Semi-Graphoid Properties of Variants of Epistemic Independence Based on Regular Conditioning

Fabio Gagliardi Cozman　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　FGCOZMAN@USP.BR
Escola Politécnica – Universidade de São Paulo, Brazil

Abstract

Graphoid properties attempt to capture the most important features of abstract "independence". We examine which semi-graphoid properties are satisfied by various concepts of independence for credal sets; we focus on variants of epistemic, confirmational, and type-5 independence that are based on regular conditioning.

Keywords: Graphoid properties. Credal sets. Epistemic irrelevance and independence.

1. Introduction

The word "independence" is used in many fields with different meanings: there is independence induced by graph separation; there is logical independence; there is stochastic independence in probability theory. In an attempt to capture *the* essential features of abstract "independence", graphoid properties have been proposed in a variety of settings [9, 15, 17]. In particular, graphoid properties have been studied in connection with concepts of independence for lower previsions are related modeling tools [8, 6]; however, many questions remain open on which graphoid properties are satisfied by various concepts of independence.

In this paper we focus on concepts of independence that apply to sets of Kolmogorovian probability measures. We examine regular confirmational, regular epistemic, regular type-5, and regular type-5 epistemic irrelevance/independence. Results in this paper should be useful in comparing these concepts to each other and also in comparing sets of Kolmogorovian probability measures with other modeling tools in the literature; in particular, with approaches that deal with sets of probabilities where conditioning on events of zero probability is allowed [8, 18, 19].

In Sections 2 and 3 we review needed concepts and notation. Sections 4 and 5 respectively look at epistemic/confirmational irrelevance/independence and at type-5 irrelevance/independence. Section 6 summarizes the results and briefly comments on complete, strong, and Kuznetsov independence.

2. Basic Concepts and Notation

We assume that the space of outcomes Ω is finite, and that every subset of Ω is 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.

Some notational conventions will be used [6]. Throughout we use W, X, Y and Z to denote random variables. We adopt: w denotes a possible value of W, x denotes a possible value of X, y denotes a possible value of Y, z denotes a possible value of Z. And $\{x\}$ denotes the nonempty event $\{\omega \in \Omega : X(\omega) = x\}$; likewise for $\{w\}$, $\{y\}$ and $\{z\}$.

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.

The intersection of events G and H is written as $G \cap H$, but also as GH and as G, H. When the event $\{x\}$ appears in an intersection, we remove braces whenever possible; for instance, xG denotes the event $\{x\} \cap G$. Sometimes we add braces to enhance clarity; for instance, we may write $\{y, z\}$ instead of simply y, z when we refer to the event $\{y\} \cap \{z\}$.

When w, x, y, z appear in expressions, they are universally quantified unless explicitly noted (that is, the starting expression "for all possible values $w\ldots$" is implicit). Likewise, when functions f and g appear in expressions, they are universally quantified unless explicitly noted (that is, the expression "for all functions $f\ldots$" is implicit).

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. Given two events G and H, the conditional probability of G given H, denoted by $\mathbb{P}(G|H)$, is only defined when $\mathbb{P}(H) > 0$; if so, $\mathbb{P}(G|H) = \mathbb{P}(GH) / \mathbb{P}(H)$. *Conditional stochastic independence* of random variables X and Y given random variable Z obtains when $\mathbb{P}(x, y|z) = \mathbb{P}(x|z)\mathbb{P}(y|z)$ whenever $\mathbb{P}(z) > 0$.

We are interested in concepts of *conditional* independence. If X and Y are conditionally independent given a constant function Z, we have "unconditional" independence, and in this case we simply write "independent" without mentioning the "conditional" qualifier and without any mention of Z. Often we will just write "independence" to mean both conditional and unconditional independence.

A set of (Kolmogorovian-style) 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, 11, 14].

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 $\overline{\mathbb{E}}[f(X)] = \sup_{\mathbb{P} \in \mathbb{K}} \mathbb{E}_{\mathbb{P}}[f(X)]$, where $\mathbb{E}_{\mathbb{P}}[f(X)]$ is the expectation of $f(X)$ with respect to $\mathbb{P}$. 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 $\overline{\mathbb{P}}(H) = \sup_{\mathbb{P} \in \mathbb{K}} \mathbb{P}(H)$.

The *graphoid properties* that may apply to a ternary relation $(\cdot \perp\!\!\!\perp \cdot \,|\, \cdot)$ are [9, 15, 17]

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: $(X \perp\!\!\!\perp Y \,|\, Z) \quad \wedge \quad (X \perp\!\!\!\perp W \,|\, (Y,Z)) \quad \Rightarrow \quad (X \perp\!\!\!\perp (W,Y) \,|\, Z)$

Intersection $(X \perp\!\!\!\perp W \,|\, (Y,Z)) \quad \& \quad (X \perp\!\!\!\perp Y \,|\, (W,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \,|\, Z)$.

Conditional stochastic independence satisfies all graphoid properties except Intersection; for this reason no concept of independence in this paper satisfies Intersection, and we do not look further at Intersection. Thus this paper is in fact interested in the *semi-graphoid* properties; this is the set consisting of the graphoid properties except Intersection.

3. Conditioning and Independence

Several concepts of independence depend on conditioning. 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 $\underline{\mathbb{P}}(H) = 0$ [11]. A different concept of conditional credal set focuses on those probability measures that assign positive probability to H:

$$\mathbb{K}^{>}(X|H) = \{\mathbb{P}(\cdot|H) : \mathbb{P} \in \mathbb{K}(X) \text{ and } \mathbb{P}(H) > 0\} \atop \text{whenever } \overline{\mathbb{P}}(H) > 0, \quad (1)$$

with $\mathbb{K}^{>}(X|H)$ undefined whenever $\overline{\mathbb{P}}(H) = 0$ [22, 23]. We refer to this second conditioning strategy as *regular conditioning*, as it is closely related to the concept of regular extension [21, 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 if $\mathbb{K}(X)$ is closed [7].

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 [20] that, whenever $\overline{\mathbb{P}}(H) > 0$,

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

where I_H denotes the indicator function of H. This can be slightly generalized to the following result, used later (this result mimics Lemma 1 of Ref. [4]):

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

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

Proof We have:

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

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

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

$$\mathbb{K}^{\triangleright}(X|y,z) = \mathbb{K}^{\triangleright}(X|z) \quad \text{whenever } \underline{\mathbb{P}}(y,z) > 0. \quad (2)$$

(Recall our conventions: by implicit quantification, this equality is required for all y, z). However, by using the "$\triangleright$-conditioning" we may face the problem that no conditioning can be applied when all values of (y,z) have zero lower probability, a situation that is clearly possible.

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

$$\mathbb{K}^{>}(X|y,z) = \mathbb{K}^{>}(X|z) \quad \text{whenever } \overline{\mathbb{P}}(y,z) > 0. \quad (3)$$

---

1. In Ref. [5] the concept of confirmational irrelevance (and of epistemic irrelevance) is defined instead using the same condition employed here to define type-5 irrelevance: that is, all events B specified by Y may appear in conditioning, as opposed to events $\{Y = y\}$. It seems to make sense to differentiate the condition used here to specify confirmational irrelevance from the one used later to specify type-5 irrelevance, as they satisfy different sets of semi-graphoid properties.

Walley's concept of *epistemic irrelevance* is similar to Levi's confirmational irrelevance: Y is epistemically irrelevant to X when $\mathbb{E}[f(X)|y] = \mathbb{E}[f(X)]$ [21, Chapter 9]. Say that Y is *regular-epistemically irrelevant* to X given Z when

$$\mathbb{E}^{>}[f|y,z] = \mathbb{E}^{>}[f|z] \quad \text{whenever } \overline{\mathbb{P}}(y,z) > 0. \quad (4)$$

(Recall our conventions: f is a function of X; by implicit quantification, this equality is required for all f, for all y, z).

Both regular-confirmational irrelevance and regular-epistemic irrelevance fail Symmetry. Take the following example by Couso et al. [3, Example 3]: variables X and Y binary, with $\mathbb{P}(x_0) \in [1/2, 4/5]$ and $\mathbb{P}(y_i|x_j) \geq 3/10$ for all $i \in \{0,1\}$ and $j \in \{0,1\}$; clearly X is regular-confirmationally/epistemically irrelevant to Y, but $\mathbb{P}(x_0|y_0) \in [3/10, 28/31]$, hence Y is not regular-confirmationally/epistemically irrelevant to X.

Walley's clever response to failure of Symmetry, borrowed from the work of Keynes [12], was to "symmetrize" irrelevance to obtain independence. So, 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 [21].

We can apply Walley's idea to other non-symmetric concepts: for instance, say that X and Y are *regular-confirmationally independence* 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 [10]: they say Y is *type-5* irrelevant to X if $\mathbb{K}^{>}(X|B) = \mathbb{K}(X)$ whenever $\overline{\mathbb{P}}(B) > 0$. (Recall our convention: B is an event in the algebra generated by Y). Accordingly, say that Y is *type-5 irrelevant* to X given Z if

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

Type-5 irrelevance is an strenghtened version of confirmational irrelevance. Now we can also strenghten epistemic irrelevance, and say that Y is *type-5 epistemically irrelevant* to X given Z if

$$\mathbb{E}^{>}[f|B,z] = \mathbb{E}^{>}[f|z] \text{ whenever } \overline{\mathbb{P}}(B,z) > 0.$$

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

4. Semi-Graphoid Properties: Epistemic/Confirmational Independence

We start with the semi-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 properties. We have:

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

Proof Redundancy: we do have, whenever $\overline{\mathbb{P}}(X = x_1, X = x_2) > 0$, that $\mathbb{E}^{>}[g|X = x_1, X_2 = x_2] = \mathbb{E}^{>}[g|X = x_1]$ (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 $\mathbb{E}^{>}[f|x,y] = f(x) = \mathbb{E}^{>}[f|x]$.

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

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

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

Contraction: if $\mathbb{E}^{>}[f|w,y,z] = \mathbb{E}^{>}[f|y,z]$ whenever $\overline{\mathbb{P}}(w, y, z) > 0$ and $\mathbb{E}^{>}[f|y,z] = \mathbb{E}^{>}[f|z]$ whenever $\overline{\mathbb{P}}(y, z) > 0$, then, by transitivity, $\mathbb{E}^{>}[f|w,y,z] = \mathbb{E}^{>}[f|z]$ whenever $\overline{\mathbb{P}}(w, y, z) > 0$, as desired. ∎

The graphoid properties of epistemic irrelevance have been studied before, both ignoring the possibility of zero lower probabilities [4] and assuming that lower probabilities may be zero but adopting the framework of lower previsions, where conditioning is allowed on events of zero lower probability [8]. The latter study showed that, with such conditioning, direct and reverse Redundancy, direct Decomposition and reverse Contraction are satisfied by epistemic irrelevance even if lower probabilities are zero, while direct Weak Union and reverse versions of Decomposition and Weak Union are satisfied when appropriate lower probabilities are larger than zero (with positivity conditions, these properties also hold for regular-epistemic irrelevance). Thus Theorem 2 indicates a difference (that is, direct Weak Union) between regular-epistemic independence and Walley's original concept of independence.

All other forms of semi-graphoid properties one can think of fail even with positivity conditions (that is, forms that are not mentioned in Theorem 2). For instance, consider Decomposition: other than the "direct" form in Theorem 2, there are three other possibilities:
- If $(X \mathrel{\mathsf{IR}} W, Y \mid Z)$, then $(Y \mathrel{\mathsf{IR}} X \mid Z)$;
- If $(W, Y \mathrel{\mathsf{IR}} X \mid Z)$, then $(X \mathrel{\mathsf{IR}} Y \mid Z)$;

	w_0y_0	w_0y_1	w_1y_0	w_1y_1
x_0	$\frac{1-\alpha}{2}$	$\frac{1-\alpha}{2}$	0	0
x_1	$\alpha/2$	$\alpha/2$	0	0

	w_0y_0	w_0y_1	w_1y_0	w_1y_1
x_0	0	0	$\frac{1-\alpha}{2}$	$\alpha/2$
x_1	0	0	$\alpha/2$	$\frac{1-\alpha}{2}$

	w_0y_0	w_0y_1	w_1y_0	w_1y_1
x_0	0	0	$\frac{1-\alpha}{2}$	$\alpha/2$
x_1	0	0	$\frac{1-\alpha}{2}$	$\alpha/2$

Table 1: Tables employed in Example 1.

● 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 next example demonstrates that the last version also fails; the example is used later to show failure of other properties.

Example 1 *Consider three binary variables W, X, and Y. For some selected $\beta \in (0,1/2)$, build six distributions, using the values in Table 1 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 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. □

Now consider Weak Union: 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 version fail due to failure of Symmetry. The last example demonstrates that the last version also fails.

Finally, consider Contraction: there are eight possible versions; six of them lead to failure of Symmetry by taking either $W = Z = 1$ or $W = Y = 1$, and a reverse version is proved in Theorem 2. The remaining version is a direct one:

	w_0y_0	w_0y_1	w_1y_0	w_1y_1
x_0	$\alpha(1-\alpha)^2$	$\alpha(1-\alpha)/2$	$(1-\alpha)^3$	$\alpha(1-\alpha)/2$
x_1	$\alpha^2(1-\alpha)$	$\alpha^2/2$	$\alpha(1-\alpha)^2$	$\alpha^2/2$

	w_0y_0	w_0y_1	w_1y_0	w_1y_1
x_0	$\alpha/4$	$\alpha/4$	$(1-\alpha)/4$	$(1-\alpha)/4$
x_1	$(1-\alpha)/4$	$\alpha/4$	$\alpha/4$	$(1-\alpha)/4$

Table 2: Tables employed in Example 2.

● $(X \text{ IR } Y \mid Z) \wedge (X \text{ IR } W \mid (Y,Z)) \Rightarrow (X \text{ IR } (W,Y) \mid Z)$.
Ref. [4] presents a numerical example that violates this property, while Ref. [8] goes over several constructions that also lead to failure of direct Contraction. The next example is hopefully simpler to grasp.

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

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 $\alpha = 1-\beta$. Finally build a third joint probability distribution by using the bottom table with $\alpha = \beta$. And build the joint credal set that is the convex hull of these three 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 posible 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 $\underline{\mathbb{P}}(w_0,y_0|x_0) < \underline{\mathbb{P}}(w_0,y_0)$.

As for regular-epistemic independence, Symmetry clearly holds, and Theorem 2 implies Redundancy. As noted in previous work [8], Decomposition and Weak Union hold when appropriate lower probabilities are larger

than zero; this result holds here as Walley's form of conditioning is identical to regular conditioning when lower probabilities are larger than zero. However, both Decomposition and Weak Union may fail when lower probabilities are allowed to be zero, as illustrated by Example 1. And Contraction fails in Example 2, where lower probabilities are even positive.

We now move to regular-confirmattional irrelevance. We start by proving the counterpart of Theorem 2 (note the absence of assumptions concerning closure/convexity/positivity).

Theorem 3 *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 for* IR.

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, 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, $\mathbb{K}_1 = \mathbb{K}_2 \implies \{F(p) : p \in \mathbb{K}_1\} = \{F(p) : p \in \mathbb{K}_2\}$.

Proof Redundancy: we do have, whenever $\overline{\mathbb{P}}(X = x_1, X = x_2) > 0$, that $\mathbb{K}^>(Y | X = x_1, X = x_2) = \mathbb{K}^>(Y | X = x_1)$ (true because if $x_1 = x_2$, then trivially $\overline{\mathbb{P}}(X = x_1, X = x_2) = \{X = x_1\}$); also, whenever $\overline{\mathbb{P}}(x, y) > 0$, we have $\mathbb{K}^>(X | X = x, Y = y) = \mathbb{K}^>(X | X = x)$ 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 $\{x, z\}$ can be represented by a vector p containing the probability value of each $\{w, y\}$. 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 the marginalization function $F(p)$ that yields a vector containing, for each value y, the value of the sum $\sum_w \mathbb{P}(w, y | x, z)$. 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: For every p whose encoded distribution satisfies $\mathbb{P}(w | x, z) > 0$, define the conditioning function $F(p)$ that yields a vector containing, for each value w, the value of the ratio $\mathbb{P}(w, y | x, z) / \sum_{y'} \mathbb{P}(w, y' | x, z)$. So, the equality $\mathbb{K}^>(W, Y | x, z) = \mathbb{K}^>(W, Y | z)$ whenever $\overline{\mathbb{P}}(x, z) > 0$ implies the equality $\mathbb{K}^>(Y | x, w, z) = \mathbb{K}^>(Y | w, z)$ whenever $\overline{\mathbb{P}}(x, w, z) > 0$ (by applying $F(\cdot)$ 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 1 and arguments given right after that example. And all other versions of Contraction fail for regular-confirmational irrelevance given Example 2 and arguments given right before that example.

Example 1 also shows that Decomposition and Weak Union fail for regular-confirmational independence, and Example 2 shows that Contraction fails for regular-confirmational independence. Hence we only have Symmetry and Redundancy for regular-confirmational independence.

If all credal sets are closed and convex, and all lower probabilities are larger than zero, then Decomposition and Weak Union hold — in this case all credal sets are exactly represented by their associated lower expectations, and the same arguments used to show Decomposition and Weak Union under positivity assumptions for regular-epistemic independence hold.

We might also ask whether Decomposition and Weak Union hold when credal sets are convex and probabilities are larger than zero (even if *lower* probabilities may be zero). But consider:

Example 3 *Take the interior of the joint credal set in Example 1; 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 $\{[1/2, 1/2]\}$. 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-convex credal sets [5]. That example used non-connected credal sets; here is a perhaps easier to grasp 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 4 *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 $\mathbb{P}_1^\alpha(X = x|w,y) = A^\alpha$ and $\mathbb{P}_1^\alpha(w,y) = 1/4$, and another set of distributions such that $\mathbb{P}_2^\alpha(X = x|w,y) = B^\alpha$ and $\mathbb{P}_2^\alpha(w,y) = 1/4$. Continue by defining an additional sets of distributions parameterized by $\alpha \in [0,1]$, as follows:

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

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

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

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

However,

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

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

Now suppose we require both convexity and positive lower probabilities, but drop the assumption of closure. It is an open question whether Decomposition and Weak Union hold in this case.

These examples suggest that regular-confirmational and regular-epistemic independence are rather weak concepts: we cannot keep Decomposition and Weak Union when lower probabilities may be zero.

5. Semi-Graphoid Properties: Type-5 and Type-5 Epistemic Independence

Now consider type-5 and type-5 epistemic irrelevance. For type-5 epistemic irrelevance we have:
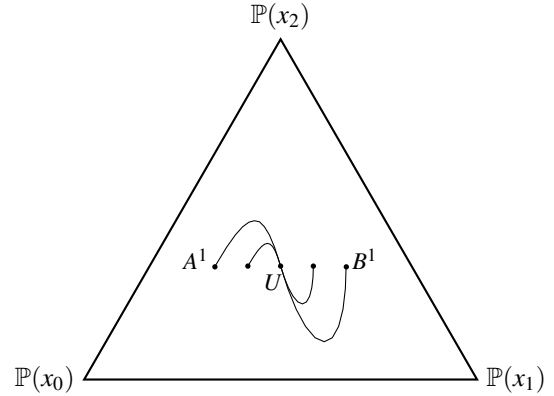

Figure 1: Credal sets from Example 4, in barycentric coordinates. The marginal credal set $\mathbb{K}(X)$ appears as the longer curve from A^1 to B^1, going through the uniform distribution U. The credal sets $\mathbb{K}(X|y_0)$ and $\mathbb{K}(X|y_1)$ include the shorter curve that goes from $(A^1 + U)/2$ to $(B^1 + U)/2$.

Theorem 4 *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 Redundancy: we do have, whenever $\overline{\mathbb{P}}(X = x, A) > 0$, that $\underline{\mathbb{E}}^>[g|X = x, A] = \underline{\mathbb{E}}^>[g|X = x]$ (true because if $x \in A$, then trivially $\{X = x, A\} = \{X = x\}$); also, whenever $\overline{\mathbb{P}}(x, B) > 0$, we have $\underline{\mathbb{E}}^>[f|x, B] = f(x) = \underline{\mathbb{E}}^>[f|x]$.

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

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

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

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

	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	0	$\alpha/2$	$\alpha/2$	0
x_1	0	$(1-\alpha)/2$	$(1-\alpha)/2$	0

	$w_0 y_0$	$w_0 y_1$	$w_1 y_0$	$w_1 y_1$
x_0	$\alpha/2$	0	0	$(1-\alpha)/2$
x_1	$(1-\alpha)/2$	0	0	$\alpha/2$

Table 3: Tables employed in Example 5.

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

All other versions of Decomposition and Weak Union fail for type-5 epistemic irrelevance given the "symmetry-based" arguments given right after Example 1. The same arguments there cover all other versions of Contraction except direct Contraction (fails in Example 2) and reverse Contraction. The latter property fails in the next example

Example 5 *Consider three binary variables W, X and Y. For Table 3 top, $\mathbb{P}(x_0) = \mathbb{P}(x_0|y) = \alpha$ for each possible y, and $\mathbb{P}(x_0|w,y) = \alpha$ whenever $\mathbb{P}(w,y) > 0$. For the bottom table we have $\mathbb{P}(x_0) = 1/2$, $\mathbb{P}(x_0|y_0) = \alpha$, $\mathbb{P}(x_0|y_1) = 1-\alpha$, $\mathbb{P}(x_0|w_0,y_0) = \alpha$ and $\mathbb{P}(x_0|w_1,y_1) = 1-\alpha$.*

Select $\beta \in (0, 1/2)$. First build two joint probability distributions for (W,X,Y) by using the top table respectively with $\alpha = \beta$ and $\alpha = 1-\beta$. Then build two additional joint probability distributions by using the bottom table respectively 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, because $\mathbb{K}^{>}(X|\{w_0,y_0\} \cup \{w_1,y_1\})$ is a singleton containing the probability distribution such that $\mathbb{P}(x_0|\{w_0,y_0\} \cup \{w_1,y_1\}) = 1/2$. □

For type-5 epistemic independence we get, using Theorem 4:[2]

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

---

2. This result corrects a mistake in Ref. [5], where it is stated that type-5 epistemic independence fails Decomposition and Weak Union.

Clearly type-5 epistemic independence fails Contraction and Intersection.

Finally, consider type-5 irrelevance. We have the counterpart of Theorem 4:

Theorem 6 *If $(Y \mathbin{\text{IR}} X \mid Z)$ denotes type-5 irrelevance of Y to X given Z, then the same properties listed in Theorem 4 hold.*

Proof Redundancy: There are two equalities to prove:

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

(true because if $\{A,x\} = \emptyset$, then $\overline{\mathbb{P}}(A,x) = 0$; otherwise $\{A,x\} = \{x\}$, then trivially $\mathbb{K}^{>}(X|A,x) = \mathbb{K}^{>}(X|x)$), and

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

(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 $\{A,z\}$ can be represented by a vector p containing the probability value of each $\{w,y\}$. 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 sum $\sum_w \mathbb{P}(w,y|A,z)$. 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.

Direct Weak Union: For every p whose encoded distribution satisfies $\mathbb{P}(w|A,z) > 0$, define the conditioning function $F(p)$ that yields a vector containing, for each value w, the value of the ratio $\mathbb{P}(w,y|A,z)/\sum_{y'}\mathbb{P}(w,y'|A,z)$. 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 $F(\cdot)$ 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 $\overline{\mathbb{P}}(B,z) > 0$, 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 $\overline{\mathbb{P}}(B,w,z) > 0$; 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. ∎

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

Theorem 7 *Type-5 independence satisfies Symmetry, Redundancy, Decomposition and Weak Union.*

6. Conclusion

We have examined the behavior of variants of epistemic irrelevance and independence with respect to semi-graphoid properties. All variants adopt regular conditioning and thus are arguably counterparts to the standard definition of stochastic independence. 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. 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 all of them fail 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.

To conclude, we note that a popular way to express the (standard, Kolmogorovian) independence of X and Y given Z is to require $\mathbb{P}(x,y|z) = \mathbb{P}(x|z)\mathbb{P}(y|z)$ whenever $\mathbb{P}(z) > 0$. 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* iff each probability distribution in $\mathbb{K}(X,Y)$ satisfy stochastic independence; the conditional version is simply produced by conditioning everything on some Z. And say that X and Y are *strongly independent* iff $\mathbb{K}(X,Y)$ is the convex hull of a credal set that satisfies complete independence. A natural scheme to define conditional strong independence, suggested by Walley [20], is to say that X and Y are *strongly independent* given Z iff $\mathbb{K}^{>}(X,Y|z)$ is the convex hull of a credal set satisfying complete independence of X and Y, for all z such that $\overline{\mathbb{P}}(z) > 0$. The last notable concept of independence we mention for credal sets is due to Kuznetsov [13]: 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.

Complete independence satisfies all semi-graphoid properties. As for strong independence, one might suspect that it satisfies all semi-graphoid properties. However, as shown by the next example, strong independence fails Contraction.[3]

---

3. This corrects a statement in Ref. [5] concerning strong independence and Contraction.

	w_0y_0	w_0y_1	w_1y_0	w_1y_1
x_0	2	3	6	1
x_1	4	6	12	2

	w_0y_0	w_0y_1	w_1y_0	w_1y_1
x_0	17	39	51	13
x_1	25	51	75	17

Table 4: Example 6. Top: the values of $36 \times \mathbb{P}_2$. Bottom: the values of $288 \times \mathbb{P}_3$.

Example 6 *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 $\mathbb{P}_3$. Take $\mathbb{P}_1(w,x,y) = 1/8$ (that is, the uniform distribution) and $\mathbb{P}_2(w,x,y) = \mathbb{P}_2(w|y)\mathbb{P}_2(x)\mathbb{P}_2(y)$ 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 $\mathbb{P}_3(w,x,y) = \mathbb{P}_2(w|y)(\mathbb{P}_1(x)\mathbb{P}_1(y)/2 + \mathbb{P}_2(x)\mathbb{P}_2(y)/2)$. The distributions $\mathbb{P}_2$ and $\mathbb{P}_3$ are shown in Table 4. 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, $\mathbb{P}_3(x,y)$ 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, $\mathbb{P}_3(x_0|w_0,y_0) = 17/42 \neq 5/12 = \mathbb{P}_3(x_0)$.* □

Finally, Kuznetsov-independence satisfies Symmetry, Redundancy and Decomposition; it fails Contraction even when all probabilities are positive [7], and it is an open question whether it satisfies Weak Union or not.

Acknowledgments

References

[1] T. Augustin, F. P. A. Coolen, G. de Cooman, and M. C. M. Troffaes. *Introduction to Imprecise Probabilities*. Wiley, 2014.

[2] M. S. Bazaraa and C. M. Shetty. *Foundations of Optimization*, volume 122 of *Mathematical Programming*. Springer, Berlin, 1976.

[3] I. Couso, S. Moral, and P. Walley. Examples of independence for imprecise probabilities. In G. de Cooman, F. G. Cozman, S. Moral, and P. Walley, editors, *International Symposium on Imprecise Probabilities and Their Applications*, pages 121–130, 1999. Imprecise Probabilities Project, Universiteit Gent.

[4] F. G. Cozman. Irrelevance and independence axioms in quasi-bayesian theory. In A. Hunter and S. Parsons, editors, *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, pages 128–136. Springer, London, England, 1999.

[5] F. G. Cozman. Sets of probability distributions, independence, and convexity. *Synthese*, 186(2):577–600, 2012.

[6] F. 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*, pages 87–98, 2013.

[7] F. G. Cozman and C. Polpo de Campos. Kuznetsov independence for interval-valued expectations and sets of probability distributions: Properties and algorithms. *International Journal of Approximate Reasoning*, 55: 666–682, 2014.

[8] F. G. Cozman and P. Walley. Graphoid properties of epistemic irrelevance and independence. *Annals of Mathematics and Artificial Intelligence*, 45:173–195, 2005.

[9] A. P. Dawid. Conditional independence in statistical theory (with discussion). *Journal of the Royal Statistical Society B*, 41:1–31, 1979.

[10] L. de Campos and S. Moral. Independence concepts for convex sets of probabilities. In P. Besnard and S. Hanks, editors, *Conference on Uncertainty in Artificial Intelligence*, pages 108–115, 1995. Morgan Kaufmann.

[11] F. J. Giron and S. Rios. Quasi-Bayesian behaviour: A more realistic approach to decision making? In J. M. Bernardo, J. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics*, pages 17–38. University Press, Valencia, Spain, 1980.

[12] J. M. Keynes. *A Treatise on Probability*. Macmillan and Co., London, 1921.

[13] V. P. Kuznetsov. *Interval Statistical Methods*. Radio i Svyaz Publ., (in Russian), 1991.

[14] I. Levi. *The Enterprise of Knowledge*. MIT Press, Cambridge, Massachusetts, 1980.

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

[16] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[17] M. Studeny. *Probabilistic Conditional Independence Structures*. Springer Verlag, 2005.

[18] B. Vantaggi. Conditional independence in a coherent finite setting. *Annals of Mathematics and Artificial Intelligence*, 32(1-4):287–313, 2001.

[19] B. Vantaggi. Graphical representation of asymmetric graphoid structures. *International Symposium on Imprecise Probabilities and Their Applications*, pages 560–574, 2003.

[20] P. Walley and T. L. Fine. Towards a frequentist theory of upper and lower probability. *Annals of Statistics*, 10(3):741–761, 1982.

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

[22] K. Weichselberger. The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2-3):149–170, 2000.

[23] K. Weichselberger, T. Augustin (assistant), and A. Wallner (assistant). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I: Intervallwahrscheinlichkeit als umfassendes Konzept*. Physica-Verlag Heidelberg, 2001.