Independence for Full Conditional Measures, Graphoids and Bayesian Networks

Fabio G. Cozman
Universidade de Sao Paulo

Teddy Seidenfeld
Carnegie Mellon University

May 17, 2007

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. Several 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. We propose a concept of "full" independence, characterize the form of a full conditional measure under full independence, and suggest how to build a theory of Bayesian networks that accommodates null events.

Keywords: Bayesian networks, graphoids, stochastic independence, full conditional measures.

1 Introduction

In this paper we wish to consider independence concepts associated with *full conditional measures* [25]. 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 [22] and his followers [11, 17], Rényi [51], Krauss [38] and Dubins [25]. A significant number of philosophers have argued for serious consideration of null events [1, 31, 42, 44, 49, 52, 53, 54], as have economists and game theorists [8, 9, 32, 33, 34, 46].

Null events, or closely related concepts, have appeared repeatedly in the literature on artificial intelligence; for example, ranking and ordinal measures [18, 30, 52, 63] have direct interpretations as "layers" of full conditional measures [11, 28, 47]. Also, some of the most general principles of default reasoning [37] can be interpreted through various types of lexicographic probabilities that include null events [6, 7]. As another example, lexicographic representations for "small" probabilities have appeared in various proposals for default and nonmonotonic reasoning [7, 27, 48]. As a final example, the combination of probabilities and logical constraints in expert systems [2, 10, 11, 23] does seem to call for some sophistication in dealing with null events. For example, one may specify a conditional probability $P(A|B)$, only to later learn from a second source that B implies C. Now, if C is a null event, then B is a null event. What should be made of the original probabilistic assessment $P(A|B)$: should it be deleted or not? The matter becomes difficult when one must deal with many variables and constraints, as it may be difficult to determinine which events are in fact null.

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 not be as defensible as others, they offer a good starting point for discussions of independence. Second, the graphoid properties are rather useful in proving results about probabilities, graphs, lattices, and other models [20]. In particular, graphoid properties elegantly connect probabilistic concepts to graphs; examples of this marriage are Markov random fields and Bayesian networks — in both cases the graphoid properties can be used to prove important soundness and closure results [29, 47].

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. We also present characterizations of independence in terms of joint full conditional measures, and examine the interplay between independence and lexicographic decisions (Section 6). We then suggest ways to build a theory of Bayesian networks that smoothly accommodates null events, by combining definitions of independence with additional assumptions (Section 7).

2 Full Conditional Measures

In this paper we *always* consider a *finite* set of states $\Omega = \{\omega_1, \ldots, \omega_N\}$; 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.

A probability measure is a set function $P : 2^\Omega \rightarrow \Re$ 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 [24].

A different theory of probability ensues if we take conditional probability as a truly primitive concept, as already advocated by Keynes [36] and de Finetti [22]. The first question is the domain of probabilistic assessments. Rényi [51] investigates the general case where $P : \mathcal{B} \times \mathcal{C} \rightarrow \Re$, where $\mathcal{B}$ is a Boolean algebra and $\mathcal{C}$ is an arbitrary subset of $\mathcal{B}$. Popper considers a similar set up [49]. Here we focus on $P : \mathcal{B} \times (\mathcal{B} \backslash \emptyset) \rightarrow \Re$, where $\mathcal{B}$ is again a Boolean algebra [38], 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 $B \neq \emptyset$ [5]. We refer to such a P as a *full conditional measure*, following Dubins [25]; there are other names in the literature, such as *conditional probability measure* [38] and *complete conditional probability system* [46]. 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 $L_0, \dots, L_K$, where $K \leq N$, such that $L_i \gg L_{i+1}$ for $i \in \{0, \dots, K-1\}$ if $K > 0$. Each event L_i is a "layer" of P. 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 $\circ(\emptyset) = \infty$ [11]. 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 [11, 38].

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 $P_0, \dots, P_K$, 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 [38] and Dubins [25], and it has been derived for partial assessments by Coletti [13, 11].

We have, for events A, B:

- $P(B|A) = P\Big(B|A \cap L_{\circ(A)}\Big)$ [5, 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 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 $\{x \in A(X) : x \cap B(Y) \neq \emptyset\}$):

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

	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 ($\alpha, \beta \in (0,1)$).

The following notation will be useful. If A is such that $\circ(A) = i$ and $P(A) = p$, 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 ($P(A) = 0$ and $P(B|A) = \beta$).

There are several decision theoretic derivations of full conditional measures. The original arguments of de Finetti concerning called-off gambles [22] have been formalized in several ways [35, 50, 61, 62]. Derivations based on axioms on preferences have also been presented, both by Myerson [46] and by Blume et al [8]. The last derivation is particularly interesting as it is based on non-Archimedean preferences and lexicographic preferences (discussed further in Section 6).

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 [29]; the term "graphoid" is usually used to mean slightly different concepts [19, 47]. In fact, the following property is often listed as a graphoid property:

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 [15], it is less important than the other properties in the context of the present paper. Finally, the following property is sometimes presented together with the previous ones [47]:

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, $(X \perp\!\!\!\perp Y \,|\, Z) \Leftrightarrow (X \backslash Z \perp\!\!\!\perp Y \backslash Z \,|\, Z)$ [47].

Graphoids offer a compact and intuitive abstraction of independence. They can be used to prove abstract results on independence; for example, 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 [29]. Several authors have employed the graphoid properties as a benchmark to evaluate concepts of independence [4, 58, 55]; 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 very concise definition of independence: events A and B are independent iff $P(A|B) = P(A)$. However, this definition is not entirely satisfying 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 his theory of probability, Keynes faced the problem of non-symmetric independence by defining first a concept of *irrelevance* and then "symmetrizing" it [36]. Thus 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 [59].[1] Levi has also proposed $P(A|B) = P(A)$ as a definition of irrelevance, without considering the symmetrized concept [42].[2] Crisma has further strenghtened Walley's definitions by requiring logical independence [17]; we later return to logical independence. 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 both epistemic and coherent irrelevance/independence can be extended to sets of events, random variables, and to concepts of conditional irrelevance/independence [59].

Definition 1 *Random variables X are epistemically irrelevant to random variables Y conditional on event A (denoted by $(X \text{ EIR } Y \mid A)$) if $P(Y = y|\{X = x\} \cap A) = P(Y = y|A)$ for all values x, y whenever these probabilities are defined (the "unconditional" version is obtained when A is just Ω).*

Definition 2 *Random variables X are epistemically irrelevant to random variables Y conditional on random variables Z (denoted by $(X \text{ EIR } Y \mid Z)$) if $P(Y = y|\{X = x\} \cap \{Z = z\}) = P(Y = y|Z = z)$ for all values x, y, z whenever these probabilities are defined.*

Epistemice independence, denoted using similar triplets with the symbol EIN, is always 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

[1] Walley's original definitions are more complex as he deals with lower and upper previsions.

[2] Levi's definition is geared towards sets of full conditional measures, but clearly it specializes 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 ($\alpha, \beta, \gamma \in (0,1)$, with $\alpha \neq \beta \neq \gamma$). The full conditional measure in the table satisfies $(X \text{ EIN } (W,Y))$ but fails $(X \text{ EIR } Y \mid W)$.

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 [16, 26, 43, 55, 58]).

The following proposition relates epistemic irrelevance/independence with the graphoid properties (several results in the proposition can be extracted from Vantaggi's results [55]).

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.*

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, 12, and 13. 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. ☐

Coletti and Scozzafava have proposed a stronger concept of independence by explicitly dealing with layers [14, 12, 55]. They define the following condition:

Definition 3 *The* Coletti-Scozzafava condition *on (B, A) holds iff whenever $B \neq \emptyset \neq B^c$ then $\circ(A|B) = \circ(A|B^c)$ and $\circ(A^c|B) = \circ(A^c|B^c)$.*

	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 and 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 $(X \text{ EIR } (W, Y))$.

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

Table 4: Failure of Coletti-Scozzafava condition.

Coletti and Scozzafava then define independence of B to A as epistemic irrelevance of B to A plus the Coletti-Scozzafava condition on (B, A). The concept is not symmetric (Table 1), so it is perhaps better to refer to it as *coherent irrelevance*. Coletti and Scozzafava argue that their condition deals adequately with some situations of logical independence [11].[3]

Example 1 Consider two events A and B, and the probability values in Table 4, where $\alpha \in (0, 1)$. Here B is epistemically irrelevant to A, but B is not coherently irrelevant to A because $\circ(A|B) = 1 \neq 2 = \circ(A|B^c)$.

It is interesting to note that the Coletti-Scozzafava condition is symmetric.[4]

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

Proof. Define $a = \circ(AB)$; $b = \circ(AB^c)$; $c = \circ(A^cB)$; $d = \circ(A^cB^c)$. Each one of these four layer levels may be finite or infinite; there are thus 16 situations to consider but only 9 are relevant (one situation is impossible, as it contains no finite layer level; six of the remaining situations violate the Coletti-Scozzafava condition on (B, A)). And six other situations do not satisfy $A \neq \emptyset \neq A^c$ (as required by the Coletti-Scozzafava condition on (A, B)). We are left with three situations: either all layer levels are finite, or just a and c are infinite, or just b and d are infinite. Suppose first 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 $\circ(AB^c) - \circ(B^c) = 0$ by

[3] The Coletti-Scozzafava condition does not attempt to block every case of logical independence: If $A^c = \emptyset$, $P(B) > 0$, $P(B^c) > 0$, then $P(A|B) = P(A|B^c) = P(A) = 1$ (epistemic irrelevance) and $\circ(A|B) = \circ(A|B^c) = 0$, $\circ(A^c|B) = \circ(A^c|B^c) = \infty$, thus B is coherently irrelevant to A even though A and B are not logically independent (third table in Table 5). The condition does block situations where A and B are both involved in deterministic constraints, not situations where one of these events is empty.

[4] Symmetry does not necessarily hold when $P(A|B) = P(A|B^c)$ and $P(B|A) = P(B|A^c)$ and A and B are logically independent (contrary to [11, Theorem 17]). Take logically independent events A and B such that $\circ(AB) = 0$, $\circ(AB^c) = 1$, $\circ(A^cB) = 2$ and $\circ(A^cB^c) = 4$; then $P(A|B) = P(A|B^c) = P(B|A) = P(B|A^c) = 1$ and $\circ(A|B) = \circ(A|B^c) = 0$, but $\circ(A^c|B) = 2 \neq 3 = \circ(A^c|B^c)$.

	B	B^c
A	0	b
A^c	c	d

	B	B^c
A	0	∞
A^c	c	∞

	B	B^c
A	0	b
A^c	∞	∞

	B	B^c
A	0	∞
A^c	∞	∞

Table 5: Symmetry of the Coletti-Scozzafava condition; each table must have a zero entry.

hypothesis. Then $c - 0 = d - b$, thus $d = c + b$, and the table is symmetric. Now, if $\circ(AB) \neq 0$, then we can always re-label rows and columns so that the top left entry is zero, and the same reasoning follows. If either only a and c are infinite, or only b and d are infinite, the result is immediate (second table in Table 5). □

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

Proposition 3 *Events A and B satisfy the Coletti-Scozzafava 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 Coletti-Scozzafava condition and a more obviously symmetric condition:[5]

Corollary 1 *Events A and B satisfy the Coletti-Scozzafava condition iff*

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

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

Vantaggi considers two concepts of coherent irrelevance for random variables X to Y conditional on Z. She uses the following generalization of Coletti-Scozzafava's condition.

Definition 4 *Variables X and Y satisfy the* conditional Coletti-Scozzafava condition *given variables Z 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

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

Vantaggi then proposes two concepts:

- Vantaggi's *weaker* concept of coherent irrelevance assumes [55, Definition 7.3]:
  - epistemic irrelevance of X to Y conditional on Z, and

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

– that X and Y satisfy the conditional Coletti-Scozzafava condition given Z.

- Vantaggi's *stronger* concept of coherent irrelevance assumes *and additionally* that coherent irrelevance fails if any conditioning event is equal to the empty set [55, Definition 7.1]. This is a very stringent concept, as coherent irrelevance then requires X and Z to be logically independent.

We use *weak coherent independence* and *strong coherent independence* to refer to "symmetrizations" of weak coherent irrelevance and strong coherent irrelevance respectively.[6] As shown by Vantaggi [55], strong coherent irrelevance fails symmetry, direct and reverse redundancy, and direct weak union. Her results imply that strong coherent independence fails redundancy and weak union (Table 2). Finally, strong coherent irrelevance fails reverse intersection, as shown by Table 3.

We present an analysis of graphoid properties of weak coherent independence in Proposition (6); before, we derive some useful results on the Coletti-Scozzafava condition.

Proposition 4 *The conditional Coletti-Scozzafava 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 Coletti-Scozzafava 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:

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

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 Coletti-Scozzafava is satisfied. For the second part, note that

$$\circ(X \neq x|B(Y) \cap \{Z = z\}) = \min_{x' \neq x} \circ(X = x'|B(Y) \cap \{Z = z\})$$

[6] It should be noted that Vantaggi's concepts of independence correspond to irrelevance in our terminology; she does not consider "symmetrized" versions. Also, Vantaggi's writing of properties is different from ours; for example, her reverse weak union property is our direct weak union property.

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

A more obviously symmetric version of the conditional Coletti-Scozzafava condition is:

Corollary 2 *The conditional Coletti-Scozzafava 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 $(X \,\textsf{cs}\, Y \mid Z)$ the fact that X and Y satisfy the conditional Coletti-Scozzafava condition given Z. It is interesting to note that this relation is a graphoid.

Proposition 5 *The relation $(X \,\textsf{cs}\, Y \mid Z)$ 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 ($(Y \perp\!\!\!\perp X \,|\, X)$): 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

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

	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 Coletti-Scozzafava condition) in the absence of logical independence. Entries in the table are layer levels; the four central entries denote empty events.

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. □

We now easily have:

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. □

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. The associated concepts of irrelevance fail several non-direct non-reverse versions of graphoid properties for irrelevance, but this is hardly surprising, given the asymmetric character of irrelevance. However, the failure of weak union leads to serious practical consequences, as discussed in Section 7.

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. In that table there is a clear disparity between the two rows, as can be 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 [32]:

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

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 5 *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:

Definition 6 *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 7 *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 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

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

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

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

if $x_1 \neq x_2$ and

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

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, 12, and 13.
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. $\square$

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 Coletti-Scozzafava condition. The first question is whether or not to strenghten the conditional Coletti-Scozzafava condition itself. We might consider the following condition:

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

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 Coletti-Scozzafava 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:

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

We propose the following definition.

Definition 8 *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.*

Full independence imposes a great deal of structure over joint distributions, as shown by the next result. Here $L(X = x|Z = z)$ is the layer of $P(X|Z = z)$ that contains $\{X = x\}$; $L(Y = y|Z = z)$ is the layer of $P(Y|Z = z)$ that contains $\{Y = y\}$;

- C_{xyz} is the set $\{Z = z\} \cap L(X = x|Z = z) \cap L(Y = y|Z = z)$;
- C_{xz} is the set $\{Z = z\} \cap L(X = x|Z = z)$;
- C_{yz} is the set $\{Z = z\} \cap L(Y = y|Z = z)$.

Theorem 3 *For random variables X, Y, and Z, $(X \text{ FIN } Y \mid Z)$ iff for all x,y,z such that $C_{xyz} \neq \emptyset$, we have both*

$$P(X = x, Y = y|C_{xyz}) = P(X = x|C_{xz}) \times P(Y = y|C_{yz}) \tag{7}$$

and

$$\circ(X = x, Y = y|Z = z) = \circ(X = x|Z = z) + \circ(Y = y|Z = z)\,. \tag{8}$$

Proof. In this proof we use $\{x\}$ for $\{X = x\}$, $\{y\}$ for $\{Y = y\}$, $\{z\}$ for $\{Z = z\}$, $\{y, z\}$ for $\{Y = y\} \cap \{Z = z\}$, and likewise for similar events.

The "only if" direction is simple, by taking $A(X) = \{x\}$, $B(Y) = \{y\}$, $C(X) = L(x|z)$, $D(Y) = L(y|z)$ in Expression (5).

To prove the "if" direction, take a set $C(X)$ such that $\{y, z\} \cap C(X) \neq \emptyset$. There are two possibilities: either $\{x\} \cap (\{z\} \cap C(X)) = \emptyset$ or $\{x\} \cap (\{z\} \cap C(X)) \neq \emptyset$.

If $\{x\} \cap (\{z\} \cap C(X)) = \emptyset$, then $P(x|\{y, z\} \cap C(X)) = P(x|\{z\} \cap C(X)) = 0$. If instead $\{x\} \cap (\{z\} \cap C(X)) \neq \emptyset$, there are two cases to consider, as $\circ(C(X)|z) = \min_{x' \in C(X)} \circ(x'|z) \leq \circ(x|z)$.

Case 1: $\{x\} \cap (\{z\} \cap C(X)) \neq \emptyset$ and $\circ(C(X)|z) < \circ(x|z)$. In this case, $\circ(\{z\} \cap C(X)) < \circ(x, z)$ (by adding $\circ(z)$ to both sides in the previous inequality), and then $P(x|\{z\} \cap C(X)) = 0$. Condition 8 yields $\circ(x|y, z) = \circ(x|z)$ and similarly $\circ(C(X)|y, z) = \min_{x' \in C(X)} \circ(x'|y, z) = \min_{x' \in C(X)} \circ(x'|z) = \circ(C(X)|z)$. Thus $\circ(C(X)|z) < \circ(x|z)$ implies $\circ(C(X)|y, z) < \circ(x|y, z)$ and then $P(x|\{y, z\} \cap C(X)) = P(x|\{z\} \cap C(X)) = 0$.

Case 2: $\{x\} \cap (\{z\} \cap C(X)) \neq \emptyset$ and $\circ(C(X)|z) = \circ(x|z)$. We have

$$\begin{aligned}
P(x|\{z\} \cap C(X)) &= P(x|\{z\} \cap C(X) \cap L(x|z)) \\
&= \frac{P(x \cap C(X)|\{z\} \cap L(x|z))}{P(C(X)|\{z\} \cap L(x|z))} \quad \text{(as } P(C(X)|\{z\} \cap L(x|z)) > 0\text{)} \\
&= \frac{P(x|\{z\} \cap L(x|z))}{\sum_{x' \in C(X)} P(x'|\{z\} \cap L(x|z))}
\end{aligned}$$

where the summation is over those x' such that $\{x', z\} \cap L(x|z) \neq \emptyset$. Note that $\{y, z\} \neq \emptyset$ and $\{x, z\} \neq \emptyset$ by previous assumptions; thus $\circ(x, y|z) = \circ(x|z) + \circ(y|z) < \infty$ and (x, y, z) is not logically impossible. We conclude that $\{y, z\} \cap L(x, y) \neq \emptyset$ and then by irrelevance of Y to X, we have $P(x|\{y, z\} \cap L(x|z)) = P(x|\{z\} \cap L(x|z))$. Using this fact,

$$\begin{aligned}
P(x|\{z\} \cap C(X)) &= \frac{P(x|\{y, z\} \cap L(x|z))}{\sum_{x' \in C(X)} P(x'|\{y, z\} \cap L(x|z))} \\
&= P(x|\{y, z\} \cap C(X) \cap L(x|z))\,,
\end{aligned}$$

We now claim that $\{y, z\} \cap L(x, y|z) = \{y, z\} \cap L(x|z)$, where $L(x, y|z)$ is the layer of $P(X, Y|z)$ that contains (x, y, z). In fact, $(x', y, z) \in \{y, z\} \cap L(x, y|z)$ iff $\circ(x', y|z) = \circ(x, y|z)$ iff $\circ(x'|z) + \circ(y|z) = \circ(x|z) + \circ(y|z)$ iff $\circ(x'|z) = \circ(x|z)$ iff $(x', y, z) \in \{y, z\} \cap L(x|z)$. Consequently,

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

	$L_0(X)$	$L_1(X)$	$\dots$	$L_{n'}(X)$
$L_0(Y)$	$p_0 q_0$	$p_1 q_0$	$\dots$	$p_{n'} q_0$
$L_1(Y)$	$p_0 q_1$	$p_1 q_1$	$\dots$	$p_{n'} q_1$
$L_2(Y)$	$p_0 q_2$	$p_1 q_2$	$\dots$	$p_{n'} q_2$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$L_{n''}(Y)$	$p_0 q_{n''}$	$p_1 q_{n''}$	$\dots$	$p_{n'} q_{n''}$

Table 7: Structure of the joint distribution for fully independent variables; each cell C_{ij} contains $P(X, Y|C_{ij})$, where $p_j = P(X|L_j(X))$ and $q_i = P(Y|L_i(Y))$.

By just interchanging X and Y, $P(y|\{x, z\} \cap D(Y)) = P(y|\{z\} \cap D(Y))$ for any $B(D)$ such that $\{x, z\} \cap D(Y) \neq \emptyset$, and this completes the proof. □

This result leads to an interesting construction. Suppose one is given full conditional measures $P(X)$ and $P(Y)$, with layers $\{L_0(X), L_1(X), \dots, L_{n'}(X)\}$ and $\{L_0(Y), L_1(Y), \dots, L_{n''}(Y)\}$. If X and Y are fully independent, what can be said of the joint distribution $P(X, Y)$? Denote by C_{ij} the set $L_i(X) \cap L_j(Y)$. Full independence of X and Y is equivalent to

$$P(X = x, Y = y|C_{ij}) = P(X = x|L_i(X))\, P(Y = y|L_j(Y)) \qquad (9)$$

and the conditional Coletti-Scozzafava condition. Denote by p_i the distribution $P(X|L_i(X))$, and by q_j the distribution $P(Y|L_j(Y))$. If we order the values of X and Y by their layer levels, we can visualize Expression (9) as in Table 7.

Full independence is sufficiently strong to fix $P(X, Y|C_{ij})$ for each C_{ij}. However, if we form a set C_k containing the C_{ij} such that $i + j = k$, we find that we have no information in the marginals on the relative probabilities of any two C_{ij} in a set C_k. For example, suppose that both C_{01} and C_{10} are not empty; then we can specify $\alpha = P(C_{01}|C_{01} \cup C_{10}))$ as any real number in $(0, 1)$. That is, if we only have the marginals $P(X)$ and $P(Y)$, full independence implies that there is a *set* of full conditional measures that satisfies our specification; the set is characterized by all distributions $P_k(C_{ij})$. In fact, note that typically we are given marginals $P(X)$ and $P(Y)$ with the corresponding layers, but we do not have the exact layer levels at our disposal. Thus there is great freedom in assigning layer levels to various layers, so as to respect condition (7), and then there is freedom assigning distributions over the various sets C_k.

It might seem that the indeterminacy of the joint distribution from the marginals is an unfortunate feature of full independence, that ought to be removed by further strenghtening the concept of independence. However, this indeterminacy goes beyond the concept of independence; it may appear even when marginal and conditional measures are specified separately. Consider the following example.

Example 2 Consider two variables X and Y with four values each, and the full conditional measure in Table 8, with $\alpha, \beta, \gamma_i, \gamma_i', \delta \in (0, 1)$. The joint measure produces the marginal and conditional probability values for $P(X)$ and $P(Y|X)$ right below Table 8. It is impossible to determine the value of δ just from $P(X)$ and $P(Y|X)$ (given these two measures, a set of joint measures is specified by $\delta \in (0, 1)$).

	x_0	x_1	x_2	x_3
y_0	$\gamma_0\alpha$	$\gamma_1(1-\alpha)$	$\lfloor\delta\gamma_2\beta\rfloor_1$	$\lfloor\delta\gamma_3(1-\beta)\rfloor_1$
y_1	$(1-\gamma_0)\alpha$	$(1-\gamma_1)(1-\alpha)$	$\lfloor\delta(1-\gamma_2)\beta\rfloor_1$	$\lfloor\delta(1-\gamma_3)(1-\beta)\rfloor_1$
y_2	$\lfloor(1-\delta)\gamma_0'\alpha\rfloor_1$	$\lfloor(1-\delta)\gamma_1'(1-\alpha)\rfloor_1$	$\lfloor\gamma_2'\beta\rfloor_2$	$\lfloor\gamma_3'(1-\beta)\rfloor_2$
y_3	$\lfloor(1-\delta)(1-\gamma_0')\alpha\rfloor_1$	$\lfloor(1-\delta)(1-\gamma_1')(1-\alpha)\rfloor_1$	$\lfloor(1-\gamma_2')\beta\rfloor_2$	$\lfloor(1-\gamma_3')(1-\beta)\rfloor_2$

$$\begin{array}{ll} P(x_0)=\alpha & P(x_1)=1-\alpha \\ P(x_2|x_2\cup x_3)=\beta & P(x_3|x_2\cup x_3)=1-\beta \\ P(y_0|x_i)=\gamma_i & P(y_1|x_i)=1-\gamma_i \\ P(y_2|x_i\cap\{y_2\cup y_3\})=\gamma_i' & P(y_3|x_i\cap\{y_2\cup y_3\})=1-\gamma_i' \end{array}$$

Table 8: Values of joint probability $P(X,Y)$ for Example 2 are entries in the table. Events $\{X = x_i\}$ and $\{Y = y_j\}$ are denoted simply by x_i and y_j respectively, for $i = 0,\ldots,3$. Below the table, marginal and conditional probabilities.

The difficulty here is that a marginal full conditional measure $P(X)$ does not encode all available information regarding variable X. It might seem that a more satisfactory picture would be obtained using general lexicographic probabilities, as we would then be able to indicate the probabilities "beneath" various layers. We briefly discuss general lexicographic probabilities in the next section.

6 Non-Archimedean preferences, lexicographic probabilities, and independence

In this section we examine whether lexicographic probabilities would easily lead to an adequate concept of independence. We first review the derivation of lexicographic probabilities by Blume et al [8]. They start from Anscombe-Aumann's axiomatization of subjective decision theory [3] and relax the usual Archimedean axiom. The result is a sequence of probability measures, such that each probability measure P_i has support L_i. If we write an act f as a function $f : \Omega \to \Re$ that is expressed in utiles, Blume et al's theory yields the following representation result. Act f is preferred over act g iff

$$\left[\sum_{\omega\in\Omega} f(\omega)P_i(\omega)\right]_{i=1}^{K} >_{\mathrm{L}} \left[\sum_{\omega\in\Omega} g(\omega)P_i(\omega),\right]_{i=1}^{K} \tag{10}$$

where $>_{\mathrm{L}}$ denotes lexicographic comparison (for $a, b \in \Re^K$, $a >_{\mathrm{L}} b$ iff $a_j > b_j$ for some $j \leq K$ and $a_i = b_i$ for $1 < i < j$). This representation is not unique; each P_i is unique up to linear combinations of $P_0,\ldots,P_i$ that assign positive weight to P_i [8, Theorem 3.1].

If a lexicographic probability is such that its layers form a partition, then it is equivalent to a full conditional measure. In fact, for any lexicographic probability, the function $P(A|B) = P_i(A|B)$, where P_i the the first measure such that $P_i(B) > 0$, is a full conditional measure. Conversely, any full conditional measure can be represented as a lexicographic probability (Table 1).

Note that not all theories of decision making with full conditional measures subscribe to Expression (10). For example, consider Myerson's theory [46] and several theories that follow de Finetti's tradition [35, 50, 61, 62]. In these theories two acts "tie" conditional on A if they have identical expected values in the first layer that assigns strictly positive probability to A. The following example illustrates the difference between these theories and Expression (10). Consider Table 1 and take an act f that is worth $(1-\alpha)$ if B obtains and $-\alpha$ if B^c obtains. Using Expression (10), act f will be strictly preferred to the zero gamble if $\beta > \alpha$. However Myerson's (and similar) theories prescribe a tie for f and the zero gamble. As another example, note that if $f(\omega) \geq g(\omega)$ for all ω, with strictly inequality for $\omega \in A$, then f is strictly preferred to g in Blume et al's theory, but this may not happen in Myerson's theory when $P(A) = 0$ (this point is also discussed by Walley [60]).

We now return to the discussion of independence. Consider first Hammond's proposal for independence, encoded by Expression (4). This concept avoids obvious drawbacks of epistemic independence; besides, it has the following connection with a concept of independence for non-Archimedean preferences proposed by Blume et al. [8]. We start from a relation $\succeq$ that stands for "weakly preferred to" and additionally we define a three-place relation for conditional preferences: thus $f \succeq g|A$ means that act f is preferred to g conditional on event A. Blume et al propose that X and Y are independent when [8]

$$[f_1(X) \succeq f_2(X)|Y = y'] \Leftrightarrow [f_1(X) \succeq f_2(X)|Y = y''],$$

and

$$[g_1(Y) \succeq g_2(Y)|X = x'] \Leftrightarrow [g_1(Y) \succeq g_2(Y)|X = x''],$$

for any bounded acts $f_1(X)$, $f_2(X)$, $g_1(Y)$ and $g_2(Y)$, and any x', x'', y', y''. We can take each one of these equivalences separately as a definition of irrelevance; then independence obtains when X is irrelevant to Y and vice-versa.

Battigali and Veronesi have shown that Blume et al.'s proposal is equivalent to Hammond's definition of independence [5]. The result is very pleasant: not only lexicographic preferences can be used as a foundation for full conditional measures, they can also be a foundation for (unconditional) h-independence.

The most natural definition for conditional preferences in this setting is: X and Y are conditionally independent given Z iff

$$[f_1(X) \succeq f_2(X)|Y = y', Z = z] \Leftrightarrow [f_1(X) \succeq f_2(X)|Y = y'', Z = z],$$

and

$$[g_1(Y) \succeq g_2(Y)|X = x', Z = z] \Leftrightarrow [g_1(Y) \succeq g_2(Y)|X = x'', Z = z],$$

for all x', x'', y', y'', z, and all bounded acts $f_1(X)$, $f_2(X)$, $g_1(Y)$ and $g_2(Y)$.

Contrary to what could be suspected from Battigali and Veronesi's results, this definition of conditional independence is *not* equivalent to conditional h-independence (Definition 7). This can be observed in Table 2. Here we have that X and Y are h-independent. However, consider the act $g(Y)$ such that $g(y_0) = (1-\alpha)$ and $g(y_1) = -\alpha$. Suppose $\beta > \alpha > \gamma$; then $g(Y)$ is preferred to the "zero gamble" (that is, to the act that always yields zero) conditional on $X = x_0$, but the zero gamble is preferred to $g(Y)$ conditional on $X = x_1$.

	w_0y_0	w_0y_1	w_1y_0	w_1y_1	w_2y_0	w_2y_1
x_0	α	$1-\alpha$	$\lfloor\alpha\rfloor_2$	$\lfloor 1-\alpha\rfloor_2$	$\lfloor\beta\rfloor_4$	$\lfloor 1-\beta\rfloor_4$
x_1	$\lfloor\alpha\rfloor_1$	$\lfloor 1-\alpha\rfloor_1$	$\lfloor\beta\rfloor_3$	$\lfloor 1-\beta\rfloor_3$	$\lfloor\beta\rfloor_5$	$\lfloor 1-\beta\rfloor_5$

Table 9: Failure of contraction for the "conditional" version of Blume et al's concept of independence. Variables X and Y have two categories and W has three categories; α, β and γ are distinct real numbers in $(0, 1)$. For this measure, X and Y are independent, and X and W are independent conditional on Y. But X fails to be irrelevant to (W, Y) because it is possible to have preferences on act $f(W, Y)$ that depend on X.

The difficulty here is that the full conditional measure $P(Y|X)$ does not carry all probabilities on finer partitions (for example, including W). Preferences depend on *all* layers and may require such finer distinctions.[7]

Unfortunately, contraction still fails for this new concept, as can be seen in Table 9. (Other difficulties faced by independence with lexicographic probabilities have been pointed out by Blume et al [8] and Hammond [32, 33].)

Note that the same measure in Table 9 shows that contraction fails for the following strenghtened version of Blume et al's concept: say that X and Y are independent conditional on Z iff

$$[f_1(X) \succeq f_2(X)||Y = y, Z = z] \Leftrightarrow [f_1(X) \succeq f_2(X)|Z = z],$$

and

$$[g_1(Y) \succeq g_2(Y)|X = x, Z = z] \Leftrightarrow [g_1(Y) \succeq g_2(Y)|Z = z],$$

for all x, y, z, and all bounded acts $f_1(X)$, $f_2(X)$, $g_1(Y)$ and $g_2(Y)$.

In response to related concerns, Hammond has proposed to deal with utilities and probabilities that have nonstandard values, and to equate independence with product of nonstandard marginals [33]. This move produces a concept of independence that satisfies all properties commonly associated with stochastic independence. However, the ensuing theory displays a non-uniqueness of representation (infinitesimals may be raised to arbitrary powers without change in representation). As discussed by LaValle and Fishburn in a series of publications [39, 40], to deal with non-Archimedean preferences in a general setting one must look at lexicographic utility *and* matrix-valued probabilities. In short, defining conditional independence for non-Archimedean preferences is an open problem that we leave for future exploration.

7 Bayesian networks with full conditional measures

In this section we return to full conditional measures; we discuss the consequences of adopting such measures on the theory of Bayesian networks. We base our discussion on the theory presented

[7]Consider another example of preferences in full conditional measures. Take binary variables W, X and Y, and probabilities $P(w_0, x_0, y_0) = \alpha^2/2$, $P(w_0, x_0, y_1) = \alpha(1-\alpha)/2$, $P(w_1, x_0, y_0) = \alpha(1-\alpha)/2$, $P(w_1, x_0, y_1) = (1-\alpha)^2/2$, $P(w_0, x_1, y_0) = \lfloor\beta\rfloor_1$, $P(w_0, x_1, y_1) = \lfloor 1-\beta\rfloor_1$, $P(w_1, x_1, y_0) = \alpha/2$, $P(w_1, x_1, y_1) = (1-\alpha)/2$ with α,$\beta \in (0, 1)$ and $\beta > \alpha$. By marginalization we obtain $P(x_0, y_0) = P(x_1, y_0) = \alpha/2$, $P(x_0, y_1) = P(x_1, y_1) = (1-\alpha)/2$. But $P(X, Y)$ is not an accurate representation of preferences: an act $f(y_0) = 1-\alpha$, $f(y_1) = -\alpha$ is preferred to the zero gamble conditional on x_1.

by Geiger et al. [29], as this theory includes a fairly detailed discussion of deterministic relations.

A summary of Geiger et al.'s theory of Bayesian networks is as follows. Suppose we have random variables ordered as $U = \{X_1, \ldots, X_n\}$. For each random variable X_i, denote by U_i the set of variables $\{X_1, \ldots, X_{i-1}\}$. We then select, for each X_i, a subset Z_i of U_i, called the set of *parents* of X_i. We have, for each X_i, either:

- a judgement that X_i is conditionally independent of $U_i \backslash Z_i$ given Z_i, or
- a judgement that X_i is conditionally independent of $U \backslash (X_i, Z_i)$ given Z_i.

The second type of judgement characterizes what Geiger et al. call *deterministic* variables, a concept we examine later. These definitions mention sets of variables that may be empty; we adopt the convention that an empty set of variables is independent of any other variable (for whatever concept of independence), and that an empty conditioning set of variables Z_i indicates an "unconditional" judgement of independence.

The set of judgements of independence described in the previous paragraph is called an *enhanced basis*. From an enhanced basis a directed acyclic graph $\mathcal{G}$ can be *generated*: every X_i is now a node of $\mathcal{G}$, with an edge from each node in Z_i to X_i. The graph $\mathcal{G}$ is a *Bayesian network*. If all independence relations in an enhanced basis L are observed in a joint distribution $P(U)$, this joint distribution *satisfies* the basis L. Any distribution that satisfies an enhanced basis also satisfies the *directed local Markov condition*: A variable X_i is conditionally independent of its nondescendants nonparents given its parents Z_i. This leads to factorization of the joint distribution:

$$P(U) = \prod_{i=1}^{n} P(X_i|Z_i)\,, \tag{11}$$

where $P(X_i|Z_i)$ is just the marginal $P(X_i)$ whenever Z_i is empty. Geiger et al. present a criterion, called D-separation, to be presented shortly, that allows one to read off of graph $\mathcal{G}$ all independence relations that *must* be satisfied by any $P(U)$ that satisfies an enhanced basis generating $\mathcal{G}$. This criterion fills the role of an inference rule for independence, due to the following soundness and completeness results [29].

Soundness If L is an enhanced basis, every D-separation in the directed acyclic graph generated by L corresponds to independence in any joint distribution satisfying L.

Closure If L is an enhanced basis, the set of D-separations in the directed acyclic graph generated by L is identical to the set of independences implied by L using only the graphoid properties.

Completeness Consider a graph $\mathcal{G}$, generated by an enhanced basis L. If variables X and variables Y are not D-separated by Z in $\mathcal{G}$, then there is a joint distribution $P(U)$ satisfying all independence relations indicated by L and such that X and Y are not conditionally independent given Z in $P(U)$.

Closure deals only with properties of graphs and graphoids; its proof by Geiger et al. is not affected by full conditional measures. Completeness also holds for the concepts we have been considering in this paper; Geiger et al. show how to construct the joint distribution required by the theorem using logical dependence, and Meek shows how to do so with joint distributions that are everywhere positive (in which case the concepts of independence in this paper collapse to stochastic independence) [45]. We do not further discuss closure and completeness.

Geiger et al. formulate the soundness of D-separation in more abstract terms, assuming that L is drawn from some relation that satisfies the graphoid properties. We cannot assume this here, as our focus in on independence concepts that do not satisfy all graphoid properties. And in fact, soundness fails for these concepts of independence. Not only soundness fails, but the very idea that a graph is to be generated from an enhanced basis fails. Before we present examples of this, we must discuss the definition of D-separation. Some concepts are required, directly from Geiger et al. [29].

Definition 9 *A* trail *in a directed acyclic graph is a sequence of edges that form a path in the underlying undirected graph. A node X is* head-to-head *with respect to a trail if either: (a) the node starts or ends the trail and $Y \rightarrow X$ belongs to the trail; or (b) $Y \rightarrow X \leftarrow Z$ belongs to the trail. A node X is* tail-to-tail *with respect to a trail if either: (a) the node starts or ends the trail and $X \rightarrow Y$ belongs to the trail; or (b) $Y \leftarrow X \rightarrow Z$ belongs to the trail.*

Definition 10 *A node X is* functionally determined *by Z if $X \in Z$ or X is a deterministic node and all its parents, if any, are functionally determined by Z. A set of variables is functionally determined by Z if each of its members is functionally determined by Z.*

Again, we postpone the definition of *deterministic* nodes.

Definition 11 *A trail is* activated *by Z if (a) every node with converging arrows in the trail either is or has a descendant in Z; and (b) every other node in the trail is outside Z; and (c) no tail-to-tail node in the trail is functionally determined by Z. A trail that is not activated by Z is* blocked *by Z.*

Definition 12 *If X, Y and Z are three disjoint subsets of variables in a directed acyclic graph $\mathcal{G}$, then X and Y are D-separated by Z if there is no trail between a node in X and a node in Y that is activated by Z.*

It remains to define deterministic nodes. Geiger et al. offer the following definition.

Definition 13 *A node X_i is* deterministic *if X_i is conditionally independent of $U \backslash Z_i$ given Z_i.*

The interesting aspect of this definition is that it attempts to reduce functional relationships to judgements of independence. However, this reduction does not work in the presence of full conditional measures: we can have a variable X_i that is deterministic in Geiger et al.'s sense and yet is not a function of its parents; this is basically a consequence of failure of intersection. Consider an example

Example 3 Take an enhanced basis for $U = \{Y, W, X\}$ such that W and X are independent conditional on Y, and X and Y are independent conditional on W. The previous conventions imply that W is determined by Y, and lead to a graph with a root Y, an edge from Y to W, and an edge from W to X. However, in the measure in Table 10 the judgements of independence are satisfied, but W is clearly not a function of Y.

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

Table 10: A full conditional measure that violates Geiger et al's definition of deterministic variable.

Thus we need to use a direct definition for deterministic variable, as follows.

Definition 14 *A node X is* deterministic *if X is a function of its parents, $X = f(Z)$.*

So, an enhanced basis is a series of judgements, either:

- a judgement that X_i is independent of $U_i \backslash Z_i$ conditional on Z_i, or
- a judgement that X_i is a function of Z_i.

We now analyze the relationship between independence concepts, graph generation, and D-separation. We start with epistemic independence, where weak union fails. Note that all examples to follow do not contain deterministic variables, so they are not affected by our change in Definition 13. Note also that these examples do not violate the Coletti-Scozzafava condition.

Example 4 Consider four binary variables ordered as Z, Y, X and W, and the following pair of judgements of epistemic independence: $(X \text{ EIN } Y \mid Z)$ and $(W \text{ EIN } (X, Y) \mid Z)$. These variables and judgements form an enhanced basis. The resulting network has root Z with three children (the other variables), and no other edges. Clearly X and Y are D-separated by (W, Z). However X and Y may not be epistemically independent given (W, Z): just take Table 2 with $\alpha = \beta$ as $P(W, X, Y|Z = z_0)$, and arbitrary positive probabilities for $P(Z = z_0)$ and $P(W, X, Y|Z = z_1) = P(W|Z = z_1)\, P(X|Z = z_1)\, P(Y|Z = z_1)$.

Example 5 Suppose that one receives a directed acyclic graph with three nodes, W, X, and Y, where Y is the sole parent of W, and where X is disconnected from W and Y. The Markov condition on this graph leads to: X independent of (W, Y), Y independent of X, and W independent of X conditional on Y. These independence relations are all satisfied by the measure in Table 2, but the D-separation of X and Y given W does not imply independence of X and Y conditional on W.

It does seem difficult to build a theory of Bayesian networks in the absence of weak union.[8] Consider now full independence, a concept with arguably better properties than epistemic independence. Full independence satisfies weak union, but fails contraction and intersection. The failure of intersection is not a major concern; however the failure of contraction does create serious difficulties for a theory of Bayesian networks.

[8] Vantaggi's reaction to the failure of symmetry and weak union for concepts of coherent independence is to build a theory of graph-theoretical models that does not assume such properties [56, 57]. The resulting models are quite different from the usual Bayesian networks.

	w_0y_0	w_1y_0	w_0y_1	w_1y_1
x_0z_0	$\mu\alpha$	$\lfloor\beta\rfloor_2$	$\mu(1-\alpha)$	$\lfloor 1-\beta\rfloor_2$
x_1z_0	$\lfloor\nu\alpha\rfloor_1$	$\lfloor\gamma\rfloor_3$	$\lfloor\nu(1-\alpha)\rfloor_1$	$\lfloor 1-\gamma\rfloor_3$
x_0z_1	$(1-\mu)p_{00}q_0$	$(1-\mu)p_{10}q_0$	$(1-\mu)p_{01}q_1$	$(1-\mu)p_{11}q_1$
x_1z_1	$\lfloor(1-\nu)p_{00}q_0\rfloor_1$	$\lfloor(1-\nu)p_{10}q_0\rfloor_1$	$\lfloor(1-\nu)p_{01}q_1\rfloor_1$	$\lfloor(1-\nu)p_{11}q_1\rfloor_1$

Table 11: Failure of soundness of d-separation with full independence; α, β, γ, μ, ν, p_{00}, p_{11} and q_0 are distinct real numbers in $(0, 1)$, and $q_1 = 1 - q_0$, $p_{10} = 1 - p_{00}$ and $p_{01} = 1 - p_{11}$.

Example 6 Consider four binary variables ordered as Z, Y, X and W, and the following enhanced basis: $(X \text{ FIN } Y \mid Z)$ and $(W \text{ FIN } X \mid (Y, Z))$. The resulting network has a root Z with three children (the other variables); there is only one other edge from Y to W. Clearly X and (W, Y) are D-separated by Z; however X and (W, Y) may not be fully independent given Z: just take Table 2 as $P(W, X, Y|Z = z_0)$ and arbitrary positive probabilities for $P(Z = z_0)$ and $P(W, X, Y|Z = z_1) = P(W|Y, Z = z_1)\,P(Y|Z = z_1)\,P(X|Z = z_1)$.

Example 7 Suppose one receives a directed acyclic graph with four nodes, where X is the only root, the only parent of Z is X, the only parent of Y is Z, and W has both Y and Z as parents. The Markov condition on this graph leads to: W and X are independent conditional on (Y, Z); Y and X are independent conditional on Z. The measure in Table 11 satisfies these judgements of full independence, but the D-separation of X and (W, Y) given Z does not imply full independence of X and (W, Y) conditional on Z.

We have so far build a negative view of full conditional measures when applied to the standard theory of Bayesian networks, based on enhanced bases and D-separation. In the remainder of this section we examine two different strategies. We focus on concepts of indepedence discussed in Section 5; that is, concepts that fail contraction. Note that these concepts fail only “half” of contraction; that is, their non-symmetric irrelevance counterparts fail direct contraction. We do not consider concepts that fail weak union, for two reasons. First, weak union does seem to be a most desirable property of independence, and it is hard to imagine sensible ways to justify its failure. Second, the concepts that fail weak union do so by failing to “look into” the possibly many layers of full conditional probability (consider Table 2). Such a miss clearly goes against the spirit of full conditional measures itself.

7.1 Enhancing the enhanced basis

Suppose that, instead of starting with an enhanced basis, one starts with the following set of independence relations.

Definition 15 *Given an ordered set of variables $\{X_i\}$, a* full basis *is a sequence of judgements of independence such that for each X_i,*

- *either $(X_i \text{ FIN } U_i \backslash Z_i \mid Z_i)$ and*

$$(A_i \text{ FIN } B_i \mid C_i) \,\&\, (X_i \text{ FIN } B_i \mid (A_i, C_i)) \Rightarrow ((X_i, A_i) \text{ FIN } B_i \mid C_i) \tag{12}$$

for any disjoint sets of variables A_i, B_i and C_i in Ui,

- *or X_i is a function of Z_i.*

The difference between a full basis and an enhanced basis is that, whenever a new variable is considered, the contraction property is assumed to hold for that variable. This is enough to obtain soundness of D-separation, and consequently to obtain contraction for arbitrary sets of variables:

Theorem 4 *If L is a full basis, every D-separation in the directed acyclic graph generated by L corresponds to a full independence in any joint distribution satisfying the independence judgements expressed by L.*

As the proof of this theorem is long and basically follows the proof of soundness by Geiger et al., it is left to Appendix B.

The move from enhanced to full bases recovers D-separation without changes in definitions of independence, bases, graphoids, and graphs. The whole idea can be easily applied to other concepts that fail direct contraction; for example, epistemic independence for sets of positive probability distributions [16]. However, this strategy has two drawbacks. First, the definition of full basis is rather complex, and any associated "Markov condition" is certain to be equally complex. Second, this strategy does not lead to a clear factorization for joint distributions such as Expression (11).

7.2 Adopting a factorization

In practice, the factorization (11). is an essential element of the theory of Bayesian networks. Consider then a direct strategy where we assume a particular form of factorization. Suppose, as a start, that we specify a directed acyclic graph $\mathcal{G}_i$ for each layer of the joint full conditional measure. We might then specify $P(X_i|Z_i)$ for each graph separately. However, this is not possible if we are interested in preserving some sort of Markov condition. For instance, suppose the graph $X_1 \rightarrow X_2 \rightarrow X_3$ is adopted both as $\mathcal{G}_0$ and $\mathcal{G}_1$ (graphs for the layers 0 and 1). Suppose that all probabilities associated with $\mathcal{G}_0$ are positive except that $P(X_3 = x_3|X_2 = x_2) = 0$. In this case, $P(X_1|x_2)$ is computed on the zeroth layer, while $P(X_1|x_2, x_3)$ is computed on the next layer. To satisfy a sensible Markov condition on such a "Bayesian network", we would like $P(X_1|X_2, X_3)$ and $P(X_1|X_2)$ to be equal. This equality may be violated if $P_0(X_1) \neq P_1(X_1)$ or $P_0(X_2|X_1) \neq P_1(X_2|X_1)$. In short, there are constraints across layers that must be respected.

So, instead of trying to capture joint measures in all their possible detail, we propose a direct factorization given a single graph, as follows. Consider variables $U = \{X_1, \ldots, X_n\}$ and a directed acyclic graph $\mathcal{G}$ where each node is a variable in U. Each variable X_i is then associated with a full conditional measure $P(X_i|Z_i)$, where Z_i are the parents of X_i in $\mathcal{G}$. Define layer levels and probability values as:

$$\circ(U) = \circ(X_1, \ldots, X_n) = \sum_{i=1}^{n} \circ(X_i|Z_i)\,, \tag{13}$$

$$\lfloor P(U) \rfloor_{\circ(U)} = \mu(\circ(U))^{-1} \prod_{i=1}^{n} \lfloor P(X_i|Z_i) \rfloor_{\circ(X_i|Z_i)}\,, \tag{14}$$

where $\mu(\circ(U))$ is a normalization constant for the whole layer of U:

$$\mu(\circ(U)) = \sum_{U':\circ(U')=\circ(U)} \prod_{i=1}^{n} \lfloor P(X_i|Z_i) \rfloor_{\circ(X_i|Z_i)} .$$

This definition of a joint full conditional measure may seem complicated, but the basic idea is simple. First, classify all values of U into layers, using Expression (13). Second, collect the values in a layer, apply the usual Markov condition (that is, produce a factorization) and normalize them. Note that a layer at level k may consist of combinations of several marginal and conditional layers.

We can formulate a separation criterion for such "full Bayesian networks" that replaces D-separation. Suppose we are interested in evaluating whether Z separates X and Y. We must first examine in which layer Z is located. To do so, we just run over the graph, applying Expression (13) to determine the lowest layer level of any value of Z. Once we have $\circ(Z)$, we run D-separation in an auxiliary graph built as follows. If $\circ(Z) = 0$, then the auxiliary graph is $\mathcal{G}$ itself. Otherwise, we must write down a logical formula that characterizes all values of U for all layers with level smaller than $\circ(Z)$. We thus have a graph $\mathcal{G}$ and a logical formula ϕ specifying those configurations that are "forbidden" at level $\circ(Z)$. Now we simply run D-separation for the query $(X \perp\!\!\!\perp Y \,|\, Z, \phi)$ as described by Dechter and Mateescu [23].

By adopting a "factorization" outright, we produce a compact representation and we have a separation criterion at our disposal. However, one might argue that Expression (14) assumes too much: layers must be based on identical graphs and several distributions are implicitly assumed uniform (these are the distributions over the "diagonals" in Table 7). One might also argue that assumptions of factorization are less elegant than assumptions of independence.[9]

8 Conclusion

In this paper we have examined properties of full conditional measures, focusing on concepts of irrelevance and independence. We have presented:

- The analysis of Hammond's concepts of irrelevance and independence with respect to graphoid properties (Theorem 1). In fact, Hammond and others have not attempted to study conditional irrelevance and independence (our Definitions 6 and 7). We have also indicated the similarities and differences between conditional versions of Hammond's concepts and Blume et al.'s independence.

- The definition of full irrelevance and independence, and the analysis of their graphoid properties (Definition 8 and Theorem 2). We have also investigated the structure of joint full conditional measures under full independence (Theorem 3).

- The analysis of Bayesian networks associated with full conditional measures. We have shown the challenges that must be faced in such a theory (Examples 4, 5, 6, 7). We have

[9] Note that factorizations may not exist when countable additivity and continuous domains are handled [21]; whether or not some form of factorization can be defined extended to continuous domains through finitely additive full conditional measures is an open problem.

also suggested two ways to remove some of these difficulties, either by moving to full bases (Definitions 14 and 15, and Theorem 4), or by imposing a factorization scheme.

Two secondary contributions should be mentioned:

- The study of the Coletti-Scozzafava condition (Propositions 2, 3, 4), and in particular the proof of its graphoid properties (Proposition 5).
- The analysis of weak coherent irrelevance/independence with respect to graphoid properties. (Propositions 1 and 6). Some of these results can be found in Vantaggi's previous work [55, 56, 57], often with different assumptions concerning logical independence.

Future work should investigate whether other concepts of independence can be defined so that they satisfy all graphoid properties even in the presence of null events. Another line of future work that merits attention is the study of Bayesian networks and other graphical models associated with full conditional measures, either by adopting full bases, specialized factorizations, or other techniques.

Acknowledgements

We thank Matthias Troffaes for many valuable discussions and for suggesting the symmetric form of Coletti-Scozzafava's condition.

A Counterexamples to graphoid properties

Tables 2, 3, 12, 13 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 14 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 12. 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).

B Proof of soundness of D-separation for full basis

Theorem 4 states soundness of D-separation given a full basis. We prove here a result that implies that theorem, with weaker conditions on the basis.

Theorem 5 *Suppose L is an enhanced basis such that for each X_i,*

- *either we have*
$$(X_i \text{ FIN } U_i \backslash Z_i \mid Z_i)$$
and, for any disjoint sets of variables A_i, B_i and C_i in Ui,
$$(B_i \text{ FIR } A_i \mid C_i) \,\&\, (B_i \text{ FIR } X_i \mid (A_i, C_i)) \Rightarrow (B_i \text{ FIR } (X_i, A_i) \mid C_i)\,, \quad (15)$$

	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 12: 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).

	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 13: 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 $(Y \text{ EIN } X)$ and $(W \text{ EIR } X \mid Y)$, but it fails $(X \text{ EIR } (W,Y))$ (two versions of contraction). The full conditional measure in the bottom table satisfies $(X \text{ EIN } Y)$ and $(X \text{ EIR } W \mid Y)$, but it fails $((W,Y) \text{ EIR } X)$ (two versions of contraction).

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

Non-direct/non-reverse properties of irrelevance	Epistemic/Coherent/H-/Full
$(X \perp\!\!\!\perp (W,Y) \mid Z) \Rightarrow (Y \perp\!\!\!\perp X \mid Z)$	12 (top)
$((W,Y) \perp\!\!\!\perp X \mid Z) \Rightarrow (X \perp\!\!\!\perp Y \mid Z)$	12 (bottom)
$(X \perp\!\!\!\perp (W,Y) \mid Z) \Rightarrow (Y \perp\!\!\!\perp X \mid (W,Z))$	12 (top)
$((W,Y) \perp\!\!\!\perp X \mid Z) \Rightarrow (X \perp\!\!\!\perp Y \mid (W,Z))$	12 (bottom)
$(Y \perp\!\!\!\perp X \mid Z)$ & $(X \perp\!\!\!\perp W \mid (Y,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \mid Z)$	12 (bottom)
$(X \perp\!\!\!\perp Y \mid Z)$ & $(W \perp\!\!\!\perp X \mid (Y,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \mid Z)$	13 (top)
$(Y \perp\!\!\!\perp X \mid Z)$ & $(W \perp\!\!\!\perp X \mid (Y,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \mid Z)$	12 (bottom)
$(X \perp\!\!\!\perp Y \mid Z)$ & $(X \perp\!\!\!\perp W \mid (Y,Z)) \Rightarrow ((W,Y) \perp\!\!\!\perp X \mid Z)$	12 (top)
$(Y \perp\!\!\!\perp X \mid Z)$ & $(X \perp\!\!\!\perp W \mid (Y,Z)) \Rightarrow ((W,Y) \perp\!\!\!\perp X \mid Z)$	13 (bottom)
$(X \perp\!\!\!\perp Y \mid Z)$ & $(W \perp\!\!\!\perp X \mid (Y,Z)) \Rightarrow ((W,Y) \perp\!\!\!\perp X \mid Z)$	12 (top)
$(W \perp\!\!\!\perp X \mid (Y,Z))$ & $(X \perp\!\!\!\perp Y \mid (W,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \mid Z)$	3
$(X \perp\!\!\!\perp W \mid (Y,Z))$ & $(Y \perp\!\!\!\perp X \mid (W,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \mid Z)$	3
$(W \perp\!\!\!\perp X \mid (Y,Z))$ & $(Y \perp\!\!\!\perp X \mid (W,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \mid Z)$	3
$(X \perp\!\!\!\perp W \mid (Y,Z))$ & $(X \perp\!\!\!\perp Y \mid (W,Z)) \Rightarrow ((W,Y) \perp\!\!\!\perp X \mid Z)$	12 (top)
$(W \perp\!\!\!\perp X \mid (Y,Z))$ & $(X \perp\!\!\!\perp Y \mid (W,Z)) \Rightarrow ((W,Y) \perp\!\!\!\perp X \mid Z)$	12 (top)
$(X \perp\!\!\!\perp W \mid (Y,Z))$ & $(Y \perp\!\!\!\perp X \mid (W,Z)) \Rightarrow ((W,Y) \perp\!\!\!\perp X \mid Z)$	12 (top)

Table 14: 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.

- *or X_i is a function of Z_i.*

Then every D-separation in the directed acyclic graph generated by L corresponds to a full independence in any joint distribution satisfying the independence judgements expressed by L.

The proof simply adapts the steps in Geiger et al.'s proof of soundness of D-separation for enhanced bases. To simplify the terminology, we refer to "independence" instead of "full independence" throughout (in fact the same result applies to h-independence).

The proof is by induction on the number of variables. For a single variable, the graph with a single node satisfies the required independences (that is, none). Now suppose that bases with less than k variables satisfy the theorem. Consider a basis L with k variables and a joint distribution $P(U)$ satisfying L. Denote by $\mathcal{G}$ the graph generated by L, by $\mathcal{D}$ the set of D-separations in $\mathcal{G}$, and by $\mathcal{M}$ the set of independence relations in $P(U)$. We wish to show that any D-separation in $\mathcal{D}$ implies a corresponding independence in $\mathcal{M}$.

Denote by W the last variable in L, by $L[W]$ the basis without the judgements associated with W, by $\mathcal{G}[W]$ the graph formed from $\mathcal{G}$ by removing W and all edges pointing to W, by $\mathcal{D}[W]$ the set of D-separations in $\mathcal{G}[W]$, by $\mathcal{M}[W]$ the set of independence relations in $\mathcal{M}$ that do not contain W. Then $L[W]$ is a full basis for $U\backslash W$ and it generates $\mathcal{G}[W]$; thus by the induction hypothesis every D-separation in $\mathcal{G}[W]$ implies an independence in $\mathcal{M}[W]$.

We denote by $(X, Y|Z)$ either the D-separation of X and Y by Z, or $(X \text{ FIN } Y \mid Z)$, depending on context. Each triplet $T \in \mathcal{D}$ falls into one of three categories: (1) W does not appear in T; (2) W appears on the first or second entries of T; (3) W appears on the third entry of T. For each case, we show that $T \in \mathcal{D}$ implies $T \in \mathcal{M}$, thus proving that any D-separation in the graph $\mathcal{G}$ must necessarily correspond to an independence.

Case 1. This case can be proved without resort to contraction, so the proof is identical to Geiger et al.'s. If W does not appear in T, then we show that $T \in \mathcal{D}[W]$; given that $\mathcal{D}[W] \subseteq \mathcal{M}[W]$ by the induction hypothesis and that $\mathcal{M}[W] \subseteq \mathcal{M}$, we obtain $\mathcal{D}[W] \subseteq \mathcal{M}$ and thus $T \in \mathcal{M}$. To show that $T \in \mathcal{D}[W]$, suppose $T = (X, Y|Z)$ where X, Y and Z are disjoint subsets of U that do not contain W. Suppose $T \notin \mathcal{D}[W]$ because there is a trail between a node in X and a node in Y that is activated by Z. Any trail activated in $\mathcal{G}[W]$ remains activated in $\mathcal{G}$, so T would be activated in $\mathcal{G}$, contradicting the fact that $T \in \mathcal{D}$.

Case 2. Suppose $T = ((W, X), Y|Z)$ for disjoint subsets of variables X, Y and Z (none of which contain W). Consider the subsets of variables indicated in Figure 1; the parents of W are $\{B_0, B_X, B_Y, B_Z\}$.

First we show that that $((B_0, X), Y|Z) \in \mathcal{M}$. This statement can be proved without resort to contraction, so the proof is identical to Geiger et al.'s. First, $(B_0, Y|Z) \in \mathcal{D}$; otherwise there would be an activated trail between a node in B_0 and a node in Y, that could be augmented to form a trail activated by Z between W and a node in Y (by using the edge that connects a variable in B_0 to W), thus contradicting the fact that $(W, Y|Z) \in \mathcal{D}$ (by decomposition from $((W, X), Y|Z) \in \mathcal{D}$). Decomposition also implies $(X, Y|Z) \in \mathcal{D}$. Now, two sets are D-separated iff each of their individual elements are D-separated; thus $(B_0, Y|Z) \in \mathcal{D}$ and $(X, Y|Z) \in \mathcal{D}$ imply $((B_0, X), Y|Z) \in \mathcal{D}$. The last triplet does not contain W so it belongs to $\mathcal{M}$ as desired, using the argument in Case 1.

There are two possibilities (W is deterministic or not). Suppose first that W is associated with a triplet $(W, R|B)$ (the last element of L). The condition $(W \text{ FIN } R \mid B)$ on W im-

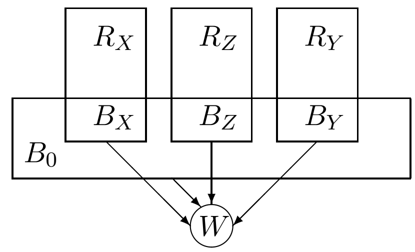

Figure 1: Subsets: $X = (B_X, R_X)$, $Y = (B_Y, R_Y)$, $Z = (B_Z, R_Z)$, $B_0 = B \backslash (B_X, B_Y, B_Z)$.

plies, by decomposition and weak union, that $(W, R_Y | (B_0, X, Z)) \in \mathcal{M}$. Now for B_Y to be D-separated from W, either it is empty or it is functionally determined by Z. In either case, $(W, R_Y | (B_0, X, Z)) \in \mathcal{M}$ implies $(W, Y | (B_0, X, Z)) \in \mathcal{M}$. We have proved $((B_0, X), Y | Z) \in \mathcal{M}$ and $(W, Y | (B_0, X, Z)) \in \mathcal{M}$. Reverse contraction on $\mathcal{M}$ and condition (15) imply the triplet $((B_0, W, X), Y | Z) \in \mathcal{M}$ and then $((W, X), Y | Z) \in \mathcal{M}$ by decomposition.

Now suppose W is deterministic and $W = f(B)$ for some function $f(B)$. Again, B_Y is either empty or functionally determined by Z. Thus $P(B_0, W, X | Y, Z)$ is equal to 0 if $W \neq f(B_0, B_X, B_Y, B_Z)$ and equal to $P(B_0, X | Y, Z)$ otherwise; the last probability is equal to $P(B_0, X | Z)$ because $((B_0, X), Y | Z) \in \mathcal{M}$ (assuming that $(Y, Z) \neq \emptyset$). And $P(Y | B_0, W, X, Z)$ is equal to $P(Y | B_0, X, Z)$ whenever $W = f(B_0, B_X, B_Y(Z), B_Z)$ (otherwise $(B_0, W, X, Z) = \emptyset$). Consequently, $((B_0, W, X), Y | Z) \in \mathcal{M}$ and by decomposition, $((W, X), Y | Z) \in \mathcal{M}$.

If W appears in the second entry in T (that is, $T = (Y, (W, X) | Z)$), then by symmetry of D-separation $((W, X), Y | Z) \in \mathcal{D}$, and $((W, X), Y | Z) \in \mathcal{M}$. Finally, $(Y, (W, X) | Z) \in \mathcal{M}$ by symmetry.

Case 3. This case can be proved without resort to contraction, so the proof is identical to Geiger et al.'s. Suppose $T = (X, Y | (W, Z))$ for disjoint subsets X, Y and Z (none of which contain W). Then $(X, Y | Z) \in \mathcal{D}$; otherwise, there would be a trail between a node in X and a node in Y activated by Z, and this trail would remain activated by (W, Z) because W is a head-to-head node in that trail, contradicting the assumption that $(X, Y | (W, Z)) \in \mathcal{D}$. The fact that $(X, Y | Z) \in \mathcal{D}$ and $(X, Y | (W, Z)) \in \mathcal{D}$ imply that either $(W, X | Z)$ or $(W, Y | Z)$ belong to $\mathcal{D}$ (using a property of D-separation called "weak-transitivity" [29, 47]. The definition of D-separation then implies that either $((W, X), Y | Z)$ or $(X, (W, Y) | Z)$ belong to $\mathcal{D}$. Case 2 applies, so either $((W, X), Y | Z)$ or $(X, (W, Y) | Z)$ belong to $\mathcal{M}$. And by weak union, $(X, Y | (W, Z)) \in \mathcal{M}$. $\square$

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.

[3] F. J. Anscombe and R. J. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34:199–205, 1963.

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

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

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

[7] 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.

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

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

[10] 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.

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

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

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

[14] 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.

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

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

[17] 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.

[18] 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.

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

[20] 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.

[21] A. P. Dawid and M. Studený. Conditional products: An alternative approach to conditional independence. In D. Heckerman and J. Whittaker, editors, *Artificial Intelligence and Statistics*, pages 32–40. Morgan Kaufmann, 1999.

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

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

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

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

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

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

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

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

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

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

[32] 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.

[33] 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.

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

[35] S. Holzer. On coherence and conditional prevision. *Boll. Un. Mat. Ital. Serie VI*, 1:441–460, 1985.

[36] J. M. Keynes. *A treatise on probability*. Macmillan and Co., London, 1921.

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

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

[39] I. H. LaValle and P. C. Fishburn. On the varieties of matrix probabilities in nonarchimedean decision theory. *Journal of Mathematical Economics*, 25:33–54, 1996.

[40] I. H. LaValle and P. C. Fishburn. Structuring and assessing linear matrix-probability distributions. *Journal of Risk and Uncertainty*, 13:125–146, 1996.

[41] D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55:1–60, 1992.

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

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

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

[45] C. Meek. Strong completeness and faithfulness in Bayesian networks. In *Conference on Uncertainty in Artificial Intelligence*, pages 411–418, 1995.

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

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

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

[49] K. R. Popper. *The logic of scientific discovery*. Hutchinson, London, 1975.

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

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

[52] 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.

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

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

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

[56] B. Vantaggi. Graphical models for conditional independence structures. In *Second International Symposium on Imprecise Probabilities and Their Applications*, pages 332–341. Shaker, 2001.

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

[58] 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.

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

[60] P. Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.

[61] 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.

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

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