big) \cup \big(PN(s) \cap P_2(K)\big).$$

Since we have that $(P(K) \setminus PN(s)) \cup (PN(s) \cap P_2(K)) = P_2(K) \cup (P(K) \setminus PN(s))$ and $K_2 \cup (K_1 \setminus PN(s)) = K_2 \cup (K \setminus PN(s))$, this is in turn equivalent to:

$$K_2 \cup \big(K \setminus PN(s)\big) = K \cap D\big(P_2(K) \cup \big(P(K) \setminus PN(s)\big)\big).$$

We will prove this statement instead of the original one.

We start by proving that $K_2 \cup (K \setminus PN(s)) \subseteq K \cap D(P_2(K) \cup (P(K) \setminus PN(s)))$. Consider any $q \in K_2 \cup (K \setminus PN(s))$. On the one hand, if $q \in K_2$, then $q \in D(P_2(K))$. Since $q \notin P_2(K) \cup (P(K) \setminus PN(s))$, because K and $P_2(K) \cup (P(K) \setminus PN(s)) \subseteq P(K)$ are disjoint, we can infer that indeed $q \in K \cap D(P_2(K) \cup (P(K) \setminus PN(s)))$. On the other hand, if $q \in K \setminus PN(s)$, then $q \notin PN(s)$ and therefore $q \in \{s\} \cup D(s)$. Since $s \in PN(K)$ and therefore due to Lemma 20(ii), $s \notin K$, we know from $q \in K$ that $q \neq s$. We

[20] Here too, we allow that one of the sets $P_1(K)$ or $P_2(K)$ may be empty.

can therefore infer that $q \in D(s)$, implying the existence of a directed path $s = r_1, \dots, r_n = q$, $n > 1$. Now let j be the first index in $\{1, \dots, n\}$ for which $r_j \in K$. Since $q \in K$, such an index exists, and since $s \notin K$, $j > 1$, and therefore we can consider the node r_{j-1}. Since $r_{j-1} \in P(r_j)$, $r_j \in K$ and $r_{j-1} \notin K$, we infer that $r_{j-1} \in P(K)$. Since $s \sqsubseteq r_{j-1}$, we have $r_{j-1} \in \{s\} \cup D(s)$ and therefore $r_{j-1} \notin PN(s)$, whence $r_{j-1} \in P(K) \setminus PN(s) \subseteq P_2(K) \cup (P(K) \setminus PN(s))$. Since $q \notin P_2(K) \cup (P(K) \setminus PN(s))$ because $q \in K$, and since $r_{j-1} \sqsubseteq q$, we obtain that $q \in K \cap D(P_2(K) \cup (P(K) \setminus PN(s)))$.

We now prove that $K_2 \cup (K \setminus PN(s)) \supseteq K \cap D(P_2(K) \cup (P(K) \setminus PN(s)))$. Consider any $q \in K \cap D(P_2(K) \cup (P(K) \setminus PN(s)))$. Then there is some $r \in P_2(K) \cup (P(K) \setminus PN(s))$ such that $r \sqsubseteq q$. On the one hand, if $r \in P_2(K)$, then $q \in D(P_2(K))$ because $q \in K$ and therefore $q \notin P(K)$, whence indeed $q \in K \cap D(P_2(K)) = K_2 \subseteq K_2 \cup (K \setminus PN(s))$. On the other hand, if $r \in P(K) \setminus PN(s)$, then since $r \notin PN(s)$, $r \in \{s\} \cup D(s)$ and therefore $s \sqsubseteq r$, implying that $s \sqsubseteq q$. We thus find that $q \in \{s\} \cup D(s)$, or equivalently, that $q \notin PN(s)$. Combined with $q \in K$, we obtain that $q \in K \setminus PN(s) \subseteq K_2 \cup (K \setminus PN(s))$. $\square$

Proof of Theorem 10. Fix any closed set $K \subseteq G$, $x_{P(K)} \in \mathcal{X}_{P(K)}$, $g \in \mathcal{G}(\mathcal{X}_{N(K)})_{>0}$ and $f \in \mathcal{G}(\mathcal{X}_K)$. Clearly, without loss of generality, we can assume f to be non-zero.

We first prove the converse implication: $g\mathbb{I}_{\{x_{P(K)}\}} f \in \mathcal{D}^{\mathrm{irr}}_G \Leftarrow f \in \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$.

Due to the coherence of $\mathcal{D}^{\mathrm{irr}}_G$ and the definition of $\mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ and $\mathcal{G}(\mathcal{X}_{N(K)})_{>0}$, we can assume, without loss of generality, that $f \in \mathcal{A}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ and $g = \mathbb{I}_{\{z_{N(K)}\}}$, with $z_{N(K)} \in \mathcal{X}_{N(K)}$. Since $f \in \mathcal{A}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$, $f = \mathbb{I}_{\{z_{PN_K(s)}\}} f'$ for some $s \in K$, $f' \in \mathcal{D}_{s\rfloor z_{P(s)}}$ and $z_{PN_K(s)} \in \mathcal{X}_{PN_K(s)}$ with $z_{P(s)\cap P(K)} = x_{P(s)\cap P(K)}$. We need to prove that $\mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)}\}}\mathbb{I}_{\{z_{PN_K(s)}\}} f' \in \mathcal{D}^{\mathrm{irr}}_G$. Due to Lemma 19,

$$\mathbb{I}_{\{z_{P(s)}\}} = \mathbb{I}_{\{z_{P(K)\cap P(s)}\}}\mathbb{I}_{\{z_{P_K(s)}\}} = \mathbb{I}_{\{x_{P(K)\cap P(s)}\}}\mathbb{I}_{\{z_{P_K(s)}\}},$$

and therefore

$$\begin{aligned}
\mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)}\}}\mathbb{I}_{\{z_{PN_K(s)}\}} f' = \mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)}\}}\mathbb{I}_{\{z_{P_K(s)}\}}\mathbb{I}_{\{z_{N_K(s)}\}} f' &= \mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)\setminus P(s)}\}}\mathbb{I}_{\{x_{P(K)\cap P(s)}\}}\mathbb{I}_{\{z_{P_K(s)}\}}\mathbb{I}_{\{z_{N_K(s)}\}} f' \\
&= \mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)\setminus P(s)}\}}\mathbb{I}_{\{z_{P(s)}\}}\mathbb{I}_{\{z_{N_K(s)}\}} f' \\
&= \mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)\setminus P(s)}\}}\mathbb{I}_{\{z_{N_K(s)}\}}\mathbb{I}_{\{z_{P(s)}\}} f' = g'\mathbb{I}_{\{z_{P(s)}\}} f',
\end{aligned}$$

in which $g' := \mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)\setminus P(s)}\}}\mathbb{I}_{\{z_{N_K(s)}\}}$. So we are left to prove that $g'\mathbb{I}_{\{z_{P(s)}\}} f' \in \mathcal{D}^{\mathrm{irr}}_G$. We have already explained, both in Section 4.2 and the proof of Proposition 1, that since $f' \in \mathcal{D}_{s\rfloor z_{P(s)}}$, $\mathbb{I}_{\{y_{N(s)}\}}\mathbb{I}_{\{z_{P(s)}\}} f' \in \mathcal{D}^{\mathrm{irr}}_G$ for any $y_{N(s)} \in \mathcal{X}_{N(s)}$. Therefore, the proof follows from the coherence of $\mathcal{D}^{\mathrm{irr}}_G$, since due to $N(K)$, $P(K) \setminus P(s)$ and $N_K(s)$ being disjoint subsets of $N(s)$ by Lemma 21(iv), g' is a finite (and non-empty) sum of indicators $\mathbb{I}_{\{y_{N(s)}\}}$, $y_{N(s)} \in \mathcal{X}_{N(s)}$.

We now turn to the proof of the direct implication: $g\mathbb{I}_{\{x_{P(K)}\}} f \in \mathcal{D}^{\mathrm{irr}}_G \Rightarrow f \in \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$.

The proof is rather involved and uses ideas similar to the proof of Proposition 5. First of all, by assumption, $f'' := g\mathbb{I}_{\{x_{P(K)}\}} f \in \mathcal{D}^{\mathrm{irr}}_G$ and therefore also, due to Proposition 3

$$f'' = \sum_{s\in G} \sum_{z_{PN(s)}\in\mathcal{X}_{PN(s)}} \mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}}, \tag{A.1}$$

where every $f_{s,z_{PN(s)}}$ is an element of $\mathcal{D}_{s\rfloor z_{P(s)}} \cup \{0\}$ and at least one of them is non-zero.

Now assume *ex absurdo* that $f \notin \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$. We will show that this allows us to construct a probability mass function p_G on $\mathcal{X}_G$ such that the corresponding expectation operator E_G on $\mathcal{G}(\mathcal{X}_G)$ yields both $E_G(f'') > 0$ and $E_G(f'') < 0$. Since this is a contradiction, we find that $f \in \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$, which concludes the proof.

As shown in the proof of Proposition 5, it is possible to find, for all $s \in G$ and all $z_{P(s)} \in \mathcal{X}_{P(s)}$, a local mass function $p_s(\cdot|z_{P(s)})$ on $\mathcal{X}_s$ with expectation operator $E_s(\cdot|z_{P(s)})$ on $\mathcal{G}(\mathcal{X}_s)$, such that $p_s(z_s|z_{P(s)}) > 0$ for all $z_s \in \mathcal{X}_s$, and $E_s(f_{s,z_{PN(s)}}|z_{P(s)}) > 0$ for all $z_{N(s)} \in \mathcal{X}_{N(s)}$ for which $f_{s,z_{PN(s)}} \neq 0$. We will now use these local mass functions to create, for specific closed subsets S of G, Bayesian networks that have a graphical structure corresponding to this closed subset S. By an argument similar to the one for local sets of desirable gambles in Section 5, we see that in order to do so, all that is needed is for us to instantiate the value of $X_{P(S)}$. Every choice of $y_{P(S)} \in \mathcal{X}_{P(S)}$ then yields, for all $s \in S$ and $z_{P_S(s)} \in \mathcal{X}_{P_S(s)}$, a conditional local mass function $p_s(\cdot|z_{P_S(s)})$ and expectation operator $E_s(\cdot|z_{P_S(s)})$ obtained by identifying them with $p_s(\cdot|z_{P(s)})$ and $E_s(\cdot|z_{P(s)})$, where we let $z_{P(s)\setminus P_S(s)} = y_{P(s)\setminus P_S(s)}$. We denote the mass function of the resulting Bayesian network by $p_S(\cdot|y_{P(S)})$ and its corresponding expectation operator by $E_S(\cdot|y_{P(S)})$. In order to explicitly recall the specific choice of $y_{P(S)} \in \mathcal{X}_{P(S)}$ also in the notation used for the local models, we will also write $p_s(\cdot|z_{P_S(s)}, y_{P(S)}) := p_s(\cdot|z_{P_S(s)})$ and $E_s(\cdot|z_{P_S(s)}, y_{P(S)}) := E_s(\cdot|z_{P_S(s)})$. For every fixed $y_{P(S)} \in \mathcal{X}_{P(S)}$, this expectation operator has a number of useful properties.

A first and trivial property is that $E_S(1|y_{P(S)}) = 1$.

Secondly, consider any $s \in S$. S is a closed subset of G and therefore, due to Lemma 20(ix), $P(S) \subseteq PN(s)$. It then holds for all $z_{PN(s)} \in \mathcal{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$ and $z_{P(S)} = y_{P(S)}$, that $E_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}} f_{s,z_{PN(s)}}|y_{P(S)}) > 0$. To see why, first notice that because S is closed, $PN_S(s) = PN(s) \cap S$ due to Lemma 21(iii). It then follows from the conditional independence properties of Bayesian networks that indeed

$$\begin{aligned}E_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}}f_{s,z_{PN(s)}}|y_{P(S)}) = E_S(\mathbb{I}_{\{z_{PN_S(s)}\}}f_{s,z_{PN(s)}}|y_{P(S)}) &= p_S(z_{PN_S(s)}|y_{P(S)})E_S(f_{s,z_{PN(s)}}|z_{PN_S(s)}, y_{P(S)})\\ &= p_S(z_{PN_S(s)}|y_{P(S)})E_s(f_{s,z_{PN(s)}}|z_{P_S(s)}, y_{P(S)})\\ &= p_S(z_{PN_S(s)}|y_{P(S)})E_s(f_{s,z_{PN(s)}}|z_{P(s)}) > 0,\end{aligned}$$

where the inequality holds because $E_s(f_{s,z_{PN(s)}}|z_{P(s)})$ and $p_S(z_{PN_S(s)}|y_{P(S)})$ are strictly positive. For $E_s(f_{s,z_{PN(s)}}|z_{P(s)})$, this is true by construction, and for $p_S(z_{PN_S(s)}|y_{P(S)})$, this holds because all local probabilities are by construction strictly positive and therefore the global ones are too.

Thirdly, fix $s \in G \setminus S$ and $z_{PN(s)} \in \mathcal{X}_{PN(s)}$ such that $z_{P(S)\cap PN(s)} = y_{P(S)\cap PN(s)}$. By applying the factorisation and conditional independence properties of the resulting Bayesian network, we find that

$$\begin{aligned}E_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}}|y_{P(S)}) &= \sum_{\substack{w_S\in\mathcal{X}_S\\ w_{PN(s)\cap S}=z_{PN(s)\cap S}}} p_S(w_S|y_{P(S)}) = \sum_{\substack{w_S\in\mathcal{X}_S\\ w_{PN(s)\cap S}=z_{PN(s)\cap S}}} \prod_{k\in S} p_k(w_k|w_{P_S(k)}, y_{P(S)})\\ &= \prod_{k\in PN(s)\cap S} p_k(z_k|z_{P_S(k)}, y_{P(S)}) \sum_{\substack{w_S\in\mathcal{X}_S\\ w_{PN(s)\cap S}=z_{PN(s)\cap S}}} \prod_{k\in S\setminus PN(s)} p_k(w_k|w_{P_S(k)}, y_{P(S)})\\ &= \prod_{k\in PN(s)\cap S} p_k(z_k|z_{P_S(k)}, y_{P(S)}) = \prod_{k\in PN(s)\cap S} p_k(z_k|z_{P(k)}).\end{aligned}$$

To understand the third equality, notice that since S is closed, Lemma 21(v) implies, for all $k \in PN(s)\cap S$, that $P_S(k) \subseteq PN(s)\cap S$. For the fifth equality, it suffices to apply Lemma 21(vi) to find that for all $k \in PN(s)\cap S$, $P(k)\setminus P_S(k) \subseteq PN(s)\cap P(S)$. The fourth equality is a bit more complicated. It is trivial if $S\setminus PN(s) = \emptyset$, so suppose that $S \setminus PN(s) \neq \emptyset$. Pick any leaf ℓ from the sub-DAG that corresponds to the nodes in $S\setminus PN(s)$; this is possible because $S\setminus PN(s) \neq \emptyset$ and a DAG always has at least one leaf. We then find that

$$\begin{aligned}&\sum_{\substack{w_S\in\mathcal{X}_S\\ w_{PN(s)\cap S}=z_{PN(s)\cap S}}} \prod_{k\in S\setminus PN(s)} p_k(w_k|w_{P_S(k)}, y_{P(S)})\\ &\quad= \sum_{\substack{w_{S\setminus\{\ell\}}\in\mathcal{X}_{S\setminus\{\ell\}}\\ w_{PN(s)\cap S}=z_{PN(s)\cap S}}} \sum_{w_\ell\in\mathcal{X}_\ell} \prod_{k\in(S\setminus PN(s))} p_k(w_k|w_{P_S(k)}, y_{P(S)})\\ &\quad= \sum_{\substack{w_{S\setminus\{\ell\}}\in\mathcal{X}_{S\setminus\{\ell\}}\\ w_{PN(s)\cap S}=z_{PN(s)\cap S}}} \prod_{k\in(S\setminus PN(s))\setminus\{\ell\}} p_k(w_k|w_{P_S(k)}, y_{P(S)}) \sum_{w_\ell\in\mathcal{X}_\ell} p_\ell(w_\ell|w_{P_S(\ell)}, y_{P(S)})\\ &\quad= \sum_{\substack{w_{S\setminus\{\ell\}}\in\mathcal{X}_{S\setminus\{\ell\}}\\ w_{PN(s)\cap S}=z_{PN(s)\cap S}}} \prod_{k\in(S\setminus PN(s))\setminus\{\ell\}} p_k(w_k|w_{P_S(k)}, y_{P(S)}).\end{aligned}$$

The first equality holds because $\ell \notin PN(s)\cap S$, and the second one because $\ell \notin P_S(k)$ for $k \in (S\setminus PN(s))\setminus\{\ell\}$, since ℓ was assumed to be a leaf of $S\setminus PN(s)$. By repeating this argument for the sub-DAG that corresponds to the nodes in $(S\setminus PN(s))\setminus\{\ell\}$, we can remove yet another node, and if we go on in this way until no node remains, we eventually obtain that indeed

$$\sum_{\substack{w_S\in\mathcal{X}_S\\ w_{PN(s)\cap S}=z_{PN(s)\cap S}}} \prod_{k\in S\setminus PN(s)} p_k(w_k|w_{P_S(k)}, y_{P(S)}) = 1.$$

Hence, for any $s \in G\setminus S$ and $z_{PN(s)} \in \mathcal{X}_{PN(s)}$ such that $z_{P(S)\cap PN(s)} = y_{P(S)\cap PN(s)}$:

$$E_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}}|y_{P(S)}) = \prod_{k\in PN(s)\cap S} p_k(z_k|z_{P(k)}).$$

We can derive two additional things from this result. First of all, $E_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}}|y_{P(S)})$ is strictly positive because all local probabilities are strictly positive by construction. And secondly, $E_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}}|y_{P(S)})$ does not depend on the particular value of $y_{P(S)\setminus PN(s)}$ because for all $k \in PN(s)\cap S$, $P(k) \subseteq PN(s)$ due to Lemma 20(iv).

If we now no longer consider a fixed value of $y_{P(S)} \in \mathcal{X}_{P(S)}$, then the results mentioned above have a number of immediate consequences. First of all, the gamble $E_S(1|X_{P(S)})$ is constant and equal to 1. Secondly, for all $s \in S$ and $z_{PN(s)} \in \mathcal{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$, $E_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}}f_{s,z_{PN(s)}}|z_{P(S)}) > 0$. And thirdly, for all $s \in G\setminus S$ and $z_{PN(s)} \in \mathcal{X}_{PN(s)}$ the gamble

$$E_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}}|z_{P(S)\cap PN(s)}, X_{P(S)\setminus PN(s)})$$

is constant and strictly positive: it is equal to $\prod_{k\in PN(s)\cap S} p_k(z_k|z_{P(k)})$.

We start, in a first stage, with the Bayesian networks that correspond with the subset $S := D(K)$ of G, closed due to Lemma 20(iii). Every $y_{P(D(K))} \in \mathcal{X}_{P(D(K))}$ yields a Bayesian network, and thus a mass function $p_{D(K)}(\cdot|y_{P(D(K))})$ on $\mathcal{X}_{D(K)}$ and an associated expectation operator $E_{D(K)}(\cdot|y_{P(D(K))})$ on $\mathcal{G}(\mathcal{X}_{D(K)})$. If we do not fix the value of $y_{P(D(K))} \in \mathcal{X}_{P(D(K))}$, then $E_{D(K)}(\cdot|X_{P(D(K))})$ satisfies a number of properties, which have already been proven above for general closed subsets S of G. First of all, the gamble $E_{D(K)}(1|X_{P(D(K))})$ is constant and equal to 1. Secondly, for all $s \in D(K)$ and $z_{PN(s)} \in \mathcal{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$, $E_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} f_{s,z_{PN(s)}}|z_{P(D(K))}) > 0$. And thirdly, for all $s \in G \setminus D(K)$ and $z_{PN(s)} \in \mathcal{X}_{PN(s)}$ the gamble

$$E_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}|z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)})$$

is constant and strictly positive: it is equal to $\prod_{k\in PN(s)\cap D(K)} p_k(z_k|z_{P(k)})$.

Next, in the second stage, we consider the (single) Bayesian network that corresponds with the subset $S := PN(K)$ of G, closed due to Lemma 20(iii). Since K is closed, we infer from Lemma 20(v) that $P(PN(K)) = \emptyset$. Therefore, we do not need to instantiate $X_{P(PN(K))}$ since it is deterministic. We thus obtain an unconditional mass function $p_{PN(K)}$ on $\mathcal{X}_{PN(K)}$ and a corresponding expectation operator $E_{PN(K)}$ on $\mathcal{G}(\mathcal{X}_{PN(K)})$. Again, we derive a number of properties that will be needed further on in this proof.

First of all, since all local probabilities are strictly positive, we find that, for all $z_{PN(K)} \in \mathcal{X}_{PN(K)}$, $p_{PN(K)}(z_{PN(K)}) > 0$.

Secondly, for all $s \in PN(K)$ and $z_{PN(s)} \in \mathcal{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$, we have $E_{PN(K)}(\mathbb{I}_{\{z_{PN(s)\cap PN(K)}\}} f_{s,z_{PN(s)}}) > 0$. To see why, first recall that, due to Lemma 20(iii), $PN(K)$ is closed. Therefore $PN_{PN(K)}(s) = PN(s) \cap PN(K)$ due to Lemma 21(iii), and $P_{PN(K)}(s) = P(s)$ because of Lemma 19 and $P(PN(K)) = \emptyset$. It then follows from the conditional independence properties of Bayesian networks that

$$\begin{aligned} E_{PN(K)}(\mathbb{I}_{\{z_{PN(s)\cap PN(K)}\}} f_{s,z_{PN(s)}}) = E_{PN(K)}(\mathbb{I}_{\{z_{PN_{PN(K)}(s)}\}} f_{s,z_{PN(s)}}) &= p_{PN(K)}(z_{PN_{PN(K)}(s)}) E_{PN(K)}(f_{s,z_{PN(s)}}|z_{PN_{PN(K)}(s)}) \\ &= p_{PN(K)}(z_{PN_{PN(K)}(s)}) E_s(f_{s,z_{PN(s)}}|z_{P_{PN(K)}(s)}) \\ &= p_{PN(K)}(z_{PN_{PN(K)}(s)}) E_s(f_{s,z_{PN(s)}}|z_{P(s)}) > 0, \end{aligned}$$

where the inequality holds because $E_s(f_{s,z_{PN(s)}}|z_{P(s)})$ and $p_{PN(K)}(z_{PN_{PN(K)}(s)})$ are strictly positive. For $E_s(f_{s,z_{PN(s)}}|z_{P(s)})$, this is true by construction, and for $p_{PN(K)}(z_{PN_{PN(K)}(s)})$, this holds because all local probabilities are by construction strictly positive and therefore the global ones are too.

Thirdly, we have $E_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}}) > 0$. Indeed, since $g\mathbb{I}_{\{x_{P(K)}\}} \in \mathcal{G}(\mathcal{X}_{PN(K)_{>0}})$, we can derive that $E_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}})$ is a positive linear combination of probabilities $p_{PN(K)}(z_{PN(K)})$, with $z_{PN(K)} \in \mathcal{X}_{PN(K)}$, for which we have already shown that they are strictly positive.

For the third stage, we start by associating with the gamble f the following collection of gambles on $\mathcal{X}_K$:

$$\mathcal{A}^f_{K\rfloor x_{P(K)}} := \left\{\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}} : s \in K,\ z_{PN(s)} \in \mathcal{X}_{PN(s)},\ z_{P(s)\setminus P_K(s)} = x_{P(s)\setminus P_K(s)},\ P(s)\cap K \subseteq K_1 \subseteq K,\ f_{s,z_{PN(s)}} \neq 0\right\},$$

which is a finite subset of $\mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}} := \mathrm{posi}(\mathcal{A}^{\mathrm{irr}}_{K\rfloor x_{P(K)}})$. To see why, first notice that because $PN_K(s) = PN(s) \cap K$ due to Lemma 21(iii), $\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}$ is clearly the (finite) sum of all indicators $\mathbb{I}_{\{y_{PN_K(s)}\}}$ such that $y_{PN_K(s)} \in \mathcal{X}_{PN_K(s)}$ and $y_{PN(s)\cap K_1} = z_{PN(s)\cap K_1}$. By definition of the posi operator, we are now left to show that for any $y_{PN_K(s)} \in \mathcal{X}_{PN_K(s)}$ such that $y_{PN(s)\cap K_1} = z_{PN(s)\cap K_1}$, we have $\mathbb{I}_{\{y_{PN_K(s)}\}} f_{s,z_{PN(s)}} \in \mathcal{A}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$. By construction of $\mathcal{A}^f_{K\rfloor x_{P(K)}}$, $z_{P(s)\setminus P_K(s)} = x_{P(s)\setminus P_K(s)}$, and it therefore suffices to show that $y_{P_K(s)} = z_{P_K(s)}$. To see why this last equality holds, first notice that $P_K(s) = P(s) \cap K$ due to Lemma 19. Also, $P(s) \cap K \subseteq PN(s) \cap K_1$ because $P(s) \cap K \subseteq K_1$ by construction of $\mathcal{A}^f_{K\rfloor x_{P(K)}}$ and $P(s) \cap K \subseteq PN(s)$ by definition of $PN(s)$. Therefore, we find that $P_K(s) \subseteq PN(s) \cap K_1$, implying that $y_{P_K(s)} = z_{P_K(s)}$ is a direct consequence of $y_{PN(s)\cap K_1} = z_{PN(s)\cap K_1}$.

Due to the coherence of $\mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ and the assumption that the non-zero $f \notin \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$, $0 \notin \mathrm{posi}(\{-f\} \cup \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}})$. To see why this holds, assume *ex absurdo* that $0 \in \mathrm{posi}(\{-f\} \cup \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}})$, then it follows from the coherence of $\mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ that we can find $\lambda_1, \lambda_2 > 0$ and $h \in \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ such that $\lambda_1(-f) + \lambda_2 h = 0$ and therefore $f = (\lambda_2/\lambda_1)h \in \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$, contradicting $f \notin \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$.

Since $\mathcal{A}^f_{K\rfloor x_{P(K)}}$ and $\mathcal{G}(\mathcal{X}_K)_{>0}$ are both subsets of $\mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$, we can infer from $0 \notin \mathrm{posi}(\{-f\} \cup \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}})$ that $0 \notin \mathrm{posi}(\{-f\} \cup \mathcal{A}^f_{K\rfloor x_{P(K)}} \cup \mathcal{G}(\mathcal{X}_K)_{>0}) =: \mathcal{E}(\{-f\} \cup \mathcal{A}^f_{K\rfloor x_{P(K)}})$. We also know that $\{-f\} \cup \mathcal{A}^f_{K\rfloor x_{P(K)}}$ is a finite subset of $\mathcal{G}(\mathcal{X}_K)$ and therefore, we can apply Lemma 18. This provides us with a mass function $p_{K\rfloor x_{P(K)}}$ on $\mathcal{X}_K$ with expectation operator $E_{K\rfloor x_{P(K)}}$ on $\mathcal{G}(\mathcal{X}_K)$ for which $p_{K\rfloor x_{P(K)}}(z_K) > 0$ for all $z_K \in \mathcal{X}_K$, $E_{K\rfloor x_{P(K)}}(h) > 0$ for all $h \in \mathcal{A}^f_{K\rfloor x_{P(K)}}$ and $E_{K\rfloor x_{P(K)}}(f) < 0$.

Using the mass function $p_{K\rfloor x_{P(K)}}$ on $\mathcal{X}_K$ and the local mass functions, we will now construct, for every instantiation $y_{P(K)} \in \mathcal{X}_{P(K)}$, a conditional mass function $p_K(\cdot|y_{P(K)})$ on $\mathcal{X}_K$. So consider any $y_{P(K)} \in \mathcal{X}_{P(K)}$. If $y_{P(K)} = x_{P(K)}$, we define $p_K(\cdot|x_{P(K)}) := p_{K\rfloor x_{P(K)}}$. If $y_{P(K)}$ is such that $y_k \neq x_k$ for all $k \in P(K)$, then $p_K(\cdot|y_{P(K)})$ is constructed in the same way as we have done several times before: we use the local mass functions and the instantiation $y_{P(K)}$ of the parent variables $X_{P(K)}$ to construct a Bayesian network that has a graphical structure corresponding to the subset K of G. Unlike the one in the preceding paragraphs, this construction does not take into account the gamble f. In all other cases, we need a more complex construction that includes the previous two as a special case.

Let us denote by $P_1(K)$ the largest subset of $P(K)$ such that $y_k = x_k$ for all $k \in P_1(K)$ and let $P_2(K) := P(K) \setminus P_1(K)$. We also let $K_2 := K \cap D(P_2(K))$ and $K_1 := K \setminus D(P_2(K))$. These sets depend on $y_{P(K)}$, but we have not reflected this in the notation to avoid cluttering up the formulas and because $y_{P(K)}$ is fixed in this part of the proof.

It should be clear by now that for any $z_{P(K_2)} \in \mathcal{X}_{P(K_2)}$, we can construct a strictly positive mass function $p_{K_2}(\cdot|z_{P(K_2)})$ on $\mathcal{X}_{K_2}$ and an associated expectation operator $E_{K_2}(\cdot|z_{P(K_2)})$ on $\mathcal{G}(\mathcal{X}_{K_2})$, by using the local mass functions to construct a Bayesian network that has a graphical structure corresponding to the subset K_2 of G. Since we know that K_2 is a closed subset of G due to Lemma 22(i), we have for all $s \in G$ and $z_{PN(s)} \in \mathcal{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$ that $E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} f_{s,z_{PN(s)}}|z_{P(K_2)}) > 0$ if $s \in K_2$ and that the gamble $E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}|z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)})$ is constant, strictly positive and equal to $\prod_{k \in PN(s)\cap K_2} p_k(z_k|z_{P(k)})$ if $s \notin K_2$. The proof for both these properties follows from the discussion above for general closed subsets S, by taking $S = K_2$.

Next, we define the mass function $p_{K_1}(\cdot|x_{P_1(K)})$ on $\mathcal{X}_{K_1}$ as the marginalisation of $p_{K\rfloor x_{P(K)}}$ to $\mathcal{X}_{K_1}$: for all $z_{K_1} \in \mathcal{X}_{K_1}$, we let

$$p_{K_1}(z_{K_1}|x_{P_1(K)}) := \sum_{\substack{w_K \in \mathcal{X}_K \\ w_{K_1} = z_{K_1}}} p_{K\rfloor x_{P(K)}}(w_K).$$

Since all the terms in this sum are strictly positive by construction, we have that $p_{K_1}(z_{K_1}|x_{P_1(K)}) > 0$ for all $z_{K_1} \in \mathcal{X}_{K_1}$.

For the corresponding expectation operator $E_{K_1}(\cdot|x_{P_1(K)})$ on $\mathcal{G}(\mathcal{X}_{K_1})$, we get that $E_{K_1}(h|x_{P_1(K)}) = E_{K\rfloor x_{P(K)}}(h)$ for all $h \in \mathcal{G}(\mathcal{X}_{K_1})$.

We can now construct the mass function $p_K(\cdot|y_{P(K)})$ by defining, for all $z_K \in \mathcal{X}_K$:

$$\begin{aligned} p_K(z_K|y_{P(K)}) &:= p_{K_1}(z_{K_1}|y_{P_1(K)}) p_{K_2}(z_{K_2}|y_{P(K_2)\cap P(K)}, z_{P(K_2)\setminus P(K)}) \\ &= p_{K_1}(z_{K_1}|x_{P_1(K)}) p_{K_2}(z_{K_2}|y_{P(K_2)\cap P(K)}, z_{P(K_2)\setminus P(K)}), \end{aligned}$$

which makes sense because $P(K_2) \setminus P(K) \subseteq K_1$. It should be clear that for all $z_K \in \mathcal{X}_K$, we have that $p_K(z_K|y_{P(K)}) > 0$.

For the corresponding expectation operator $E_K(\cdot|y_{P(K)})$, the law of iterated expectation yields for all $h \in \mathcal{G}(\mathcal{X}_K)$ that

$$E_K(h|y_{P(K)}) = E_{K_1}\big(E_{K_2}(h|y_{P(K_2)\cap P(K)}, X_{P(K_2)\setminus P(K)})|x_{P_1(K)}\big). \tag{A.2}$$

This expectation operator has two useful properties that we will need further on in this proof.

For the first property of $E_K(\cdot|y_{P(K)})$, consider any $s \in K$, implying that $P(K) \subseteq PN(s)$ due to Lemma 20(ix). It then holds for all $z_{PN(s)} \in \mathcal{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$ and $z_{P(K)} = y_{P(K)}$, that $E_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}}|z_{P(K)}) > 0$. To see why, consider two distinct cases: $s \in K_2$ and $s \in K_1$.

If $s \in K_2$, then because $K_1 \subseteq PN(s)$ due to Lemma 22(iv), we get, using Eq. (A.2):

$$\begin{aligned} E_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}}|y_{P(K)}) &= E_{K_1}\big(E_{K_2}(\mathbb{I}_{\{z_{K_1}\}}\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} f_{s,z_{PN(s)}}|y_{P(K_2)\cap P(K)}, X_{P(K_2)\setminus P(K)})|x_{P_1(K)}\big) \\ &= E_{K_1}\big(\mathbb{I}_{\{z_{K_1}\}} E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} f_{s,z_{PN(s)}}|y_{P(K_2)\cap P(K)}, z_{P(K_2)\setminus P(K)})|x_{P_1(K)}\big) \\ &= E_{K_1}(\mathbb{I}_{\{z_{K_1}\}}|x_{P_1(K)}) E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} f_{s,z_{PN(s)}}|y_{P(K_2)\cap P(K)}, z_{P(K_2)\setminus P(K)}) \\ &= p_{K_1}(z_{K_1}|x_{P_1(K)}) E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} f_{s,z_{PN(s)}}|z_{P(K_2)}) > 0, \end{aligned}$$

where the final expression is strictly positive because both factors have been proved above to be strictly positive.

If $s \in K_1$, then because $P(K) \subseteq PN(s)$, we get, using Eq. (A.2):

$$\begin{aligned} E_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}}|y_{P(K)}) &= E_{K_1}\big(E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}}\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}|z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)})|x_{P_1(K)}\big) \\ &= E_{K_1}\big(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}} E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}|z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)})|x_{P_1(K)}\big) \\ &= E_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}}|x_{P_1(K)}) E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}|z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)}) > 0. \end{aligned}$$

The third equality holds because $E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}|z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)})$ has been shown to be a constant gamble earlier on and the final expression is strictly positive since the two constituting factors are strictly positive. For $E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}|z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)})$, this has already been proved. For $E_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}}|x_{P_1(K)})$, this follows from

$$E_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}}|x_{P_1(K)}) = E_{K\rfloor x_{P(K)}}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}}) > 0,$$

where the final inequality is due to $\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}}$ being an element of $\mathcal{A}^f_{K\rfloor x_{P(K)}}$, which is in turn true because $P(s) \cap K = P(s) \cap K_1 \subseteq K_1 \subseteq K$ due to Lemma 22(v), and because $P(s) \cap P(K) = P(s) \cap P(K_1) \subseteq P(K_1) \subseteq P_1(K)$ due to Lemma 22(vi) and (ii) and therefore $z_{P(s)\cap P(K)} = x_{P(s)\cap P(K)}$, implying that $z_{P(s)\setminus P_K(s)} = x_{P(s)\setminus P_K(s)}$ due to Lemma 19.

The second property of $E_K(\cdot|y_{P(K)})$ is that for all $s \in PN(K)$ and $z_{PN(s)} \in \mathcal{X}_{PN(s)}$ such that $z_{P(K)\cap PN(s)} = y_{P(K)\cap PN(s)}$, $E_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}}|y_{P(K)})$ is strictly positive and does not depend on the particular value of $y_{P(K)\setminus PN(s)}$. To prove this, we start by recalling from the discussion above that, because $s \notin K_2$, $E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}|z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)})$ is a constant, strictly positive gamble that is furthermore equal to $\prod_{k\in PN(s)\cap K_2} p_k(z_k|z_{P(k)})$. Hence, we also have that $E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}|y_{P(K_2)\cap P(K)}, z_{(P(K_2)\setminus P(K))\cap PN(s)}, X_{(P(K_2)\setminus P(K))\setminus PN(s)})$ is a constant, strictly positive gamble that is equal to $\prod_{k\in PN(s)\cap K_2} p_k(z_k|z_{P(k)})$ because

$$\begin{aligned}
&(y_{P(K_2)\cap P(K)}, z_{(P(K_2)\setminus P(K))\cap PN(s)}, X_{(P(K_2)\setminus P(K))\setminus PN(s)})\\
&\quad= (y_{(P(K_2)\cap P(K))\setminus PN(s)}, y_{P(K_2)\cap P(K)\cap PN(s)}, z_{(P(K_2)\setminus P(K))\cap PN(s)}, X_{(P(K_2)\setminus P(K))\setminus PN(s)})\\
&\quad= (y_{(P(K_2)\cap P(K))\setminus PN(s)}, z_{P(K_2)\cap P(K)\cap PN(s)}, z_{(P(K_2)\setminus P(K))\cap PN(s)}, X_{(P(K_2)\setminus P(K))\setminus PN(s)})\\
&\quad= (y_{(P(K_2)\cap P(K))\setminus PN(s)}, z_{P(K_2)\cap PN(s)}, X_{(P(K_2)\setminus P(K))\setminus PN(s)})\\
&\quad= (y_{(P(K_2)\setminus PN(s))\cap P(K)}, z_{P(K_2)\cap PN(s)}, X_{(P(K_2)\setminus PN(s))\setminus P(K)}).
\end{aligned}$$

We therefore get that, using Eq. (A.2),

$$\begin{aligned}
E_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}}|y_{P(K)}) &= E_{K_1}\big(E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}|y_{P(K_2)\cap P(K)}, X_{P(K_2)\setminus P(K)})|x_{P_1(K)}\big)\\
&= E_{K_1}\big(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}E_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}|y_{P(K_2)\cap P(K)}, z_{(P(K_2)\setminus P(K))\cap PN(s)}, X_{(P(K_2)\setminus P(K))\setminus PN(s)})|x_{P_1(K)}\big)\\
&= E_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}|x_{P_1(K)}) \prod_{k\in PN(s)\cap K_2} p_k(z_k|z_{P(k)}),
\end{aligned}$$

where the second equality is a consequence of $P(K_2)\setminus P(K)\subseteq K_1$.

The property that we are trying to prove will therefore follow if we can show that both $E_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}|x_{P_1(K)})$ and $\prod_{k\in PN(s)\cap K_2} p_k(z_k|z_{P(k)})$ are strictly positive and do not depend on the particular value of $y_{P(K)\setminus PN(s)}\in\mathcal{X}_{P(K)\setminus PN(s)}$.

We start with $E_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}|x_{P_1(K)})$. It is by definition equal to $E_{K\rfloor x_{P(K)}}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}})$ and therefore strictly positive because $p_{K\rfloor x_{P(K)}}$ is a strictly positive mass function. Since $s\in PN(K)$, we can use Lemma 22(vii) to infer that $PN(s)\cap K_1 = K\setminus D((P(K)\setminus PN(s))\cup(PN(s)\cap P_2(K)))$. We therefore find that $PN(s)\cap K_1$ does not depend on the particular value of $y_{P(K)\setminus PN(s)}$ in $\mathcal{X}_{P(K)\setminus PN(s)}$ because $PN(s)\cap P_2(K)$ is fully determined by $y_{P(K)\cap PN(s)}$. Hence, $E_{K\rfloor x_{P(K)}}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}})$ does not depend on $y_{P(K)\setminus PN(s)}$ either.

For $\prod_{k\in PN(s)\cap K_2} p_k(z_k|z_{P(k)})$, we start by noticing that $PN(s)\cap K_2$ does not depend on the particular choice of $y_{P(K)\setminus PN(s)}$ in $\mathcal{X}_{P(K)\setminus PN(s)}$ because, as we have shown in the previous paragraph, $PN(s)\cap K_1$ does not and because $PN(s)\cap K_2 = (PN(s)\cap K)\setminus(PN(s)\cap K_1)$. Next, for all $k\in PN(s)\cap K_2$, the factor $p_k(z_k|z_{P(k)})$ will not depend on the particular value of $y_{P(K)\setminus PN(s)}$ in $\mathcal{X}_{P(K)\setminus PN(s)}$ because $P(k)\subseteq PN(s)$ due to Lemma 20(iv) and we therefore find that $\prod_{k\in PN(s)\cap K_2} p_k(z_k|z_{P(k)})$ does not depend on the particular value of $y_{P(K)\setminus PN(s)}$ in $\mathcal{X}_{P(K)\setminus PN(s)}$.

If we now no longer consider a fixed value of $y_{P(K)}\in\mathcal{X}_{P(K)}$, then the results mentioned above have two immediate consequences. First, for all $s\in K$ and $z_{PN(s)}\in\mathcal{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}}\neq 0$, $E_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}}f_{s,z_{PN(s)}}|z_{P(K)})>0$. And second, for all $s\in PN(K)$ and $z_{PN(s)}\in\mathcal{X}_{PN(s)}$, the gamble $E_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}}|z_{P(K)\cap PN(s)}, X_{P(K)\setminus PN(s)})$ is constant and strictly positive.

We are now ready to define the mass function p_G on $\mathcal{X}_G$ that we have been after all along. For all $z_G\in\mathcal{X}_G$, we let

$$p_G(z_G) := p_{PN(K)}(z_{PN(K)})p_K(z_K|z_{P(K)})p_{D(K)}(z_{D(K)}|z_{P(D(K))}),$$

where all three factors (mass functions) have been defined in earlier parts of this proof. Since $PN(K)$, K and $D(K)$ constitute a partition of G due to Lemma 20(ii), and since $P(K)\subseteq PN(K)$ and $P(D(K))\subseteq PN(K)\cup K$, we see that p_G is indeed a mass function on $\mathcal{X}_G$. For the corresponding expectation operator E_G, we know from the law of iterated expectation, and, again, $P(K)\subseteq PN(K)$ and $P(D(K))\subseteq PN(K)\cup K$, that for all $h\in\mathcal{G}(\mathcal{X}_G)$:

$$E_G(h) = E_{PN(K)}\big(E_K\big(E_{D(K)}(h|X_{P(D(K))})|X_{P(K)}\big)\big).$$

Taking the expectation of the gamble $f'' = g\mathbb{I}_{\{x_{P(K)}\}}f$, and recalling that $g\in\mathcal{G}(\mathcal{X}_{N(K)})_{>0}$ and $f\in\mathcal{G}(\mathcal{X}_K)$, we find that

$$\begin{aligned}
E_G\big(f''\big) &= E_{PN(K)}\big(E_K\big(E_{D(K)}\big(f''\big|X_{P(D(K))}\big)\big|X_{P(K)}\big)\big)\\
&= E_{PN(K)}\big(g\mathbb{I}_{\{x_{P(K)}\}}E_K\big(fE_{D(K)}(1|X_{P(D(K))})|x_{P(K)}\big)\big)\\
&= E_{PN(K)}\big(g\mathbb{I}_{\{x_{P(K)}\}}E_K(f|x_{P(K)})\big) = E_K(f|x_{P(K)})E_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}}) < 0,
\end{aligned}$$

where the last inequality holds because $E_K(f|x_{P(K)})<0$ and $E_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}})>0$. For $E_K(f|x_{P(K)})$, this is true by construction since $E_K(f|x_{P(K)}) = E_{K\rfloor x_{P(K)}}(f)<0$ and for $E_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}})$, this has been shown earlier on in this proof.

All that is now left to do, is to show that also $E_G(f'')>0$. Since $f_{s,z_{PN(s)}}\neq 0$ for at least one $s\in G$ and $z_{PN(s)}\in\mathcal{X}_{PN(s)}$, Eq. (A.1) tells us that it suffices to show that $E_G(\mathbb{I}_{\{z_{PN(s)}\}}f_{s,z_{PN(s)}})>0$ for all $s\in G$ and $z_{PN(s)}\in\mathcal{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}}\neq 0$. So let us fix any such $s\in G$ and $z_{PN(s)}\in\mathcal{X}_{PN(s)}$ and show that $E_G(\mathbb{I}_{\{z_{PN(s)}\}}f_{s,z_{PN(s)}})>0$. We consider three exhaustive and mutually exclusive cases: $s\in D(K)$, $s\in K$ and $s\in PN(K)$. The fact that these cases are indeed exhaustive and mutually exclusive follows from $D(K)$, K and $PN(K)$ constituting a partition of G, due to Lemma 20(ii).

If $s\in D(K)$, then, because $D(K)$ is closed due to Lemma 20(iii) and because $PN(D(K)) = PN(K)\cup K$ due to Lemma 20(vi), we can use Lemma 20(x) to infer that $PN(K)\cup K\subseteq PN(s)$ and therefore

$$
\begin{aligned}
E_G(\mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}}) &= E_G(\mathbb{I}_{\{z_{PN(K)}\}}\mathbb{I}_{\{z_K\}}\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} f_{s,z_{PN(s)}}) \\
&= p_{PN(K)}(z_{PN(K)})p_K(z_K|z_{P(K)})E_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} f_{s,z_{PN(s)}}|z_{P(D(K))}) > 0,
\end{aligned}
$$

where the inequality holds because all three constituting factors have been shown to be strictly positive earlier on.

If $s \in K$, then since $PN(K) \subseteq PN(s)$ due to Lemma 20(x), we get

$$
\begin{aligned}
&E_G(\mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}}) \\
&= E_G(\mathbb{I}_{\{z_{PN(K)}\}}\mathbb{I}_{\{z_{PN(s)\cap K}\}}\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} f_{s,z_{PN(s)}}) \\
&= p_{PN(K)}(z_{PN(K)})E_K\big(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}} E_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}|z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)})|z_{P(K)}\big) \\
&= p_{PN(K)}(z_{PN(K)})E_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}}|z_{P(K)})E_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}|z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)}) > 0.
\end{aligned}
$$

The third equality holds because $E_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}|z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)})$ has been shown to be a constant gamble earlier on (observe that $s \notin D(K)$ because $s \in K$), and the final expression is strictly positive since all three constituting factors have been shown to be strictly positive earlier on as well.

Finally, if $s \in PN(K)$, then

$$
\begin{aligned}
&E_G(\mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}}) \\
&= E_G(\mathbb{I}_{\{z_{PN(s)\cap PN(K)}\}}\mathbb{I}_{\{z_{PN(s)\cap K}\}}\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} f_{s,z_{PN(s)}}) \\
&= E_{PN(K)}\big(\mathbb{I}_{\{z_{PN(s)\cap PN(K)}\}} f_{s,z_{PN(s)}} \\
&\qquad E_K\big(\mathbb{I}_{\{z_{PN(s)\cap K}\}} E_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}|z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)})|z_{P(K)\cap PN(s)}, X_{P(K)\setminus PN(s)}\big)\big) \\
&= E_{PN(K)}(\mathbb{I}_{\{z_{PN(s)\cap PN(K)}\}} f_{s,z_{PN(s)}}) \\
&\qquad E_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}}|z_{P(K)\cap PN(s)}, X_{P(K)\setminus PN(s)})E_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}|z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)}) > 0,
\end{aligned}
$$

where the third equality is a consequence of the earlier proven fact that both $E_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}}|z_{P(K)\cap PN(s)}, X_{P(K)\setminus PN(s)})$ and $E_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}|z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)})$ are constant gambles. The inequality is again due to the three constituting factors being strictly positive, which was also shown earlier on in this proof. $\square$

Proof of Corollary 11. Fix any closed $K \subseteq G$, $x_{P(K)} \in \mathcal{X}_{P(K)}$, non-empty $A_{N(K)} \subseteq \mathcal{X}_{N(K)}$ and $f \in \mathcal{G}(\mathcal{X}_K)$. Since $\mathbb{I}_{A_{N(K)}}$ is an element of $\mathcal{G}(\mathcal{X}_{N(K)})_{>0}$, we know from Eq. (1) and Theorem 10 that

$$
f \in \mathrm{marg}_K\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor\{x_{P(K)}\} \times A_{N(K)}\big) \quad \Leftrightarrow \quad \mathbb{I}_{A_{N(K)}}\mathbb{I}_{\{x_{P(K)}\}} f \in \mathcal{D}_G^{\mathrm{irr}} \quad \Leftrightarrow \quad f \in \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}. \qquad \square
$$

Proof of Corollary 12. Fix any closed $K \subseteq G$, and $x_{P(K)} \in \mathcal{X}_{P(K)}$ and any non-empty $A_{N(K)} \subseteq \mathcal{X}_{N(K)}$. Then due to Corollary 11, $\mathrm{marg}_K(\mathcal{D}_G^{\mathrm{irr}}\rfloor\{x_{P(K)}\} \times A_{N(K)}) = \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$. The special case $A_{N(K)} = \mathcal{X}_{N(K)}$ yields

$$
\mathrm{marg}_K\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor x_{P(K)}\big) = \mathrm{marg}_K\big(\mathcal{D}_G^{\mathrm{irr}}\rfloor\{x_{P(K)}\} \times \mathcal{X}_{N(K)}\big) = \mathcal{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}. \qquad \square
$$

Proof of Proposition 13. Consider any partition $G_1, \dots, G_n$ of G such that $P(G_i) = \emptyset$ for all $i \in \{1, \dots, n\}$. Then, clearly, the sets G_i are disconnected from one another. Indeed, assume *ex absurdo* that there is an arrow from a node $s_i \in G_i$ to a node $s_j \in G_j$, with $i, j \in \{1, \dots, n\}$ and $i \neq j$. Then s_i is an element of $P(G_j)$, which contradicts our assumption that $P(G_j) = \emptyset$. Consequently, for any $i \in \{1, \dots, n\}$ and all $s \in G_i$, we have that $P_{G_i}(s) = P(s) \subseteq G_i$, $D(s) \subseteq G_i$ and therefore also $G \setminus G_i \subseteq N(s)$ and $PN(s) = PN_{G_i}(s) \cup (G \setminus G_i)$.

Now let $\mathcal{D}_G^{\mathrm{irr}}$ be the irrelevant natural extension of the complete network, as given by Eq. (4), and, for all $i \in \{1, \dots, n\}$, let $\mathcal{D}_{G_i}^{\mathrm{irr}}$ be the irrelevant natural extension of the network that has the sub-DAG associated with G_i as its graphical structure, as given by Eq. (7). The result then follows from

$$
\begin{aligned}
\bigotimes_{i=1}^{n} \mathcal{D}_{G_i}^{\mathrm{irr}} &= \mathrm{posi}\big(\big\{\mathbb{I}_{\{z_{G\setminus G_i}\}} f : i \in \{1, \dots, n\},\ z_{G\setminus G_i} \in \mathcal{X}_{G\setminus G_i},\ f \in \mathcal{D}_{G_i}^{\mathrm{irr}}\big\}\big) \\
&= \mathrm{posi}\big(\big\{\mathbb{I}_{\{z_{G\setminus G_i}\}} f : i \in \{1, \dots, n\},\ z_{G\setminus G_i} \in \mathcal{X}_{G\setminus G_i}, \\
&\qquad\qquad f \in \mathrm{posi}\big(\{\mathbb{I}_{\{z_{PN_{G_i}(s)}\}} f : s \in G_i,\ z_{PN_{G_i}(s)} \in \mathcal{X}_{PN_{G_i}(s)},\ f \in \mathcal{D}_{s\rfloor z_{P(s)}}\}\big)\big\}\big) \\
&= \mathrm{posi}\big(\big\{\mathbb{I}_{\{z_{G\setminus G_i}\}}\mathbb{I}_{\{z_{PN_{G_i}(s)}\}} f : i \in \{1, \dots, n\},\ z_{G\setminus G_i} \in \mathcal{X}_{G\setminus G_i},\ s \in G_i,\ z_{PN_{G_i}(s)} \in \mathcal{X}_{PN_{G_i}(s)},\ f \in \mathcal{D}_{s\rfloor z_{P(s)}}\big\}\big) \\
&= \mathrm{posi}\big(\{\mathbb{I}_{\{z_{PN(s)}\}} f : s \in G,\ z_{PN(s)} \in \mathcal{X}_{PN(s)},\ f \in \mathcal{D}_{s\rfloor z_{P(s)}}\}\big) = \mathcal{D}_G^{\mathrm{irr}}.
\end{aligned}
$$

The first equality in this derivation follows from Eq. (9). The second one follows from Eq. (7) and the earlier proven fact that, for all $i \in \{1, \dots, n\}$ and every $s \in G_i$, $P_{G_i}(s) = P(s)$. The third equality is due to the definition of the posi operator and the fourth equality holds because $G_1, \dots, G_n$ is a partition of G and because we have already shown that, for all $i \in \{1, \dots, n\}$ and every $s \in G_i$, $PN(s) = PN_{G_i}(s) \cup (G \setminus G_i)$. The final equality follows from the definition of $\mathcal{D}_G^{\mathrm{irr}}$, as given by Eq. (4). □

Proof of Theorem 14. Fix $i \in \{1, \dots, n\}$ and consider any $s \in P(K_i)$. Then there is some $q \in K_i$ such that $s \in P(q)$ and $s \notin K_i$. Due to our assumption that $P_K(K_i) = \emptyset$, we find that $s \notin K$ and therefore $s \in P(K)$. Hence, we have for all $i \in \{1, \dots, n\}$ that $P(K_i) \subseteq P(K)$.

The rest of the proof is now a direct consequence of Proposition 13. It suffices to apply Proposition 13 to a credal network that has the sub-DAG associated with K as its graphical structure and whose local models are given by Eq. (6). □

Proof of Theorem 15. *Direct* and *reverse redundancy* follow from the consideration that every path from I to O is blocked by I in its first node and by O in its last node.

To prove *direct decomposition*, notice that every path from I to O is also a path from I to $O \cup S$. It is therefore blocked by C due to $\mathrm{AD}(I, O \cup S|C)$. *Reverse decomposition* is proved analogously.

To verify *direct weak union*, consider any path from $i \in I$ to $o \in O$. Since this is also a path from I to $O \cup S$, we know from $\mathrm{AD}(I, O \cup S|C)$ that it is blocked by C. Now let s be the first node in the path for which that path is blocked by C in s. If the path from i to o is blocked by C in s using condition B1, B2 or B4, then $s \in C \subseteq C \cup S$, implying that the path is blocked by $C \cup S$ in s and concluding the proof.

So suppose that the path is blocked by C in s using condition B3. We then have that $s \notin C$ and $D(s) \cap C = \emptyset$. If $s \notin S$ and $D(s) \cap S = \emptyset$, then $s \notin C \cup S$ and $D(s) \cap (C \cup S) = \emptyset$, implying that the path is blocked by $C \cup S$ in s and concluding the proof.

So suppose, and this is the only remaining possibility, that there is some node $t \in \{s\} \cup D(s)$ for which $t \in S$. In that case there is a directed path from s to t and one can concatenate the section from i to s with this directed path from s to t, obtaining a path from $i \in I$ to $t \in O \cup S$. This however leads to a contradiction with $\mathrm{AD}(I, O \cup S|C)$ because this path from i to t is not blocked by C. To see why, first consider all the nodes in the part from i to s, excluding s. The path from i to t cannot be blocked by C in these nodes, because s was the first node in the original path from i to o for which this path was blocked by C in s. It also cannot be blocked by C in the nodes in the part from s to t because this part is directed, $s \notin C$ and $D(s) \cap C = \emptyset$. This means that this possibility cannot occur, which concludes the proof of direct weak union.

Reverse weak union has a similar proof. Every path from $i \in I$ to $o \in O$ is also a path from $I \cup S$ to O and is thus blocked by C. Let s be the last node in the path for which that path is blocked by C in s. If the path from i to o is blocked by C in s using condition B1, B2 or B4, then $s \in C \subseteq C \cup S$, implying that the path is blocked by $C \cup S$ in s and concluding the proof.

So suppose that the path is blocked by C in s by condition B3. We then have that $s \notin C$ and $D(s) \cap C = \emptyset$. If $s \notin S$ and $D(s) \cap S = \emptyset$, then $s \notin C \cup S$ and $D(s) \cap (C \cup S) = \emptyset$, implying that the path is blocked by $C \cup S$ in s and concluding the proof.

So suppose, and this is the only remaining possibility, that there is some node $t \in \{s\} \cup D(s)$ for which $t \in S$. In that case there is a directed path from s to t and one can understand it as a reverse directed path from t to s and concatenate it with the section from s to o, obtaining a path from $t \in I \cup S$ to $o \in O$. This however leads to a contradiction with $\mathrm{AD}(I \cup S, O|C)$ because this path from t to o is not blocked by C. To see why, first consider all the nodes in the part from s to o, excluding s. The path from t to o cannot be blocked by C in these nodes, because s was the last node in the original path from i to o for which this path was blocked by C in s. It also cannot be blocked by C in the nodes in the part from t to s because this part is a reverse directed path, $s \notin C$ and $D(s) \cap C = \emptyset$. This means that this possibility cannot occur, which concludes the proof of reverse weak union.

To prove *direct contraction*, consider any path from $i \in I$ to $s \in O \cup S$. We need to show that it is blocked by C. If $s \in O$, this follows directly from $\mathrm{AD}(I, O|C)$, so we can assume that $s \in S$, implying that the path from i to s is blocked by $C \cup O$ because of $\mathrm{AD}(I, S|C \cup O)$. Let t be one of the nodes for which the path from i to s is blocked by $C \cup O$ in t. If $t \in C$ or $t \notin C \cup O$, then the path from i to s is blocked by C in t, concluding the proof. If $t \in O$, and this is the only remaining possibility, then $\mathrm{AD}(I, O|C)$ implies that the path from i to t must be blocked by C, from which one can also infer that the path from i to s is blocked by C.

Reverse contraction has a similar proof. Take any path from $s \in I \cup S$ to $o \in O$. We need to show that it is blocked by C. If $s \in I$, this follows directly from $\mathrm{AD}(I, O|C)$, so we can assume that $s \in S$, implying that the path from i to s is blocked by $C \cup I$ because of $\mathrm{AD}(S, O|C \cup I)$. Let t be one of the nodes for which the path from s to o is blocked by $C \cup I$ in t. If $t \in C$ or $t \notin C \cup I$, then the path from s to o is blocked by C in t, concluding the proof. If $t \in I$, and this is the only remaining possibility, then $\mathrm{AD}(I, O|C)$ implies that the path from t to o is blocked by C, from which one can also infer that the path from s to o is blocked by C.

For the verification of *direct intersection*, consider any path from $i \in I$ to $s \in O \cup S$. We need to show that it is blocked by C. Due to the symmetry of the problem, we can assume without loss of generality that $s \in S$, implying that the path from i to s is blocked by $C \cup O$ because of $\mathrm{AD}(I, S|C \cup O)$. Now let t be the first node in the path from i to s for which this path is blocked by $C \cup O$ in t. If $t \in C$ or $t \notin C \cup O$, then the path from i to s is blocked by C in t, concluding the proof. If $t \in O$, and this is the only remaining possibility, then this implies that $t \neq s$, since $O \cap S = \emptyset$ by assumption. It also implies that the path from i to t is blocked by $C \cup S$ because of $\mathrm{AD}(I, O|C \cup S)$. If it is blocked by some q for which $q \in C$

or $q \notin C \cup S$, then the path from i to s is blocked by C in q, concluding the proof. If $q \in S$, then this implies that $q \neq t$, since $O \cap S = \emptyset$ by assumption. It also implies that the path from i to q is blocked by $C \cup O$ because of $\mathrm{AD}(I, S|C \cup O)$. However, this would in turn imply that the path from i to s is blocked by $C \cup O$ in some node of the path from i to q, contradicting the earlier assumption that t is the first node for which this is the case.

Finally, to prove *reverse intersection*, consider any path from $s \in I \cup S$ to $o \in O$. We need to show that it is blocked by C. Due to the symmetry of the problem, we can assume without loss of generality that $s \in S$, implying that the path from s to o is blocked by $C \cup I$ because of $\mathrm{AD}(S, O|C \cup I)$. Now let t be the last node in the path from s to o for which this path is blocked by $C \cup I$ in t. If $t \in C$ or $t \notin C \cup I$, then the path from s to i is blocked by C in t, concluding the proof. If $t \in I$, and this is the only remaining possibility, then this implies that $t \neq s$, since $I \cap S = \emptyset$ by assumption. It also implies that the path from t to o is blocked by $C \cup S$ because of $\mathrm{AD}(I, O|C \cup S)$. If it is blocked by some q for which $q \in C$ or $q \notin C \cup S$, then the path from s to o is blocked by C in q, concluding the proof. If $q \in S$, then this implies that $q \neq t$, since $I \cap S = \emptyset$ by assumption. It also implies that the path from q to o is blocked by $C \cup I$ because of $\mathrm{AD}(S, O|C \cup O)$. However, this would in turn imply that the path from s to o is blocked by $C \cup I$ in some node of the path from q to o, contradicting the earlier assumption that t is the last node for which this is the case. $\square$

Our proof for Theorem 16 and Corollary 17 makes use of the following two subsets of G. For all $I, C \subseteq G$, we define

$$B(I,C) := \big\{k \in G: \mathrm{AD}\big(I, \{k\}|C\big)\big\}$$

and

$$BB(I,C) := \big\{r \in B(I,C): \big(\exists k \in B(I,C) \setminus C\big) k \sqsubseteq r\big\}.$$

Proposition 23. *For all subsets I and C of G, it holds that $P(BB(I,C)) \subseteq B(I,C) \setminus BB(I,C) \subseteq C$, $B(I,C) \setminus C \subseteq BB(I,C)$, $D(BB(I,C)) \cap C = \emptyset$ and $I \setminus C \subseteq N(BB(I,C)) \setminus C$. Furthermore, $BB(I,C)$ is closed.*

Proof. We start by showing that any node $r \in B(I,C) \setminus BB(I,C)$ is also an element of C, thus proving $B(I,C) \setminus BB(I,C) \subseteq C$. Indeed, suppose $r \notin C$, then $k = r$ is an element of $B(I,C)$ such that $k \sqsubseteq r$ and $k \notin C$, contradicting the assumption that $r \notin BB(I,C)$.

That $B(I,C) \setminus C \subseteq BB(I,C)$, is a direct consequence of the result above. Indeed, $B(I,C) \setminus BB(I,C) \subseteq C$ implies $B(I,C) \setminus C \subseteq B(I,C) \setminus (B(I,C) \setminus BB(I,C)) = BB(I,C)$, where the last equality follows from $BB(I,C) \subseteq B(I,C)$.

For the proof of $P(BB(I,C)) \subseteq B(I,C) \setminus BB(I,C)$, consider any node $p \in P(BB(I,C))$ and let s be (one of) the child(ren) of p for which $s \in BB(I,C) \subseteq B(I,C)$. Since p is by its definition not an element of $BB(I,C)$, we only have to prove that $p \in B(I,C)$. So let us assume *ex absurdo* that $p \notin B(I,C)$, implying the existence of a path from some $i \in I$ to p that is not blocked by C. If $s \notin C$, then the concatenation of the path from i to p with the node s, yields a path from $i \in I$ to $s \in B(I,C)$ that is not blocked by C, a contradiction. If $s \in C$, then $s \in BB(I,C)$ implies the existence of some $k \in B(I,C)$ such that $k \sqsubset s$ and $k \notin C$. Since $k \sqsubset s$, we can now construct a directed path from k to s, yielding a reverse directed path from s to k. If we concatenate the path from i to p with this reverse directed path from s to k, we obtain a path from $i \in I$ to $k \in B(I,C)$ that is not blocked by C, a contradiction.

To verify that $D(BB(I,C)) \cap C = \emptyset$, assume *ex absurdo* that $D(BB(I,C)) \cap C \neq \emptyset$, and choose any $s \in D(BB(I,C)) \cap C$. Then $s \in C$ implies that $s \in B(I,C)$, because any path from any $i \in I$ to s is blocked by $s \in C$. On the other hand, $s \in D(BB(I,C))$ implies the existence of some $r \in BB(I,C)$ such that $r \sqsubseteq s$. Since $r \in BB(I,C)$ in turn implies the existence of some $k \in B(I,C)$ such that $k \sqsubseteq r \sqsubseteq s$ and $k \notin C$, we obtain from $s \in B(I,C)$ that $s \in BB(I,C)$, contradicting $s \in D(BB(I,C))$.

$I \setminus C \subseteq N(BB(I,C)) \setminus C$ follows trivially from $I \setminus C \subseteq N(BB(I,C))$, so it suffices to prove the latter statement. Consider any $i \in I \setminus C$, implying that $i \notin C$. Since $i \in G = N(BB(I,C)) \cup P(BB(I,C)) \cup BB(I,C) \cup D(BB(I,C))$, it suffices to prove that $i \notin P(BB(I,C))$, $i \notin BB(I,C)$ and $i \notin D(BB(I,C))$. First, i cannot be an element of $P(BB(I,C))$ because $P(BB(I,C)) \subseteq C$ then yields a contradiction with $i \notin C$. Second, i is not an element of $BB(I,C)$ because then $i \in B(I,C)$, implying that the trivial path from i to i should be blocked by C, again yielding a contradiction with $i \notin C$. Third, suppose *ex absurdo* that $i \in D(BB(I,C))$. We have shown in the proof of $D(BB(I,C)) \cap C = \emptyset$ that this would imply the existence of some $k \in B(I,C)$ such that $k \sqsubseteq i$ and $k \notin C$. Since $k \sqsubseteq i$, we can now construct a directed path from k to i, yielding a reverse directed path from i to k that is not blocked by C [because it is a reverse directed path, and because neither i nor k belong to C], contradicting $k \in B(I,C)$.

Finally, let us prove that $BB(I,C)$ is a closed subset of G. Fix $s, t \in BB(I,C)$ and $r \in G$ such that $s \sqsubseteq r \sqsubseteq t$ and assume *ex absurdo* that $r \notin BB(I,C)$. Since $s \in BB(I,C)$, we can infer the existence of some $k \in B(I,C)$ such that $k \notin C$ and $k \sqsubseteq s \sqsubseteq r$, implying that $r \notin B(I,C)$ because $r \in B(I,C)$ would imply $r \in BB(I,C)$, contradicting $r \notin BB(I,C)$. So we now know that there is a path from some $i \in I$ to r that is not blocked by C. Since $k \sqsubseteq r$, we also have a directed path from k to r and thus a reverse directed path from r to k. Concatenating the path both from i to r and the reverse directed path from r to k, we obtain a path from $i \in I$ to k, which should be blocked by C since $k \in B(I,C)$. The only way for this to be possible is if $r \notin C$ and $D(r) \cap C = \emptyset$. However, then the path from $i \in I$ to t, formed by concatenating the path from i to r and a directed path from r to t, is not blocked by C, contradicting $t \in BB(I,C) \subseteq B(I,C)$. $\square$

Proof of Theorem 16. Fix $I, O, C \subseteq G$ such that $\mathrm{AD}(I, O|C)$, $x_C \in \mathcal{X}_C$, $g \in \mathcal{G}(\mathcal{X}_I)_{>0}$ and $f \in \mathcal{G}(\mathcal{X}_O)$. Since we know from Proposition 23 that $BB(I,C)$ is closed and that $D(BB(I,C)) \cap C = \emptyset$, it follows from Lemma 20(ii) that

$$C = \big(C \cap N\big(BB(I,C)\big)\big) \cup \big(C \cap P\big(BB(I,C)\big)\big) \cup \big(C \cap BB(I,C)\big) = C_1 \cup C_2 \cup C_3,$$

where $C_1 := C \cap N(BB(I,C))$, $C_2 := C \cap P(BB(I,C))$ and $C_3 := C \cap BB(I,C)$ are disjoint. Now let $g' := g(\cdot, x_{I\cap C})\mathbb{I}_{\{x_{C_1}\}}$ and $f' := \mathbb{I}_{\{x_{C_3}\}} f(\cdot, x_{O\cap C})$ and notice that $C_2 = P(BB(I,C))$ because $P(BB(I,C)) \subseteq C$ due to Proposition 23. Then $g\mathbb{I}_{\{x_C\}} f = g(\cdot, x_{I\cap C})\mathbb{I}_{\{x_C\}} f(\cdot, x_{O\cap C}) = g'\mathbb{I}_{\{x_{P(BB(I,C))}\}} f'$ and $\mathbb{I}_{\{x_C\}} f = \mathbb{I}_{\{x_C\}} f(\cdot, x_{O\cap C}) = \mathbb{I}_{\{x_{C_1}\}}\mathbb{I}_{\{x_{P(BB(I,C))}\}} f'$. Since, again, $BB(I,C)$ is a closed subset of G due to Proposition 23, the proof will follow directly from Theorem 10 if we can show that $\mathbb{I}_{\{x_{C_1}\}} \in \mathcal{G}(\mathcal{X}_{N(BB(I,C))})_{>0}$, $g' \in \mathcal{G}(\mathcal{X}_{N(BB(I,C))})_{>0}$ and $f' \in \mathcal{G}(\mathcal{X}_{BB(I,C)})$.

That $\mathbb{I}_{\{x_{C_1}\}} \in \mathcal{G}(\mathcal{X}_{N(BB(I,C))})_{>0}$ is trivial, and $g' \in \mathcal{G}(\mathcal{X}_{N(BB(I,C))})_{>0}$ because $I \setminus C \subseteq N(BB(I,C)) \setminus C$ due to Proposition 23. Finally, we know that $O \subseteq B(I,C)$ due to $\mathrm{AD}(I,O|C)$, implying that $O \setminus C \subseteq B(I,C) \setminus C \subseteq BB(I,C)$, where the last inclusion follows from Proposition 23. Therefore, $f(\cdot, x_{O\cap I})$ and also f' are elements of $\mathcal{G}(\mathcal{X}_{BB(I,C)})$. $\square$

Proof of Corollary 17. Fix $I, O, C \subseteq G$ such that $\mathrm{AD}(I,O|C)$. As shown in the proof of Theorem 16, $O \setminus C \subseteq BB(I,C)$ and we know from Proposition 23 that $I \setminus C \subseteq N(BB(I,C))$. Therefore, $I \setminus C$ and $O \setminus C$ are disjoint and the subset-irrelevance statement $\mathrm{SIR}(I,O|C)$ is thus well defined and by definition equivalent to $\mathrm{SIR}(I \setminus C, O \setminus C|C)$.

To prove that $\mathrm{SIR}(I \setminus C, O \setminus C|C)$, we choose $x_C \in \mathcal{X}_C$, non-empty $A_{I\setminus C} \subseteq \mathcal{X}_{I\setminus C}$ and $f \in \mathcal{G}(\mathcal{X}_{O\setminus C})$. Since $\mathbb{I}_{A_{I\setminus C}}$ is an element of $\mathcal{G}(\mathcal{X}_{I\setminus C})_{>0}$, we know from Eq. (1) and Theorem 16 that

$$f \in \mathrm{marg}_{O\setminus C}\big(\mathcal{D}_G^{\mathrm{irr}} \rfloor \{x_C\} \times A_{I\setminus C}\big) \quad \Leftrightarrow \quad \mathbb{I}_{A_{I\setminus C}}\mathbb{I}_{\{x_C\}} f \in \mathcal{D}_G^{\mathrm{irr}} \quad \Leftrightarrow \quad \mathbb{I}_{\{x_C\}} f \in \mathcal{D}_G^{\mathrm{irr}} \quad \Leftrightarrow \quad f \in \mathrm{marg}_{O\setminus C}\big(\mathcal{D}_G^{\mathrm{irr}} \rfloor x_C\big),$$

concluding the proof for the first part of this corollary. The last part is a direct consequence of Theorem 15. $\square$

References

[1] A. Benavoli, M. Zaffalon, E. Miranda, Robust filtering through coherent lower previsions, IEEE Trans. Autom. Control 56 (7) (2011) 1567–1581.
[2] G. Corani, A. Antonucci, M. Zaffalon, Bayesian networks with imprecise probabilities: theory and application to classification, in: D.E. Holmes, L.C. Jain (Eds.), Data Mining: Foundations and Intelligent Paradigms, 2012, pp. 49–93, Ch. 4.
[3] I. Couso, S. Moral, Sets of desirable gambles: conditioning, representation, and precise probabilities, Int. J. Approx. Reason. 52 (7) (2011) 1034–1055.
[4] F.G. Cozman, Credal networks, Artif. Intell. 120 (2000) 199–233.
[5] F.G. Cozman, Separation properties of sets of probability measures, in: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, 2000, pp. 107–114.
[6] F.G. Cozman, Graphical models for imprecise probabilities, Int. J. Approx. Reason. 39 (2–3) (2005) 167–184.
[7] F.G. Cozman, P. Walley, Graphoid properties of epistemic irrelevance and independence, Ann. Math. Artif. Intell. 45 (1–2) (2005) 173–195.
[8] A.P. Dawid, Conditional independence, in: Encyclopedia of Statistical Sciences, vol. 2, Wiley, New York, 1998, pp. 146–155.
[9] J. De Bock, G. de Cooman, An efficient algorithm for estimating state sequences in imprecise hidden Markov models, J. Artif. Intell. Res. 50 (2014) 189–233.
[10] J. De Bock, G. de Cooman, Credal networks under epistemic irrelevance using sets of desirable gambles, in: Proceedings of 8th International Symposium on Imprecise Probabilities: Theories and Applications, 2013, pp. 99–108.
[11] J. De Bock, G. de Cooman, Allowing for probability zero in credal networks under epistemic irrelevance, in: Proceedings of the 8th International Symposium on Imprecise Probabilities: Theories and Applications, 2013, pp. 109–118.
[12] C.P. de Campos, F.G. Cozman, Computing lower and upper expectations under epistemic independence, Int. J. Approx. Reason. 44 (2007) 244–260.
[13] G. de Cooman, Belief models: an order-theoretic investigation, Ann. Math. Artif. Intell. 45 (2005) 5–34.
[14] G. de Cooman, F. Hermans, A. Antonucci, M. Zaffalon, Epistemic irrelevance in credal nets: the case of imprecise Markov trees, Int. J. Approx. Reason. 51 (2010) 1029–1052.
[15] G. de Cooman, E. Miranda, M. Zaffalon, Independent natural extension, Artif. Intell. 175 (12–13) (2011) 1911–1950.
[16] G. de Cooman, E. Miranda, Irrelevant and independent natural extension for sets of desirable gambles, J. Artif. Intell. Res. 45 (2012) 601–640.
[17] G. de Cooman, E. Quaeghebeur, Exchangeability and sets of desirable gambles, Int. J. Approx. Reason. 53 (3) (2012) 363–395, Special issue in honour of Henry E. Kyburg, Jr.
[18] D. Geiger, T. Verma, J. Pearl, Identifying independence in Bayesian networks, Networks 20 (1990) 507–534.
[19] D. Geiger, T. Verma, J. Pearl, d-separation: from theorems to algorithms, in: Proceedings of the 5th Conference on Uncertainty in Artificial Intelligence, 1990, pp. 139–148.
[20] D. Koller, N. Friedman, Probabilistic Graphical Models: Principles and Techniques, MIT Press, 2009.
[21] I. Levi, The Enterprise of Knowledge, MIT Press, London, 1980.
[22] D.D. Mauá, C.P. de Campos, A. Benavoli, A. Antonucci, On the complexity of strong and epistemic credal networks, in: Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence, 2013, pp. 391–400.
[23] S. Moral, Epistemic irrelevance on sets of desirable gambles, Ann. Math. Artif. Intell. 45 (2005) 197–214.
[24] E. Miranda, A survey of the theory of coherent lower previsions, Int. J. Approx. Reason. 48 (2) (2008) 628–658.
[25] J. Pearl, T. Verma, The logic of representing dependencies by directed graphs, in: Proceedings of the AAAI'87 Conference, AAAI/MIT Press, 1987, pp. 347–379.
[26] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.
[27] A. Piatti, A. Antonucci, M. Zaffalon, Building knowledge-based expert systems: a tutorial, in: A.R. Baswell (Ed.), Adv. Math. Res., vol. 11, Nova Science Publishers, New York, 2010.
[28] O. Pourret, P. Naim, B. Marcot, Bayesian Networks: A Practical Guide to Applications, Wiley, Chichester, 2008.
[29] E. Quaeghebeur, The CONEstrip algorithm, Adv. Intell. Sys. Comput. 190 (2013) 45–54.
[30] E. Quaeghebeur, Desirability, in: F. Coolen, T. Augustin, G. de Cooman, M.C.M. Troffaes (Eds.), An Introduction to Imprecise Probabilities, John Wiley & Sons, 2014.
[31] E. Quaeghebeur, G. de Cooman, F. Hermans, Accept & reject statement-based uncertainty models, submitted for publication, arXiv:1208.4462v3 [math.PR].
[32] M.C.M. Troffaes, G. de Cooman, Lower Previsions, Wiley, 2014.
[33] B. Vantaggi, Graphical models for conditional independence structures, in: Proceedings of the 2nd International Symposium on Imprecise Probabilities and Their Applications, 2001, pp. 332–341.

[34] B. Vantaggi, The L-separation criterion for description of cs-independence models, Int. J. Approx. Reason. 29 (2002) 291–316.
[35] B. Vantaggi, Graphical representation of asymmetric graphoid structures, in: Proceedings of the 3th International Symposium on Imprecise Probabilities and Their Applications, 2003, pp. 560–574.
[36] B. Vantaggi, Conditional independence structures and graphical models, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 11 (5) (2003) 545–571.
[37] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.
[38] P. Walley, Towards a unified theory of imprecise probability, Int. J. Approx. Reason. 24 (2000) 125–148.
[39] P.M. Williams, Notes on conditional previsions, Technical report, School of Mathematical and Physical Science, University of Sussex, UK, 1975, revised journal version: [40].
[40] P.M. Williams, Notes on conditional previsions, Int. J. Approx. Reason. 44 (2007) 366–383, revised journal version of [39].
[41] N. Wilson, S. Moral, A logical view of probability, in: ECAI, 1994, pp. 386–390.
[42] M. Zaffalon, E. Miranda, Notes on desirability and conditional lower previsions, Ann. Math. Artif. Intell. 66 (3–4) (2010) 251–309.