International Journal of Approximate Reasoning 131 (2021) 56–79

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Graphoid properties of concepts of independence for sets of probabilities

Fabio Gagliardi Cozman

Escola Politécnica, Universidade de São Paulo, Brazil

ARTICLE INFO

Article history:
Received 2 April 2020
Received in revised form 25 November 2020
Accepted 14 December 2020
Available online 28 December 2020

Keywords:
Graphoid properties
Epistemic irrelevance and independence
Credal sets
Lexicographic probabilities
Sets of desirable gambles

ABSTRACT

We examine several concepts of independence associated with (1) credal sets, understood as sets of probability measures, (2) sets of full conditional probabilities, (3) sets of lexicographic probabilities, and (4) sets of desirable gambles. Concepts of independence are evaluated with respect to the graphoid properties they satisfy, as these properties capture important abstract features of "independence". We emphasize the analysis of sets of probability measures as this is a popular formalism, looking at versions of epistemic, confirmational, and type-5 independence that are based on regular conditioning, as well as complete and strong independence. We then examine analogous concepts of independence for sets of full conditional probabilities, sets of lexicographic probabilities, and sets of desirable gambles.

1. Introduction

In this paper we examine several concepts of conditional independence for random variables, all of which associated with formalisms that resort to sets of probabilities of various kinds. We consider sets of probability measures, sets of full conditional probabilities, sets of lexicographic probabilities, and sets of desirable gambles. These formalisms share the desire to represent indeterminacy and imprecision in probability values [1,37,55], and they offer various distinct ways to handle conditioning. One notable feature of some of these formalisms is that they do allow conditioning to proceed even when the conditioning event may receive zero probability [15,51,53].

We evaluate existing and novel concepts of (conditional) independence by their graphoid properties. These properties try to capture the intuitive content behind concepts of independence, and there has been interest in verifying which properties are satisfied by which concepts [17,43,49]. In fact graphoid properties have already been studied in connection with sets of probabilities [13,15], but many questions remain open.

Results in this paper are useful both in comparing the concepts of independence with each other and in comparing the formalisms themselves with each other. Moreover, results in this paper should add to our understanding about the adequacy of graphoid properties for the purposes they were designed.

In Section 2 we quickly review needed terminology. Then in Section 3 we focus on concepts of independence that apply to sets of probability measures. Given that such sets are often employed in practical settings, we provide a detailed analysis of their associated concepts of independence: in Section 3 we examine regular confirmational, regular epistemic, regular type-5, and regular type-5 epistemic irrelevance/independence, as well as complete and strong independence. In Section 4 we examine a number of formalisms where conditional probability is defined so that a conditioning event can have zero

E-mail address: fgcozman@usp.br.

https://doi.org/10.1016/j.ijar.2020.12.015
0888-613X/

Table 1
Failure of Intersection: three binary variables W, X, Y (take Z to be a constant); $(X \perp\!\!\!\perp W \,|\, (Y, Z))$ and $(X \perp\!\!\!\perp Y \,|\, (W, Z))$ hold but $(X \perp\!\!\!\perp (W, Y) \,|\, Z)$ fails.

$\mathbb{P}$	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	1/6	0	0	1/3
x_1	1/3	0	0	1/6

probability: sets of full conditional probabilities, sets of lexicographic probabilities, and sets of desirable gambles. Some final comments are collected in Section 5.

2. Basics: probability measures and graphoid properties

We assume throughout that the possibility space Ω is finite, so there are no issues of measurability to complicate matters. We take every subset of Ω to be an event. If an event is nonempty, it is a *possible* event. We take the topology induced by Euclidean distance throughout. A random variable, or simply a variable, is a function from Ω to the real numbers.

For any event H, the indicator function of H is denoted by $\mathbb{I}_H$. The intersection of events G and H is written as $G \cap H$, but also as G, H when it appears in a conditional probability.

Some notational conventions will be adopted throughout [13]. We use W, X, Y and Z to denote random variables. We always use w to denote a possible value of W, x to denote a possible value of X, y to denote a possible value of Y, z to denote a possible value of Z. And $\{x\}$ denotes the nonempty event $\{\omega \in \Omega : X(\omega) = x\}$; likewise for $\{w\}$, $\{y\}$ and $\{z\}$. We write event $\{\omega \in \Omega : X(\omega) = x\} \cap \{\omega \in \Omega : Y(\omega) = y\}$ as $\{X = x, Y = y\}$ and, when it appears in a conditional probability, we write simply x, y as usual in the literature (and likewise for similar events).

The letter A will always denote nonempty events in the algebra generated by random variable X (that is, A is a set of values of X). Similarly, the letter B will always denote nonempty events in the algebra generated by Y.

The letter f will always denote a function of X and the letter g will always denote a function of Y.

A probability measure $\mathbb{P}$ is an additive set-function that assigns a non-negative real number to each event, and such that Ω has probability 1. When we deal with probability measures, we have that the conditional probability of event G given event H, denoted by $\mathbb{P}(G|H)$, is only defined when $\mathbb{P}(H) > 0$; if so, then $\mathbb{P}(G|H) = \mathbb{P}(G \cap H)/\mathbb{P}(H)$. If a conditioning event H is equal to Ω, we simply write $\mathbb{P}(G)$ instead of $\mathbb{P}(G|\Omega)$. The expectation of $f(X)$ with respect to $\mathbb{P}$ is denoted by $\mathbb{E}_{\mathbb{P}}[f(X)]$, while the expectation of $f(X)$ given H with respect to $\mathbb{P}$ is denoted by $\mathbb{E}_{\mathbb{P}}[f(X)|H]$ whenever $\mathbb{P}(H) > 0$.

Conditional stochastic independence of random variables X and Y given random variable Z, with respect to a probability measure $\mathbb{P}$, is obtained when $\mathbb{P}(x, y|z) = \mathbb{P}(x|z)\,\mathbb{P}(y|z)$ whenever $\mathbb{P}(z) > 0$. Equivalently, conditional stochastic independence is obtained when we have $\mathbb{P}(x|y, z) = \mathbb{P}(x|z)$ whenever $\mathbb{P}(y, z) > 0$. Often we just write "independence" to mean both conditional and unconditional independence.

We now summarize a few facts about the *graphoid properties* [17,43,49]. These properties apply to an abstract ternary relation denoted by $(\cdot \perp\!\!\!\perp \cdot \,|\, \cdot)$; the idea is that $(X \perp\!\!\!\perp Y \,|\, Z)$ indicates "independence" (in some specified sense) of X and Y given Z. The properties are:

Symmetry: $(X \perp\!\!\!\perp Y \,|\, Z) \Rightarrow (Y \perp\!\!\!\perp X \,|\, Z)$
Redundancy: $(X \perp\!\!\!\perp Y \,|\, X)$
Decomposition: $(X \perp\!\!\!\perp (W, Y) \,|\, Z) \Rightarrow (X \perp\!\!\!\perp Y \,|\, Z)$
Weak union: $(X \perp\!\!\!\perp (W, Y) \,|\, Z) \Rightarrow (X \perp\!\!\!\perp Y \,|\, (W, Z))$
Contraction: $(X \perp\!\!\!\perp Y \,|\, Z) \wedge (X \perp\!\!\!\perp W \,|\, (Y, Z)) \Rightarrow (X \perp\!\!\!\perp (W, Y) \,|\, Z)$
Intersection $(X \perp\!\!\!\perp W \,|\, (Y, Z))$ & $(X \perp\!\!\!\perp Y \,|\, (W, Z)) \Rightarrow (X \perp\!\!\!\perp (W, Y) \,|\, Z)$.

The *semi-graphoid properties* are the first five properties above; that is, all graphoid properties except Intersection.

Conditional stochastic independence satisfies all graphoid properties except Intersection (that is, conditional stochastic independence satisfies all semi-graphoid properties). An example that depicts failure of Intersection for conditional stochastic independence is shown in Table 1. The fact that some events have probability zero is crucial in this example: if the probability distribution is strictly positive, then Intersection holds for conditional stochastic independence [43]. Hence conditional stochastic independence satisfies all graphoid properties when all probabilities are positive.

3. Credal sets

A set of probability measures is referred to as a *credal set*. We do not assume credal sets to be convex; we do not assume credal sets to be closed either. A credal set can represent imprecision about probability values, perhaps due to incomplete elicitation, perhaps due to disagreement amongst decision-makers [1,27,37]. In this section we examine concepts of independence for credal sets and their graphoid properties.

3.1. Conditioning and independence for credal sets

Denote by $\mathbb{K}(X)$ a set of probability distributions for variable X (that is, a credal set). Given a function $f(X)$, its lower and upper expectations are, respectively $\underline{\mathbb{E}}[f(X)] = \inf_{\mathbb{P}\in\mathbb{K}} \mathbb{E}_{\mathbb{P}}[f(X)]$ and $\overline{\mathbb{E}}[f(X)] = \sup_{\mathbb{P}\in\mathbb{K}} \mathbb{E}_{\mathbb{P}}[f(X)]$. Similarly, given an event H, its lower and upper probabilities are respectively $\underline{\mathbb{P}}(H) = \inf_{\mathbb{P}\in\mathbb{K}} \mathbb{P}(H)$ and $\overline{\mathbb{P}}(H) = \sup_{\mathbb{P}\in\mathbb{K}} \mathbb{P}(H)$.

Given a credal set $\mathbb{K}(X)$, a possible way to define a conditional credal set is:

$$\mathbb{K}^{\triangleright}(X|H) = \{\mathbb{P}(\cdot|H) : \mathbb{P} \in \mathbb{K}(X)\} \quad \text{whenever } \underline{\mathbb{P}}(H) > 0,$$

with $\mathbb{K}^{\triangleright}(X|H)$ undefined whenever $\underline{\mathbb{P}}(H) = 0$ [27]. A different concept of conditional credal set focuses on those probability measures that assign positive probability to H:

$$\mathbb{K}^{>}(X|H) = \{\mathbb{P}(\cdot|H) : \mathbb{P} \in \mathbb{K}(X) \text{ and } \mathbb{P}(H) > 0\} \quad \text{whenever } \overline{\mathbb{P}}(H) > 0,$$

with $\mathbb{K}^{>}(X|H)$ undefined whenever $\overline{\mathbb{P}}(H) = 0$ [57,58]. We refer to this second conditioning strategy as *regular conditioning*, as it is closely related to the concept of regular extension [55, Appendix J]. Note that when $\mathbb{K}(X)$ is convex, the set $\mathbb{K}^{>}(X|H)$ is convex whenever it is defined, but $\mathbb{K}^{>}(X|H)$ may be open even when $\mathbb{K}(X)$ is closed [14].

Define $\underline{\mathbb{E}}^{>}[f(X)|H] = \inf_{\mathbb{P}(\cdot|H)\in\mathbb{K}^{>}(X|H)} \mathbb{E}_{\mathbb{P}}[f(X)|H]$ and $\overline{\mathbb{E}}^{>}[f(X)|H] = \sup_{\mathbb{P}(\cdot|H)\in\mathbb{K}^{>}(X|H)} \mathbb{E}_{\mathbb{P}}[f(X)|H]$, whenever $\overline{\mathbb{P}}(H) > 0$. It is known [54] that, whenever $\overline{\mathbb{P}}(H) > 0$,

$$\underline{\mathbb{E}}^{>}[f(X)|H] = \sup\left(\alpha : \underline{\mathbb{E}}[(f(X) - \alpha)\mathbb{I}_H] \geq 0\right).$$

This can be slightly generalized to the following result, used later (this result mimics Lemma 1 of Ref. [9]):

Theorem 1. *If* $\overline{\mathbb{P}}(G, H) > 0$, *then*

$$\underline{\mathbb{E}}^{>}[f(X)|G, H] = \sup\left(\alpha : \underline{\mathbb{E}}^{>}[(f(X) - \alpha)\mathbb{I}_G|H] \geq 0\right).$$

Proof. We have:

$$\begin{aligned}
\underline{\mathbb{E}}^{>}[f(X)|G, H] &= \inf_{\mathbb{P}\in\mathbb{K}:\mathbb{P}(G,H)>0} \frac{\mathbb{E}_{\mathbb{P}}[f(X)\mathbb{I}_G\mathbb{I}_H]}{\mathbb{P}(G, H)} \\
&= \inf_{\mathbb{P}\in\mathbb{K}:\mathbb{P}(H)>0,\mathbb{P}(G|H)>0} \frac{\mathbb{E}_{\mathbb{P}}[f(X)\mathbb{I}_G|H]\,\mathbb{P}(H)}{\mathbb{P}(G|H)\,\mathbb{P}(H)} \\
&= \inf_{\mathbb{P}(\cdot|H)\in\mathcal{A}} \frac{\mathbb{E}_{\mathbb{P}}[f(X)\mathbb{I}_G|H]}{\mathbb{P}(G|H)} \\
&= \sup\left(\alpha : \left(\inf_{\mathbb{P}(\cdot|H)\in\mathcal{A}} \frac{\mathbb{E}_{\mathbb{P}}[f(X)\mathbb{I}_G|H]}{\mathbb{P}(G|H)}\right) \geq \alpha\right) \\
&= \sup\left(\alpha : \left(\inf_{\mathbb{P}(\cdot|H)\in\mathcal{A}} \frac{\mathbb{E}_{\mathbb{P}}[f(X)\mathbb{I}_G|H] - \alpha\mathbb{P}(G|H)}{\mathbb{P}(G|H)}\right) \geq 0\right) \\
&= \sup\left(\alpha : \left(\inf_{\mathbb{P}(\cdot|H)\in\mathcal{A}} \mathbb{E}_{\mathbb{P}}[(f(X) - \alpha)\mathbb{I}_G|H]\right) \geq 0\right) \\
&= \sup\left(\alpha : \underline{\mathbb{E}}^{>}[(f(X) - \alpha)\mathbb{I}_G|H] \geq 0\right),
\end{aligned}$$

where $\mathcal{A}$ is the set obtained by collecting all measures from $\mathbb{K}^{>}(\cdot|H)$ such that $\mathbb{P}(G|H) > 0$. □

We now present a few concepts of independence that directly depend on conditioning. A more detailed historical review can be found in a previous survey [11].

Levi says that Y is *confirmationally irrelevant* to X when beliefs about X are not affected by observation of Y. We might take this definition of confirmational irrelevance to mean[1]

$$\mathbb{K}^{\triangleright}(X|y, z) = \mathbb{K}^{\triangleright}(X|z) \quad \text{for all } (y, z) \text{ such that } \underline{\mathbb{P}}(Y = y, Z = z) > 0. \tag{1}$$

We are surely flirting with disaster with Expression (1), because it is not difficult to have a variable Z with a nontrivial $\mathbb{K}(Z)$ for which every value of Z has zero lower probability; for any such variable the credal sets in Expression (1) are undefined, and confirmational irrelevance obtains trivially!

[1] In Ref. [11] the concept of confirmational irrelevance is defined using all events B specified by Y in conditioning, as opposed just to events $\{Y = y\}$. In this paper we differentiate both conditions; conditioning on events B is used later to define type-5 irrelevance. Similarly, the concept of epistemic irrelevance in Ref. [11] is akin to type-5 epistemic irrelevance as defined in this paper.

A more reasonable strategy to define confirmational irrelevance employs regular conditioning: say that Y is *regular-confirmationally irrelevant* to X given Z when

$$\mathbb{K}^{>}(X|y,z) = \mathbb{K}^{>}(X|z) \quad \text{for all } (y,z) \text{ such that } \overline{\mathbb{P}}(Y=y, Z=z) > 0. \tag{2}$$

Walley's concept of *epistemic irrelevance* is similar to Levi's confirmational irrelevance: Y is epistemically irrelevant to X when $\underline{\mathbb{E}}[f(X)|y] = \underline{\mathbb{E}}[f(X)]$ for every function $f(X)$ and every possible y [55, Chapter 9]. Say that Y is *regular-epistemically irrelevant* to X given Z when, for every function $f(X)$,

$$\underline{\mathbb{E}}^{>}[f(X)|y,z] = \underline{\mathbb{E}}^{>}[f(X)|z] \quad \text{for all } (y,z) \text{ such that } \overline{\mathbb{P}}(Y=y, Z=z) > 0. \tag{3}$$

Both regular-confirmational irrelevance and regular-epistemic irrelevance fail Symmetry. Take the following example by Couso et al. [7, Example 3]:

Example 1. Consider two binary variables X and Y with $\mathbb{P}(x_0) \in [1/2, 4/5]$ and $\mathbb{P}\left(y_i|x_j\right) \geq 3/10$ for all $i \in \{0,1\}$ and all $j \in \{0,1\}$; then X is regular-confirmationally and regular-epistemically irrelevant to Y, but $\mathbb{P}(x_0|y_0) \in [3/10, 28/31]$, thus Y is not regular-confirmationally nor regular-epistemically irrelevant to X. □

Walley's clever response to failure of Symmetry, borrowed from the work of Keynes [31], was to "symmetrize" irrelevance to obtain independence. Inspired by that, say that X and Y are *regular-epistemically independent* given Z when X is regular-epistemically irrelevant to Y given Z, and Y is regular-epistemically irrelevant to X given Z [55].

We can apply Walley's idea to other settings: for instance, say that X and Y are *regular-confirmationally independent* given Z when both X is regular-confirmationally irrelevant to Y given Z, and Y is regular-confirmationally irrelevant to X given Z.

Yet another concept of independence has been proposed for credal sets by de Campos and Moral [21]: they say Y is type-5 irrelevant to X if $\mathbb{K}^{>}(X|B) = \mathbb{K}(X)$ whenever $\overline{\mathbb{P}}(B) > 0$, where B is any event in the algebra generated by Y. Accordingly, say that Y is *type-5 irrelevant* to X given Z if

$$\mathbb{K}^{>}(X|B,\{z\}) = \mathbb{K}^{>}(X|z) \quad \text{for all } (B,\{z\}) \text{ such that } \overline{\mathbb{P}}(B, \{Z=z\}) > 0,$$

where B denotes an event in the algebra generated by Y. Type-5 irrelevance is an strengthened version of confirmational irrelevance. Now we can similarly strengthen epistemic irrelevance: say that Y is *type-5 epistemically irrelevant* to X given Z if

$$\underline{\mathbb{E}}^{>}[f(X)|B,\{z\}] = \underline{\mathbb{E}}^{>}[f(X)|z] \quad \text{for all } (B,\{z\}) \text{ such that } \overline{\mathbb{P}}(B, \{Z=z\}) > 0.$$

We can symmetrize type-5 irrelevance and type-5 epistemic irrelevance to get corresponding concepts of independence.

3.2. Graphoid properties of regular-epistemic irrelevance and independence

We start with the graphoid properties of regular-epistemic irrelevance. As noted already, this concept of irrelevance does not satisfy Symmetry, hence we can contemplate several versions of the graphoid properties. We have:

Theorem 2. *If $(Y \text{ IR } X \mid Z)$ denotes regular-epistemic irrelevance of Y to X given Z, then:*
- *$(X \text{ IR } Y \mid X)$ and $(Y \text{ IR } X \mid X)$ ("direct" and "reverse" forms of Redundancy);*
- *If $(X \text{ IR } (W,Y) \mid Z)$, then $(X \text{ IR } Y \mid Z)$ (a "direct" form of Decomposition);*
- *If $(X \text{ IR } (W,Y) \mid Z)$, then $(X \text{ IR } Y \mid (W,Z))$ (a "direct" form of Weak Union);*
- *If $(Y \text{ IR } X \mid Z)$ and $(W \text{ IR } X \mid (Y,Z))$, then $((W,Y) \text{ IR } X \mid Z)$ (a "reverse" form of Contraction).*

Proof. Redundancy: we do have, whenever $\overline{\mathbb{P}}(X=x_1, X=x_2) > 0$, that $\underline{\mathbb{E}}^{>}[g(Y)|X=x_1, X_2=x_2] = \underline{\mathbb{E}}^{>}[g(Y)|X=x_1]$ (true because if $x_1 = x_2$, then trivially $\{X=x_1, X=x_2\} = \{X=x_1\}$); also, whenever $\overline{\mathbb{P}}(x,y) > 0$, we have $\underline{\mathbb{E}}^{>}[f(X)|x,y] = f(x) = \underline{\mathbb{E}}^{>}[f(X)|x]$.

Decomposition: if X is regular-epistemically irrelevant to (W,Y) given Z, then $\underline{\mathbb{E}}^{>}[g(Y)|x,z] = \underline{\mathbb{E}}^{>}[g(Y)|z]$ whenever $\overline{\mathbb{P}}(X=x, Z=z) > 0$ as any $g(Y)$ is obviously a function of (W,Y).

Weak Union: by assumption we have $\underline{\mathbb{E}}^{>}[h(W,Y)|x,z] = \underline{\mathbb{E}}^{>}[h(W,Y)|z]$ whenever $\overline{\mathbb{P}}(X=x, Z=z) > 0$; hence, using Theorem 1, if $\overline{\mathbb{P}}(w,x,z) > 0$,

$$\begin{aligned}\underline{\mathbb{E}}^{>}[g(Y)|w,x,z] &= \sup\left(\alpha : \underline{\mathbb{E}}^{>}[(g(Y)-\alpha)\mathbb{I}_w(W)|x,z] \geq 0\right) \\ &= \sup\left(\alpha : \underline{\mathbb{E}}^{>}[(g(Y)-\alpha)\mathbb{I}_w(W)|z] \geq 0\right) \\ &= \underline{\mathbb{E}}^{>}[g(Y)|w,z].\end{aligned}$$

Table 2
Tables employed in Example 2.

$\mathbb{P}_1, \mathbb{P}_2$	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	$\frac{1-\alpha}{2}$	$\frac{1-\alpha}{2}$	0	0
x_1	$\alpha/2$	$\alpha/2$	0	0
$\mathbb{P}_3, \mathbb{P}_4$	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	0	0	$\frac{1-\alpha}{2}$	$\alpha/2$
x_1	0	0	$\alpha/2$	$\frac{1-\alpha}{2}$
$\mathbb{P}_5, \mathbb{P}_6$	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	0	0	$\frac{1-\alpha}{2}$	$\alpha/2$
x_1	0	0	$\frac{1-\alpha}{2}$	$\alpha/2$

Contraction: if equality $\underline{\mathbb{E}}^{>}[f(X)|w, y, z] = \underline{\mathbb{E}}^{>}[f(X)|y, z]$ holds whenever $\overline{\mathbb{P}}(W = w, Y = y, Z = z) > 0$ and $\underline{\mathbb{E}}^{>}[f(X)|y, z] = \underline{\mathbb{E}}^{>}[f(X)|z]$ whenever $\overline{\mathbb{P}}(Y = y, Z = z) > 0$, then, by transitivity, we have $\underline{\mathbb{E}}^{>}[f(X)|w, y, z] = \underline{\mathbb{E}}^{>}[f(X)|z]$ whenever $\overline{\mathbb{P}}(W = w, Y = y, Z = z) > 0$, as desired. □

The graphoid properties of epistemic irrelevance have been studied before within the framework of lower previsions, where conditioning is allowed on events of zero lower probability using Walley's natural extensions [15]. The latter study showed that direct and reverse Redundancy, direct Decomposition and reverse Contraction are satisfied by Walley's notion of epistemic irrelevance even if lower probabilities are zero. Thus Theorem 2 indicates a difference (that is, direct Weak Union) between regular-epistemic independence and Walley's concept of epistemic independence.

The graphoid properties of epistemic irrelevance have also been studied before under the assumption of positivity for lower probabilities [9]; indeed, appropriate positivity conditions guarantee that regular-epistemic independence satisfies reverse Decomposition and reverse Weak Union[2]:

Theorem 3. *Suppose* $\inf_{\mathbb{P}:\mathbb{P}(z)>0} \mathbb{P}(y|z) > 0$ *for all values* y *of* Y *whenever* $\overline{\mathbb{P}}(Z = z) > 0$*; then, using the same conventions of Theorem 2,*
- *If* $((W, Y) \text{ IR } X \mid Z)$, *then* $(Y \text{ IR } X \mid Z)$ *(a "reverse" form of Decomposition);*
- *If* $((W, Y) \text{ IR } X \mid Z)$, *then* $(W \text{ IR } X \mid (Y, Z))$ *(a "reverse" form of Weak Union).*

Proof. The proof of this theorem is in essence identical to the proof of Theorem 1 from Ref. [15]; thus we present a heavily abbreviated argument. Recall that f is a function of X, w is a value of W, y is a value of Y, and z is a value of Z.

Decomposition: Note first that $\underline{\mathbb{E}}^{>}[f|y, z] \geq \min_w \underline{\mathbb{E}}^{>}[f|w, y, z]$, so we have $\underline{\mathbb{E}}^{>}[f|y, z] \geq \min_w \underline{\mathbb{E}}^{>}[f|z] = \underline{\mathbb{E}}^{>}[f|z]$ given the assumed irrelevance of (W, Y) to X. Using this latter equality and the fact that $\underline{\mathbb{E}}^{>}[f|z] \geq \underline{\mathbb{E}}^{>}\left[\underline{\mathbb{E}}^{>}[f|Y, z]\,|z\right]$, we obtain $\underline{\mathbb{E}}^{>}[f|z] \geq \underline{\mathbb{E}}^{>}\left[\underline{\mathbb{E}}^{>}[f|Y, z]\,|z\right] \geq \underline{\mathbb{E}}^{>}\left[\underline{\mathbb{E}}^{>}[f|z]\,|z\right] = \underline{\mathbb{E}}^{>}[f|z]$, implying $\underline{\mathbb{E}}^{>}\left[\underline{\mathbb{E}}^{>}[f|Y, z]\,|z\right] = \underline{\mathbb{E}}^{>}[f|z]$. The assumed condition that $\mathbb{P}(Y = y|Z = z) > 0$ for each possible distribution then implies $\underline{\mathbb{E}}^{>}[f|y, z] = \underline{\mathbb{E}}^{>}[f|z]$ as desired.

Weak Union: By hypothesis we have $\underline{\mathbb{E}}^{>}[f|w, y, z] = \underline{\mathbb{E}}^{>}[f|z]$; using the reverse Decomposition property we have that $\underline{\mathbb{E}}^{>}[f|y, z] = \underline{\mathbb{E}}^{>}[f|z]$. Combining both equalities we obtain $\underline{\mathbb{E}}^{>}[f|w, y, z] = \underline{\mathbb{E}}^{>}[f|y, z]$ as desired. □

All other forms of graphoid properties that are not mentioned in Theorem 2 fail. For instance, consider Decomposition: other than the direct form in Theorem 2, there are three other possibilities:
- If $(X \text{ IR } (W, Y) \mid Z)$, then $(Y \text{ IR } X \mid Z)$;
- If $((W, Y) \text{ IR } X \mid Z)$, then $(X \text{ IR } Y \mid Z)$;
- If $((W, Y) \text{ IR } X \mid Z)$, then $(Y \text{ IR } X \mid Z)$.

If we take $W = Z = 1$, then the first two of these versions fail due to failure of Symmetry. The last version is the reverse Decomposition property that was proved with a positivity condition in Theorem 3; in general it can fail as the next example shows (the example is used later to show failure of other properties).

Example 2. Consider three binary variables W, X, and Y. For some selected $\beta \in (0, 1/2)$, build six distributions, using the values in Table 2 both for $\alpha = \beta$ and for $\alpha = 1 - \beta$. Take the credal set $\mathbb{K}$ that is the convex hull of these six joint distributions. We have that $\mathbb{K}^{>}(X|w, y)$ is the convex hull of $\{[\beta, 1 - \beta], [1 - \beta, \beta]\}$ for every possible w, y, where each vector denotes $[\mathbb{P}(x_0|w, y), \mathbb{P}(x_1|w, y)]$. Also, we have that $\mathbb{K}(X)$ is the convex hull of $\{[\beta, 1 - \beta], [1 - \beta, \beta]\}$, so (W, Y) is regular-epistemically irrelevant to X. The credal sets $\mathbb{K}(W, Y)$, $\mathbb{K}^{>}(W, Y|x_0)$ and $\mathbb{K}^{>}(W, Y|x_1)$ are identical: each of them is the convex hull of the three distributions $[1/2, 1/2, 0, 0]$, $[0, 0, \beta, 1 - \beta]$ and $[0, 0, 1 - \beta, \beta]$, where each vector

[2] These are the same conditions used to guarantee that Walley's epistemic independence satisfies direct Weak Union and reverse versions of Decomposition and Weak Union [15, Theorem 1].

Table 3
Tables employed in Example 3.

$\mathbb{P}_1$	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	$\alpha(1-\alpha)^2$	$\alpha(1-\alpha)/2$	$(1-\alpha)^3$	$\alpha(1-\alpha)/2$
x_1	$\alpha^2(1-\alpha)$	$\alpha^2/2$	$\alpha(1-\alpha)^2$	$\alpha^2/2$

$\mathbb{P}_2$	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	$\alpha/4$	$\alpha/4$	$(1-\alpha)/4$	$(1-\alpha)/4$
x_1	$(1-\alpha)/4$	$\alpha/4$	$\alpha/4$	$(1-\alpha)/4$

contains probabilities for values (w_0, y_0), (w_0, y_1), (w_1, y_0) and (w_1, y_1). Hence X and (W, Y) are regular-epistemically independent.

However, we have that $\mathbb{K}^>(X|w_0)$ is the convex hull of $\{[\beta, 1-\beta], [1-\beta, \beta]\}$ but $\mathbb{K}^>(X|w_1)$ is the singleton $\{[1/2, 1/2]\}$. Hence W is not regular-epistemically irrelevant to X, a failure of Decomposition. And Y is not regular-epistemically irrelevant to X given W, a failure of Weak Union. Note that $\underline{\mathbb{P}}(W = w_0) = \underline{\mathbb{P}}(W = w_1) = 0$, thus there is no contradiction with Theorem 3. □

Now consider Weak Union under regular-epistemic irrelevance. Other than the direct form in Theorem 2, there are three other possibilities:

- If $(X \text{ IR } (W, Y) \mid Z)$, then $(Y \text{ IR } X \mid (W, Z))$;
- If $((W, Y) \text{ IR } X \mid Z)$, then $(X \text{ IR } Y \mid (W, Z))$;
- If $((W, Y) \text{ IR } X \mid Z)$, then $(Y \text{ IR } X \mid (W, Z))$.

If we take $W = Z = 1$, then the first two of these versions fail due to failure of Symmetry. Example 2 shows that the third version fails.

Consider Contraction under regular-epistemic irrelevance. There are eight possible versions; a reverse version is proved in Theorem 2. As for the other seven versions, six of them lead to failure of Symmetry by taking either $W = Z = 1$ or $Y = Z = 1$.[3] The remaining version is a direct one:

- $(X \text{ IR } Y \mid Z) \wedge (X \text{ IR } W \mid (Y, Z)) \Rightarrow (X \text{ IR } (W, Y) \mid Z)$.

Ref. [9] presents a numerical example that violates this property, while Ref. [15] goes over several constructions that also lead to failure of direct Contraction. The next example is hopefully simpler to grasp.

Example 3. Consider three binary variables W, X and Y. For Table 3 top, $\mathbb{P}_1(x_0) = 1 - \alpha = \mathbb{P}_1(x_0|y)$ for each possible y and $\mathbb{P}_1(x_0|w, y) = 1 - \alpha$ for each possible (w, y) and $\mathbb{P}_1(y_0) = 1 - \alpha = \mathbb{P}_1(y_0|x)$ for each possible x; also $\mathbb{P}_1(w_0|x, y_0) = \mathbb{P}_1(w_0|y_0) = \alpha$ and $\mathbb{P}_1(w_0|x, y_1) = \mathbb{P}_1(w_0|y_1) = 1/2$ for each possible x. For Table 3 bottom, $\mathbb{P}_2(x_0) = \mathbb{P}_2(x_0|y) = 1/2$ for each possible y and $\mathbb{P}_2(y_0) = \mathbb{P}_2(y_0|x) = 1/2$ for each possible x; also $\mathbb{P}_2(x_0|w_0, y_0) = \alpha$ and $\mathbb{P}_2(x_0|w_1, y_0) = 1 - \alpha$ for each possible y, $\mathbb{P}_2(w_0|x, y_1) = \mathbb{P}_2(w_0|y_1) = \alpha$ for each possible x, and $\mathbb{P}_2(w_0|x_0, y_0) = \alpha$, $\mathbb{P}_2(w_0|x_1, y_0) = 1 - \alpha$, $\mathbb{P}_2(w_0|y_0) = 1/2$. Note also that $\mathbb{P}_1(w_0, y_0)$ is $\alpha(1-\alpha)$, while $\mathbb{P}_2(w_0, y_0)$ is $1/4$. And $\mathbb{P}_1(w_0, y_0|x_0)$ is $\alpha(1-\alpha)$, while $\mathbb{P}_2(w_0, y_0|x_0)$ is $\alpha/2$.

Select $\beta \in (0, 1/2)$. First build a joint probability distribution for (W, X, Y) by using $\mathbb{P}_1$ with $\alpha = \beta$. Then build a second joint probability distribution by using $\mathbb{P}_1$ with $\alpha = 1 - \beta$. Finally build a third/fourth joint probability distribution by using $\mathbb{P}_2$ with $\alpha = \beta$ and $\alpha = 1 - \beta$. And build the joint credal set that is the convex hull of these joint probability distributions. For this credal set we have that all of $\mathbb{P}(x_0)$, $\mathbb{P}(x_0|y)$ for each possible y, $\mathbb{P}(y_0)$, $\mathbb{P}(y_0|x)$ for each possible x, $\mathbb{P}(x_0|w, y)$ for each possible (w, y), $\mathbb{P}(w_0|x, y)$ for each possible (x, y), and $\mathbb{P}(w_0|y)$ for each possible y, vary within the interval $[\beta, 1-\beta]$. Thus we have that X is regular-epistemically irrelevant to Y, and X is regular-epistemically irrelevant to W given Y. However, X is not regular-epistemically irrelevant to (W, Y). To see this, consider that $\underline{\mathbb{P}}(w_0, y_0) = \beta(1-\beta)$ (the minimum between $\beta(1-\beta)$, obtained from the top table, and $1/4$, obtained from the bottom table, as $\beta(1-\beta) < 1/4$ for $\beta < 1/2$), and $\underline{\mathbb{P}}(w_0, y_0|x_0) = \beta/2$ (the minimum between $\beta(1-\beta)$, obtained from the top table, and $\beta/2$, obtained from the bottom table); as $\beta < 1/2$, we have $\underline{\mathbb{P}}(w_0, y_0|x_0) < \underline{\mathbb{P}}(w_0, y_0)$. □

Finally, consider Intersection under regular-epistemic irrelevance. There are eight versions; all of them fail *even* when all lower probabilities are larger than zero. Six of these versions of Intersection fail Symmetry by taking either $W = Z = 1$ or

[3] These six versions of Contraction are:
$(X \text{ IR } Y \mid Z) \wedge (X \text{ IR } W \mid (Y, Z)) \Rightarrow ((W, Y) \text{ IR } X \mid Z)$ (fails when $W = Z = 1$);
$(X \text{ IR } Y \mid Z) \wedge (W \text{ IR } X \mid (Y, Z)) \Rightarrow (X \text{ IR } (W, Y) \mid Z)$ (fails when $Y = Z = 1$);
$(X \text{ IR } Y \mid Z) \wedge (W \text{ IR } X \mid (Y, Z)) \Rightarrow ((W, Y) \text{ IR } X \mid Z)$ (fails when $W = Z = 1$);
$(Y \text{ IR } X \mid Z) \wedge (X \text{ IR } W \mid (Y, Z)) \Rightarrow (X \text{ IR } (W, Y) \mid Z)$ (fails when $W = Z = 1$);
$(Y \text{ IR } X \mid Z) \wedge (X \text{ IR } W \mid (Y, Z)) \Rightarrow ((W, Y) \text{ IR } X \mid Z)$ (fails when $Y = Z = 1$);
$(Y \text{ IR } X \mid Z) \wedge (W \text{ IR } X \mid (Y, Z)) \Rightarrow (X \text{ IR } (W, Y) \mid Z)$ (fails when $W = Z = 1$).

Table 4
Tables employed in Example 4.

$\mathbb{P}_1$	w_0y_0	w_0y_1	w_1y_0	w_1y_1
x_0	$(1-\alpha)^2/2$	$\alpha(1-\alpha)/2$	$\alpha(1-\alpha)/2$	$\alpha^2/2$
x_1	$\alpha(1-\alpha)/2$	$\alpha^2/2$	$(1-\alpha)^2/2$	$\alpha(1-\alpha)/2$

$\mathbb{P}_2$	w_0y_0	w_0y_1	w_1y_0	w_1y_1
x_0	$\alpha(1-\alpha)/2$	$\alpha^2/2$	$(1-\alpha)^2/2$	$\alpha(1-\alpha)/2$
x_1	$\alpha^2/2$	$\alpha(1-\alpha)/2$	$\alpha(1-\alpha)/2$	$(1-\alpha)^2/2$

$Y = Z = 1$.[4] As for the other two versions, Ref. [15] presents examples that display failure of Intersection even when all lower probabilities are positive; as they require significant computations, the next example is hopefully simpler to grasp.

Example 4. Consider three binary variables W, X and Y. For Table 4 top, $\mathbb{P}_1(x_0) = \mathbb{P}_1(x_0|y) = \mathbb{P}_1(w_0|y) = 1/2$ for each possible y, $\mathbb{P}_1(x_0|w_0) = \mathbb{P}_1(x_0|w_0, y) = 1-\alpha$ and $\mathbb{P}_1(x_0|w_1) = \mathbb{P}_1(x_0|w_1, y) = \alpha$ for each possible y, also $\mathbb{P}_1(y_0|w) = 1-\alpha$ for each possible w and $\mathbb{P}_1(y_0|w, x) = 1-\alpha$ for each possible (w, x), $\mathbb{P}_1(w_0|x_0, y) = 1-\alpha$ and $\mathbb{P}_1(w_0|x_1, y) = \alpha$ for each possible y. For Table 4 bottom, $\mathbb{P}_2(x_0) = \mathbb{P}_2(x_0|w) = \mathbb{P}_2(y_0|w) = 1/2$ for each possible w, $\mathbb{P}_2(x_0|y_0) = 1-\alpha$, $\mathbb{P}_2(x_0|y_1) = \alpha$, $\mathbb{P}_2(x_0|w_0, y) = 1-\alpha$ and $\mathbb{P}_2(x_0|w_1, y) = \alpha$ for each possible y, also $\mathbb{P}_2(y_0|x_0, w) = 1-\alpha$ and $\mathbb{P}_2(y_0|x_1, w) = \alpha$ for each possible w, $\mathbb{P}_2(w_0|y) = \alpha$ for each possible y, and $\mathbb{P}_2(w_0|x, y) = \alpha$ for each possible (x, y).

Select $\beta \in (0, 1/2)$. First build a joint probability distribution for (W, X, Y) by using the top table with $\alpha = \beta$. Then build a second joint probability distribution by using the top table with $\alpha = 1-\beta$. Build a third joint probability distribution by using the bottom table with $\alpha = \beta$. And finally build a fourth probability distribution by using the bottom table with $\alpha = 1-\beta$. And build the joint credal set that is the convex hull of these four joint probability distributions. For this credal set we have that all of $\mathbb{P}(x_0|y)$ and $\mathbb{P}(w_0|y)$ for each possible y, $\mathbb{P}(x_0|w)$ and $\mathbb{P}(y_0|w)$ for each possible w, $\mathbb{P}(w_0|x, y)$ for each possible (x, y), $\mathbb{P}(x_0|w, y)$ for each possible (w, y), $\mathbb{P}(y_0|w, x)$ for each possible (w, x), vary within the interval $[\beta, 1-\beta]$. Thus we have that W is regular-epistemically irrelevant to X given Y, Y is regular-epistemically irrelevant to X given W, X is regular-epistemically irrelevant to Y given W, X is regular-epistemically irrelevant to W given Y. However, despite the two first irrelevances, we do not have that (W, Y) is regular-epistemically irrelevant to X (the probability of $\mathbb{P}(x_0)$ is exactly 1/2). And X is not regular-epistemically irrelevant to (W, Y) either: to see that, note that

$$\mathbb{P}(\{W = w_0, Y = y_0\} \cup \{W = w_1, Y = y_1\}) = 1/2$$

but $\mathbb{P}(\{W = w_0, Y = y_0\} \cup \{W = w_1, Y = y_1\}|x)$ vary in the interval $[2\beta(1-\beta), \beta^2 + (1-\beta)^2]$ for each possible value of x. Finally, the irrelevances indicated above imply that X and Y are regular-epistemically independent given W and X and W are regular-epistemically independent given Y; however, X and (W, Y) are clearly not regular-epistemically independent. □

As for regular-epistemic independence, Symmetry holds by definition, and Theorem 2 implies Redundancy. Using conditions indicated in Theorem 3, Decomposition and Weak Union hold when some lower probabilities are larger than zero. However, both Decomposition and Weak Union fail when lower probabilities are allowed to be zero, as illustrated by Example 2. And Contraction and Intersection fail in Examples 3 and 4 respectively, where lower probabilities are even positive.

In short: without further conditions, regular-epistemic independence satisfies only Symmetry and Redundancy.

3.3. Graphoid properties of regular-confirmational irrelevance and independence

We now examine the graphoid properties of regular-confirmational irrelevance, starting with the counterpart of Theorem 2 (note the absence of assumptions concerning closure/convexity/positivity).

Theorem 4. *If $(Y \text{ IR } X \mid Z)$ denotes regular-confirmational irrelevance of Y to X given Z, then the same properties listed in Theorem 2 hold.*

To prove direct Decomposition and direct Weak Union we use the following fact. Suppose we have a probability distribution for variables (W, Y), conditional on some event H. This distribution can be represented by a vector p,

[4] These six versions of Intersection are:
$(X \text{ IR } W \mid (Y, Z)) \wedge (X \text{ IR } Y \mid (W, Z)) \Rightarrow ((W, Y) \text{ IR } X \mid Z)$ (fails when $W = Z = 1$);
$(X \text{ IR } W \mid (Y, Z)) \wedge (Y \text{ IR } X \mid (W, Z)) \Rightarrow (X \text{ IR } (W, Y) \mid Z)$ (fails when $W = Z = 1$);
$(X \text{ IR } W \mid (Y, Z)) \wedge (Y \text{ IR } X \mid (W, Z)) \Rightarrow ((W, Y) \text{ IR } X \mid Z)$ (fails when $Y = Z = 1$);
$(W \text{ IR } X \mid (Y, Z)) \wedge (X \text{ IR } Y \mid (W, Z)) \Rightarrow (X \text{ IR } (W, Y) \mid Z)$ (fails when $Y = Z = 1$);
$(W \text{ IR } X \mid (Y, Z)) \wedge (X \text{ IR } Y \mid (W, Z)) \Rightarrow ((W, Y) \text{ IR } X \mid Z)$ (fails when $W = Z = 1$);
$(W \text{ IR } X \mid (Y, Z)) \wedge (Y \text{ IR } X \mid (W, Z)) \Rightarrow (X \text{ IR } (W, Y) \mid Z)$ (fails when $Y = Z = 1$).

containing values $\mathbb{P}(w, y|H)$. Suppose we have a function $F(\cdot)$ that takes p and returns another vector q. If we have two credal sets $\mathbb{K}_1$ and $\mathbb{K}_2$ that are equal, the elementwise application of $F(\cdot)$ generates two identical sets. That is, $\mathbb{K}_1 = \mathbb{K}_2 \quad \Rightarrow \quad \{F(p) : p \in \mathbb{K}_1\} = \{F(p) : p \in \mathbb{K}_2\}$.

Proof. Recall that w is a value of W, x is a value of X, and so on.

Redundancy: whenever $\overline{\mathbb{P}}(X = x_1, X = x_2) > 0$, then we have the equality $\mathbb{K}^>(Y|X = x_1, X = x_2) = \mathbb{K}^>(Y|X = x_1)$ (because if $x_1 = x_2$, then trivially $\{X = x_1, X = x_2\} = \{X = x_1\}$); also, whenever $\overline{\mathbb{P}}(x, y) > 0$, we have the equality $\mathbb{K}^>(X|X = x, Y = y) = \mathbb{K}^>(X|X = x)$ as both credal sets $\mathbb{K}^>(X|X = x, Y = y)$ and $\mathbb{K}^>(X|X = x)$ contain exactly the distribution that assigns probability one to $\{X = x\}$.

For the next two paragraphs, note that any probability distribution for (W, Y) given $\{X = x, Z = z\}$ can be represented by a vector p containing the probability value of each possible (w, y). Also, suppose X is regular-confirmationally irrelevant to (W, Y) given Z; that is, $\mathbb{K}^>(W, Y|x, z) = \mathbb{K}^>(W, Y|z)$ whenever $\overline{\mathbb{P}}(x, z) > 0$.

Decomposition: Define a (marginalization) function $F(p)$ that yields a vector containing, for each value y, the value of the summation $\sum_w \mathbb{P}(w, y|x, z)$. So, the equality $\mathbb{K}^>(W, Y|x, z) = \mathbb{K}^>(W, Y|z)$ whenever $\overline{\mathbb{P}}(x, z) > 0$ implies $\mathbb{K}^>(Y|x, z) = \mathbb{K}^>(Y|z)$ whenever $\overline{\mathbb{P}}(x, z) > 0$, as desired.

Weak Union: Given a value w of W, define a (conditioning) function $F_w(p)$ that yields a vector containing, for each value y, either: the value of the ratio $\mathbb{P}(w, y|x, z)/\mathbb{P}(w|x, z)$ where $\mathbb{P}(w|x, z) = \sum_k \mathbb{P}(w, k|x, z)$ if $\mathbb{P}(w|x, z) > 0$, or 0 if $\mathbb{P}(w|x, z) = 0$. So, the equality $\mathbb{K}^>(W, Y|x, z) = \mathbb{K}^>(W, Y|z)$ whenever $\overline{\mathbb{P}}(x, z) > 0$ implies the equality $\mathbb{K}^>(Y|x, w, z) = \mathbb{K}^>(Y|w, z)$ whenever $\overline{\mathbb{P}}(x, w, z) > 0$ (by applying $F_w(\cdot)$ to the relevant distributions and discarding the others), as desired.

Contraction: We have $\mathbb{K}^>(X|w, y, z) = \mathbb{K}^>(X|y, z)$ whenever $\overline{\mathbb{P}}(w, y, z) > 0$ and $\mathbb{K}^>(X|y, z) = \mathbb{K}^>(X|z)$ whenever $\overline{\mathbb{P}}(y, z) > 0$, hence by transitivity $\mathbb{K}^>(X|w, x, z) = \mathbb{K}^>(X|z)$ whenever $\overline{\mathbb{P}}(w, y, z) > 0$ as desired. □

All other versions of Decomposition and Weak Union fail for regular-confirmational irrelevance given Example 2 and arguments given around that example. And all other versions of Contraction fail for regular-confirmational irrelevance given Example 3 and arguments given right before that example. All versions of Intersection fail for regular-confirmational irrelevance given Example 4 and arguments given right before that example.

Example 2 also shows that Decomposition and Weak Union fail for regular-confirmational independence. Examples 3 and 4 show that both Contraction and Intersection fail for regular-confirmational independence. Hence we only have Symmetry and Redundancy for regular-confirmational independence when no additional conditions are adopted.

We might ask whether Decomposition and Weak Union hold when credal sets are convex and probabilities are larger than zero, but closure is not required — hence it may happen still that *lower* probabilities are zero while *all* probabilities are larger than zero. The following example demonstrates that Decomposition and Weak Union fail in such circumstances:

Example 5. Take the interior of the joint credal set in Example 2; this is an open credal set where each distribution assigns positive probability to every possible event. Decomposition and Weak Union still fail for regular-confirmational independence: $\mathbb{P}(x_0|w, y)$ belongs to the open interval $(\beta, 1 - \beta)$ for every possible (w, y), and also $\mathbb{P}(x_0)$ belongs to the same open interval; similarly, $\mathbb{K}(W, Y)$, $\mathbb{K}^>(W, Y|x_0)$ and $\mathbb{K}^>(W, Y|x_1)$ are identical open sets — but $\mathbb{K}^>(X|w_1)$ is the singleton $\{[1/2, 1/2]\}$, so Decomposition and Weak Union fail. This means that we cannot simply require probabilities to be larger than zero to obtain Decomposition and Weak Union for regular-confirmational independence; we must assume that *lower* probabilities are larger than zero. □

Another possibility is to assume positive lower probabilities but to drop convexity. Failure of Decomposition and Weak Union has been shown before for non-connected (hence non-convex) credal sets [11]; we now present an example that displays failure of Decomposition and Weak Union in connected non-convex sets. Note that all lower probabilities are positive in this example.

Example 6. Consider three variables W, X and Y. While W and Y are binary, X has three values. Denote by U the point $[1/3, 1/3, 1/3]$; we will interpret such a point as the distribution where $\mathbb{P}(x_0) = 1/3$, $\mathbb{P}(x_1) = 1/3$, and $\mathbb{P}(x_2) = 1/3$. Denote by A^α the point

$$[\alpha/2 + (1 - \alpha)/3, \alpha/6 + (1 - \alpha)/3, 1/3 + \alpha(1 - \alpha)]/k_1,$$

where $k_1 = 1 + \alpha(1 - \alpha)$. Note that as α varies from 0 to 1 we obtain a curve in the three-dimensional simplex. Note also that $A^0 = U$, and $A^1 = [1/2, 1/6, 1/3]$. Denote by B^α the point

$$[\alpha/6 + (1 - \alpha)/3, \alpha/2 + (1 - \alpha)/3, 1/3 - \alpha(1 - \alpha)]/k_2,$$

where $k_2 = 1 - \alpha(1 - \alpha)$.

Now define a set of distributions parameterized by $\alpha \in [0, 1]$ such that, for every value (w, y) of (W, Y), $\mathbb{P}_1^\alpha(X = x|w, y)$ is the point A^α and $\mathbb{P}_1^\alpha(w, y) = 1/4$, and another set of distributions such that $\mathbb{P}_2^\alpha(X = x|w, y)$ is the point B^α and $\mathbb{P}_2^\alpha(w, y) = 1/4$. Continue by defining an additional set of distributions parameterized by $\alpha \in [0, 1]$, as follows:

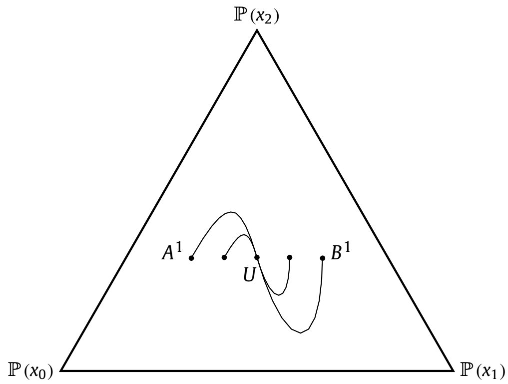

Fig. 1. Credal sets from Example 6, in barycentric coordinates. The marginal credal set $\mathbb{K}(X)$ appears as the longer curve from A^1 to B^1, going through the uniform distribution U. The credal sets $\mathbb{K}(X|y_0)$ and $\mathbb{K}(X|y_1)$ include the shorter curve that goes from $(A^1+U)/2$ to $(B^1+U)/2$.

$$\begin{aligned}\mathbb{P}_3^\alpha(X=x|W=w_0,Y=y_0)&=A^\alpha,\\ \mathbb{P}_3^\alpha(X=x|W=w_0,Y=y_1)&=B^\alpha,\\ \mathbb{P}_3^\alpha(X=x|W=w_1,Y=y_0)&=U,\\ \mathbb{P}_3^\alpha(X=x|W=w_1,Y=y_1)&=U,\end{aligned}$$

and $\mathbb{P}_3^\alpha(w_0,y_0)=k_1/4$, $\mathbb{P}_3^\alpha(w_0,y_1)=k_2/4$, $\mathbb{P}_3^\alpha(w_1,y_0)=1/4$, $\mathbb{P}_3^\alpha(w_1,y_1)=1/4$.

By taking the union of distributions $\mathbb{P}_1^\alpha$, $\mathbb{P}_2^\alpha$ and $\mathbb{P}_3^\alpha$ as $\alpha\in[0,1]$, we obtain $\mathbb{K}^>(X|w,y)=\mathbb{K}^>(X)$; consequently, (W,Y) is regular-confirmationally irrelevant to X. However, this does not guarantee that X is also regular-confirmationally irrelevant to (W,Y); in fact the credal set consisting of the convex hull of $\mathbb{P}_1^\alpha$, $\mathbb{P}_2^\alpha$ and $\mathbb{P}_3^\alpha$ does not lead to this regular-confirmational irrelevance.

To obtain regular-confirmationally irrelevance of X to (W,Y), we build four additional sets of distributions that guarantee the necessary equalities. We build these distributions as follows: take $\mathbb{P}_i^\alpha(x|w,y)=U$ for $i\in\{4,5,6,7\}$ and

$$\begin{aligned}\mathbb{P}_4^\alpha(w,y)&=\mathbb{P}_3^\alpha(w,y);\\ \mathbb{P}_5^\alpha(w,y)&=\mathbb{P}_3^\alpha(w,y|x_0);\\ \mathbb{P}_6^\alpha(w,y)&=\mathbb{P}_3^\alpha(w,y|x_1);\\ \mathbb{P}_7^\alpha(w,y)&=\mathbb{P}_3^\alpha(w,y|x_2).\end{aligned}$$

We then have that $\mathbb{K}^>(W,Y|x)=\mathbb{K}^>(W,Y)$; hence (W,Y) and X are regular-confirmationally independent.

However, for every value y of Y,

$$\mathbb{K}^>(X|y)=\cup_{\alpha\in[0,1]}\{A^\alpha,B^\alpha,(A^\alpha+U)/2,(B^\alpha+U)/2\},$$

so $\mathbb{K}^>(X|w,y)\neq\mathbb{K}^>(X|y)$ (failure of Weak Union) for every possible (w,y), and $\mathbb{K}^>(X|y)\neq\mathbb{K}^>(X)$ (failure of Decomposition). Fig. 1 depicts some of the geometry behind this example. □

Now if we require both convexity and positive lower probabilities, but drop the assumption of closure, we do obtain reverse Decomposition and reverse Weak Union:

Theorem 5. *Suppose that all lower probabilities are positive and that credal sets $\mathbb{K}^>(W,X,Y|z)$ are convex. If $(Y \;\textsc{ir}\; X \mid Z)$ denotes regular-confirmational irrelevance of Y to X given X, then:*
- *If $((W,Y) \;\textsc{ir}\; X \mid Z)$, then $(Y \;\textsc{ir}\; X \mid Z)$ (a "reverse" form of Decomposition);*
- *If $((W,Y) \;\textsc{ir}\; X \mid Z)$, then $(W \;\textsc{ir}\; X \mid (Y,Z))$ (a "reverse" form of Weak Union),*

where $(Y \;\textsc{ir}\; X \mid Z)$ denotes regular-confirmational irrelevance of Y to X given Z.

Proof. Again, recall that w is a value of W, x is a value of X, and so on.

Because all lower probabilities are positive, there is no need to indicate whether we use $\mathbb{K}^\vartriangleright$ or $\mathbb{K}^>$; we simply use $\mathbb{K}$ for all credal sets. In this proof we treat elements of credal sets as "points" regardless of the conditioning events; that is, a distribution is viewed as a vector containing probability values.

We abuse notation by writing $p\in\mathbb{K}$ when p is a vector indexed by the relevant random variables. For instance, if $p\in\mathbb{K}(X)$, then p is indexed by values of X; if $p\in\mathbb{K}(Y|z)$, then p is indexed by values of Y for a given z that is fixed in the background.

Assume $\mathbb{K}(X|w,y,z)=\mathbb{K}(X|z)$ for every (w,y), with the value z fixed throughout.

First we show that if a point p' encoding a distribution of X is not in $\mathbb{K}(X|z)$, then this point p' cannot encode a distribution in $\mathbb{K}(X|y,z)$. Take point $p'\notin\mathbb{K}(X|z)$; then by the assumed regular-confirmational irrelevance, $p'\notin\mathbb{K}(X|w,y,z)$. To obtain a contradiction, suppose that $p' \in \mathbb{K}(X|y,z)$: that would imply the convex combination $\sum_w p(w|y,z)\,p(x|w,y,z)$ to produce the elements of p' indexed by x. As each distribution of X given $\{W=w, Y=y, Z=z\}$ also belongs to $\mathbb{K}(X|z)$ by the assumed regular-confirmational irrelevance, convexity of $\mathbb{K}(X|z)$ would imply $p' \in \mathbb{K}(X|z)$. Hence if $p' \notin \mathbb{K}(X|z)$, then $p' \notin \mathbb{K}(X|y,z)$; that is, $p' \in \mathbb{K}(X|y,z)$ implies $p' \in \mathbb{K}(X|z)$, and so $\mathbb{K}(X|y,z) \subseteq \mathbb{K}(X|z)$.

Note that using the same summation and reasoning in the previous paragraph we conclude that $\mathbb{K}(X|y,z) \subseteq \mathbb{K}(X|w,y,z)$ for all (w,z).

Now suppose there is $\mathbb{K}(X|y^*,z)$ that is a strict subset of $\mathbb{K}(X|z)$ for some value y^*. Note that for any $p \in \mathbb{K}(X|z)$, if we write $p(x)$ to refer to the element of p indexed by x, then we must have $p(x|z) = \sum_y p(y|z)\,p(x|y,z)$ for each x, hence $p(x|z) = p(y^*|z)\,p(x|y^*,z) + (1 - p(y^*|z))q(x)$ where $q(x) = \sum_{y \neq y^*} p(y|z)\,/(1 - p(y^*|z))p(x|y)$. Thus $q(X)$ is a convex combination of distributions from $\mathbb{K}(X|w,y,z)$; $q(X)$ is thus constrained to be within $\mathbb{K}(X|w,y,z)$ and thus within $\mathbb{K}(X|z)$. But it is impossible to write down every point of $\mathbb{K}(X|z)$ as a convex combination, where the mixing coefficient is necessarily away from zero by assumption, of a point in $\mathbb{K}(X|z)$ and a point in a strict subset of $\mathbb{K}(X|z)$.[5] Thus we must have $\mathbb{K}(X|y,z) = \mathbb{K}(X|z)$ for all (y,z) and we obtain reverse Decomposition. Using transitivity to get $\mathbb{K}(X|w,y,z) = \mathbb{K}(X|z) = \mathbb{K}(X|y,z)$, we obtain reverse Weak Union. $\square$

The latter result shows that convexity and positive lower probabilities suffice to guarantee Symmetry, Redundancy, Decomposition and Weak Union for regular-confirmational independence.

These results and examples suggest that regular-epistemic and regular-confirmational independence are rather weak concepts without further conditions: we cannot keep Decomposition and Weak Union when lower probabilities may be zero, and Contraction and Intersection fail even with positive lower probabilities. We are left with Symmetry, a property that is enforced by definition, and Redundancy, a rather weak property.

3.4. Graphoid properties of type-5 and type-5 epistemic irrelevance and independence

For type-5 epistemic irrelevance we have:

Theorem 6. *If $(Y \text{ IR } X \mid Z)$ denotes type-5 epistemic irrelevance of Y to X given Z, then:*
- $(X \text{ IR } Y \mid X)$ *and* $(Y \text{ IR } X \mid X)$ *("direct" and "reverse" forms of Redundancy);*
- *If* $(X \text{ IR } (W,Y) \mid Z)$*, then* $(X \text{ IR } Y \mid Z)$ *(a "direct" form of Decomposition);*
- *If* $(X \text{ IR } (W,Y) \mid Z)$*, then* $(X \text{ IR } Y \mid (W,Z))$ *(a "direct" form of Weak Union);*
- *If* $((W,Y) \text{ IR } X \mid Z)$*, then* $(Y \text{ IR } X \mid Z)$ *(a "reverse" form of Decomposition);*
- *If* $((W,Y) \text{ IR } X \mid Z)$*, then* $(Y \text{ IR } X \mid (W,Z))$ *(a "reverse" form of Weak Union).*

Proof. Recall that f is a function of X, g is a function of Y, w is a value of W, x is a value of X, and so on.

Redundancy: whenever $\overline{\mathbb{P}}(A,\{x\}) > 0$ for an event A in the algebra generated by X, the equality $\underline{\mathbb{E}}^{>}[g(Y)|A,\{x\}] = \underline{\mathbb{E}}^{>}[g(Y)|X=x]$ holds for every function $g(Y)$ (because if $x \in A$, then trivially $A \cap \{x\} = \{x\}$); also, whenever $\overline{\mathbb{P}}(B,\{x\}) > 0$ for an event B in the algebra generated by Y, we have $\underline{\mathbb{E}}^{>}[f(X)|B,\{x\}] = f(x) = \underline{\mathbb{E}}^{>}[f(X)|x]$ for every function $f(X)$.

Direct Decomposition: by assumption X is type-5 epistemically irrelevant to (W,Y) given Z; thus we have, for any function $g(Y)$, that $\underline{\mathbb{E}}^{>}[g(Y)|A,\{z\}] = \underline{\mathbb{E}}^{>}[g(Y)|z]$ whenever $\overline{\mathbb{P}}(A,\{z\}) > 0$ for an event A in the algebra generated by X, by hypothesis as any $g(Y)$ is obviously also a function of (W,Y).

Direct Weak Union: by assumption we have the equality $\underline{\mathbb{E}}^{>}[h(W,Y)|A,\{z\}] = \underline{\mathbb{E}}^{>}[h(W,Y)|z]$ whenever $\overline{\mathbb{P}}(A,\{z\}) > 0$ for an event A in the algebra generated by X; hence, using Theorem 1, if $\overline{\mathbb{P}}(A,\{w\},\{z\}) > 0$, then

$$
\begin{aligned}
\underline{\mathbb{E}}^{>}[g(Y)|A,\{w\},\{z\}] &= \sup\left(\alpha : \underline{\mathbb{E}}^{>}[(g(Y)-\alpha)\mathbb{I}_w(W)|A,\{z\}] \geq 0\right) \\
&= \sup\left(\alpha : \underline{\mathbb{E}}^{>}[(g(Y)-\alpha)\mathbb{I}_w(W)|z] \geq 0\right) \\
&= \underline{\mathbb{E}}^{>}[g(Y)|w,z].
\end{aligned}
$$

Reverse Decomposition: by assumption (W,Y) is type-5 epistemically irrelevant to X given Z; thus, for any function $f(X)$, $\underline{\mathbb{E}}[f(X)|B,\{x\}] = \underline{\mathbb{E}}[f(X)|z]$ for an event B in the algebra generated by Y, because any event B in the algebra generated by Y is also an event in the algebra generated by (W,Y).

Reverse Weak Union: by assumption (W,Y) is type-5 epistemically irrelevant to X given Z; hence, for any function $f(X)$, $\underline{\mathbb{E}}^{>}[f(X)|B,\{w\},\{z\}] = \underline{\mathbb{E}}^{>}[f(X)|z]$ whenever $\overline{\mathbb{P}}(B,\{w\},\{z\}) > 0$ for an event B in the algebra generated by Y; also we have $\underline{\mathbb{E}}^{>}[f(X)|w,z] = \underline{\mathbb{E}}^{>}[f(X)|z]$ as a consequence of reverse Decomposition; thus $\underline{\mathbb{E}}^{>}[f(X)|B,\{w\},\{z\}] = \underline{\mathbb{E}}^{>}[f(X)|z] = \underline{\mathbb{E}}^{>}[f(X)|w,z]$ whenever $\overline{\mathbb{P}}(B,\{w\},\{z\}) > 0$ (that is, Y is type-5 epistemically irrelevant to X given (W,Z)). $\square$

[5] For suppose we have two convex sets A and B such that $A \subset B$. Then there must be an extreme point p of the closure of B that is not in A (otherwise the two sets would be identical). If this point p is in B and all of its neighboring points in B are in A, then the point p cannot be produced by a nontrivial combination of a point in B and a point in A. So suppose there is a ball around p that is not in A; there must be a smaller ball containing the points of B that cannot be attained by a convex combination of a point close to B and a point in A.

Table 5
Tables employed in Example 7.

$\mathbb{P}_1$	w_0y_0	w_0y_1	w_1y_0	w_1y_1
x_0	0	$\alpha/2$	$\alpha/2$	0
x_1	0	$(1-\alpha)/2$	$(1-\alpha)/2$	0

$\mathbb{P}_2$	w_0y_0	w_0y_1	w_1y_0	w_1y_1
x_0	$\alpha/2$	0	0	$(1-\alpha)/2$
x_1	$(1-\alpha)/2$	0	0	$\alpha/2$

All other versions of Decomposition and Weak Union fail for type-5 epistemic irrelevance given the "symmetry-based" arguments around Example 2. The same arguments there cover all other versions of Contraction except direct Contraction and reverse Contraction. While direct Contraction fails in Example 3 (where all lower probabilities are positive), reverse Contraction fails in the next example:

Example 7. Consider three binary variables W, X and Y. For Table 5 top, $\mathbb{P}_1(x_0) = \mathbb{P}_1(x_0|y) = \alpha$ for each possible y, and $\mathbb{P}_1(x_0|w, y) = \alpha$ whenever $\mathbb{P}_1(w, y) > 0$. For the bottom table we have $\mathbb{P}_2(x_0) = 1/2$, $\mathbb{P}_2(x_0|y_0) = \alpha$, $\mathbb{P}_2(x_0|y_1) = 1-\alpha$, $\mathbb{P}_2(x_0|w_0, y_0) = \alpha$ and $\mathbb{P}_2(x_0|w_1, y_1) = 1-\alpha$.

Select $\beta \in (0, 1/2)$. First build two joint probability distributions for (W, X, Y) by using $\mathbb{P}_1$ with $\alpha = \beta$ and $\alpha = 1-\beta$. Then build two additional joint probability distributions by using $\mathbb{P}_2$ with $\alpha = \beta$ and $\alpha = 1-\beta$. Build the joint credal set consisting of the convex hull of these four joint probability distributions. For this credal set we have that all of $\mathbb{P}(x_0)$, $\mathbb{P}(x_0|y)$ for each possible y vary within the interval $[\beta, 1-\beta]$. Moreover, if we discard the distributions for which $\mathbb{P}(w_0, y_0) = 0$, then $\mathbb{P}(x_0|w_0, y_0)$ varies within the interval $[\beta, 1-\beta]$ as well. Likewise, $\mathbb{P}(x_0|w_0, y_1)$, $\mathbb{P}(x_0|w_1, y_0)$ and $\mathbb{P}(x_0|w_1, y_1)$ vary within the same interval. Thus we have that Y is type-5 epistemically irrelevant to X, and W is type-5 epistemically irrelevant to X given Y. However, (W, Y) is not type-5 epistemically irrelevant to X, as $\mathbb{K}^>(X|\{W = w_0, Y = y_0\} \cup \{W = w_1, Y = y_1\})$ is a singleton containing the probability distribution such that $\mathbb{P}(x_0|\{W = w_0, Y = y_0\} \cup \{W = w_1, Y = y_1\}) = 1/2$. □

However, reverse Contraction holds when lower probabilities are positive:

Theorem 7. *If $(Y \text{ IR } X \mid Z)$ denotes type-5 epistemic irrelevance of Y to X given Z, and all lower probabilities are positive, then:*
- *If $(Y \text{ IR } X \mid Z)$ and $(W \text{ IR } X \mid (Y, Z))$, then $((W, Y) \text{ IR } X \mid Z)$ (a "reverse" form of Contraction).*

Proof. Take D to be a set containing values (w, y) of (W, Y). For each value y of Y that appears in D, denote by H_y the set of pairs (w, y) that belong to $D \cap \{Y = y\}$. That is, H_y is a "slice" of D such that Y is set to y. Thus $D = \cup_y H_y$. For any function $f(X)$, use Theorem 1 to obtain:

$$\begin{aligned}
\underline{\mathbb{E}}^>[f(X)|D, z] &= \sup\left(\alpha : \underline{\mathbb{E}}^>[(f(X)-\alpha)\mathbb{I}_D|z] \geq 0\right) \\
&= \sup\left(\alpha : \underline{\mathbb{E}}^>\left[(f(X)-\alpha)\mathbb{I}_{\bigcup_y H_y}|z\right] \geq 0\right) \\
&= \sup\left(\alpha : \underline{\mathbb{E}}^>\left[(f(X)-\alpha)\sum_y \mathbb{I}_{H_y}|z\right] \geq 0\right);
\end{aligned}$$

and then we use the fact that $\underline{\mathbb{E}}^>[\sum_i f_i|z] \geq \sum_i \underline{\mathbb{E}}^>[f_i|z]$ for any f_i, and the assumptions of independence to obtain:

$$\begin{aligned}
\underline{\mathbb{E}}^>[f(X)|D, z] &\geq \sup\left(\alpha : \sum_y \underline{\mathbb{E}}^>\left[(f(X)-\alpha)\mathbb{I}_{H_y}|z\right] \geq 0\right) \\
&\geq \min_y \sup\left(\alpha : \underline{\mathbb{E}}^>\left[(f(X)-\alpha)\mathbb{I}_{H_y}|z\right] \geq 0\right) \\
&= \min_y \underline{\mathbb{E}}^>\left[f(X)|H_y, \{z\}\right] = \min_y \underline{\mathbb{E}}^>[f(X)|y, z] \\
&= \min_y \underline{\mathbb{E}}[f(X)|z] = \underline{\mathbb{E}}[f(X)|z],
\end{aligned}$$

where we use both the fact that $H_y \cap \{z\}$ is the intersection of a set of values of W with $\{y, z\}$ and the assumed irrelevances.

If we repeat this reasoning with the complement of D^c instead of D, we obtain both $\underline{\mathbb{E}}^>[f(X)|D, \{z\}] \geq \underline{\mathbb{E}}^>[f(X)|z]$ and $\underline{\mathbb{E}}^>[(X)f|D^c, \{z\}] \geq \underline{\mathbb{E}}[f(X)|z]$. These inequalities, and the fact that

$$\mathbb{E}[f(X)|z] = \mathbb{P}(D|z)\,\mathbb{E}[f(X)|D, \{z\}] + (1-\mathbb{P}(D|z))\mathbb{E}\left[f(X)|D^c, \{z\}\right]$$

for any probability distribution, imply $\underline{\mathbb{E}}^>[f(X)|D,\{z\}] = \underline{\mathbb{E}}^>[f(X)|D^c,\{z\}] = \underline{\mathbb{E}}^>[f(X)|z]$ (as the distribution that attains the latter lower expectation must attain the other two as well). □

As for Intersection, all versions fail for type-5 epistemic irrelevance either using "symmetry-based" arguments or Example 4.

Hence, for type-5 epistemic independence we get, using Theorem 6[6]:

Theorem 8. *Type-5 epistemic independence satisfies Symmetry, Redundancy, Decomposition and Weak Union.*

All versions of Contraction and Intersection fail for type-5 epistemic independence.

Finally, consider type-5 irrelevance. We have the counterpart of Theorem 6:

Theorem 9. *If $(Y \text{ IR } X \mid Z)$ denotes type-5 irrelevance of Y to X given Z, then the same properties listed in Theorem 6 hold.*

Proof. Again, recall that w is a value of W, x is a value of X, and so on.

Redundancy: There are two equalities to prove: first,

$$\mathbb{K}^>(Y|A,\{x\}) = \mathbb{K}^>(Y|x) \text{ whenever } \overline{\mathbb{P}}(A,\{x\}) > 0$$

for any event A in the algebra generated by X (true because if $A \cap \{x\} = \emptyset$, then $\overline{\mathbb{P}}(A,\{x\}) = 0$; otherwise $A \cap \{x\} = \{x\}$, then trivially $\mathbb{K}^>(Y|A,\{x\}) = \mathbb{K}^>(Y|x)$); second,

$$\mathbb{K}^>(X|B,\{x\}) = \mathbb{K}^>(X|x) \text{ whenever } \overline{\mathbb{P}}(B,\{x\}) > 0$$

for any event B in the algebra generated by Y (true because both credal sets $\mathbb{K}^>(X|B,\{x\})$ and $\mathbb{K}^>(X|x)$ contain exactly the distribution that assigns probability one to $\{x\}$ whenever $\overline{\mathbb{P}}(B,\{x\}) > 0$).

For the next two paragraphs, note that any probability distribution for (W,Y) given $A \cap \{z\}$, where A is an event in the algebra generated by X, can be represented by a vector p containing the probability value of each possible (w,y). Also, suppose X is type-5 irrelevant to (W,Y) given Z; that is, $\mathbb{K}^>(W,Y|A,\{z\}) = \mathbb{K}^>(W,Y|z)$ whenever $\overline{\mathbb{P}}(A,\{z\}) > 0$.

Direct Decomposition: Define the (marginalization) function $F(p)$ that yields a vector containing, for each value y, the value of the summation $\sum_w \mathbb{P}(w,y|A,\{z\})$ where the event A in the algebra generated by X is fixed. So, the equality $\mathbb{K}^>(W,Y|A,\{z\}) = \mathbb{K}^>(W,Y|z)$ whenever $\overline{\mathbb{P}}(A,\{z\}) > 0$ implies $\mathbb{K}^>(Y|A,\{z\}) = \mathbb{K}^>(Y|z)$ whenever $\overline{\mathbb{P}}(A,\{z\}) > 0$, as desired.

Weak Union: Given a value w of W, define a (conditioning) function $F_w(p)$ that yields a vector containing, for each value y, either: the value of the ratio $\mathbb{P}(w,y|A,\{z\})/\mathbb{P}(w|A,\{z\})$ where $\mathbb{P}(w|A,\{z\}) = \sum_k \mathbb{P}(w,k|A,\{z\})$ if $\mathbb{P}(w|A,\{z\}) > 0$, or 0 if $\mathbb{P}(w|A,\{z\}) = 0$. So, the equality $\mathbb{K}^>(W,Y|A,\{z\}) = \mathbb{K}^>(W,Y|z)$ whenever $\overline{\mathbb{P}}(A,\{z\}) > 0$ implies the equality $\mathbb{K}^>(Y|A,\{w\},\{z\}) = \mathbb{K}^>(Y|w,z)$ whenever $\overline{\mathbb{P}}(A,\{w\},\{z\}) > 0$ (by applying $F_w(\cdot)$ to the relevant distributions and discarding the others), as desired.

Reverse Decomposition: Suppose (W,Y) is type-5 irrelevant to X given Z; then $\mathbb{K}^>(X|B,\{z\}) = \mathbb{K}^>(X|z)$ whenever $\overline{\mathbb{P}}(B,\{z\}) > 0$, for any event B in the algebra generated by Y, given that any event B in the algebra generated by Y is also an event in the algebra generated by (W,Y).

Reverse Weak Union: by assumption (W,Y) is type-5 irrelevant to X given Z; hence $\mathbb{K}^>(X|B,\{w\},\{z\}) = \mathbb{K}^>(X|z)$ whenever $\overline{\mathbb{P}}(B,\{w\},\{z\}) > 0$ for any event B in the algebra generated by Y; also we have $\mathbb{K}^>(X|w,z) = \mathbb{K}^>(X|z)$ whenever $\overline{\mathbb{P}}(w,z) > 0$ as a consequence of reverse Decomposition; hence $\mathbb{K}^>(X|B,\{w\},\{z\}) = \mathbb{K}^>(X|z) = \mathbb{K}^>(X|w,z)$ whenever $\overline{\mathbb{P}}(B,\{w\},\{z\}) > 0$, as desired. □

When all lower probabilities are positive and credal sets are convex and closed, then the arguments in Theorem 7 show that reverse Contraction holds. If we take the interior of the credal set built in Example 7, we note that reverse Contraction fails if we require convexity and positive probabilities but do not require closedness (in which case we may have positive probabilities while still zero lower probabilities).[7]

All other versions of Decomposition, Weak Union, Contraction and Intersection fail for type-5 irrelevance given previous arguments and examples. As for type-5 independence, we can use the previous theorem to state the counterpart of Theorem 8:

Theorem 10. *Type-5 independence satisfies Symmetry, Redundancy, Decomposition and Weak Union.*

[6] Theorem 8 corrects a mistake in Ref. [11], where it is stated that type-5 epistemic independence fails Decomposition and Weak Union.

[7] Two other interesting questions are left for future work. First, whether reverse Contraction holds for type-5 irrelevance when all lower probabilities are positive and credal sets are convex (but may fail to be closed). Second, whether reverse Contraction holds for type-5 irrelevance when all lower probabilities are positive and credal sets are not necessarily convex.

3.5. Graphoid properties of complete and strong independence

A popular way to express stochastic independence of X and Y given Z is to require $\mathbb{P}(x, y|z) = \mathbb{P}(x|z)\,\mathbb{P}(y|z)$ for any value (x, y) of (X, Y) whenever $\mathbb{P}(Z = z) > 0$. There are several proposals in the literature to mimic this latter expression in the context of credal sets. For instance, say that X and Y are *completely independent* if and only if each probability distribution in $\mathbb{K}(X, Y)$ satisfies stochastic independence; the conditional version is simply produced by conditioning everything on some Z. We might refer to complete independence as "elementwise independence" as it operates on individual probability distributions.[8] Note that convexity is in general not satisfied by a credal set $\mathbb{K}(X, Y)$ when X and Y are completely independent [36].

Complete independence satisfies all semi-graphoid properties.

Theorem 11. *Complete independence satisfies Symmetry, Redundancy, Decomposition, Weak Union and Contraction; when all probabilities are positive, then complete independence satisfies Intersection.*

Proof. Each individual probability distribution in a credal set satisfies Symmetry, Redundancy, Decomposition and Weak Union, so they hold for credal sets under complete independence. Concerning Contraction: suppose that, for each probability measure in the credal set, X and Y are stochastically independent given Z, and for any probability measure in the credal set X and W are stochastically independent given (Y, Z); then for any probability measure in the credal set we have both independence relations and then we obtain that X and (W, Y) are stochastically independent given Z for any probability measure in the credal set. The same sort of reasoning leads to the result on Intersection. □

Walley introduced, in early work [54, Appendix], the following concept of independence. Variables X and Y are strongly independent with respect to a closed convex credal set $\mathbb{K}(X, Y)$ in case two conditions are satisfied. First, X and Y must be stochastically independent with respect to each extreme point of $\mathbb{K}(X, Y)$. Second, it must be the case that $\underline{\mathbb{P}}(G) = 0$ implies $\overline{\mathbb{P}}(G) = 0$ for any event specified either by values of X or by values of Y. This second rather severe condition was added by Walley to guarantee that his concept of strong independence implies his concept of regular epistemic independence [54, Appendix, Lemma 1].[9]

Walley's first condition corresponds to the concept of independence that has been most popular in the literature, often under the label of "strong independence".[10] To accommodate credal sets that may not be closed, say that X and Y are *strongly independent* given Z if and only if $\mathbb{K}^{>}(X, Y|z)$ is the convex hull of a credal set satisfying complete independence of X and Y, for all z such that $\overline{\mathbb{P}}(Z = z) > 0$. The idea behind strong independence is to stay close to stochastic independence while assuming convexity, given that imposing stochastic independence over a set of probability measures may generate a nonconvex set of measures. Strong independence can be derived in some cases from assumptions of exchangeability together with a requirement on extreme points [11] or, in a more satisfying manner, assumptions of exchangeability together with epistemic irrelevance [20].

One might suspect that strong independence satisfies all semi-graphoid properties. Concerning Redundancy, if $\mathbb{K}^{>}(X, Y|X = x)$ is a convex set, then X and Y are strongly independent given X (because any individual probability distribution satisfies Redundancy); if instead $\mathbb{K}^{>}(X, Y|X = x)$ is not convex, strong independence fails (somewhat trivially). One might think that, in the context of strong independence, convexity is always assumed, thus leading to Redundancy.

We also have:

Theorem 12. *Strong independence satisfies Symmetry, Decomposition and Weak Union.*

Proof. Symmetry is immediate.

To prove Decomposition and Weak Union, note that both properties have a common assumption; for both of them, the credal set $\mathbb{K}(W, X, Y|z)$, for every z such that $\overline{\mathbb{P}}(Z = z) > 0$, is the set of convex combinations of a set of distributions all of which factorize in a certain way. That is, $\mathbb{K}(W, X, Y|z)$ contains probability distributions that satisfy $\mathbb{P}(w, x, y|z) = \sum_i \alpha_i \mathbb{P}_i(w, x, y|z)$, where $\alpha_i > 0$ and $\sum_i \alpha_i = 1$, and where each indexed probability distribution inside the summation factorizes as $\mathbb{P}_i(w, x, y|z) = \mathbb{P}_i(x|z)\mathbb{P}_i(w, y|z)$.

[8] The term "complete independence" is due to Teddy Seidenfeld. Levi referred to a similar concept as "strong confirmational irrelevance" [37, Section 10.6] in his pioneering work on indeterminate probabilities.

[9] Such condition may be used to avoid some perplexing scenarios based on zero/one probabilities [54, Appendix, Example 1]. Consider the following example due to de Campos and Moral [21]: Take binary variables X and Y, and $\mathbb{K}(X, Y)$ as the convex hull of two distributions, one that assigns probability one to (x_0, y_0) and another that assigns probability one to (x_1, y_1); even though the extreme points of $\mathbb{K}(X, Y)$ satisfy stochastic independence, X and Y fail to be regular-confirmational/regular-epistemic/type-5/type-5-epistemic independent! This example highlights a feature of regular-epistemic irrelevance: the convex hull of distributions that satisfy regular-epistemic irrelevance may fail to satisfy regular-epistemic irrelevance (contrary to results obtained within Walley's theory of lower previsions [23, Proposition 12]).

[10] Levi also discusses a similar strategy where "independence" requires relevant credal sets to be convex hull of sets satisfying his strong confirmational irrelevance [37]. Variants on these concepts have been proposed, with names such as *type-1 product*, *type-2 product*, *type-2 independence*, *independence in the selection*, *repetition independence* [11,39].

Table 6
Tables employed in Example 8. Top: the values of $36 \times \mathbb{P}_2$. Bottom: the values of $288 \times \mathbb{P}_3$.

	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	2	3	6	1
x_1	4	6	12	2

	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	17	39	51	13
x_1	25	51	75	17

The strategy of the proof for both Decomposition and Weak Union will be to start with an assumed stochastic independence and then to show that relevant operations (marginalization and conditioning) led to other desired stochastic independence relations.

Decomposition: The credal set $\mathbb{K}(X, Y|z)$ is obtained, for each z such that $\overline{\mathbb{P}}(Z = z) > 0$, by collecting distributions such that $\mathbb{P}(x, y|z) = \sum_w \mathbb{P}(w, x, y|z)$ where each distribution inside the summation comes from $\mathbb{K}(W, X, Y|z)$. Hence we have, using the assumed strong independence relation,

$$\begin{aligned}\mathbb{P}(x, y|z) &= \sum_w \sum_i \alpha_i \mathbb{P}_i(x|z)\mathbb{P}_i(w, y|z) = \sum_i \alpha_i \mathbb{P}_i(x|z) \sum_w \mathbb{P}_i(w, y|z) \\ &= \sum_i \alpha_i \mathbb{P}_i(x|z)\mathbb{P}_i(y|z),\end{aligned}$$

thus proving strong independence of X and Y given Z.

Weak Union: The credal set $\mathbb{K}(X, Y|w, z)$ is obtained, for each (w, z) such that $\overline{\mathbb{P}}(W = w, Z = z) > 0$, by collecting distributions such that $\mathbb{P}(x, y|w, z) = \mathbb{P}(w, x, y|z)/(\sum_{x,y} \mathbb{P}(w, x, y|z))$ where each distribution in the right-hand side comes from $\mathbb{K}(W, X, Y|z)$. Hence we have, using the assumed strong independence relation,

$$\begin{aligned}\mathbb{P}(x, y|w, z) &= \frac{\sum_i \alpha_i \mathbb{P}_i(x|z)\mathbb{P}_i(w, y|z)}{\sum_{x,y}\sum_i \alpha_i \mathbb{P}_i(x|z)\mathbb{P}_i(w, y|z)} = \frac{\sum_i \alpha_i \mathbb{P}_i(x|z)\mathbb{P}_i(w, y|z)}{\sum_i \alpha_i \mathbb{P}_i(w|z)} \\ &= \sum_i \beta_i \frac{\mathbb{P}_i(x|z)\mathbb{P}_i(w, y|z)}{\mathbb{P}_i(w|z)},\end{aligned}$$

where $\beta_i = \frac{\alpha_i \mathbb{P}_i(w|z)}{\sum_i \alpha_i \mathbb{P}_i(w|z)}$. Thus $\beta_i \geq 0$ and $\sum_i \beta_i = 1$. Consequently,

$$\mathbb{P}(x, y|w, z) = \sum_i \beta_i \mathbb{P}_i(x|z)\mathbb{P}_i(y|w, z) = \sum_i \beta_i \mathbb{P}_i(x|w, z)\mathbb{P}_i(y|w, z),$$

where we use the fact that $\mathbb{P}_i(x|w, z) = \mathbb{P}_i(x|z)$ (by Decomposition), thus proving strong independence of X and Y given (W, Z). □

As shown by the next example, strong independence fails Contraction.[11]

Example 8. Consider three binary variables W, X, and Y, and a joint credal set that is the convex hull of three distributions $\mathbb{P}_1$, $\mathbb{P}_2$ and $\mathbb{P}_3$ as follows.

Distribution $\mathbb{P}_1$ is uniform (that is, $\mathbb{P}_1(w, x, y) = 1/8$ for every (w, x, y)). To build distribution $\mathbb{P}_2$, take $\mathbb{P}_2(w, x, y) = \mathbb{P}_2(w|y)\mathbb{P}_2(x)\mathbb{P}_2(y)$ where $\mathbb{P}_2(x_0) = 1/3$, $\mathbb{P}_2(y_0) = 2/3$, $\mathbb{P}_2(w_0|y_0) = 1/4$ and $\mathbb{P}_2(w_0|y_1) = 3/4$. Finally, take a third distribution $\mathbb{P}_3(w, x, y) = \mathbb{P}_2(w|y)(\mathbb{P}_1(x)\mathbb{P}_1(y)/2 + \mathbb{P}_2(x)\mathbb{P}_2(y)/2)$. The distributions $\mathbb{P}_2$ and $\mathbb{P}_3$ are shown in Table 6.

Note that $\mathbb{P}_3$ is not a convex combination of $\mathbb{P}_1$ and $\mathbb{P}_2$ as there is no $\alpha \in [0, 1]$ such that $\mathbb{P}_3 = \alpha\mathbb{P}_1 + (1 - \alpha)\mathbb{P}_2$ (for instance, there is no $\alpha \in [0, 1]$ such that $\alpha/8 + 4(1 - \alpha)/36 = 25/288$). However, the marginal $\mathbb{P}_3(x, y)$ is the convex combination $\mathbb{P}_1(x)\mathbb{P}_1(y)/2 + \mathbb{P}_2(x)\mathbb{P}_2(y)/2$, so the marginal credal set $\mathbb{K}(X, Y)$ is the convex hull of two product distributions that factorize as $\mathbb{P}_1(x)\mathbb{P}_1(y)$ and $\mathbb{P}_2(x)\mathbb{P}_2(y)$. Also, the credal set $\mathbb{K}(W, X|y)$ is the convex hull of three distributions; one satisfies $\mathbb{P}(w, x|y) = \mathbb{P}_1(w)\mathbb{P}_1(x)$; another satisfies $\mathbb{P}(w, x|y) = \mathbb{P}_2(w|y)\mathbb{P}_2(x)$; the third satisfies $\mathbb{P}(w, x|y) = \mathbb{P}_2(w|y)\mathbb{P}_3(x|y)$.

Hence X and Y are strongly independent, and X and W are strongly independent given Y. However, we do not have strong independence of X and (W, Y): for instance, $\mathbb{P}_3(x_0|w_0, y_0) = 17/42 \neq 5/12 = \mathbb{P}_3(x_0)$. □

As for Intersection, we know that it fails for strong independence when lower probabilities are zero (Table 1). The next example shows that Intersection fails for strong independence even when all lower probabilities are positive.

[11] This corrects a statement in Ref. [11] concerning strong independence and Contraction.

Example 9. Consider three binary variables W, X, and Y, and a joint credal set that is the convex hull of three distributions $\mathbb{P}_1$, $\mathbb{P}_2$ and $\mathbb{P}_3$ as follows.

Distribution $\mathbb{P}_1$ is uniform (that is, $\mathbb{P}_1(w,x,y) = 1/8$ for every (w,x,y)). Distribution $\mathbb{P}_2$ is defined by $\mathbb{P}_2(x,x,y) = \mathbb{P}_2(w)\mathbb{P}_2(x)\mathbb{P}_2(y)$ where $\mathbb{P}_2(w_0) = \mathbb{P}_2(x_0) = \alpha$ and $\mathbb{P}_2(y_0) = 1/2$, with $\alpha \in (0, 1/2)$. And distribution $\mathbb{P}_3$ is specified by

$$\mathbb{P}_3(w,x,y) = \left(\frac{\mathbb{P}_1(w,x)}{2} + \frac{\mathbb{P}_2(w,x)}{2}\right)\mathbb{P}_3(y),$$

where $\mathbb{P}_3(y) = 1 - \alpha$. The latter distribution is *not* a convex combination of $\mathbb{P}_1$ and $\mathbb{P}_2$ (for instance, there is no β such that we can simultaneously satisfy $\mathbb{P}_3(w_0,x_0,y_0) = \beta\mathbb{P}_1(w_0,x_0,y_0) + (1-\beta)\mathbb{P}_2(w_0,x_0,y_0)$ and $\mathbb{P}_3(w_0,x_0,y_1) = \beta\mathbb{P}_1(w_0,x_0,y_1) + (1-\beta)\mathbb{P}_2(w_0,x_0,y_1)$).

Note that $\mathbb{P}_3(w,x|y) = \mathbb{P}_1(w)\mathbb{P}_1(x)/2 + \mathbb{P}_2(w)\mathbb{P}_2(x)/2$. Thus we have that $\mathbb{K}(W,X|y)$ is, for every possible y, the convex hull of two marginal distributions obtained from $\mathbb{P}_1$ and $\mathbb{P}_2$ respectively, both of which satisfy stochastic independence of W and X given Y.

And note that $\mathbb{P}_3(y|w,x) = \mathbb{P}_3(y)$ for every possible (w,y), hence Y and (W,X) are stochastically independent and consequently X and Y are stochastically independent given W (by Weak Union). As X and Y are stochastically independent given W for $\mathbb{P}_1$, $\mathbb{P}_2$, and $\mathbb{P}_3$, we have that X and Y are strongly independent given W.

However, X and (W,Y) are not strongly independent; for instance, we have $\mathbb{P}_3(x_0) = \alpha/2 + 1/4 \neq (1+4\alpha^2)/(2+4\alpha) = \mathbb{P}_3(x_0|w_0,y_0)$ (the inequality holds whenever $\alpha \neq 1/2$). □

Complete and strong independence have a flavor that is distinct from concepts of independence examined in previous sections, such as regular-epistemic independence. While the latter are based on conditional probability, complete and strong independence mimic the factorization property that is usually employed to define stochastic independence.

In this section we have only dealt with notions of stochastic independence that focus on the factorization condition — a condition that is, in many ways, too weak [6]. Additional conditions can be added to factorization so as to prevent independence of events that have logical connection; in doing so, arguably better concepts are produced [2,29,51]. It seems impossible to do justice to this strategy without enlarging too much this investigation; we therefore leave for the future a more detailed study of conditions that prevent logical connections amongst variables.

To conclude this section, we mention another notable concept of independence for credal sets that is based on factorization, due to Kuznetsov [35]: X and Y are Kuznetsov-independent if

$$\mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)] \boxtimes \mathbf{E}[g(Y)]$$

for all functions $f(X)$ and $g(Y)$, where $\mathbf{E}[\cdot]$ denotes the interval from lower to upper expectations, and $\boxtimes$ denotes interval multiplication. Kuznetsov-independence satisfies Symmetry, Redundancy and Decomposition; it fails Contraction even when all probabilities are positive, and it is an open question whether it satisfies Weak Union or not [14].

4. Full credal sets, sets of lexicographic probabilities, sets of desirable gambles

One of the main drawbacks of probability measures is that conditional probability is a secondary concept that appears only after (unconditional) probability is discussed. This problem is eliminated when one adopts *full conditional probabilities*, as these objects take conditional probability as the primary notion of interest [26]. There are other names for full conditional probabilities in the literature, such as *conditional probabilities* [34] and *complete conditional probability systems* [41]; there are also direct connections with theories advocating *coherent probabilities* [6] and *relative probabilities* [33], and with influential theories proposed by Popper [45] and Renyi [46], among others.

Full conditional probabilities have found applications in several fields, notably economy, philosophy, and statistics [6, 25,33,38,42,45,46,50]. In Section 4.1 we summarize the relevant results concerning full conditional probabilities, including existing concepts of independence. And in Section 4.2 we investigate the graphoid properties of concepts of independence associated with sets of full conditional probabilities.

There are other formalisms that go even beyond full conditional probabilities in stripping usual probability measures from its most rigid assumptions regarding precision of assessments and conditioning; two of them are sets of lexicographic probabilities and sets of desirable gambles. We examine both of them in Section 4.3; our main goal there is to adapt some of the results obtained for full conditional probabilities to other contexts. Certainly a much more detailed study of independence concepts for sets of lexicographic probabilities and sets of desirable gambles is possible; we leave that to future work.

4.1. Full conditional probabilities

A *full conditional probability* [26] $\mathbb{P} : \mathcal{B} \times (\mathcal{B}\backslash\{\emptyset\}) \to \Re$, where $\mathcal{B}$ is a Boolean algebra, is a two-place set-function such that for every event $H \neq \emptyset$:
(1) $\mathbb{P}(H|H) = 1$;
(2) $\mathbb{P}(G|H) \geq 0$ for all G;

Table 7
Full distributions for binary variables X and Y.

	y_0	y_1
x_0	$\lfloor 1\rfloor_0$	$\lfloor 1-\alpha\rfloor_1$
x_1	$\lfloor \alpha\rfloor_1$	$\lfloor 1\rfloor_2$

	y_0	y_1
x_0	$\lfloor 1\rfloor_0$	$\lfloor 1\rfloor_1$
x_1	$\lfloor 1\rfloor_2$	$\lfloor 1\rfloor_3$

	y_0	y_1
x_0	$\lfloor 1\rfloor_0$	$\lfloor 1\rfloor_2$
x_1	$\lfloor 1\rfloor_1$	$\lfloor 1\rfloor_3$

Table 8
Full distribution for stochastically independent binary variables, where $\mathbb{P}(x_0) = 1 \neq 0 = \mathbb{P}(x_0|y_1)$.

	y_0	y_1
x_0	$\lfloor 1\rfloor_0$	$\lfloor 1\rfloor_3$
x_1	$\lfloor 1\rfloor_1$	$\lfloor 1\rfloor_2$

(3) $\mathbb{P}(G_1 \cup G_2|H) = \mathbb{P}(G_1|H) + \mathbb{P}(G_2|H)$ whenever $G_1 \cap G_2 = \emptyset$;
(4) $\mathbb{P}(G_1, G_2|H) = \mathbb{P}(G_1|G_2, H) \times \mathbb{P}(G_2|H)$ whenever $G_2H \neq \emptyset$.

These axioms imply that $\mathbb{P}(\Omega|H) = 1$; if the latter equality is assumed for every nonempty event H, then the fourth axiom can be replaced by: $\mathbb{P}(G_1|H) = \mathbb{P}(G_1|G_2)\,\mathbb{P}(G_2|H)$ when $G_1 \subseteq G_2 \subseteq H$ and $G_2 \neq \emptyset$ [16, Section 2].

Define the "unconditional" probability $\mathbb{P}(G)$ of an event G to be $\mathbb{P}(G|\Omega)$. That is, whenever the conditioning event H is equal to Ω, we suppress it and write the "unconditional" probability $\mathbb{P}(G)$. For a full conditional probability $\mathbb{P}$, the expected value $\mathbb{E}_{\mathbb{P}}[f|H]$ is given by the usual expected value with respect to $\mathbb{P}(\cdot|H)$.

We can partition Ω into events $L_0, \ldots, L_K$ as follows. First, take L_0 to be the set of elements of Ω that have positive unconditional probability. Then take L_1 to be the set of elements of Ω that have positive probability conditional on $\Omega\backslash L_0$. And then take L_i, for $i \in \{2, \ldots, K\}$, to be the set of elements of Ω that have positive probability conditional on $\Omega\backslash \cup_{j=0}^{i-1} L_j$. The events L_i are called the *layers* of the full probability. Note that some authors use a different terminology, using instead the sequence $\cup_{j=i}^{K} L_j$ rather than L_i [6,34].

Any full conditional probability can be represented by a sequence of probability measures $\mathbb{P}_0, \ldots, \mathbb{P}_K$, where $\mathbb{P}_i$ is positive over L_i [4,6,29,34].

For nonempty H, denote by L_H the first layer such that $\mathbb{P}(H|L_H) > 0$, and refer to it as the *layer of* H. We then have $\mathbb{P}(G|H) = \mathbb{P}(G|H \cap L_H)$ [2, Lemma 2.1a].

We often write $\lfloor \alpha \rfloor_i$ to denote a probability value α that belongs to the ith layer L_i. Table 7 depicts three full conditional probabilities using this compact notation; we only show probability values and their layers for individual values of (X, Y) as they characterize the whole full conditional probability. We refer to functions such as the ones depicted in Table 7 as *full distributions*.

Given a full conditional probability and a nonempty event H, the two-place function $\mathbb{P}(\cdot|\cdot \cap H)$ is also a full conditional probability from which a partition of H consisting of layers $L_{0|H}, L_{1|H}, \ldots, L_{K|H}$ can be built. Given an event G such that $G \cap H \neq \emptyset$, denote by $L_{G|H}$ the first layer of $\mathbb{P}(\cdot|\cdot \cap H)$ such that $\mathbb{P}\big(G|L_{G|H}\big) > 0$.

For a nonempty event G, the index i of the first layer L_i of the full conditional probability $\mathbb{P}$ such that $\mathbb{P}(G|L_i) > 0$ is the *layer number* of G. Layer numbers have been studied by Coletti and Scozzafava [6], who refer to them as *zero-layers*. The layer number of G is denoted by $\circ(G)$. Inspired by Coletti and Scozzafava [6], we define the layer number of G given nonempty H as $\circ(G|H) = \circ(G \cap H) - \circ(H)$, and we adopt $\circ(\emptyset) = \infty$.

Stochastic independence is rather weak when applied to full conditional probabilities: it may happen that X and Y are stochastically independent and yet $\mathbb{P}(A|B) \neq \mathbb{P}(A)$ when $\mathbb{P}(B) = 0$ (Table 8 shows an extreme example). To avoid this embarrassment, more stringent notions of independence have been proposed for full conditional probabilities [4,6,29,51]. Note that some of these concepts have names that have appeared previously in Section 3 in connection with sets of probability measures, due to the fact that they are essentially the same concepts applied to different objects.

First, say that Y is *epistemically irrelevant* to X given Z if $\mathbb{P}(x|y,z) = \mathbb{P}(x|z)$ whenever $\{Y = y, Z = z\} \neq \emptyset$, and then say that X and Y are *epistemically independent* given Z if X is epistemically irrelevant to Y given Z and vice-versa.

As proposed by Hammond [29], say that Y is *h-irrelevant* to X given Z when

$$\mathbb{P}(C|A, B, \{z\}) = \mathbb{P}(C|A, \{z\}) \text{ whenever } A \cap B \cap \{Z = z\} \neq \emptyset$$

for any A and C in the algebra generated by X, and any B in the algebra generated by Y. Say that X and Y are *h-independent* given Z when X is h-irrelevant to Y given Z and vice-versa.

Hammond has shown that, if X and Y are h-independent given Z, then

$$\mathbb{P}(A, B|C, D, \{z\}) = \mathbb{P}(A|C, \{z\})\,\mathbb{P}(B|D, \{z\}) \quad (4)$$

whenever $C \cap D \cap \{Z = z\} \neq \emptyset$, for any A and C in the algebra generated by X, and any B and D in the algebra generated by Y.

Table 9
Full distribution for W, X, Y, with distinct $\alpha \in (0, 1)$, $\beta \in (0, 1)$, $\gamma \in (0, 1)$.

	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	$\lfloor \alpha \rfloor_0$	$\lfloor 1-\alpha \rfloor_0$	$\lfloor \beta \rfloor_2$	$\lfloor 1-\beta \rfloor_2$
x_1	$\lfloor \alpha \rfloor_1$	$\lfloor 1-\alpha \rfloor_1$	$\lfloor \gamma \rfloor_3$	$\lfloor 1-\gamma \rfloor_3$

Coletti and Scozzafava [6] have proposed conditions on zero-layers to characterize independence. Say that event H is *cs-irrelevant* to event G, where $H \neq \emptyset \neq H^c$, if $\mathbb{P}(G|H) = \mathbb{P}(G|H^c)$, $\circ(G|H) = \circ(G|H^c)$, and $\circ(G^c|H) = \circ(G^c|H^c)$. To understand the motivation for these conditions on layer numbers, suppose that $G \cap H$, $G \cap H^c$, $G^c \cap H$ are nonempty, but $G^c \cap H^c = \emptyset$. Hence observation of H^c does provide information about G. However, the indicator functions of G and H can be epistemically/h-independent! Coletti and Scozzafava eliminate such difficulties using their conditions on layer numbers (other authors, such as Hammond [29] and Battigalli [2], explicitly require the possibility space to be the product of the possibility spaces for each of the variables).

Vantaggi [51,52] has extended Coletti and Scozzafava conditions to independence of variables. Say that Y is *cs-irrelevant* to X given Z when event $\{Y = y\}$ is cs-irrelevant to event $\{X = x\}$ given event $\{Z = z\}$, whenever $\{Y = y, Z = z\} \neq \emptyset \neq \{Y \neq y, Z = z\}$ [51, Definition 7.3]. Refer to the symmetrized concept as *cs-independence* of X and Y given Z.

The conditions on layer numbers imposed by cs-independence can be written as [10, Corollary 2]: for each possible (x, y),

$$\circ(x, y|z) = \circ(x|z) + \circ(y|z) \quad \text{whenever } \{Z = z\} \neq \emptyset. \tag{5}$$

Condition (5) can be used to generate additional concepts of independence. For instance, say that Y is *hcs-irrelevant*[12] to X given Z if Y is h-irrelevant to X given Z and if they satisfy Condition (5); and say that X and Y are *hcs-independent* given Z if they are h-independent given Z and satisfy Condition (5) [10].

A different concept of independence has been proposed by Kohlberg and Reny [33], essentially as follows. Say that X and Y are *kr-independent* given Z when both the following conditions hold:

- if $\{X = x, Z = z\} \neq \emptyset$ and $\{Y = y, Z = z\} \neq \emptyset$, then $\{X = x, Y = y, Z = z\} \neq \emptyset$;
- if, for all pairs of values (x, y) and (x', y') of (X, Y) such that conditioning events are nonempty, we have
$$\frac{\mathbb{P}\left(x, y|L_{x,y|z} \cup L_{x',y'|z}\right)}{\mathbb{P}\left(x', y'|L_{x,y|z} \cup L_{x',y'|z}\right)} = \lim_{n\to\infty} \frac{\mathbb{P}_n(x|z)\mathbb{P}_n(y|z)}{\mathbb{P}_n(x'|z)\mathbb{P}_n(y'|z)} \tag{6}$$
for some sequence of product probability measures $\mathbb{P}_n(\cdot|z)$ (the limit applies to each pair of values whenever the left denominator is larger than zero; otherwise, the ratio is assumed to have the value ∞ and the limit is required to be ∞).

Finally, the concept of *layer independence* requires factorization across layers of the full conditional probability [12]. More precisely: X and Y are *layer independent* given Z if, for each layer L_i of the underlying full conditional probability $\mathbb{P}$, and each z such that $L_i \cap \{z\} \neq \emptyset$, we have both

$$\mathbb{P}(x, y|L_i \cap \{z\}) = \mathbb{P}(x|L_i \cap \{z\})\, \mathbb{P}(y|L_i \cap \{z\})$$

and

$$\circ(x, y|z) = \circ(x|z) + \circ(y|z)$$

for every possible (x, y).

Concerning the semi-graphoid properties, the following facts hold for full conditional probabilities. Epistemic independence satisfies Symmetry, Redundancy, Decomposition and Contraction, but it fails Weak Union [10, Proposition 4.2]. The full conditional probability in Table 9 displays failure of Weak Union for epistemic independence. And h-independence satisfies Symmetry, Redundancy, Decomposition and Weak Union, but it fails Contraction [10, Theorem 5.4]. The full distribution in Table 9 displays failure of Contraction for h-independence. Besides Symmetry, cs-independence satisfies Redundancy, Decomposition and Contraction, and it fails Weak Union [51, Section 9]. And hcs-independence satisfies Symmetry, Redundancy, Decomposition and Weak Union, but it fails Contraction [10, Theorem 5.7]. Table 9 displays failure of Contraction for hcs-independence. And kr-Independence satisfies Symmetry, Redundancy, Decomposition and Weak Union, and it fails Contraction as can be seen in Table 9 [12]. Layer independence satisfies Symmetry, Redundancy, Decomposition, Weak Union

[12] The notions of hcs-irrelevance/independence have been called "full irrelevance" and "full independence" previously [12], but it is perhaps confusing to overload the word "full" in that manner.

Table 10
Table employed in Example 10.

	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	$\lfloor 1 \rfloor_0$	$\lfloor 1 \rfloor_1$	$\lfloor 1 \rfloor_3$	$\lfloor 1 \rfloor_6$
x_1	$\lfloor 1 \rfloor_2$	$\lfloor 1 \rfloor_5$	$\lfloor 1 \rfloor_4$	$\lfloor 1 \rfloor_7$

and Contraction; in fact, this seems to be the only known concept of independence for full conditional probabilities that satisfies all these five properties. We have listed here only the semi-graphoid properties of independence concepts; clearly we might also consider their associated concepts of irrelevance. As Symmetry fails for epistemic-/h-/cs-/hcs-irrelevance, these concepts satisfy varied lists of direct and reverse versions of semi-graphoid properties; we refrain from repeating them here [10].

Now consider the Intersection property. Given that epistemic/h-/cs-/hcs-/layer independence do not make demands on which events must be nonempty, the same example depicted by Table 1 fails Intersection for all of them if we consider the four cells that are marked with 0 as impossible (empty) events.[13] The remaining concept is kr-independence: as it does require events to be nonempty, one might hope that Intersection would hold. Alas, this is not the case:

Example 10. Consider three binary variables W, X, and Y, and the full conditional probability given in Table 10.

If we focus on the probabilities given the event $\{W = w_0\}$, we obtain a full conditional probability that can be represented as the middle table in Table 7. Do we have kr-independence of X and Y given $\{W = w_0\}$? To obtain a positive answer to this question, we must find a sequence of distributions $\mathbb{P}_n(x|w_0)$ and $\mathbb{P}_n(y|w_0)$ such that Expression (6) holds whatever the pairs (x, y) and (x', y'). The middle table in Table 7 can be generated as the limit of a product distribution $\mathbb{P}_n(x, y|w_0) = \mathbb{P}_n(x|w_0)\mathbb{P}_n(y|w_0)$ where $\mathbb{P}_n(x_0|w_0) = 1-1/(2n)^2$ and $\mathbb{P}_n(y_0|w_0) = 1-1/(2n)$ for $n > 0$. Alas, there does not seem to exist an automatic way to find such sequences; another possibility is $\mathbb{P}_n(x_0|w_0) = 1-2^{-2n}$ and $\mathbb{P}_n(y_0|w_0) = 1-2^{-n}$ for $n > 0$.

If we instead focus on the probabilities given the event $\{W = w_1\}$, we obtain a full conditional probability that can be represented as the right table in Table 7. This full conditional probability can be generated as the limit of a product distribution $\mathbb{P}_n(x, y|w_0) = \mathbb{P}_n(x|w_0)\mathbb{P}_n(y|w_0)$ where $\mathbb{P}_n(x_0|w_0) = 1 - 1/(2n)$ and $\mathbb{P}_n(y_0|w_0) = 1 - 1/(2n)^2$ for $n > 0$.

Likewise, if we focus on the probabilities given the event $\{Y = y_0\}$, we obtain the right table in Table 7, with the proviso that y_0 and y_1 are now replaced by w_0 and w_1 respectively. And the very same table is obtained if we focus on the probabilities given the event $\{Y = y_1\}$.

Hence, we have that X and Y are kr-independent given W and X and W are kr-independent given Y. However, X and (W, Y) are not kr-independent. If they were, we should have

$$\frac{\mathbb{P}(w_0, x_0, y_1|H_0)}{\mathbb{P}(w_1, x_0, y_0|H_0)} = \infty = \lim_{n\to\infty} \frac{\mathbb{P}_n(x_0)\mathbb{P}_n(w_0, y_1)}{\mathbb{P}_n(x_0)\mathbb{P}_n(w_1, y_0)} = \lim_{n\to\infty} \frac{\mathbb{P}_n(w_0, y_1)}{\mathbb{P}_n(w_1, y_0)},$$

where $H_0 = \{W = w_0, X = x_0, Y = y_1\} \cup \{W = w_1, X = x_0, Y = y_0\}$; also we should have

$$\frac{\mathbb{P}(w_0, x_1, y_1|H_1)}{\mathbb{P}(w_1, x_1, y_0|H_1)} = 0 = \lim_{n\to\infty} \frac{\mathbb{P}_n(x_1)\mathbb{P}_n(w_0, y_1)}{\mathbb{P}_n(x_1)\mathbb{P}_n(w_1, y_0)} = \lim_{n\to\infty} \frac{\mathbb{P}_n(w_0, y_1)}{\mathbb{P}_n(w_1, y_0)},$$

where $H_1 = \{W = w_0, X = x_1, Y = y_1\} \cup \{W = w_1, X = x_1, Y = y_0\}$. Clearly these equalities are incompatible. □

4.2. Full credal sets and independence

We now focus on sets of full conditional probabilities, and investigate the graphoid properties of several concepts of independence. We refer to such sets as *full credal sets*; we do not assume the sets to be convex nor closed. In fact, it is difficult to deal with convexity in the context of full conditional probabilities, because a convex combination of two functions $\mathbb{P}_1(\cdot|\cdot)$ and $\mathbb{P}_2(\cdot|\cdot)$ may not satisfy the required axioms even when all probabilities are positive.[14] Refer to these as “adapted” versions of confirmational and epistemic irrelevance.

So, consider first confirmational and epistemic irrelevance as defined in Section 3, but applied to full credal sets that are not necessarily convex.[15] That is, we take Expressions (2) and (3) and remove the subscripts “>” as there is no need for regular conditioning (a full credal set simply collects the full conditional probabilities with respect to the relevant conditioning events).

[13] Vantaggi has considered versions of cs-irrelevance that do require various events to be nonempty, thus obtaining Intersection [10,51].

[14] Thanks to Angelo Gilio for pointing this out (personal communication). Consider a simple example where all probabilities are positive. Take variable X with three values and assessments $\mathbb{P}_1(x_0) = \mathbb{P}_1(x_1) = \mathbb{P}_2(x_2) = 1/3$ and $\mathbb{P}_2(x_0)/2 = \mathbb{P}_2(x_1) = \mathbb{P}_2(x_2) = 1/4$. The two-place set function $\mathbb{P}_1(\cdot|\cdot)/2 + \mathbb{P}_2(\cdot|\cdot)/2$ does not qualify as a full conditional probability. For instance, $\mathbb{P}_1(x_0|X \neq x_2)/2 + \mathbb{P}_2(x_0|X \neq x_2)/2 = 7/12$, different from the value consistent with Bayes rule, $(\mathbb{P}_1(\{x_0\}, \{X \neq x_2\}) + \mathbb{P}_2(\{x_0\}, \{X \neq x_2\}))/(\mathbb{P}_1(X \neq x_2) + \mathbb{P}_2(X \neq x_2)) = 10/17$.

[15] These concepts were originally proposed by Levi [37] and by Walley [55] within theories that are tightly connected with full conditional probabilities and that extensively use convexity.

Table 11
Full conditional distributions in Example 11 (left) and their marginal values for (X, Y) (right).

$\mathbb{P}_1$	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	$\lfloor \frac{\alpha}{2} \rfloor_0$	$\lfloor \frac{\alpha}{2} \rfloor_0$	$\lfloor \frac{1-\alpha}{2} \rfloor_0$	$\lfloor \frac{1-\alpha}{2} \rfloor_0$
x_1	$\lfloor \frac{\alpha}{2} \rfloor_1$	$\lfloor \frac{1-\alpha}{2} \rfloor_1$	$\lfloor \frac{1-\alpha}{2} \rfloor_1$	$\lfloor \frac{\alpha}{2} \rfloor_1$

$\mathbb{P}_2$	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	$\lfloor \frac{1-\alpha}{2} \rfloor_0$	$\lfloor \frac{1-\alpha}{2} \rfloor_0$	$\lfloor \frac{\alpha}{2} \rfloor_0$	$\lfloor \frac{\alpha}{2} \rfloor_0$
x_1	$\lfloor \frac{\alpha}{2} \rfloor_1$	$\lfloor \frac{1-\alpha}{2} \rfloor_1$	$\lfloor \frac{\alpha}{2} \rfloor_1$	$\lfloor \frac{1-\alpha}{2} \rfloor_1$

$\mathbb{P}_1$	y_0	y_1
x_0	$\lfloor \frac{1}{2} \rfloor_0$	$\lfloor \frac{1}{2} \rfloor_0$
x_1	$\lfloor \frac{1}{2} \rfloor_1$	$\lfloor \frac{1}{2} \rfloor_1$

$\mathbb{P}_2$	y_0	y_1
x_0	$\lfloor \frac{1}{2} \rfloor_0$	$\lfloor \frac{1}{2} \rfloor_0$
x_1	$\lfloor \alpha \rfloor_1$	$\lfloor 1-\alpha \rfloor_1$

Previous work has shown that the resulting version of confirmational independence fails Decomposition, Weak Union and Contraction even when lower probabilities are positive [11]; this can also be verified in Examples 3 and 6. Intersection also fails for the adapted version of confirmational independence (Example 4). And the adapted version of epistemic independence fails Decomposition and Weak Union when applied to full credal sets, as can be seen in the next example (this example is perhaps simpler than the ones discussed in Ref. [15]):

Example 11. Consider the two full distributions in Table 11(left) by selecting α to be some $\beta \in (0, 1/2)$; then build two additional full distributions by selecting α to be $1 - \beta$. Build the set of all full distributions that consist of two layers, where the first layer carries a convex combination of the first layer of these four full distributions, and the second layer carries a convex combination of the second layer of these four full distributions (because these layers have identical support, this procedure always generates full distributions).

We have $\mathbb{P}_1(w_0) = \alpha$ and $\mathbb{P}_2(w_0) = 1 - \alpha$, so $\mathbb{P}(w_0) \in [\beta, 1 - \beta]$ for every full distribution in the full credal set. Similarly, we have $\mathbb{P}(w_0|x, y) \in [\beta, 1 - \beta]$ for all possible (x, y): for instance, for (x_0, y_0), we have $\mathbb{P}_1(w_0|x_0, y_0) = \alpha$ and $\mathbb{P}_2(w_0|x_0, y_0) = 1 - \alpha$. Note that $\mathbb{P}_2(w_0|x_1, y_0) = \mathbb{P}_2(w_0|x_1, y_1) = 1/2$, still within the interval $[\beta, 1 - \beta]$. Thus (X, Y) is epistemically irrelevant to W.

Now consider computing $\mathbb{P}(x, y)$ for the full conditional distributions in the full credal set. By marginalization we obtain the probability values in Table 11(right). Thus for any marginal for (X, Y) the first layer carries probability $1/2$ both to (x_0, y_0) and to (x_0, y_1), while the second layer carries a convex combination of $[\beta, 1 - \beta]$ and $[1 - \beta, \beta]$. If we instead restrict our attention to conditional probabilities given $\{W = w_0\}$, we obtain the same structure: two layers where the first layer assigns probability $1/2$ both to (x_0, y_0) and to (x_0, y_1), while the second layer is a convex combination of $[\beta, 1-\beta]$ and $[1 - \beta, \beta]$. And likewise if we restrict our attention to conditional probabilities given $\{W = w_1\}$. Hence W is epistemically irrelevant to (X, Y).

Thus we have epistemic independence of W and (X, Y). However, $\mathbb{P}(w_0|x_1) = 1/2$ for every full distribution in the full credal set; consequently, X is not epistemically irrelevant to W (Decomposition fails), and Y is not epistemically irrelevant to W given X (Weak Union fails). □

Finally, the adapted version of epistemic independence fails Contraction and Intersection even when all lower probabilities are positive (Examples 3 and 4).

So, at least from the point of view of graphoid properties, both confirmational and epistemic independence fare rather poorly. We should strengthen these concepts perhaps using ideas behind type-5 and h-independence.

So we extend h-independence to full credal sets as follows. First, say that Y is *h-irrelevant* to X given Z if

$$\underline{\mathbb{E}}[f(X)|A, B, \{z\}] = \underline{\mathbb{E}}[f(X)|A, \{z\}] \quad \text{whenever } A \cap B \cap \{Z = z\} \neq \emptyset,$$

for any event A in the algebra generated by X and any event B in the algebra generated by Y. Then say that X and Y are *h-independent* given Z when X is h-irrelevant to Y given Z and vice-versa.[16]

For h-independence, with respect to full credal sets, we have:

Theorem 13. *H-independence satisfies Symmetry, Redundancy, Decomposition, and Weak Union.*

Proof. Symmetry holds by definition; Redundancy is trivial using the same arguments employed in results about credal sets. To prove Decomposition, reason as follows. From the assumed h-independence of X and (W, Y), for any A in the algebra generated by X and any B in the algebra generated by Y we obtain the equalities $\underline{\mathbb{E}}[f(X)|A, B, \{z\}] = \underline{\mathbb{E}}[f(X)|A, \{z\}]$, as any B in the algebra generated by Y is also an event in the algebra generated by (W, Y), and $\underline{\mathbb{E}}[g(Y)|A, B, \{z\}] = \underline{\mathbb{E}}[g(Y)|B, \{z\}]$, as any $g(Y)$ is also a function of (W, Y). Weak Union follows from $\underline{\mathbb{E}}[g(Y)|A, B, \{w\}, \{z\}] = \underline{\mathbb{E}}[g(Y)|B, \{w\}, \{z\}]$, and then,

16 This definition was first advocated in Ref. [13] and has been employed by De Bock [18,19].

Table 12
Marginal probabilities from Table 9.

	y_0	y_1
x_0	$\lfloor\alpha\rfloor_0$	$\lfloor 1-\alpha\rfloor_0$
x_1	$\lfloor\alpha\rfloor_1$	$\lfloor 1-\alpha\rfloor_1$

Table 13
Lexicographic marginal probabilities from Table 9.

	y_0	y_1
x_0	$\lfloor\alpha\rfloor_0, \lfloor\beta\rfloor_2$	$\lfloor 1-\alpha\rfloor_0, \lfloor 1-\beta\rfloor_2$
x_1	$\lfloor\alpha\rfloor_1, \lfloor\gamma\rfloor_3$	$\lfloor 1-\alpha\rfloor_1, \lfloor 1-\gamma\rfloor_3$

using Decomposition, we have $\underline{\mathbb{E}}[f(X)|A,\{w\},\{z\}] = \underline{\mathbb{E}}[f(X)|A,\{z\}] = \underline{\mathbb{E}}[f(X)|A,B,\{w\},\{z\}]$ (again A and B are events in appropriate algebras as before). □

From this proof one can see that both direct and reverse versions of Decomposition and Weak Union are satisfied by h-irrelevance. Note also that h-independence fails Contraction (Tables 3, 5, and 9) and Intersection (Table 4).

Having examined concepts of irrelevance/independence that depend on equalities between conditional objects, we now investigate possible concepts that mimic complete independence. First note that a concept of independence that relies on product factorizations is too weak in the context of full conditional probabilities. Indeed we have that the singleton full credal set defined by the full distribution in Table 8 satisfies complete independence!

In any case, complete independence can be adapted to full credal sets as follows. Define *elementwise epistemic/h-/cs-/hcs-/kr-/layer independence* of X and Y given Z to hold when we have epistemic/h-/cs-/hcs-/kr-/layer independence whenever $\{Z=z\} \neq \emptyset$ for each element of the full credal set $\mathbb{K}(X,Y|z)$.[17] Elementwise versions of irrelevance concepts could be contemplated, together with the related direct and reverse versions of graphoid properties; we do not dwell further on such "elementwise irrelevance" notions in this paper.[18]

Given the results mentioned in the previous section, we have:

Proposition 1. *Elementwise epistemic/cs-independence satisfy Symmetry, Redundancy, Decomposition and Contraction (and fail Weak Union). Elementwise h-/hcs-/kr-independence satisfy Symmetry, Redundancy, Decomposition and Weak Union (and fail Contraction). Elementwise layer independence satisfies Symmetry, Redundancy, Decomposition, Weak Union and Contraction.*

One might be interested in extending the concepts of independence in this latter result in the same way that complete independence is extended to strong independence; that is, by only requiring that some credal set is the convex hull of a set satisfying elementwise independence. We do not pursue such a path here, because, as noted before, the reliance on convexity seems misplaced in the context of full conditional probabilities.

4.3. Lexicographic probabilities and sets of desirable gambles

Consider again Table 9. For this full distribution we have X and Y epistemic/h-/cs-/full/kr-/layer independent. One might argue that there is something strange about this "independence". For take a function $g(Y)$ such that $g(y_0) = -(1-\alpha)$ and $g(y_1) = \alpha$; this function has expected utility

$$-(1-\alpha)\alpha + \alpha(1-\alpha) = 0.$$

But if $\beta < \alpha$ one might argue that g is better than the zero function; after all, if $\{w_1\}$ happens to be observed, then the expected value of g given $\{w_1\}$ is $\alpha - \beta$, and g should then be considered better than the zero function.

One way to understand this example is to look at the marginal full conditional probability for (X,Y), shown in Table 12. Note that when the full conditional probability in Table 9 is marginalized over W, the content of layers L_2 and L_3 disappear: in Table 12 one sees neither β nor γ. Preferences about g that might depend on deeper layers can only be exposed by observing W. In a sense, the direct marginalization of Table 9 loses important information about the full conditional probability. It would make more sense to say that the marginal probabilities obtained from Table 9 should be given by the overlapping layers in Table 13, so as to conclude that X and Y are *not* independent.

[17] Note that Coletti and Scozzafava's concept of independence for lower probabilities [5, Definition 6], extended to variables by Vantaggi [52, Definition 7], is quite similar to elementwise cs-independence.
[18] A challenge left for future work is to justify these concepts of irrelevance and independence from behavioral or decision-theoretic arguments, perhaps in the same way that complete independence has been derived using choice functions [11,47].

Table 14
Lexicographic distribution for W, X, Y, with distinct $\alpha \in (0,1)$, $\beta \in (0,1)$, $\gamma \in (0,1)$.

	$w_0 y_0$	$w_1 y_0$	$w_0 y_1$	$w_1 y_1$
x_0	$\lfloor \alpha \rfloor_0$	$\lfloor \beta \rfloor_2$	$\lfloor 1-\alpha \rfloor_0$	$\lfloor 1-\beta \rfloor_2$
x_1	$\lfloor \alpha \rfloor_1$	$\lfloor \beta \rfloor_3$, $\lfloor \gamma \rfloor_4$	$\lfloor 1-\alpha \rfloor_1$	$\lfloor 1-\beta \rfloor_3$, $\lfloor 1-\gamma \rfloor_4$

We are then moving into *lexicographic probabilities* that assign probability measures to various layers with possibly overlapping support. We omit detailed background on lexicographic probabilities, and refer the reader to the work of Blume et al. [4] for extended discussion and justification. We assume their axiomatization of the non-Archimedean preference relation $\succeq$, and use the fact that this preference relation can be represented by a sequence of probability measures over Ω; each one of these measures is a "layer" of the lexicographic probability [4, Corollary 3.1]. Two functions $f_1(X)$ and $f_2(X)$ are compared with respect to a lexicographic rule in the sense that $f_1 \succeq f_2$ if and only if

$$\left[\sum_x f_1(x)\mathbb{P}_i(x)\right]_{i=0}^K \geq_{\mathrm{L}} \left[\sum_x f_2(x)\mathbb{P}_i(x)\right]_{i=0}^K,$$

(for $a, b \in \Re^{K+1}$, $a \geq_{\mathrm{L}} b$ if and only if whenever $b_j > a_j$, there exists a $k < j$ such that $a_k > b_k$). These probabilities are unique only up to linear transformations, so there is some intrinsic non-uniqueness associated with lexicographic probabilities.

Conditional lexicographic probabilities given nonempty event H are obtained by conditioning every layer of the lexicographic probability on H, after discarding those layers that do not intersect H. These conditional probabilities encode the preferences $f_1(X)\mathbb{I}_H \succeq f_2(X)\mathbb{I}_H$ [4, Theorem 4.3], denoted by $[f_1(X) \succeq f_2(X)|H]$.

The close proximity between full conditional probabilities and lexicographic probabilities is apparent. A full conditional probability can be represented by a lexicographic probability with disjoint layers [28,29]. And for any lexicographic probability, the function $\mathbb{P}(A|B) = P_i(A|B)$, where P_i the first measure such that $P_i(B) > 0$, is a full conditional probability. However, as indicated by the discussion of marginalization concerning Tables 9, 12 and 13, full conditional probabilities and lexicographic probabilities do not behave identically when it comes to marginalization.

Now consider defining a concept of independence for lexicographic probabilities. We might try to define a "product" for lexicographic probabilities. Here difficulties abound due to non-uniqueness. First, probabilities in various layers can be linearly combined at will, possibly breaking any factorization. Additionally, probability values are not tied to specific layer numbers. For instance, if we have a lexicographic probability with three overlapping layers, each with probability measures p_0, p_1 and p_2, we can generate an *equivalent* representation with four layers p_0, p_0, p_1 and p_2. Therefore a condition such as layer factorization seems rather fragile as we cannot control layer numbers just by looking at marginal lexicographic probabilities.

Indeed the difficulties with product lexicographic probabilities have already been discussed by several authors [4,29,30]. Solutions based on factorization of nonstandard measures have been advanced by some authors; the interpretation and the manipulation of such concepts do not seem easy, so we leave that avenue to future work.

Hence we are led, in our study of lexicographic probabilities, to concepts of independence that rely on conditioning. Blume, Brandenburger, and Dekel [4] say that X and Y are independent when we have both

$$[f_1(X) \succeq f_2(X)|Y = y_1] \Leftrightarrow [f_1(X) \succeq f_2(X)|Y = y_2]$$

for all $f_1(X)$, $f_2(X)$, y_1 and y_2 such that conditioning events are nonempty, and

$$[g_1(Y) \succeq g_2(Y)|X = x_1] \Leftrightarrow [g_1(Y) \succeq g_2(Y)|X = x_2]$$

for all $g_1(Y)$, $g_2(Y)$, x_1 and x_2 such that conditioning events are nonempty. Say that X and Y are *bbd-independent* given Z when the expressions above are satisfied conditional on any $\{Z = z\}$ such that conditioning events are nonempty.

Even though Table 9 no longer fails Contraction if we use bbd-independence (because X and Y are not bbd-independent), consider Table 14. The distributions for (X, Y), for (X, W) given $\{Y = y_0\}$, and for (X, W) given $\{Y = y_1\}$ are shown in Table 15. Here X and Y are bbd-independent and X and W are bbd-independent given Y; yet X and (W, Y) are not bbd-independent. Contraction fails. The fourth layer "vanishes" when one marginalizes out W as preferences are decided already at the third layer.

Now suppose we have a *set* of lexicographic probabilities, where preference is given by unanimity amongst lexicographic comparisons [48]. Example 11 shows that Decomposition and Weak Union can fail for bbd-independence (just consider each full conditional probability a lexicographic probability).

We suggest that a promising concept of independence for sets of lexicographic probabilities is obtained by starting with the following concept: Y is *lex-irrelevant* to X given Z when

$$[f_1(X) \succeq f_2(X)|A, B, \{z\}] \Leftrightarrow [f_1(X) \succeq f_2(X)|A, \{z\}],$$

Table 15
Marginal (left) and conditional (right) lexicographic probabilities from Table 14.

	y_0	y_1
x_0	$\lfloor \alpha \rfloor_0$, $\lfloor \beta \rfloor_2$	$\lfloor 1-\alpha \rfloor_0$, $\lfloor 1-\beta \rfloor_2$
x_1	$\lfloor \alpha \rfloor_1$, $\lfloor \beta \rfloor_3$	$\lfloor 1-\alpha \rfloor_1$, $\lfloor 1-\beta \rfloor_3$

$\mathbb{P}(W, X|Y = y)$

	w_0	w_1
x_0	$\lfloor 1 \rfloor_0$	$\lfloor 1 \rfloor_2$
x_1	$\lfloor 1 \rfloor_1$	$\lfloor 1 \rfloor_3$

for any event A in the algebra generated by X and any event B in the algebra generated by Y, for all functions $f_1(X)$ and $f_2(X)$, whenever conditioning events are nonempty. And X and Y are *lex-independent* given Z when Y is lex-irrelevant to X given Z and vice-versa. We submit this is an appropriate approach even for single lexicographic probabilities as it better captures the idea behind independence.

We have:

Theorem 14. *Lex-independence satisfies Symmetry, Redundancy, Decomposition, and Weak Union.*

Proof. As before, w is a value of W, x is a value of X, and so on.

Redundancy obtains because, regardless of A in the algebra generated by X and B in the algebra generated by Y,

$$[f_1(X) \succeq f_2(X)|A, B, \{x\}] \Leftrightarrow f_1(x) \geq f_2(x) \Leftrightarrow [f_1(X) \succeq f_2(X)|A, \{x\}];$$

also, for the same A and B,

$$[g_1(Y) \succeq g_2(Y)|A, B, \{x\}] \Leftrightarrow [g_1(Y) \succeq g_2(Y)|B, \{x\}]$$

given that $A \cap \{x\} = \{x\}$ whenever it is not empty.

Decomposition holds because any event B in the algebra generated by Y also belongs to the algebra generated by (W, Y), and any function $g(Y)$ is also a function of (W, Y) (hence lex-independence of X and (W, Y) given Z implies lex-independence of X and Y given Z).

Weak Union holds because, assuming X and (W, Y) lex-independent given Z, we have, for any A in the algebra generated by X and B in the algebra generated by Y,

$$[g_1(Y) \succeq g_2(Y)|A, B, \{w\}, \{z\}] \Leftrightarrow [g_1(Y) \succeq g_2(Y)|B, \{w\}, \{z\}];$$

also, using Decomposition,

$$[f_1(X) \succeq f_2(X)|A, \{w\}, \{z\}] \Leftrightarrow [f_1(X) \succeq f_2(X)|A, \{z\}]$$
$$\Leftrightarrow [f_1(X) \succeq f_2(X)|A, B, \{w\}, \{z\}]. \quad \square$$

Note that lex-independence fails Contraction (Table 14).

Sets of lexicographic probabilities are equivalent, from the point of view of preference representations, to *sets of desirable gambles*, a representation that has received considerable attention [8,22,24,40,56]. Indeed the derivation of lexicographic representations for sets of desirable gambles appears already in the work of Seidenfeld et al. [48], who show that a partially ordered set of preferences (that encodes a set of desirable gambles) can be represented by a set of complete orderings, each one of which can be represented by a lexicographic probability (either using results by Klee [32] or the more direct results by Blume et al. [4]). In recent work, Couso and Moral [8] have studied the representation of sets of desirable gambles through lexicographic probabilities. A detailed analysis of the isomorphism between sets of lexicographic probabilities and sets of desirable gambles has been produced by Benavoli et al. [3].

A set of desirable gambles $\mathbb{D}$ is a set of variables not containing the zero function and containing all nonnegative variables that are different from zero, and such that $\lambda X \in \mathbb{D}$ if $X \in \mathbb{D}$ and $\lambda > 0$, and $X + Y \in \mathbb{D}$ if $X, Y \in \mathbb{D}$ [22, Definition 1]. The set of desirable gambles conditional on event H, denoted by $[\mathbb{D}|H]$, contains all of those variables X in $\mathbb{D}$ such that $X\mathbb{I}_H = X$ [24, Section 3.2]. Following notation by Moral [40], denote by $\mathbb{D}^{\downarrow X}$ the set of variables in $\mathbb{D}$ that are functions of X (that is, $\mathbb{D}^{\downarrow X}$ is the "marginal" set of gambles with respect to X).

A natural concept of independence for sets of desirable gambles is [22, Definition 3]: Y is irrelevant to X given Z if

$$[\mathbb{D}|y, z]^{\downarrow X} = [\mathbb{D}|z]^{\downarrow X} \text{ whenever } \{Y = y, Z = z\} \neq \emptyset.$$

And then: X and Y are independent given Z if X is irrelevant to Y given Z and vice-versa. (Note that there are other concepts of independence for sets of desirable gambles in the literature [40].)

Mimicking our proposal for (sets of) lexicographic probabilities, consider instead the following definition of independence for sets of desirable gambles: Y is *dg-irrelevant* to X given Z if

$$[\mathbb{D}|A, B, \{z\}]^{\downarrow X} = [\mathbb{D}|A, \{z\}]^{\downarrow X} \text{ whenever } A \cap B \cap \{Z = z\} \neq \emptyset,$$

for any A in the algebra generated by X and B in the algebra generated by Y. And then define *dg-independence* of X and Y given Z by symmetrizing this concept of irrelevance. Once more we obtain Symmetry, Redundancy, Decomposition and Weak Union by exploring the isomorphism between sets of desirable gambles and sets of lexicographic probabilities.

5. Conclusion

We have examined the behavior of concepts of independence, studied from the perspective of graphoid properties, for formalisms that extend probability theory in several directions. In doing so we obtained results that were divided in two parts. In Section 3 we examined in detail credal sets (understood here as sets of probability measures) coupled with concepts of independence that mimic, to the extent that it is possible, stochastic independence. In Section 4 we considered various formalisms that depart from probability measures: sets of full conditional probabilities, sets of lexicographic probabilities, sets of desirable gambles.

Most of Section 3 is devoted to variants of epistemic irrelevance and epistemic independence. All variants adopt regular conditioning as it is arguably the counterpart to the standard definition of conditioning. We have shown that regular-epistemic irrelevance and regular-confirmational irrelevance satisfy a few versions of semi-graphoid properties (both forms of Redundancy, "direct" forms of Decomposition and Weak Union, and a "reverse" form of Contraction). Their corresponding concepts of independence only satisfy Symmetry and Redundancy; to obtain Decomposition and Weak Union, convexity and positive lower probabilities seem necessary. We have also shown that type-5 epistemic irrelevance and type-5 irrelevance in addition satisfy "reverse" forms of Decomposition and Weak Union, and consequently type-5 epistemic independence and type-5 independence satisfy Symmetry, Redundancy, Decomposition and Weak Union. All of these concepts of independence fail Contraction and Intersection. The failure of Decomposition and Weak Union for regular-epistemic and regular-confirmational independence is a frustrating result, as one would expect these concepts to deal smoothly with zero probabilities. The more stringent conditions of the "type-5 family" may be needed in any practical circumstances. In Section 3 we also considered complete and strong independence, showing that the former satisfies all semi-graphoid properties while the latter satisfies all such properties except Contraction.

In Section 4 we first looked at full credal sets (understood here as sets of full conditional probabilities), and then we extended our analysis to sets of lexicographic probabilities and sets of desirable gambles. To summarize, we first noted that versions of epistemic/confirmational independence are rather weak in the context of full credal sets, and investigated the stronger and more satisfying concept of h-independence for full credal sets. This concept satisfies Symmetry, Redundancy, Decomposition and Weak Union. We then investigated the graphoid properties of "elementwise" versions of independence for full credal sets (where an existing concept of independence is imposed on every element of the full credal set), with varying levels of compliance as far as graphoid properties are concerned. Finally, we transferred the definition of h-independence to sets of lexicographic probabilities and to sets of desirable gambles. Much more can be said about lexicographic probabilities and desirable gambles; our purpose here was in essence to extend to them some of relevant insights we obtained with respect to full credal sets.

Overall, it seems reasonable to assume that a concept of independence should satisfy Symmetry, Redundancy, Decomposition and Weak Union. The Contraction property seems more debatable. The "reverse" version of Contraction certainly has appeal: if one judges W irrelevant to X given some Y that is itself irrelevant to X, then both W and Y should be irrelevant to X [44]. But "direct" Contraction, the version that often fails for non-symmetric notions of irrelevance, strangely demands that, if X is irrelevant to W given some Y that X is irrelevant to, then X should be irrelevant to both W and Y. Similarly, Intersection hardly seems to be a mandatory property; it fails even for stochastic independence, so it is a property that is not likely to be missed by other formalisms.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The author has been partially supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), grant 312180/2018-7. The work was also supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), grant 2019/07665-4, and also by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - finance code 001. The work was also supported in part by the Center for Data Science (C2D) with funding by Itaú Unibanco (project FUSP3268). Part of the work was done at the Center for Artificial Intelligence (c4ai.inova.usp.br - C4AI).

References

[1] Thomas Augustin, Frank P.A. Coolen, Gert de Cooman, Matthias C.M. Troffaes, Introduction to Imprecise Probabilities, Wiley, 2014.
[2] Pierpaolo Battigalli, Pietro Veronesi, A note on stochastic independence without Savage-null events, J. Econ. Theory 70 (1) (1996) 235–248.

[3] Alessio Benavoli, Alessandro Facchini, Marco Zaffalon, José Vicente-Pérez, A polarity theory for sets of desirable gambles, in: International Symposium on Imprecise Probability: Theories and Applications, in: PMLR, vol. 62, 2017, pp. 37–48.
[4] Lawrence Blume, Adam Brandenburger, Eddie Dekel, Lexicographic probabilities and choice under uncertainty, Econometrica 58 (1) (January 1991) 61–79.
[5] G. Coletti, R. Scozzafava, Stochastic independence in a coherent setting, Ann. Math. Artif. Intell. 35 (2002) 151–176.
[6] Giulianella Coletti, Romano Scozzafava, Probabilistic Logic in a Coherent Setting, Trends in Logic, vol. 15, Kluwer, Dordrecht, 2002.
[7] I. Couso, S. Moral, P. Walley, Examples of independence for imprecise probabilities, in: G. de Cooman, Fabio G. Cozman, Serafín Moral, Peter Walley (Eds.), International Symposium on Imprecise Probabilities and Their Applications, Ghent, Belgium, 1999, pp. 121–130, Imprecise Probabilities Project, Universiteit Gent.
[8] Inés Couso, Serafín Moral, Sets of desirable gambles: conditioning, representation, and precise probabilities, Int. J. Approx. Reason. 52 (2011) 1034–1055.
[9] Fabio G. Cozman, Irrelevance and independence axioms in quasi-Bayesian theory, in: Anthony Hunter, Simon Parsons (Eds.), European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU, Springer, London, England, 1999, pp. 128–136.
[10] Fabio G. Cozman, Teddy Seidenfeld, Independence for full conditional measures and their graphoid properties, in: Benedikt Lowe, Eric Pacuit, Jan-Willem Romeijn (Eds.), Reasoning About Probabilities and Probabilistic Reasoning, in: Foundations of the Formal Sciences VI, vol. 16, College Publications, London, 2009, pp. 1–29.
[11] Fabio G. Cozman, Sets of probability distributions, independence, and convexity, Synthese 186 (2) (2012) 577–600.
[12] Fabio G. Cozman, Independence for full conditional probabilities: structure, factorization, non-uniqueness, and Bayesian networks, Int. J. Approx. Reason. 54 (2013) 1261–1278.
[13] Fabio G. Cozman, Independence for sets of full conditional measures, sets of lexicographic probabilities, and sets of desirable gambles, in: International Symposium on Imprecise Probability: Theories and Applications, 2013, pp. 87–98.
[14] Fabio G. Cozman, Cassio Polpo de Campos, Kuznetsov independence for interval-valued expectations and sets of probability distributions: properties and algorithms, Int. J. Approx. Reason. 55 (2014) 666–682.
[15] Fabio G. Cozman, Peter Walley, Graphoid properties of epistemic irrelevance and independence, Ann. Math. Artif. Intell. 45 (2005) 173–195.
[16] Ákos Császár, Sur la structure des espaces de probabilité conditionnelle, Acta Math. Acad. Sci. Hung. 6 (3) (1955) 337–361.
[17] A.P. Dawid, Conditional independence in statistical theory (with discussion), J. R. Stat. Soc. B 41 (1979) 1–31.
[18] Jasper de Bock, Reintroducing credal networks under epistemic irrelevance, in: Proceedings of PGM 2016, in: Proceedings of Machine Learning Research, vol. 52, 2016, pp. 123–135.
[19] Jasper de Bock, Credal networks under epistemic irrelevance, Int. J. Approx. Reason. 85 (2017) 107–138.
[20] Jasper de Bock, Gert de Cooman, A behavioural justification for using sensitivity analysis in imprecise multinomial processes, J. Math. Anal. Appl. 468 (1) (2018) 513–546.
[21] L. de Campos, S. Moral, Independence concepts for convex sets of probabilities, in: Phillippe Besnard, Steve Hanks (Eds.), XI Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, California, United States, 1995, pp. 108–115.
[22] Gert de Cooman, Enrique Miranda, Independent natural extension for sets of desirable gambles, J. Artif. Intell. Res. 45 (2012) 601–640.
[23] Gert de Cooman, Enrique Miranda, Marco Zaffalon, Independent natural extension, Artif. Intell. 175 (2011) 1911–1950.
[24] Gert de Cooman, Erik Quaeghebeur, Exchangeability and sets of desirable gambles, Int. J. Approx. Reason. 53 (2012) 363–395.
[25] Bruno de Finetti, Theory of Probability, vol. 1-2, Wiley, New York, 1974.
[26] Lester E. Dubins, Finitely additive conditional probability, conglomerability and disintegrations, Ann. Stat. 3 (1) (1975) 89–99.
[27] F.J. Giron, S. Rios, Quasi-Bayesian behaviour: a more realistic approach to decision making?, in: J.M. Bernardo, J.H. DeGroot, D.V. Lindley, A.F.M. Smith (Eds.), Bayesian Statistics, University Press, Valencia, Spain, 1980, pp. 17–38.
[28] Joseph Y. Halpern, Lexicographic probability, conditional probability, and nonstandard probability, Games Econ. Behav. 68 (2010) 155–179.
[29] Peter J. Hammond, Elementary non-Archimedean representations of probability for decision theory and games, in: P. Humphreys (Ed.), Patrick Suppes: Scientific Philosopher, vol. 1, Kluwer, Dordrecht, The Netherlands, 1994, pp. 25–59.
[30] Peter J. Hammond, Non-Archimedean subjective probabilities in decision theory and games, Technical Report Working Paper No. 97-038, Stanford University Department of Economics, 1997.
[31] John Maynard Keynes, A Treatise on Probability, Macmillan and Co., London, 1921.
[32] V.L. Klee, The structure of semispaces, Math. Scand. 4 (1956) 54–64.
[33] Elon Kohlberg, Philip J. Reny, Independence on relative probability spaces and consistent assessments in game trees, J. Econ. Theory 75 (1997) 280–313.
[34] Peter Krauss, Representation of conditional probability measures on Boolean algebras, Acta Math. Acad. Sci. Hung. 19 (3–4) (1968) 229–241.
[35] V.P. Kuznetsov, Interval Statistical Methods, Radio i Svyaz Publ., 1991 (in Russian).
[36] H.E. Kyburg, M. Pittarelli, Set-based Bayesianism, IEEE Trans. Syst. Man Cybern. A 26 (3) (May 1996) 324–339.
[37] Isaac Levi, The Enterprise of Knowledge, MIT Press, Cambridge, Massachusetts, 1980.
[38] Vann McGee, Learning the impossible, in: E. Bells, B. Skyrms (Eds.), Probability and Conditionals, Cambridge University Press, 1994, pp. 179–199.
[39] S. Moral, A. Cano, Strong conditional independence for credal sets, Ann. Math. Artif. Intell. 35 (2002) 295–321.
[40] Serafín Moral, Epistemic irrelevance on sets of desirable gambles, Ann. Math. Artif. Intell. 45 (1–2) (October 2005) 197–214.
[41] Roger B. Myerson, Multistage games with communication, Econometrica 54 (2) (1986) 323–358.
[42] Roger B. Myerson, Game Theory: Analysis of Conflict, Harvard University Press, Cambridge, MA, 1991.
[43] Judea Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, California, 1988.
[44] Judea Pearl, Causality: Models, Reasoning, and Inference, 2nd edition, Cambridge University Press, Cambridge, United Kingdom, 2009.
[45] Karl R. Popper, The Logic of Scientific Discovery, Hutchinson, London, 1975.
[46] Alfréd Renyi, On a new axiomatic theory of probability, Acta Math. Acad. Sci. Hung. 6 (1955) 285–335.
[47] Teddy Seidenfeld, Mark Schervish, Joseph Kadane, Coherent choice functions under uncertainty, in: International Symposium on Imprecise Probability: Theories and Applications, Prague, Czech Republic, 2007.
[48] Teddy Seidenfeld, Mark J. Schervish, Joseph B. Kadane, Decisions without ordering, in: W. Sieg (Ed.), Acting and Reflecting, Kluwer Academic Publishers, 1990, pp. 143–170.
[49] Milan Studeny, Probabilistic Conditional Independence Structures, Springer-Verlag, London, 2005.
[50] B.C. van Fraassen, Fine-grained opinion, probability, and the logic of full belief, J. Philos. Log. 24 (1995) 349–377.
[51] Barbara Vantaggi, Conditional independence in a coherent finite setting, Ann. Math. Artif. Intell. 32 (1–4) (2001) 287–313.
[52] Barbara Vantaggi, Conditional independence structures and graphical models, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 11 (5) (2003) 545–571.
[53] Barbara Vantaggi, Graphical representation of asymmetric graphoid structures, in: Third International Symposium on Imprecise Probabilities and Their Applications, Carleton Scientific, 2003, pp. 560–574.
[54] P. Walley, T.L. Fine, Towards a frequentist theory of upper and lower probability, Ann. Stat. 10 (3) (1982) 741–761.
[55] Peter Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.
[56] Peter Walley, Towards a unified theory of imprecise probability, Int. J. Approx. Reason. 24 (2000) 125–148.
[57] Kurt Weichselberger, The theory of interval-probability as a unifying concept for uncertainty, Int. J. Approx. Reason. 24 (2–3) (2000) 149–170.
[58] Kurt Weichselberger, T. Augustin (assistant), A. Wallner (assistant), Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I: Intervallwahrscheinlichkeit als umfassendes Konzept, Physica-Verlag Heidelberg, 2001.