Independence for Full Conditional Measures and their Graphoid Properties

Fabio G. Cozman
Universidade de Sao Paulo

Teddy Seidenfeld
Carnegie Mellon University

October 7, 2008

Abstract

This paper examines definitions of independence for events and variables in the context of full conditional measures; that is, when conditional probability is a primitive notion and conditioning is allowed on null events. Independence concepts are evaluated with respect to graphoid properties; we show that properties of weak union, contraction and intersection may fail when null events are present.

Keywords: Graphoids, stochastic independence, full conditional measures.

1 Introduction

In this paper we wish to consider independence concepts associated with *full conditional measures* [23]. That is, we wish to allow conditioning to be a primitive notion, defined even on *null events* (events of zero probability). The advantages of a probability theory that takes conditioning as a primitive have been explored by many authors, such as de Finetti [20] and his followers [10, 16], Rényi [46], Krauss [37] and Dubins [23]. A significant number of philosophers have argued for serious consideration of null events [1, 29, 38, 40, 44, 48, 49, 50, 51], as have economists and game theorists [7, 8, 31, 32, 33, 41]. Null events, or closely related concepts, have also appeared repeatedly in the literature on artificial intelligence: ranking and ordinal measures [17, 28, 49, 59] have direct interpretations as "layers" of full conditional measures [10, 26, 42]; some of the most general principles of default reasoning can be interpreted through various types of lexicographic probabilities [5, 6, 25, 43, 36]; and as a final example, the combination of probabilities and logical constraints in expert systems must often deal with zero probabilities [2, 9, 10, 21].

The goal of this paper is to compare concepts of independence for events and variables in the context of full conditional measures. Our strategy is to evaluate concepts of independence by the *graphoid* properties they satisfy (Section 3 reviews the theory of graphoids). This strategy is motivated by two observations. First, the graphoid properties have been often advocated as a compact set of properties that any concept of independence should satisfy. Even though some of the graphoid properties may have more limited scope than others, they offer a good starting point for discussions of independence. Second, the graphoid properties are useful in proving results about conditional probabilities, graphs, lattices, and other models [19]. In Sections 4 and 5 we analyze existing and new concepts of independence. We show that several key graphoid properties can fail due to null events.

2 Full Conditional Measures

In this paper, in order to avoid controversies about countable additivity for probability, we restrict ourselves to finite state spaces: $\Omega = \{\omega_1, \ldots, \omega_N\}$; any subset of Ω is an event. We use A, B, C to denote events and W, X, Y, Z to denote (sets of) random variables; by $A(X)$, $B(Y)$, $C(X)$ and $D(Y)$ we denote events defined either by X or by Y. Events such as $\{X = x\} \cap \{Y \neq y\} \cap \{Z = z\}$ are denoted simply as $\{X = x, Y \neq y\, Z = z\}$.

A probability measure is a set function $P : 2^\Omega \to \Re$ such that $P(\Omega) = 1$, $P(A) \geq 0$ for all A, and $P(A \cup B) = P(A) + P(B)$ for disjoint A and B. Given a probability measure, the probability of A conditional on B is usually defined to be $P(A \cap B)/P(B)$ when $P(B) > 0$; conditional probability is not defined if B is a null event. *Stochastic independence* of events A and B requires that $P(A \cap B) = P(A)\,P(B)$; or equivalently that $P(A|B) = P(A)$ when $P(B) > 0$. *Conditional stochastic independence* of events A and B given event C requires that $P(A \cap B|C) = P(A|C)\,P(B|C)$ if C is non-null. These concepts of independence can be extended to sets of events and to random variables by requiring more complex factorizations [22].

A different theory of probability ensues if we take conditional probability as a truly primitive concept, as already advocated by Keynes [35] and de Finetti [20]. The first question is the domain of probabilistic assessments. Rényi [46] investigates the general case where $P : \mathcal{B} \times \mathcal{C} \to \Re$, where $\mathcal{B}$ is a Boolean algebra and $\mathcal{C}$ is an arbitrary subset of $\mathcal{B}$ (Rényi also requires σ-additivity). Popper considers a similar set up [44]. Here we focus on $P : \mathcal{B} \times (\mathcal{B} \backslash \emptyset) \to \Re$, where $\mathcal{B}$ is again a Boolean algebra [37], such that for every event $C \neq \emptyset$:
(1) $P(C|C) = 1$;
(2) $P(A|C) \geq 0$ for all A;
(3) $P(A \cup B|C) = P(A|C) + P(B|C)$ for all disjoint A and B;
(4) $P(A \cap B|C) = P(A|B \cap C)\,P(B|C)$ for all A and B such that $B \cap C \neq \emptyset$.
This fourth axiom is often stated as $P(A|C) = P(A|B)\,P(B|C)$ when $A \subseteq B \subseteq C$ and $B \neq \emptyset$ [4]. We refer to such a P as a *full conditional measure*, following Dubins [23]; there are other names in the literature, such as *conditional probability measure* [37] and *complete conditional probability system* [41]. Whenever the conditioning event C is equal to Ω, we suppress it and write the "unconditional" probability $P(A)$ instead of $P(A|\Omega)$.

Full conditional measures place no restrictions on conditioning on null events. If B is null, the constraint $P(A \cap B) = P(A|B)\,P(B)$ is trivially true, and $P(A|B)$ must be defined separately from $P(B)$ and $P(A \cap B)$. For any two events A and B, indicate by $A \gg B$ the fact that $P(B|A \cup B) = 0$. Then we can partition Ω into events $L_0, \ldots, L_K$, where $K \leq N$, such that $L_i \gg L_{i+1}$ for $i \in \{0, \ldots, K-1\}$ if $K > 0$. Each event L_i is a "layer" of P, and the decomposition in layers always exists for any finite algebra. Coletti and Scozzafava's denote by $\circ(A)$ the index i of the first layer L_i such that $P(A|L_i) > 0$; they propose the convention $\circ(\emptyset) = \infty$ [10]. They also refer to $\circ(A)$ as the *zero-layer* of A; here we will use the term *layer level* of A for the same purpose. Note that some authors use a different terminology, where the ith "layer" is $\cup_{j=i}^{K} L_j$ rather than L_i [10, 37].

	A	A^c
B	$\lfloor \beta \rfloor_1$	α
B^c	$\lfloor 1-\beta \rfloor_1$	$1-\alpha$

Table 1: A simple full conditional measure ($\alpha, \beta \in (0,1)$) where $P(A) = 0$ and $P(B|A) = \beta$.

Coletti and Scozzafava also define the conditional layer number $\circ(A|B)$ as $\circ(A \cap B) - \circ(B)$ (defined only if $\circ(B)$ is finite).

Any full conditional measure can be represented as a sequence of strictly positive probability measures $P_0, \ldots, P_K$, where the support of P_i is restricted to L_i; that is, $P_i : 2^{L_i} \to \Re$. This result is proved assuming complete assessments in general spaces (not just finite) by Krauss [37] and Dubins [23], and it has been derived for partial assessments by Coletti [12, 10].

We have, for events A, B:

- $P(B|A) = P\Big(B|A \cap L_{\circ(A)}\Big)$ [4, Lemma 2.1a].
- $\circ(A \cup B) = \min(\circ(A), \circ(B))$.
- Either $\circ(A) = 0$ or $\circ(A^c) = 0$.

The following simple result will be useful later.

Lemma 1 *Consider two random variables X and Y, event $A(X)$ defined by X and event $B(Y)$ defined by Y such that $A(X) \cap B(Y) \neq \emptyset$. If $P(Y = y|\{X = x\} \cap B(Y)) = P(Y = y|B(Y))$ for every $x \in A(X)$ such that $\{X = x\} \cap B(Y) \neq \emptyset$, then $P(Y = y|A(X) \cap B(Y)) = P(Y = y|B(Y))$.*

Proof. We have (all summations run over $\{x \in A(X) : x \cap B(Y) \neq \emptyset\}$):

$$\begin{aligned}
P(Y = y|A(X) \cap B(Y)) &= \sum P(X = x, Y = y|A(X) \cap B(Y)) \\
&= \sum P(Y = y|\{X = x\} \cap A(X) \cap B(Y)) \times \\
&\qquad\qquad P(X = x|A(X) \cap B(Y)) \\
&= \sum P(Y = y|\{X = x\} \cap B(Y))\, P(X = x|A(X) \cap B(Y)) \\
&= \sum P(Y = y|B(Y))\, P(X = x|A(X) \cap B(Y)) \\
&= P(Y = y|B(Y)) \sum P(X = x|A(X) \cap B(Y)) \\
&= P(Y = y|B(Y))\,.\square
\end{aligned}$$

The following notation will be useful. If A is such that $\circ(A) = i$ and $P(A|L_i) = p$, we write $\lfloor p \rfloor_i$. If $\circ(A) = 0$ and $P(A) = p$, we simply write p instead of $\lfloor p \rfloor_0$. Table 1 illustrates this notation.

There are several decision theoretic derivations of full conditional measures. The original arguments of de Finetti concerning called-off gambles [20] have been formalized in several ways [34, 45, 57, 58]. Derivations based on axioms on preferences have also been presented, both by Myerson [41] and by Blume et al [7]. The last derivation is particularly interesting as it is based on non-Archimedean preferences and lexicographic preferences.

3 Graphoids

If we read $(X \perp\!\!\!\perp Y \,|\, Z)$ as "variable X is stochastically independent of variable Y given variable Z," then the following properties are true:

Symmetry: $(X \perp\!\!\!\perp Y \,|\, Z) \Rightarrow (Y \perp\!\!\!\perp X \,|\, Z)$

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

Weak union: $(X \perp\!\!\!\perp (W, Y) \,|\, Z) \Rightarrow (X \perp\!\!\!\perp W \,|\, (Y, Z))$

Contraction: $(X \perp\!\!\!\perp Y \,|\, Z)$ & $(X \perp\!\!\!\perp W \,|\, (Y, Z)) \Rightarrow (X \perp\!\!\!\perp (W, Y) \,|\, Z)$

Instead of interpreting $\perp\!\!\!\perp$ as stochastic independence, we could take this relation to indicate an abstract concept of independence. The properties just outlined are then referred to as the *graphoid properties*, and any three-place relation that satisfies these properties is called a *graphoid*. Note that we are following terminology proposed by Geiger et al's [27]; the term "graphoid" often means slightly different concepts [18, 42]. In fact, often the four properties just listed are called *semi-graphoid* properties, and the term "graphoid" is reserved to a relation that satisfies the semi-graphoid properties plus:

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

As the intersection property can already fail for stochastic independence in the presence of null events [14], it is less important than the other properties in the context of the present paper. But the intersection property is important for understanding Basu's First Theorem of statistical inference, as shown by San Martin et al [47].

Finally, the following property is sometimes presented together with the previous ones [42]:

Redundancy: $(X \perp\!\!\!\perp Y \,|\, X)$

If we interpret W, X, Y and Z as *sets* of variables, redundancy implies that any property that is valid for disjoint sets of variables is also valid in general — because given symmetry, redundancy, decomposition and contraction, $(X \perp\!\!\!\perp Y \,|\, Z) \Leftrightarrow (X \backslash Z \perp\!\!\!\perp Y \backslash Z \,|\, Z)$ as noted by Pearl [42].

Graphoids offer a compact and intuitive abstraction of independence. As an example of application, several key results in the theory of Bayesian networks can be proved just using the graphoid properties, and are consequently valid for many possible generalizations of Bayesian networks [27]. Several authors have employed the graphoid properties as a benchmark to evaluate concepts of independence [3, 55, 52]; we follow the same strategy in this paper.

4 Epistemic and coherent irrelevance and independence

Because conditional probabilities are defined even on null events, we might consider a concise definition of independence: events A and B are independent iff $P(A|B) = P(A)$. However, this definition is not entirely satisfactory because it guarantees neither $P(A|B^c) = P(A)$ nor $P(B|A) = P(B)$ (failure of symmetry can be observed in Table 1). In this section we study the graphoid properties of three concepts of irrelevance/independence that attempt to correct these deficiencies. We collect results on *epistemic* and *strong coherent* irrelevance/independence, and then we explore the new concepts of *weak coherent* irrelevance/independence.

4.1 Epistemic irrelevance/independence

Keynes faced the problem of non-symmetric independence, in his theory of probability, by defining first a concept of *irrelevance* and then "symmetrizing" it [35]: B is irrelevant to A iff $P(A|B) = P(A)$; A and B are independent iff A is irrelevant to B and B is irrelevant to A. Walley strenghtened Keynes' definitions in his theory of imprecise probabilities: B is irrelevant to A iff $P(A|B) = P(A|B^c) = P(A)$; independence is the symmetrized concept [56].[1] Note that Crisma has further strenghtened Walley's definitions by requiring logical independence [16] (we later discuss logical independence in more detail).

We follow Walley in using *epistemic irrelevance* of B to A to mean

$$P(A|B) = P(A) \text{ if } B \neq \emptyset \qquad \text{and} \qquad P(A|B^c) = P(A) \text{ if } B^c \neq \emptyset. \tag{1}$$

Epistemic independence refers to the symmetrized concept. Clearly epistemic irrelevance/independence can be extended to sets of events, random variables, and to concepts of conditional irrelevance/independence [56]. We wish to focus on:

Definition 1 *Random variables X are epistemically irrelevant to random variables Y conditional on random variables Z (denoted by $(X \text{ EIR } Y \mid Z)$) if $P(Y = y|\{X = x, Z = z\}) = P(Y = y|Z = z)$ for all values x, y, z whenever these probabilities are defined.*

Epistemice independence, denoted using similar triplets with the symbol EIN , is the symmetrized concept.

We now consider the relationship between these concepts and the graphoid properties. Because irrelevance is not symmetric, there are several possible versions of the properties that might be of interest. For example, two different versions of weak union are $(X \text{ EIR } (W, Y) \mid Z) \Rightarrow (X \text{ EIR } W \mid (Y, Z))$ and $((W, Y) \text{ EIR } X \mid Z) \Rightarrow (W \text{ EIR } X \mid (Y, Z))$, and there are two additional possible versions. Decomposition also has four versions, while contraction and intersection have eight versions each. We single out two versions for each property, which we call the *direct* and the *reverse* versions. The direct version is obtained by writing the property as initially stated in Section 3, just replacing $\perp\!\!\!\perp$ by EIR . The reverse version is obtained by switching every statement of irrelevance. Thus we have given respectively the direct and reverse versions of weak union in this paragraph (similar distinctions have appeared in the literature for various concepts of irrelevance [15, 24, 39, 52, 55]).

The following proposition relates epistemic irrelevance/independence with the graphoid properties (several results in the proposition can be extracted from Vantaggi's results [52]).

Proposition 1 *Epistemic irrelevance satisfies the graphoid properties of direct and reverse redundancy, direct and reverse decomposition, reverse weak union, and direct and reverse contraction. If W and Y are logically independent, then epistemic irrelevance satisfies reverse intersection. All other versions of the graphoid properties and intersection fail for epistemic irrelevance. Epistemic independence satisfies symmetry, redundancy, decomposition and contraction — weak union and intersection fail for epistemic independence.*

[1] Levi has also proposed $P(A|B) = P(A)$ as a definition of *irrelevance*, without considering the symmetrized concept [38]. Both Levi's and Walley's definitions are geared towards sets of full conditional measures, but clearly they specialize to a single full conditional measure.

	w_0y_0	w_1y_0	w_0y_1	w_1y_1
x_0	α	$\lfloor\beta\rfloor_2$	$1-\alpha$	$\lfloor 1-\beta\rfloor_2$
x_1	$\lfloor\alpha\rfloor_1$	$\lfloor\gamma\rfloor_3$	$\lfloor 1-\alpha\rfloor_1$	$\lfloor 1-\gamma\rfloor_3$

Table 2: Failure of direct weak union for epistemic irrelevance/independence ($\alpha, \beta, \gamma \in (0,1)$, with $\alpha \neq \beta \neq \gamma$). The full conditional measure in the table satisfies $(X \text{ EIN } (W,Y))$ but fails $(X \text{ EIR } Y \mid W)$.

	w_0y_0	w_1y_0	w_0y_1	w_1y_1
x_0	$\lfloor 1\rfloor_3$	$\lfloor\beta\rfloor_1$	$\lfloor(1-\beta)\rfloor_1$	$\lfloor 1\rfloor_5$
x_1	$\lfloor 1\rfloor_2$	α	$(1-\alpha)$	$\lfloor 1\rfloor_4$

Table 3: Failure of versions of intersection for epistemic irrelevance/independence ($\alpha, \beta \in (0,1)$, with $\alpha \neq \beta$). The full conditional measure in the table satisfies $(X \text{ EIN } W \mid Y)$ and $(X \text{ EIN } Y \mid W)$, but not $(X \text{ EIR } (W,Y))$.

Proof. For epistemic irrelevance, the proof of direct and reverse redundancy, direct and reverse decomposition, reverse weak union and reverse contraction is obtained from the proof of Theorem 1, by taking $A(X) = B(Y) = \Omega$. For direct contraction ($(X \text{ EIR } Y \mid Z)$ & $(X \text{ EIR } W \mid (Y,Z)) \Rightarrow (X \text{ EIR } (W,Y) \mid Z)$), consider that if $(X,Z) \neq \emptyset$,

$$P(W,Y|X,Z) = P(W|X,Y,Z)\, P(Y|X,Z) = P(W|Y,Z)\, P(Y|Z) = P(W,Y|Z)\,,$$

where the term $P(W|X,Y,Z)\, P(Y|X,Z)$ is only defined when (X,Y,Z) is nonempty (if it is empty, then the other equalities are valid because both sides are equal to zero). All other versions of graphoid properties fail for epistemic irrelevance, as shown by measures in Tables 2, 3, 7, and 8. For epistemic independence, symmetry is true by definition; redundancy, decomposition and contraction follow from their direct and reverse versions for epistemic irrelevance; Table 2 displays failure of weak union, and Table 3 displays failure of intersection. □

4.2 Strong coherent irrelevance/independence

Coletti and Scozzafava have proposed a concept of independence that explicitly deals with layers [13, 11, 52]. They define the following condition:

Definition 2 *The* Coletti-Scozzafava condition *on (B, A) holds iff $B \neq \emptyset \neq B^c$* and *$\circ(A|B) = \circ(A|B^c)$ and $\circ(A^c|B) = \circ(A^c|B^c)$, where these layer numbers are computed with respect to probabilities over the events $\{A \cap B, A \cap B^c, A^c \cap B, A^c \cap B^c\}$.*

The Coletti-Scozzafava condition focuses only on the four events $\{A \cap B, A \cap B^c, A^c \cap B, A^c \cap B^c\}$; consequently, it never deals with layer numbers larger than 3.[2] Situations that violate

[2] We thank Andrea Capotorti for bringing this fact to our attention.

	A	A^c
B	$\lfloor 1 \rfloor_1$	α
B^c	$\lfloor 1 \rfloor_2$	$1-\alpha$

	A	A^c
B	1	$\lfloor 1 \rfloor_2$
B^c	$\lfloor 1 \rfloor_1$	$\lfloor 1 \rfloor_4$

Table 4: Violations of the Coletti-Scozzafava condition, where $\alpha \in (0, 1)$.

the Coletti-Scozzafava condition are depicted in Table 4.[3]

Coletti and Scozzafava then define independence of B to A as: satisfaction of the Coletti-Scozzafava condition on (B, A) *plus* epistemic irrelevance of B to A. The concept is not symmetric (Table 1).

Another important aspect of the Coletti-Scozzafava condition is that if B (or B^c) is empty, then B is not deemed irrelevant to any other event. Coletti and Scozzafava argue that their condition deals adequately with logical independence, as follows. Consider a table containing layer numbers for events $\{A \cap B, A \cap B^c, A^c \cap B, A^c \cap B^c\}$. Every entry can be either finite or infinite; thus we have 16 configurations insofar as this distinction is concerned. One of these configurations is impossible (the one with four infinite entries). The other configurations can be divided in the cases depicted in Table 5. The Coletti-Scozzafava condition blocks the irrelevance of B to A in the second, third, fourth, sixth and seventh tables. Most cases of logical independence are thus removed. (However, the fifth table does not fail the Coletti-Scozzafava condition even though displays a situation of logical dependence,[4] as noted by Coletti and Scozzafava [10, Proposition 2].)

Vantaggi extends the main ideas in Coletti and Scozzafava's concept of independence by proposing a condition that can be stated as follows:

Definition 3 *The* conditional Coletti-Scozzafava condition *on variables* (Y, X) *given variables* Z *holds iff for all values* x*,* y*,* z*,*

$$\{Y = y, Z = z\} \neq \emptyset \quad \textit{and} \quad \{Y \neq y, Z = z\} \neq \emptyset,$$

and

$$\circ(X = x|Y = y, Z = z) = \circ(X = x|Y \neq y, Z = z)$$

and

$$\circ(X \neq x|Y = y, Z = z) = \circ(X \neq x|Y \neq y, Z = z)\,,$$

where these layer numbers are computed with respect to $\{\{X = x, Y = y, Z = z\}, \{X = x, Y \neq y, Z = z\}, \{X \neq x, Y = y, Z = z\}, \{X \neq x, Y \neq y, Z = z\}\}$*.*

Vantaggi then proposes a concept, here referred to as *strong coherent irrelevance*:[5] Y is strongly coherently irrelevant to X given Z if (i) Y is epistemically irrelevant to X given Z, and (ii) the

[3]The example in Table 4 (right) fails the following result by Coletti and Scozzava [10, Theorem 17]: if A and B are logically independent, then $P(A|B) = P(A|B^c)$ and $P(B|A) = P(B|A^c)$ imply $\circ(A^c|B) = \circ(A^c|B^c)$. But this result is true when we focus on the restricted set of events $\{A \cap B, A \cap B^c, A^c \cap B, A^c \cap B^c\}$.

[4]We thank Barbara Vantaggi for bringing this fact to our attention.

[5]It should be noted that Vantaggi's concepts of independence correspond to irrelevance in our terminology (she also studies special cases where symmetry holds [54]). Also, Vantaggi's writing of properties is different from ours; for example, her reverse weak union property is our direct weak union property.

conditional Coletti-Scozzafava condition holds on (Y, X) given Z [52, Definition 7.1]. This is a very stringent concept, as strong coherent irrelevance requires Y and Z to be logically independent.

As shown by Vantaggi [52], strong coherent irrelevance fails symmetry, direct and reverse redundancy, and direct weak union. Her results imply that a symmetrized concept of strong coherent independence fails redundancy and weak union (Table 2). Note also that strong coherent irrelevance fails direct intersection, as shown by Table 3.

4.3 The layer condition and weak coherent irrelevance/independence

In this sub-section we consider a condition on layer numbers that relaxes the Coletti-Scozzafava condition in two ways. First, we remove the restriction on the set of events that must be taken into account to determine layer numbers, Second, we replace conjunction with material implication in the case B (or B^c) is empty, as this replacement will lead to a host of interesting properties. In Section 6 we comment on our reasons to make these changes and on their consequences.

Definition 4 *The* layer condition *on* (B, A) *holds iff whenever* $B \neq \emptyset \neq B^c$ *then* $\circ(A|B) = \circ(A|B^c)$ *and* $\circ(A^c|B) = \circ(A^c|B^c)$.

The first thing to note is that the layer condition is symmetric.[6]

Proposition 2 *If* A *and* B *satisfy the layer condition on* (B, A)*, they satisfy the layer condition on* (A, B).

Proof. Define $a = \circ(A \cap B)$; $b = \circ(A \cap B^c)$; $c = \circ(A^c \cap B)$; $d = \circ(A^c \cap B^c)$. Each one of these four layer levels may be finite or infinite. There are thus 16 situations; one is impossible and six others always violate the layer condition. Suppose then that no layer level is infinite and assume $a = 0$, as one of the four entries must be 0. The first table in Table 5 illustrates this case. Regardless of the value of b, $\circ(B^c) = b$ because $\circ(A \cap B^c) - \circ(B^c) = 0$ by hypothesis. Then $c - 0 = d - b$, thus $d = c + b$, and the table is symmetric. Now, if $\circ(A \cap B) \neq 0$, then we can always re-label rows and columns so that the top left entry is zero, and the same reasoning follows. The remaining cases are illustrated by the fifth, sixth, and seventh tables in Table 5, where the result is immediate. □

Consequently, it is enough to indicate that two events satisfy the layer condition, without mentioning a "direction" (B, A) or (A, B). We have:

Proposition 3 *Events* A *and* B *satisfy the layer condition iff* $\circ(A|B) = \circ(A)$, $\circ(A|B^c) = \circ(A)$, $\circ(A^c|B) = \circ(A^c)$ *and* $\circ(A^c|B^c) = \circ(A^c)$ *whenever the relevant quantities are defined.*

Proof. Direct by verifying all tables in Table 5 (it may be necessary to re-label rows and columns to deal with the 9 relevant situations discussed in the proof of Proposition 2). □

The previous result directly implies equivalence of the layer condition and a more obviously symmetric condition:[7]

[6] The Coletti-Scozzavafa condition is not symmetric, as shown by the sixth table in Table 5.

[7] This result was suggested to us by Matthias Troffaes.

	A	A^c
B	a	b
B^c	c	d

	A	A^c
B	a	∞
B^c	c	d

	A	A^c
B	a	∞
B^c	∞	d

	A	A^c
B	∞	b
B^c	c	∞

	A	A^c
B	a	∞
B^c	c	∞

	A	A^c
B	a	b
B^c	∞	∞

	A	A^c
B	a	∞
B^c	∞	∞

Table 5: Cases for the Coletti-Scozzafava and layer conditions.

Corollary 1 *Events A and B satisfy the layer condition iff*

$$\circ(A \cap B) = \circ(A) + \circ(B)\,, \qquad \circ(A \cap B^c) = \circ(A) + \circ(B^c)\,,$$

$$\circ(A^c \cap B) = \circ(A^c) + \circ(B)\,, \qquad \circ(A^c \cap B^c) = \circ(A^c) + \circ(B^c)\,.$$

We now consider a version of the layer condition for random variables, clearly similar to Vantaggi's conditional Coletti-Scozzafava condition:[8]

Definition 5 *The* conditional layer condition *on variables (Y, X) given variables Z holds iff for all values x, y, z, whenever $\{Y = y, Z = z\} \neq \emptyset$ and $\{Y \neq y, Z = z\} \neq \emptyset$, then*

$$\circ(X = x|Y = y, Z = z) = \circ(X = x|Y \neq y, Z = z)$$

and

$$\circ(X \neq x|Y = y, Z = z) = \circ(X \neq x|Y \neq y, Z = z)\,.$$

Before we propose concepts of irrelevance/independence based on the conditional layer condition, we examine some useful properties of this condition.

Proposition 4 *The conditional layer condition is equivalent to*

$$\circ(X = x|Y = y, Z = z) = \circ(X = x|Z = z) \tag{2}$$

for all x, y, z such that $\{Y = y, Z = z\} \neq \emptyset$.

Proof. Assume the conditional layer condition; using Proposition 3 for every $\{Z = z\}$, we obtain $\circ(X = x|Y = y, Z = z) = \circ(X = x|Z = z)$ for all x, y and z such that $\{Y = y, Z = z\} \neq \emptyset$. Now assume Expression (2), and denote by $B(Y)$ an event defined by Y such that $B(Y) \cap \{Z = z\} \neq \emptyset$. Then:

$$\circ(X = x|B(Y) \cap \{Z = z\}) \quad = \quad \min_{y \in B(Y)} \circ(X = x, Y = y|Z = z) - \min_{y \in B(Y)} \circ(Y = y|Z = z)$$

[8]We note that Vantaggi's Definition 7.3 [52] uses a condition that is very close to the conditional layer condition in Definition 5; the difference basically is that Vantaggi requires every layer number to be computed with respect to $\{\{X = x, Y = y, Z = z\}, \{X = x, Y \neq y, Z = z\}, \{X \neq x, Y = y, Z = z\}, \{X \neq x, Y \neq y, Z = z\}\}$.

$$
\begin{aligned}
&= \min_{y \in B(Y)} (\circ(X = x|Y = y, Z = z) + \circ(Y = y|Z = z)) - \\
&\quad \min_{y \in B(Y)} \circ(Y = y|Z = z) \\
&= \min_{y \in B(Y)} (\circ(X = x|Z = z) + \circ(Y = y|Z = z)) - \\
&\quad \min_{y \in B(Y)} \circ(Y = y|Z = z) \\
&= \circ(X = x|Z = z) + \min_{y \in B(Y)} \circ(Y = y|Z = z) - \\
&\quad \min_{y \in B(Y)} \circ(Y = y|Z = z) \\
&= \circ(X = x|Z = z) \,,
\end{aligned}
$$

where the minima are taken with respect to values of Y that are logically independent of $\{Z = z\}$. Thus the first part of the conditional layer condition is satisfied. For the second part, note that

$$
\begin{aligned}
\circ(X \neq x|B(Y) \cap \{Z = z\}) &= \min_{x' \neq x} \circ(X = x'|B(Y) \cap \{Z = z\}) \\
&= \min_{x' \neq x} \circ(X = x'|Z = z) \\
&= \circ(X \neq x|Z = z) \,. \quad \square
\end{aligned}
$$

A more obviously symmetric version of the conditional layer condition is:

Corollary 2 *The conditional layer condition is equivalent to*

$$
\circ(X = x, Y = y|Z = z) = \circ(X = x|Z = z) + \circ(Y = y|Z = z) \tag{3}
$$

for all x,y,z such that $\{Z = z\} \neq \emptyset$.

Proof. The fact that Expression (3) implies Expression (2) is immediate from the definition of $\circ(X = x, Y = y|Z = z)$. To prove the converse, consider the two possible cases. For all y, z such that $\{Y = y, Z = z\} \neq \emptyset$, Proposition 4 directly yields the result as we have $\circ(X = x, Y = y|Z = z) - \circ(Y = y|Z = z) = \circ(X = x|Z = z)$. If instead $\{Y = y, Z = z\} = \emptyset$, then $\circ(X = x, Y = y|Z = z) = \circ(Y = y|Z = z) = \infty$ and Expression (3) is satisfied regardless of $\circ(X = x|Z = z)$. $\square$

Denote by $(X \text{ LC } Y \mid Z)$ the fact that X and Y satisfy the conditional layer condition given Z. It is interesting to note that this relation is a graphoid.

Proposition 5 *The relation $(X \text{ LC } Y \mid Z)$ satisfies symmetry, redundancy, decomposition and contraction; if X and Y are logically independent given Z, then intersection is satisfied.*

Proof. Symmetry follows from Expression (3).
Redundancy ($(Y \perp\!\!\!\perp X \mid X)$): we wish to show that

$$
\circ(X = x_1|Y = y, X = x_2) = \circ(X = x_1|X = x_2) \,.
$$

If $x_1 \neq x_2$, $\circ(X = x_1|Y = y, X = x_2) = \circ(X = x_1|X = x_2) = \infty$; if $x_1 = x_2$, we obtain $\circ(X = x_1, Y = y, X = x_1) - \circ(X = x_1, Y = y) = 0 = \circ(X = x_1, X = x_2) - \circ(X = x_1)$.
Decomposition ($(X \perp\!\!\!\perp (W, Y) \,|\, Z) \Rightarrow (X \perp\!\!\!\perp W \,|\, Z)$): we have

$$\begin{aligned}
\circ(W = w, X = x|Z = z) &= \min_y \circ(W = w, X = x, Y = y|Z = z) \\
&= \min_y \circ(X = x|Z = z) + \circ(W = w, Y = y|Z = z) \\
&= \circ(X = x|Z = z) + \min_y \circ(W = w, Y = y|Z = z) \\
&= \circ(X = x|Z = z) + \circ(W = w|Z = z)\,.
\end{aligned}$$

Weak union ($((W, Y) \perp\!\!\!\perp X \,|\, Z) \Rightarrow (W \perp\!\!\!\perp X \,|\, (Y, Z))$): by hypothesis we have

$$\circ(X = x|W = w, Y = y, Z = z) = \circ(X = x|Z = z)\,,$$

and from this (by decomposition) we obtain $\circ(X = x|Y = y, Z = z) = \circ(X = x|Z = z)$; consequently $\circ(X = x|W = w, Y = y, Z = z) = \circ(X = x|Y = y, Z = z)$.
Contraction ($(Y \perp\!\!\!\perp X \,|\, Z)\ \&\ (W \perp\!\!\!\perp X \,|\, (Y, Z)) \Rightarrow ((W, Y) \perp\!\!\!\perp X \,|\, Z)$): we have

$$\circ(X = x|W = w, Y = y, Z = z) = \circ(X = x|Y = y, Z = z) = \circ(X = x|Z = z)\,.$$

Intersection ($(W \perp\!\!\!\perp X \,|\, (Y, Z))\ \&\ (Y \perp\!\!\!\perp X \,|\, (W, Z)) \Rightarrow ((W, Y) \perp\!\!\!\perp X \,|\, Z)$): we use the fact that, due to the hypothesis of logical independence,

$$\circ(X = x|W = w, Z = z) = \circ(X = x|W = w, Y = y, Z = z) = \circ(X = x|Y = y, Z = z)$$

for all (w, y). Then:

$$\begin{aligned}
\circ(X = x|Z = z) &= \min_w \circ(X = x, W = w|Z = z) \\
&= \min_w \left(\circ(X = x|W = w, Z = z) + \circ(W = w|Z = z)\right) \\
&= \min_w \left(\circ(X = x|Y = y, Z = z) + \circ(W = w|Z = z)\right) \\
&= \circ(X = x|Y = y, Z = z) + \min_w \circ(W = w|Z = z) \\
&= \circ(X = x|Y = y, Z = z) \\
&= \circ(X = x|W = w, Y = y, Z = z)
\end{aligned}$$

(because $\min_w \circ(W = w|Z = z) = 0$ and then using the hypothesis). The hypothesis of logical independence is necessary, as shown by Table 6. $\Box$

We can now define concepts of irrelevance and independence that employ the conditional layer condition:

Definition 6 *Random variables Y are* weakly coherently irrelevant *to random variables X given random variables Z, denoted by $(Y \text{ WCIR } X \mid Z)$, iff $(Y \text{ EIR } X \mid Z)$ and $(Y \text{ LC } X \mid Z)$. Random variables X and Y are* weakly coherently independent *given random variables Z, denoted by $(Y \text{ WCIN } X \mid Z)$, iff $(Y \text{ EIN } X \mid Z)$ and $(Y \text{ LC } X \mid Z)$.*

Given our previous results, we easily have:

	w_0y_0	w_1y_0	w_0y_1	w_1y_1
x_0	0	∞	∞	0
x_1	1	∞	∞	2

Table 6: Failure of the intersection property (for the layer conditions) in the absence of logical independence. Entries in the table are layer levels; the four central entries denote empty events.

Proposition 6 *Weak coherent irrelevance satisfies the graphoid properties of direct and reverse redundancy, direct and reverse decomposition, reverse weak union, and direct and reverse contraction. If W and Y are logically independent, then weak coherent irrelevance satisfies reverse intersection. All other versions of the graphoid properties and intersection fail for weak coherent irrelevance. Weak coherent independence satisfies the graphoid properties of symmetry, redundancy, decomposition and contraction — weak union and intersection fail for weak coherent independence.*

Proof. Equalities among probability values have been proved for Proposition 1, and equalities among layer levels have been proved for Proposition 5. All failures of symmetry, decomposition, weak union, contraction and intersection discussed in Proposition 1 are still valid, with the same examples. $\square$

We have thus examined three concepts of independence that fail the weak union and the intersection properties. The failure of intersection is not too surprising, as this property requires strictly positive probabilities even with the usual concept of stochastic independence. However, the failure of weak union leads to serious practical consequences. To give an example, consider the theory of Bayesian networks [42], where weak union is necessary to guarantee that a structured set of independence relations has a representation based on a graph, and to guarantee that independences can be read off a graph using the d-separation criterion. We intend to explore the relationships between Bayesian networks and full conditional measures in a future publication.

In the next section we examine concepts of irrelevance/independence that satisfy the weak union property.

5 Full irrelevance and independence

Even a superficial analysis of Table 2 suggests that epistemic and coherent independence fail to detect obvious dependences among variables: there is a clear disparity between the two rows, as revealed by conditioning on $\{W = w_1\}$. The problem is that epistemic independence is regulated by the "first active" layer, and it ignores the content of lower layers. Hammond has proposed a concept of independence that avoids this problem by requiring [31]:

$$P(A(X) \cap B(Y)|C(X) \cap D(Y)) = P(A(X)|C(X))\, P(B(Y)|D(Y))\,, \tag{4}$$

for all events $A(X)$, $C(X)$ defined by X, and all events $B(Y)$, $D(Y)$ defined by Y, such that $C(X) \cap D(Y) \neq \emptyset$. Hammond shows that this symmetric definition can be decomposed into two non-symmetric parts as follows.

Definition 7 *Random variables X are h-irrelevant to random variables Y (denoted by $(X \text{ HIR } Y)$) iff $P(B(Y)|A(X) \cap D(Y)) = P(B(Y)|D(Y))$, for all events $B(Y)$, $D(Y)$ defined by Y, and all events $A(X)$ defined by X, such that $A(X) \cap D(Y) \neq \emptyset$.*

If $(X \text{ HIR } Y)$ and $(Y \text{ HIR } X)$, then X and Y are *h-independent*. Expression (4) is equivalent to h-independence of X and Y (for one direction, take first $A(X) = C(X)$ and then $B(Y) = D(Y)$; for the other direction, note that $P(A(X) \cap B(Y)|C(X) \cap D(Y))$ is equal to the product $P(A(X)|B(Y) \cap C(X) \cap D(Y))\, P(B(Y)|C(X) \cap D(Y))$).

We can extend Hammond's definition to conditional independence (a move that has not been made by Hammond himself; Halpern mentions a conditional version of Hammond's definition under the name of *approximate independence* [30]):

Definition 8 *Random variables X are h-irrelevant to random variables Y conditional on random variables Z (denoted by $(X \text{ HIR } Y \mid Z)$) iff*

$$P(B(Y)|\{Z = z\} \cap A(X) \cap D(Y)) = P(B(Y)|\{Z = z\} \cap D(Y)),$$

for all values z, all events $B(Y)$, $D(Y)$ defined by Y, and all events $A(X)$ defined by X, such that $\{Z = z\} \cap A(X) \cap D(Y) \neq \emptyset$.

The "symmetrized" concept is:

Definition 9 *Random variables X and Y are h-independent conditional on random variables Z (denoted by $(X \text{ HIN } Y \mid Z)$) iff $(X \text{ HIR } Y \mid Z)$ and $(Y \text{ HIR } X \mid Z)$.*

This symmetric concept of conditional h-independence is equivalent to (analogously to Expression (4)):

$$\begin{aligned} P(A(X) \cap B(Y)|\{Z = z\} \cap C(X) \cap D(Y)) &= \\ & P(A(X)|\{Z = z\} \cap C(X))\, P(B(Y)|\{Z = z\} \cap D(Y)), \end{aligned} \tag{5}$$

whenever $\{Z = z\} \cap C(X) \cap D(Y) \neq \emptyset$.

The definition of h-irrelevance can be substantially simplified: for random variables X, Y, and Z, $(X \text{ HIR } Y \mid Z)$ iff

$$P(Y = y|\{X = x, Z = z\} \cap D(Y)) = P(Y = y|\{Z = z\} \cap D(Y))$$

for all x, y, z and all events $D(Y)$ defined by Y such that $\{X = x, Z = z\} \cap D(Y) \neq \emptyset$ (directly from Lemma 1).

The positive feature of h-irrelevance is that it satisfies direct weak union, and in fact h-independence satisfies weak union. Unfortunately, both concepts face difficulties with contraction.

Theorem 1 *H-irrelevance satisfies the graphoid properties of direct and reverse redundancy, direct and reverse decomposition, direct and reverse weak union, and reverse contraction. If W and Y are logically independent, then h-irrelevance satisfies reverse intersection. All other versions of the graphoid properties and intersection fail for h-irrelevance. H-independence satisfies the graphoid properties of symmetry, redundancy, decomposition and weak union — contraction and intersection fail for h-independence.*

Proof. Denote by $A(X)$, $B(Y)$ arbitrary events defined by X and Y respectively, chosen such that if they appear in conditioning, they are not logically incompatible with other events. We abbreviate the set $\{W = w\}$ by w, and likewise use x for $\{X = x\}$, y for $\{Y = y\}$, z for $\{Z = z\}$.
Symmetry fails for h-irrelevance as shown by Table 1.
Direct redundancy: $(X \text{ HIR } Y \mid X)$ holds because

$$P(Y = y|\{X = x_1, X = x_2\} \cap B(Y)) = P(Y = y|\{X = x_1\} \cap B(Y)),$$

when $x_1 = x_2$ (and $\{X = x_1, X = x_2\} = \emptyset$ otherwise).
Reverse redundancy: $(Y \text{ HIR } X \mid X)$ holds because

$$P(X = x_1|\{Y = y, X = x_2\} \cap A(X)) = P(X = x_1|\{X = x_2\} \cap A(X)) = 0$$

if $x_1 \neq x_2$ and

$$P(X = x_1|\{Y = y, X = x_2\} \cap A(X)) = P(X = x_1|\{X = x_2\} \cap A(X)) = 1$$

if $x_1 = x_2$.
Direct decomposition: $(X \text{ HIR } (W, Y) \mid Z) \Rightarrow (X \text{ HIR } Y \mid Z)$ holds as

$$\begin{aligned} P(y|\{x, z\} \cap B(Y)) &= \sum_w P(w, y|\{x, z\} \cap B(Y)) \\ &= \sum_w P(w, y|\{z\} \cap B(Y)) \\ &= P(y|\{z\} \cap B(Y)). \end{aligned}$$

Reverse decomposition: $((W, Y) \text{ HIR } X \mid Z) \Rightarrow (Y \text{ HIR } X \mid Z)$ holds because (note that summations over values of W need only include values such that $\{w, y, z\} \cap A(X) \neq \emptyset$):

$$\begin{aligned} P(x|\{y, z\} \cap A(X)) &= \sum_w P(w, x|\{y, z\} \cap A(X)) \\ &= \sum_w P(x|\{w, y, z\} \cap A(X))\, P(w|\{y, z\} \cap A(X)) \\ &= \sum_w P(x|\{z\} \cap A(X))\, P(w|\{y, z\} \cap A(X)) \\ &= P(x|\{z\} \cap A(X)) \sum_w P(w|\{y, z\} \cap A(X)) \\ &= P(x|\{z\} \cap A(X)). \end{aligned}$$

Direct weak union: $(X \text{ HIR } (W, Y) \mid Z) \Rightarrow (X \text{ EIR } Y \mid (W, Z))$ holds because $\{w\} \cap B(Y)$ is an event defined by (W, Y) and consequently:

$$\begin{aligned} P(y|\{w, x, z\} \cap B(Y)) &= P(w, y|\{x, z\} \cap (\{w\} \cap B(Y))) \\ &= P(w, y|\{z\} \cap (\{w\} \cap B(Y))) \\ &= P(y|\{w, z\} \cap B(Y)). \end{aligned}$$

Reverse weak union: $((W, Y) \text{ HIR } X \mid Z) \Rightarrow (W \text{ HIR } X \mid (Y, Z))$ holds because

$$P(x|\{w, y, z\} \cap A(X)) = P(x|\{z\} \cap A(X))$$

by hypothesis and $P(x|\{z\} \cap A(X)) = P(x|\{y,z\} \cap A(X))$ by reverse decomposition.
Reverse contraction: $(Y \text{ HIR } X \mid Z)$ & $(W \text{ HIR } X \mid (Y,Z)) \Rightarrow ((W,Y) \text{ HIR } X \mid Z)$ holds because $P(x|\{w,y,z\} \cap A(X)) = P(x|\{y,z\} \cap A(X)) = P(x|\{z\} \cap A(X))$.
Reverse intersection: $(W \text{ HIR } X \mid (Y,Z))$ & $(Y \text{ HIR } X \mid (W,Z)) \Rightarrow ((W,Y) \text{ HIR } X \mid Z)$ holds because, due to the hypothesis of logical independence,

$$P(x|\{w,z\} \cap A(X)) = P(x|\{w,y,z\} \cap A(X)) = P(x|\{y,z\} \cap A(X))$$

for all (w,y). Thus we can write

$$\begin{aligned}
P(x|\{z\} \cap A(X)) &= \sum_w P(x,w|\{z\} \cap A(X)) \\
&= \sum_w P(x|\{w,z\} \cap A(X))\, P(w|\{z\} \cap A(X)) \\
&= \sum_w P(x|\{y,z\} \cap A(X))\, P(w|\{z\} \cap A(X)) \\
&= P(x|\{y,z\} \cap A(X)) \sum_w P(w|\{z\} \cap A(X)) \\
&= P(x|\{y,z\} \cap A(X)) \\
&= P(x|\{w,y,z\} \cap A(X))\,.
\end{aligned}$$

All other versions of graphoid properties fail for h-irrelevance, as shown by measures in Tables 2, 3, 7, and 8.
Now consider the "symmetrized" concept of h-independence. Symmetry is true by definition; redundancy, decomposition and contraction come from their direct and reverse versions for h-irrelevance. Table 2 displays failure of contraction, and Table 3 displays failure of intersection for h-independence. □

Note that Table 2 is now responsible for failure of direct contraction, as now $(X \text{ HIR } Y)$ and $(X \text{ HIR } W \mid Y)$ but not $(X \text{ HIR } (W,Y))$.

It is natural to consider the strenghtening of h-independence with the conditional layer condition. The first question is whether or not to strenghten the conditional layer condition itself. We might consider the following condition:

$$\circ(X = x|\{Y = y, Z = z\} \cap A(X)) = \circ(X = x|\{Z = z\} \cap A(X)) \tag{6}$$

for all x, y, z and every event $A(X)$ defined by X such that and $\{Y = y, Z = z\} \cap A(X) \neq \emptyset$. As shown by the next result, this condition is implied by the conditional layer condition.

Proposition 7 *If Expression (2) holds for all x, y, z such that and $\{Y = y, Z = z\} \neq \emptyset$, then Expression (6) holds for all x, y, z and every event $A(X)$ defined by X such that $\{Y = y, Z = z\} \cap A(X) \neq \emptyset$.*

Proof. If $x \notin A(X)$, then the relevant layer levels are both equal to infinity. Suppose then that $x \in A(X)$. Using the abbreviations adopted in the proof of Theorem 2 for events such as $\{X = x\}$, we have:

$$\circ(x|\{y,z\} \cap A(X)) \;=\; \circ(\{x\} \cap A(X)|y,z) - \circ(A(X)|y,z)$$

$$
\begin{aligned}
&= \circ(\{x\} \cap A(X)|y,z) - \min_{x' \in A(X)} \circ(x'|y,z) \\
&= \circ(\{x\} \cap A(X)|z) - \min_{x' \in A(X)} \circ(x'|z) \\
&= \circ(x|\{z\} \cap A(X)) \,. \quad \square
\end{aligned}
$$

We propose the following definition.

Definition 10 *Random variables X are* fully *irrelevant to random variables Y conditional on random variables Z (denoted $(X \text{ FIR } Y \mid Z)$) iff*

$$P(Y = y|\{X = x, Z = z\} \cap B(Y)) = P(Y = y|\{Z = z\} \cap B(Y)) \,,$$

$$\circ(Y = y|X = x, Z = z) = \circ(Y = y|Z = z) \,,$$

for all x, y, z, and all events $B(Y)$ defined by Y such that $\{X = x, Z = z\} \cap B(Y) \neq \emptyset$.

Full independence is the symmetrized concept. Theorem 2 and Proposition 6 then imply the following result.

Theorem 2 *Full irrelevance satisfies the graphoid properties of direct and reverse redundancy, direct and reverse decomposition, direct and reverse weak union, and reverse contraction. If W and Y are logically independent, then full irrelevance satisfies reverse intersection. All other versions of the graphoid properties and intersection fail for full irrelevance. Full independence satisfies the graphoid properties of symmetry, redundancy, decomposition and weak union — contraction and intersection fail for full independence.*

We have thus examined two concepts of independence that fail the contraction and intersection properties. While failure of intersection is not surprising, the failure of contraction has important consequences — for example, contraction is needed in the theory of Bayesian networks for essentially the same reasons that weak union is needed.

It must be noted that only one of direction of contraction fails. Specifically, direct contraction fails, and direct contraction is a much less compelling than most other graphoid properties. While reverse contraction convincingly stands for the fact that "if we judge W irrelevant to X after learning some irrelevant information Y, then W must have been irrelevant [to X] before we learned Y" [42]; a similar reading of direct contraction reveals a much less intuitive property. On this account, it seems that h-irrelevance is more appropriate than epistemic irrelevance, and that full irrelevance is more appropriate than weak coherent irrelevance — and likewise for the corresponding independence concepts.

6 Conclusion

In this paper we have examined properties of full conditional measures, focusing on concepts of irrelevance and independence. We started by reviewing the concepts of epistemic and strong coherent irrelevance/independence. We then introduced the layer condition and the related concept of weak coherent irrelevance/independence (Definitions 5 and 6), and derived their graphoid properties (Propositions 5 and 6). We then moved to concepts of irrelevance/independence that satisfy the weak union property. We have:

- presented an analysis of Hammond's concepts of irrelevance and independence with respect to graphoid properties (Theorem 1) — in fact, we note that Hammond and others have not attempted to study *conditional* irrelevance and independence (our Definitions 8 and 9);

- introduced the definition of full irrelevance and independence, and presented the analysis of their graphoid properties (Definition 10 and Theorem 2).

The results in this paper show that there are subtle challenges in combining full conditional measures with statistical models that depend on graphoid properties. We intend to report in detail on the consequences of our results for the theory of Bayesian networks in a future publication. We note that future work should develop a theory of Bayesian networks that effectively deals with full conditional measures, either by adopting new factorizations, new concepts of independence, or new separation conditions (perhaps even for the concepts that fail symmetry [53]). Future work should also investigate whether other concepts of independence can be defined so that they satisfy all graphoid properties even in the presence of null events.

In closing, we should appraise the convenience of the layer condition and its use in defining weak coherent and full irrelevance/independence. The conditional Coletti-Scozzafava condition is apparently motivated by a desire to connect epistemic irrelevance with logical independence. Indeed, the Coletti-Scozzafava condition on (B, A) promptly blocks all cases of logical independence in Table 5, except the fifth one, and a symmetrization of the condition blocks all cases. However, the condition is perhaps too strong when directly extended to conditional probabilities, as the conditional Coletti-Scozzafava blocks all logical dependences between the fixed conditioning events and the other relevant variables. We have thus preferred to consider a relaxed condition where logical independence is not always automatically blocked (specifically, in the fifth, sixth and seventh cases in Table 5), and the resulting layer condition does seem to have pleasant properties; for example, it generates a graphoid. Still, one might ask why should the layer condition be adopted at all given that we are not stressing too much the issue of logical dependence. There are two reasons to adopt the layer condition.

First, the layer condition does prevent pathological cases (as depicted in Table 4) that are not related to logical dependence. The characteristic of such examples is that the "relative nullity" of events is not captured just by conditional probability values. We believe that a most promising path in producing a concept of independence for full conditional measures that does satisfy the graphoid properties is to somehow preserve the "distance" in layers for various events as we deal with probabilities. Possibly the machinery of lexicographic probabilities will be required in order to retain information on layer numbers [7, 31]. The layer condition is a step in that direction.

The second reason to adopt the layer condition, now in the particular case of full independence, is that this condition seems to be necessary in order to produce representations of "product" measures that contain some resemblance of factorization. We intend to report on this issue in a future publication.

Finally, there is yet another difference between the layer condition and the Coletti-Scozzafava condition. In the latter the layer numbers are computed with respect to a restricted set of events, whereas in the former we have removed this requirement. Such a difference is again attributed to our desire to keep layer numbers "permanently" attached to events, so that mere marginalization or conditioning does not change these numbers.

	w_0y_0	w_1y_0	w_0y_1	w_1y_1
x_0	$\lfloor\gamma\beta\rfloor_1$	$\lfloor\gamma(1-\beta)\rfloor_1$	$\alpha\beta$	$\alpha(1-\beta)$
x_1	$\lfloor(1-\gamma)\beta\rfloor_1$	$\lfloor(1-\gamma)(1-\beta)\rfloor_1$	$(1-\alpha)\beta$	$(1-\alpha)(1-\beta)$

	w_0y_0	w_1y_0	w_0y_1	w_1y_1
x_0	$\lfloor\gamma\beta\rfloor_1$	$\lfloor\gamma(1-\beta)\rfloor_1$	$\lfloor(1-\gamma)\beta\rfloor_1$	$\lfloor(1-\gamma)(1-\beta)\rfloor_1$
x_1	$\alpha\beta$	$\alpha(1-\beta)$	$(1-\alpha)\beta$	$(1-\alpha)(1-\beta)$

Table 7: Failure of versions of decomposition, weak union, contraction, and intersection for epistemic irrelevance ($\alpha, \beta, \gamma \in (0,1)$, with $\alpha \neq \beta \neq \gamma \neq \alpha$). The full conditional measure in the top table satisfies $(X \text{ EIR } (W,Y))$ but it fails $(Y \text{ EIR } X)$ (version of decomposition) and it fails $(Y \text{ EIR } X \mid W)$ (version of weak union); it satisfies $(X \text{ EIR } Y)$ and $(W \text{ EIN } X \mid Y)$ but it fails $((W,Y) \text{ EIR } X)$ (two versions of contraction); it also satisfies $(X \text{ EIR } Y \mid W)$ (two versions of intersection, and by switching W and Y, another version of intersection). The full conditional measure in the bottom table satifies $((W,Y) \text{ EIR } X)$ but it fails $(X \text{ EIR } Y)$ (version of decomposition) and it fails $(X \text{ EIR } Y \mid W)$ (version of weak union); it satisfies $(Y \text{ EIR } X)$ and $(W \text{ EIN } X \mid Y)$, but it fails $(X \text{ EIR } (W,Y))$ (two versions of contraction).

Acknowledgements

We thank Matthias Troffaes for many valuable discussions and for suggesting the symmetric form of Coletti-Scozzafava's condition. Thanks also to Andrea Capotorti and Barbara Vantaggi for valuable comments on an earlier version. We acknowledge support from FAPESP (www.fapesp.br) at various stages of this research.

A Counterexamples to graphoid properties

Tables 2, 3, 7, 8 present violations of decomposition, weak union, contraction and intersection for epistemic irrelevance/independence, coherent irrelevance/independence, h-irrelevance/h-independence and full irrelevance/independence. Table 9 summarizes these counterexamples.

Note that some counterexamples for h-/full irrelevance depend on the fact that X is h-/fully irrelevant to (W,Y) in the top table of Table 7. To verify that this is true, it is necessary to verify the equality $P(W,Y|x, A(W,Y)) = P(W,Y|A(W,Y))$ for $x = \{x_0, x_1\}$ and for every nonempty subset $A(W,Y)$ of $\{w_0y_0, w_1y_0, w_0y_1, w_1y_1\}$ (there are 15 such subsets).

References

[1] E. W. Adams. *A Primer of Probability Logic*. CSLI Publications, Stanford, CA, 2002.

[2] D. Allen and A. Darwiche. New advances in inference by recursive conditioning. In *Conference on Uncertainty in Artificial Intelligence*, pages 2–10, 2003.

	w_0y_0	w_1y_0	w_0y_1	w_1y_1
x_0	$\lfloor\gamma\alpha\rfloor_1$	$\lfloor\alpha(1-\gamma)\rfloor_1$	$\lfloor(1-\alpha)\gamma\rfloor_1$	$\lfloor(1-\gamma)(1-\alpha)\rfloor_1$
x_1	$\beta\alpha$	$\alpha(1-\beta)$	$(1-\alpha)\beta$	$(1-\beta)(1-\alpha)$

	w_0y_0	w_1y_0	w_0y_1	w_1y_1
x_0	$\alpha\beta$	$\lfloor\alpha\gamma\rfloor_1$	$\lfloor(1-\alpha)\gamma\rfloor_1$	$(1-\alpha)\beta$
x_1	$\alpha(1-\beta)$	$\lfloor\alpha(1-\gamma)\rfloor_1$	$\lfloor(1-\alpha)(1-\gamma)\rfloor_1$	$(1-\alpha)(1-\beta)$

Table 8: Failures of versions of contraction ($\alpha, \beta, \gamma \in (0, 1)$, with $\alpha \neq \beta \neq \gamma$). The full conditional measure in the top table satisfies (Y EIN X) and (W EIR $X \mid Y$), but it fails (X EIR (W, Y)) (two versions of contraction). The full conditional measure in the bottom table satisfies (X EIN Y) and (X EIR $W \mid Y$), but it fails ((W, Y) EIR X) (two versions of contraction).

[3] A. Balke and J. Pearl. Counterfactual probabilities: Computational methods, bounds and applications. *Conference on Uncertainty in Artificial Intelligence*, pages 46–54, 1994.

[4] P. Battigalli and P. Veronesi. A note on stochastic independence without Savage-null events. *Journal of Economic Theory*, 70(1):235–248, 1996.

[5] S. Benferhat, D. Dubois, and H. Prade. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92(1-2):259–276, 1997.

[6] V. Biazzo, A. Gilio, T. Lukasiewicz, and G. Sanfilippo. Probabilistic logic under coherence, model-theoretic probabilistic logic, and default reasoning in system P. *Journal of Applied Non-Classical Logics*, 12(2):189–213, 2002.

[7] L. Blume, A. Brandenburger, and E. Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica*, 58(1):61–79, January 1991.

[8] L. Blume, A. Brandenburger, and E. Dekel. Lexicographic probabilities and equilibrium refinements. *Econometrica*, 58(1):81–98, January 1991.

[9] A. Capotorti, L. Galli, and B. Vantaggi. How to use locally strong coherence in an inferential process based on upper-lower probabilities. *Soft Computing*, 7(5):280–287, 2003.

[10] G. Coletti and R. Scozzafava. *Probabilistic logic in a coherent setting*. Trends in logic, 15. Kluwer, Dordrecht, 2002.

[11] G. Coletti and R. Scozzafava. Stochastic independence in a coherent setting. *Annals of Mathematics and Artificial Intelligence*, 35:151–176, 2002.

[12] G. Coletti. Coherent numerical and ordinal probabilistic assessments. *IEEE Transactions on Systems, Man and Cybernetics*, 24(12):1747–1753, 1994.

[13] G. Coletti and R. Scozzafava. Zero probabilities in stochastic independence. In B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh, editors, *Information, Uncertainty, Fusion*, pages 185–196. Kluwer, Dordrecht, 2000.

Direct properties of irrelevance and independence	Epistemic/Coherent	H-/Full
$(X \perp\!\!\!\perp (W,Y) \,\vert\, Z) \Rightarrow (X \perp\!\!\!\perp W \,\vert\, (Y,Z))$	2	-
$(X \perp\!\!\!\perp Y \,\vert\, Z)$ & $(X \perp\!\!\!\perp W \,\vert\, (Y,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \,\vert\, Z)$	-	2
$(X \perp\!\!\!\perp W \,\vert\, (Y,Z))$ & $(X \perp\!\!\!\perp Y \,\vert\, (W,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \,\vert\, Z)$	3	3

Non-direct/non-reverse properties of irrelevance	Epistemic/Coherent/H-/Full
$(X \perp\!\!\!\perp (W,Y) \,\vert\, Z) \Rightarrow (Y \perp\!\!\!\perp X \,\vert\, Z)$	7 (top)
$((W,Y) \perp\!\!\!\perp X \,\vert\, Z) \Rightarrow (X \perp\!\!\!\perp Y \,\vert\, Z)$	7 (bottom)
$(X \perp\!\!\!\perp (W,Y) \,\vert\, Z) \Rightarrow (Y \perp\!\!\!\perp X \,\vert\, (W,Z))$	7 (top)
$((W,Y) \perp\!\!\!\perp X \,\vert\, Z) \Rightarrow (X \perp\!\!\!\perp Y \,\vert\, (W,Z))$	7 (bottom)
$(Y \perp\!\!\!\perp X \,\vert\, Z)$ & $(X \perp\!\!\!\perp W \,\vert\, (Y,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \,\vert\, Z)$	7 (bottom)
$(X \perp\!\!\!\perp Y \,\vert\, Z)$ & $(W \perp\!\!\!\perp X \,\vert\, (Y,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \,\vert\, Z)$	8 (top)
$(Y \perp\!\!\!\perp X \,\vert\, Z)$ & $(W \perp\!\!\!\perp X \,\vert\, (Y,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \,\vert\, Z)$	7 (bottom)
$(X \perp\!\!\!\perp Y \,\vert\, Z)$ & $(X \perp\!\!\!\perp W \,\vert\, (Y,Z)) \Rightarrow ((W,Y) \perp\!\!\!\perp X \,\vert\, Z)$	7 (top)
$(Y \perp\!\!\!\perp X \,\vert\, Z)$ & $(X \perp\!\!\!\perp W \,\vert\, (Y,Z)) \Rightarrow ((W,Y) \perp\!\!\!\perp X \,\vert\, Z)$	8 (bottom)
$(X \perp\!\!\!\perp Y \,\vert\, Z)$ & $(W \perp\!\!\!\perp X \,\vert\, (Y,Z)) \Rightarrow ((W,Y) \perp\!\!\!\perp X \,\vert\, Z)$	7 (top)
$(W \perp\!\!\!\perp X \,\vert\, (Y,Z))$ & $(X \perp\!\!\!\perp Y \,\vert\, (W,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \,\vert\, Z)$	3
$(X \perp\!\!\!\perp W \,\vert\, (Y,Z))$ & $(Y \perp\!\!\!\perp X \,\vert\, (W,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \,\vert\, Z)$	3
$(W \perp\!\!\!\perp X \,\vert\, (Y,Z))$ & $(Y \perp\!\!\!\perp X \,\vert\, (W,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \,\vert\, Z)$	3
$(X \perp\!\!\!\perp W \,\vert\, (Y,Z))$ & $(X \perp\!\!\!\perp Y \,\vert\, (W,Z)) \Rightarrow ((W,Y) \perp\!\!\!\perp X \,\vert\, Z)$	7 (top)
$(W \perp\!\!\!\perp X \,\vert\, (Y,Z))$ & $(X \perp\!\!\!\perp Y \,\vert\, (W,Z)) \Rightarrow ((W,Y) \perp\!\!\!\perp X \,\vert\, Z)$	7 (top)
$(X \perp\!\!\!\perp W \,\vert\, (Y,Z))$ & $(Y \perp\!\!\!\perp X \,\vert\, (W,Z)) \Rightarrow ((W,Y) \perp\!\!\!\perp X \,\vert\, Z)$	7 (top)

Table 9: Summary of counterexamples. The properties are written using $\perp\!\!\!\perp$; this symbol must be replaced by the concept of interest (epistemic/coherent/h-/full irrelevance/independence). All entries indicate the number of a table containing a counterexample. The top table lists failures of properties for independence and failures of direct properties for irrelevance; the bottom table lists failures of properties of irrelevance that are neither "direct" nor "reverse" versions of graphoid properties. Note that reverse intersection may fail for all concepts in the absence of logical independence.

[14] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer-Verlag, New York, 1999.

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

[16] L. Crisma. The notion of stochastic independence in the theory of coherent probability. *Quad. Dip. Mat. Appl. Sci. Ec. St. Att. B. de Finetti*, 8/91, 1999.

[17] A. Darwiche and M. Goldszmidt. On the relation between kappa calculus and probabilistic reasoning. In *Conference on Uncertainty in Artificial Intelligence*, pages 145–153, San Francisco, California, 1994. Morgan Kaufmann.

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

[19] A. P. Dawid. Conditional independence. In S. Kotz, C. B. Read, and D. L. Banks, editors, *Encyclopedia of Statistical Sciences, Update Volume 2*, pages 146–153. Wiley, New York, 1999.

[20] B. de Finetti. *Theory of Probability, vol. 1-2*. Wiley, New York, 1974.

[21] R. Dechter and R. Mateescu. Mixtures of deterministic-probabilistic networks and their AND/OR search space. In Max Chickering and Joseph Halpern, editors, *Conference on Uncertainty in Artificial Intelligence*, pages 120–129. AUAI Press, 2004.

[22] M. H. DeGroot. *Probability and Statistics*. Addison-Wesley, 1986.

[23] L. E. Dubins. Finitely additive conditional probability, conglomerability and disintegrations. *Annals of Statistics*, 3(1):89–99, 1975.

[24] D. Galles and J. Pearl. Axioms of causal relevance. *Artificial Intelligence*, 97:9–43, 1997.

[25] H. Geffner. *Default Reasoning: Causal and Conditional Theories*. MIT Press, 1992.

[26] H. Geffner. High probabilities, model preference and default arguments. *Mind and Machines*, 2:51–70, 1992.

[27] D. Geiger, T. Verma, and J. Pearl. Identifying independence in Bayesian networks. *Networks*, 20:507–534, 1990.

[28] M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84:57–112, 1996.

[29] A. Hajek. What conditional probability could not be. *Synthese*, 137:273–323, 2003.

[30] J. Y. Halpern. Lexicographic probability, conditional probability, and nonstandard probability. *Eighth Conference on Theoretical Aspects of Rationality and Knowledge*, pages 17–30, 2001.

[31] P. J. Hammond. Elementary non-Archimedean representations of probability for decision theory and games. In P. Humphreys, editor, *Patrick Suppes: Scientific Philosopher; Volume 1*, pages 25–59. Kluwer, Dordrecht, The Netherlands, 1994.

[32] P. J. Hammond. Consequentialism, non-Archimedean probabilities, and lexicographic expected utility. In C. Bicchieri, R. Jeffrey, and B. Skyrms, editors, *The Logic of Strategy*, chapter 2, pages 39–66. Oxford University Press, 1999.

[33] P. J. Hammond. Non-Archimedean subjective probabilities in decision theory and games. *Mathematical Social Sciences*, 38:139–156, 1999.

[34] S. Holzer. On coherence and conditional prevision. *Bolletino Unione Matematica Italiana Serie VI*, 1:441–460, 1985.

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

[36] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 14(1):167–207, 1990.

[37] P. Krauss. Representation of conditional probability measures on Boolean algebras. *Acta Mathematica Academiae Scientiarum Hungaricae*, 19(3-4):229–241, 1968.

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

[39] F. Matús and M. Studený. *Workshop on Conditional Independence Structures and Graphical Models*. Fields Institute for Research in Mathematical Sciences, Toronto, Canada, October 1999.

[40] V. McGee. Learning the impossible. In E. Bells and B. Skyrms, editors, *Probability and Conditionals*, pages 179–199. Cambridge University Press, 1994.

[41] R. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge, MA, 1991.

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

[43] J. Pearl. Probabilistic semantics for nonmontonic reasoning: A survey. In *Proceedings KR-89*, pages 505–516. Morgan Kaufmann, 1989.

[44] K. R. Popper. *The Logic of Scientific Discovery*. Hutchinson, London, 1975.

[45] E. Regazzini. Finitely additive conditional probability. *Rend. Sem. Mat. Fis.*, 55:69–89, 1985.

[46] A. Renyi. On a new axiomatic theory of probability. *Acta Math. Acad. Sci. Hungarian*, 6:285–335, 1955.

[47] E. San Martin, M. Mouchart, and J.-M. Rolin. Ignorable common information, null sets and Basu's first theorem. *Sankhya: The Indian Journal of Statistics*, 67:674-698, 2005.

[48] T. Seidenfeld, M.J.Schervish, and J.B.Kadane. Improper regular conditional distributions. *Annals of Probability*, 29:1612–1624, 2001.

[49] W. Spohn. Ordinal conditional functions: a dynamic theory of epistemic states. In W. L. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change, and Statistics*, volume II, pages 105–134. D. Reidel Publishing Company, 1988.

[50] R. Stalnaker. Probability and conditionals. *Philosophy of Science*, 37:64–80, 1970.

[51] B. C. van Fraassen. Fine-grained opinion, probability, and the logic of full belief. *Journal of Philosophical Logic*, 24:349–377, 1995.

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

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

[54] B. Vantaggi. Qualitative Bayesian networks with logical constraints. *European Conference on Quantitative and Symbolic Approaches to Reasoning under Uncertainty*, pages 100–112, 2003.

[55] J. Vejnarová. Conditional independence relations in possibility theory. In *First International Symposium on Imprecise Probabilities and Their Applications*, pages 343–351, Ghent, Belgium, 1999. Imprecise Probabilities Project.

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

[57] P. Walley, R. Pelessoni, and P. Vicig. Direct algorithms for checking consistency and making inferences from conditional probability assessments. *Journal of Statistical Planning and Inference*, 126(1):119–151, 2004.

[58] P. M. Williams. Coherence, strict coherence and zero probabilities. *Fifth Int. Congress of Logic, Methodology and Philos. Sci.*, VI:29–30, 1975.

[59] N. Wilson. An order of magnitude calculus. In *Conference on Uncertainty in Artificial Intelligence*, pages 548–555, 1995.