Annals of Mathematics and Artificial Intelligence (2005) 45: 173–195
DOI: 10.1007/s10472-005-9004-z

Graphoid properties of epistemic irrelevance and independence

Fabio G. Cozman [a] and Peter Walley [b]

[a] *Cidade Universitária, Av. Prof. Mello Moraes, 2231, São Paulo, SP, Brazil*
E-mail: fgcozman@usp.br
[b] *Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Granada, Spain*

This paper investigates Walley's concepts of epistemic irrelevance and epistemic independence for imprecise probability models. We study the mathematical properties of irrelevance and independence, and their relation to the graphoid axioms. Examples are given to show that epistemic irrelevance can violate the symmetry, contraction and intersection axioms, that epistemic independence can violate contraction and intersection, and that this accords with informal notions of irrelevance and independence.

Keywords: conditional independence, credal set, epistemic independence, graphoid axioms, imprecise probability, lower prevision

1. Introduction

The objective of this paper is to study the mathematical properties of irrelevance and independence for imprecise probability models. We consider properties that are based on the graphoid axioms, as these axioms have been proposed as an abstract model for irrelevance and for independence [7, 8, 22]. We define concepts of epistemic irrelevance and independence for imprecise probabilities, and investigate which of the graphoid axioms are satisfied by these concepts.

The term *imprecise probability* refers to mathematical models for uncertainty in which probabilities are not required to have sharp numerical values. These mathematical models include upper and lower probabilities, belief functions, possibility measures, Choquet capacities, qualitative probability orderings, partial preference orderings, upper and lower previsions, and many other models. Together with the many mathematical models, there are many theories of imprecise probability. This paper is based on one of these theories, the theory of *coherent lower previsions* [32].[1]

[1] This theory has recently been called, mistakenly, 'imprecise probability theory', but there are many theories of imprecise probability and there are many concepts of imprecise probability that do not satisfy the coherence axioms required in [32].

Imprecise probabilities have many applications in statistics, economics, psychology, engineering, and other fields [1, 11, 12], and there are many applications also in artificial intelligence. For example, the combination of probability and logic (propositional, first-order, and others) has always been a theme within artificial intelligence, and most approaches require manipulation of imprecise probabilities [15, 16, 20]. Qualitative models for uncertainty [9, 21, 36] and several other imprecise probability models, such as belief functions and possibility measures, have been extensively studied and applied within artificial intelligence, e.g., for measuring and propagating uncertainty in expert systems [33]. Imprecise probabilities are needed especially in applications where there is insufficient evidence to justify precise probability assessments; probabilities based on little information are highly imprecise, and they tend to become increasingly precise as the amount of relevant information increases. Imprecision can be generated also by inconsistency between several sources of information, by vagueness or non-specificity of information, or by uncertainty about the relevance of information. See [33] for an explanation of why imprecise probabilities are needed in artificial intelligence, and [32] for a more general discussion.

In this paper, we use two equivalent models for imprecise probability. The first model involves a set of probability measures, which is called a *credal set*. A mathematically equivalent model involves *coherent lower previsions*. These models are described in section 2.

The concept of *independence* is central in most theories of uncertainty. In standard probability theory, independence is usually defined as factorization of a joint probability measure into its marginal measures, which we call *stochastic independence*. Properties of stochastic independence have received much attention in the artificial intelligence literature, as it has been argued that independence judgements are necessary to make statistical models modular and computationally tractable [22]. Markov conditions involving independence relations have been proposed for a variety of statistical models, and they have also appeared as central features in causality models [23, 26]. The *graphoid axioms* are intended to capture the abstract properties that should be satisfied by any concept of independence.

In theories of imprecise probability, several concepts of independence have been proposed [2, 10, 32, 34]. See especially [2] for definitions, examples and comparisons of the various concepts of independence.

In this paper, we adopt the concepts of *epistemic irrelevance* and *epistemic independence* that were proposed by Walley [32, Chapter 9] and have been studied by Cozman [5], Vicig [31] and Moral [19]. These concepts have clear behavioral interpretations but relatively little is known about their mathematical properties. We show that these two relations can violate several of the graphoid axioms, and that such violations are supported by intuitive notions of irrelevance and independence. Our results suggest that graphoids are not sufficiently general to model the informal concept of irrelevance.

2. Credal sets and coherent lower previsions

This section introduces concepts and notation that are used throughout the paper.

A non-empty set of probability measures, all defined on the same collection of events, is called a *credal set*.[2] A credal set of probability measures for a random variable X is denoted by $\mathcal{M}(X)$.

We assume throughout the paper that all random variables have only finitely many possible values. Given a random variable X and credal set $\mathcal{M}(X)$, *lower* and *upper previsions* for any function $f(X)$ are defined by

$$\underline{E}[f(X)] = \inf_{P \in \mathcal{M}(X)} E_P[f(X)], \qquad \overline{E}[f(X)] = \sup_{P \in \mathcal{M}(X)} E_P[f(X)],$$

where $E_P[f(X)]$ is the expectation of the function $f(X)$ under the probability measure P. Upper previsions can be obtained from lower previsions through the identity $\overline{E}[f(X)] = -\underline{E}[-f(X)]$, so generally we restrict attention to lower previsions.

If the function $f(X)$ is interpreted as a random reward that has the value $f(x)$ when $X = x$, a lower prevision $\underline{E}[f(X)]$ can be regarded as a supremum price that a subject is prepared to pay to buy the random reward $f(X)$ [32]. We take two credal sets to be *equivalent* whenever they have the same convex hull, because in that case they generate the same lower and upper previsions and therefore they produce the same behaviour. We denote equivalence of $\mathcal{M}_1$ and $\mathcal{M}_2$ by $\mathcal{M}_1 \cong \mathcal{M}_2$.

Suppose that $\underline{E}[f(X)]$ is defined for a collection of functions f that forms a linear space. The functional $\underline{E}$ is called a *coherent lower prevision* when it satisfies the three axioms:

$$\begin{aligned} &\underline{E}[f(X)] \geq \inf\{f(x) : x \text{ is a possible value of } X\} \quad \text{(convexity)} \\ &\underline{E}[cf(X)] = c\underline{E}[f(X)] \text{ for all } c > 0 \quad \text{(positive homogeneity)} \\ &\underline{E}[f(X) + g(X)] \geq \underline{E}[f(X)] + \underline{E}[g(X)] \quad \text{(super-linearity).} \end{aligned}$$

This definition of coherence can be extended to conditional lower previsions [32]. In this paper we make use of several properties that are derived from the coherence axioms. We call these properties "coherence relationships"; see [32] for proofs and discussion.

There is a one-to-one correspondence between coherent lower previsions (defined on the linear space of all functions) and closed convex credal sets: lower previsions obtained from a credal set are coherent lower previsions; and given a coherent lower prevision $\underline{E}$, the credal set

$$\{P : E_P[f(X)] \geq \underline{E}[f(X)] \text{ for all functions } f\} \tag{1}$$

is closed and convex.

[2] Levi [17] used this term to refer to convex sets of probability measures, but here we do not require convexity.

The special case where the credal set $\mathcal{M}(X)$ is a singleton set corresponds to equality of the upper and lower previsions, $\overline{E}[f(X)] = \underline{E}[f(X)]$ for all functions f. In that case, we say that probabilities concerning X are *precise*.

Lower and upper probabilities for an event A, denoted by $\underline{P}(A)$ and $\overline{P}(A)$, are defined as the lower and upper previsions of the indicator function $I_A(X)$, which takes the value one if $X \in A$ and zero otherwise:

$$\underline{P}(A) = \underline{E}[I_A(X)] = \inf_{P \in \mathcal{M}(X)} P(A),$$

$$\overline{P}(A) = \overline{E}[I_A(X)] = \sup_{P \in \mathcal{M}(X)} P(A).$$

A *conditional credal set*, denoted by $\mathcal{M}(X|Z=z)$, is a set of conditional probability measures $P(\cdot|Z=z)$ for the variable X. We use $\mathcal{M}(X|Z)$ to denote the function whose value when $Z=z$ is the credal set $\mathcal{M}(X|Z=z)$. Similarly we use $\underline{E}[f(X)|Z]$ to denote the function whose value when $Z=z$ is $\underline{E}[f(X)|Z=z]$. If $\underline{P}(Y=y|Z=z) > 0$, the conditional lower prevision $\underline{E}[f(X)|Y=y, Z=z]$ is the unique solution of the following equation in λ, called the *generalized Bayes rule* [32, section 6.4]:

$$\underline{E}\big[(f(X)-\lambda)I_{\{y\}}(Y)|Z=z\big] = 0. \tag{2}$$

If the lower probability of the conditioning event $\{Y=y\}$ is zero, then the conditional lower prevision $\underline{E}[f(X)|Y=y]$ may not be uniquely determined by unconditional lower previsions. In that case, there are several ways of proceeding; see [32, section 6.10] and [6, 35]. In this paper, we try to avoid this complication by concentrating on examples where all possible joint outcomes have positive lower probability, so that conditional lower probabilities and lower previsions are uniquely determined through the generalized Bayes rule. Most of the later examples are of this type. In those few examples which do involve zero lower probabilities, the conditional credal sets were defined directly, rather than by conditioning a joint (unconditional) credal set.

3. Graphoid properties

Let W, X, Y and Z denote random variables, each of which has only finitely many possible values. When W and Y are random variables, so is the joint variable (W, Y). Except where explicitly noted, we do not make any assumptions about the set of possible values for joint variables such as (W, Y). In particular we do not assume that the set of possible values is a product space; i.e., saying that "$(W, Y) = (w, y)$ is possible" is not equivalent to saying that "$W = w$ is possible and $Y = y$ is possible". We say that W and Y are *logically independent* conditional on Z if $(W, Y, Z) = (w, y, z)$ is possible whenever both $(W, Z) = (w, z)$ and $(Y, Z) = (y, z)$ are possible; in this case, for each fixed value of z, the set of possible values of (W, Y) is a product space.

A ternary relation $(X \perp\!\!\!\perp Y \mid Z)$ is called a *graphoid* when it satisfies the following axioms [7, 14, 22, 28].[3]

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

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

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

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

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

Intersection: If W and Y are logically independent conditional on Z, then $(X \perp\!\!\!\perp W|(Y, Z))$ & $(X \perp\!\!\!\perp Y|(W, Z)) \Rightarrow (X \perp\!\!\!\perp (W, Y)|Z)$.[4]

These axioms are satisfied (under weak assumptions) by stochastic conditional independence. The most common definition of stochastic independence, through factorization of joint probabilities, is as follows. Random variables X and Y are said to be *stochastically independent* conditional on Z if

$$P(X = x, Y = y|Z = z) = P(X = x|Z = z)P(Y = y|Z = z) \qquad (3)$$

whenever both $(X, Z) = (x, z)$ and $(Y, Z) = (y, z)$ are possible.[5] In that case we write $(X \perp\!\!\!\perp Y|Z)$.

A slightly different definition is that X and Y are stochastically independent conditional on Z if

$$P(X = x|Y = y, Z = z) = P(X = x|Z = z) \qquad (4)$$

[3] In understanding the meaning of the graphoid axioms, it may help to simplify the axioms by omitting all occurrences of the conditioning variable Z. Since this is equivalent to taking Z to be a constant random variable, all the axioms and later results remain valid when Z is omitted.

[4] This axiom is slightly different from the intersection axiom in Pearl [22], which applies to graphical relations such as d-separation.

[5] Here and elsewhere in the paper, we define $P(W = w|Z = z)$ to be zero whenever $Z = z$ is possible but $(W, Z) = (w, z)$ is impossible. That is, impossible (conditional) values of a variable are taken to have (conditional) probability zero. In the case of (3), $(X, Y, Z) = (x, y, z)$ may be impossible when both $(X, Z) = (x, z)$ and $(Y, Z) = (y, z)$ are possible, and (3) may still be satisfied provided that either $P(X = x|Z = z) = 0$ or $P(Y = y|Z = z) = 0$. That is, X and Y can be stochastically independent conditional on Z even though they are logically dependent conditional on Z. This indicates a weakness of the standard definition of stochastic independence. As suggested in [32, section 9.1], it would be reasonable to strengthen the standard definition by requiring, in addition to the factorization criterion (3), that X and Y be logically independent conditional on Z. This strengthens the standard definition only when some probabilities are zero, because if $P(X = x|Z = z)$ and $P(Y = y|Z = z)$ are positive for all possible (x, z) and (y, z) then (3) implies that X and Y are logically independent conditional on Z.

whenever $(Y, Z) = (y, z)$ is possible, and

$$P(Y = y|X = x, Z = z) = P(Y = y|Z = z) \tag{5}$$

whenever $(X, Z) = (x, z)$ is possible.

The two conditions (4) and (5) say, respectively, that if the value of Z is known, then learning the value of Y is irrelevant to uncertainty about X, and learning the value of X is irrelevant to uncertainty about Y. The definition of epistemic independence for imprecise probability that we give later is a simple generalization of this second definition. The second definition of stochastic conditional independence agrees with the first definition if all the probabilities $P(X = x|Z = z)$ and $P(Y = y|Z = z)$ are positive, but the two definitions differ when some of these probabilities are zero.[6]

Under either definition, stochastic conditional independence always satisfies four of the graphoid axioms: symmetry, redundancy, decomposition and contraction. Some weak assumptions are needed to guarantee that the other two graphoid axioms (weak union and intersection) hold. For example, under either definition, it is sufficient for weak union that $P(Y = y|Z = z) > 0$ whenever (y, z) is a possible value of (Y, Z). Similar positivity conditions are sufficient for the intersection axiom. If, for every triple of variables (X, Y, Z), every joint event $(X = x, Y = y, Z = z)$ that is possible has positive probability, then the two definitions of stochastic conditional independence agree and they satisfy all the graphoid axioms.

As we consider irrelevance relations that are asymmetric, i.e., that violate the symmetry axiom, we must distinguish between different versions of the other graphoid axioms. For example, the redundancy axiom has two versions: $(X \perp\!\!\!\perp Y|X)$ and $(Y \perp\!\!\!\perp X|X)$. The decomposition and weak union axioms have four possible versions, and the contraction and intersection axioms have eight possible versions, which are obtained by interchanging the two variables on either side of each $\perp\!\!\!\perp$ symbol. We distinguish between 'direct' and 'reverse' versions of the graphoid axioms. The 'direct' versions have variables in the same order as the graphoid axioms stated earlier, and the 'reverse' versions have variables in the reverse order.[7] As an example, the contraction axiom leads to:

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

Reverse contraction: $(Y \perp\!\!\!\perp X|Z)$ & $(W \perp\!\!\!\perp X|(Y, Z)) \Rightarrow ((W, Y) \perp\!\!\!\perp X|Z)$.

There are six other forms of the contraction property, obtained by interchanging variables in each pair. We focus our attention on the reverse and direct forms of the

[6] Some authors, such as Dawid [7], use a slightly different definition of stochastic conditional independence in which the factorization condition (3) is required to hold almost everywhere but may fail on a set of probability zero. Similarly Pearl [22, p. 83] requires (4) to hold only when $P(Y = y, Z = z) > 0$. With these weaker definitions, stochastic conditional independence satisfies all the graphoid axioms without further conditions.

[7] The names 'direct' and 'reverse' are somewhat arbitrary, because the forms in which we have written the graphoid axioms are arbitrary.

graphoid axioms; the examples in section 6 show that all other forms of the graphoid axioms can fail for epistemic irrelevance.

A similar situation is faced by Galles and Pearl [13] when presenting axioms for *causal irrelevance*. Because causal irrelevance is asymmetric in their system, they introduce 'left' and 'right' versions of the graphoid properties. Several researchers have recently considered asymmetric concepts of independence in connection with imprecise probability [18, 29, 30].

4. Epistemic irrelevance

The informal meaning of "Y is irrelevant to X" is that learning the value of variable Y would not change our uncertainty about the value of variable X. Our uncertainty about X can be expressed by a collection of lower previsions of the form $\underline{E}[f(X)]$, and our uncertainty about X given an event $\{Y = y\}$ can be expressed by a collection of conditional lower previsions of the form $\underline{E}[f(X)|Y = y]$. Consequently, we say that Y is *epistemically irrelevant* to X when $\underline{E}[f(X)|Y = y] = \underline{E}[f(X)]$ for all possible values (y) of Y and all functions (f) of X [32].

Epistemic irrelevance can also be defined in terms of credal sets [2, 3]. Let $\mathcal{M}(X)$ and $\mathcal{M}(X|Y)$ denote respectively a marginal credal set and a collection of conditional credal sets. We say that Y is epistemically irrelevant to X when the convex hull of $\mathcal{M}(X|Y = y)$ is equal to the convex hull of $\mathcal{M}(X)$ for all possible values $y : \mathcal{M}(X|Y = y) \cong \mathcal{M}(X)$. Recall that two credal sets with the same convex hull are equivalent, because they generate the same upper and lower previsions.

This definition of epistemic irrelevance can be generalized to allow judgements of irrelevance that are conditional on further variables. We say that Y is *epistemically irrelevant* to X conditional on Z, written as $(Y \,\text{IR}\, X|Z)$, when

$$\underline{E}[f(X)|Y = y, Z = z] = \underline{E}[f(X)|Z = z] \tag{6}$$

for all possible values (y, z) of (Y, Z) and all functions $f(X)$.[8]

For imprecise probabilities, epistemic irrelevance is an asymmetric relation. For precise probabilities, assuming that $P(X = x|Z = z) > 0$ and $P(Y = y|Z = z) > 0$ for all

[8] It is assumed here that $\overline{P}(X = x|Y = y, Z = z) = 0$ if $X = x$ is impossible conditional on $(Y, Z) = (y, z)$. Again, it would be reasonable to strengthen the definition of epistemic irrelevance by requiring also that X and Y be logically independent conditional on Z, but that is not required in this paper. The examples later in the paper, in which the variables are logically independent, would be unaffected by such a change. Epistemic irrelevance could be defined in a different way, by requiring only that the terms $\underline{E}[f(X)|Y = y, Z = z]$ in equation (6) do not depend on y. In that case, it is necessary for coherence that $\underline{E}[f(X)|Y = y, Z = z] \leq \underline{E}[f(X)|Z = z]$, but strict inequality is possible. It can be verified that the graphoid properties of this modified definition differ from the properties listed in Theorem 1: the modified definition satisfies the reverse intersection axiom, but it can violate the symmetry, reverse decomposition, direct and reverse contraction, and direct intersection axioms.

possible values (x, z) and (y, z), epistemic irrelevance agrees with stochastic independence (using either definition from section 3) and is symmetric.[9]

Epistemic irrelevance should not be interpreted as *causal irrelevance*, as used for example by Galles and Pearl [13]. Epistemic irrelevance is an attempt to model what Galles and Pearl call *informational irrelevance*. The probabilistic and deterministic concepts of causal irrelevance in [13] do not satisfy the symmetry property.

The next theorem summarises the most important properties of epistemic irrelevance. Some of these properties require the following *positivity assumption*:

$$\underline{P}(Y = y|Z = z) > 0, \qquad \text{whenever } (Y, Z) = (y, z) \text{ is possible.} \tag{7}$$

Theorem 1. Epistemic irrelevance satisfies the following properties, for all random variables W, X, Y, Z.

Direct redundancy (DR): $(X \,\text{IR}\, Y|X)$

Reverse redundancy (RR): $(Y \,\text{IR}\, X \,|\, X)$

Direct decomposition (DD): $(X \,\text{IR}\, (W, Y) \mid Z) \Rightarrow (X \,\text{IR}\, Y \,|\, Z)$

Reverse contraction (RC): $(Y \,\text{IR}\, X \,|\, Z) \& (W \,\text{IR}\, X \,|\, (Y, Z)) \Rightarrow ((W, Y) \,\text{IR}\, X \,|\, Z)$.

Under the positivity assumption (7),[10] epistemic irrelevance also satisfies:

Reverse decomposition (RD): $((W, Y) \,\text{IR}\, X \mid Z) \Rightarrow (Y \,\text{IR}\, X \,|\, Z)$

Direct weak union (DWU): $(X \,\text{IR}\, (W, Y) \,|\, Z) \Rightarrow (X \,\text{IR}\, W \,|\, (Y, Z))$

Reverse weak union (RWU): $((W, Y) \,\text{IR}\, X \mid Z) \Rightarrow (W \,\text{IR}\, X \,|\, (Y, Z))$.

Proof.

DR: The proof is trivial since $\underline{E}[h(Y) \,|\, X = x, X = x] = \underline{E}[h(Y)|X = x]$.

RR: It is necessary for coherence that $\underline{E}[f(X)|X = x] = f(x) = \underline{E}[f(X)|X = x, Y = y]$, hence $(Y \,\text{IR}\, X \,|\, X)$.

[9] Pearl [24] discusses examples of asymmetric irrelevance relations of the following type. Consider two variables X and Y. Suppose we have a single conditional distribution $P(Y \,|\, X)$ that is not independent of X, but we are totally ignorant about X so the credal set $\mathcal{M}(X)$ is vacuous. If we condition on Y, we obtain a vacuous credal set for $\mathcal{M}(X|Y)$. So, Y is irrelevant to X, but X is not irrelevant to Y. Pearl seems to argue that these are the only situations where asymmetry can occur, but section 6 contains examples where every atom of a model is assigned positive lower probability, and still symmetry fails. Another example where all conditioning events have positive lower probability can be found in [4].

[10] Example 1 of section 6 shows that, without the positivity assumption (7), RD and RWU can fail. Recall that similar positivity assumptions are needed to ensure that stochastic conditional independence satisfies the weak union and intersection axioms.

DD: Any function, h, of Y can be regarded as a function of Y and W, so $\underline{E}[h(Y)|X,Z] = \underline{E}[h(Y)\,|\,Z]$ by hypothesis. Consequently $(X \text{ IR } Y\,|\,Z)$.

RC: Suppose that (w, y, z) is a possible value of (W, Y, Z). Let g be any function of X. Applying the hypotheses $(W \text{ IR } X|(Y,Z))$ and $(Y \text{ IR } X|Z)$ gives $\underline{E}[g(X)|W = w, Y = y, Z = z] = \underline{E}[g(X)|Y = y, Z = z] = \underline{E}[g(X)|Z = z]$, which establishes that $((W,Y) \text{ IR } X\,|\,Z)$.

RD: Let g be any function of X. It follows from coherence of the conditional lower previsions and $((W,Y) \text{ IR } X \mid Z)$ that

$$\underline{E}[g(X)|Y = y, Z = z] \geq \min \underline{E}[g(X)|W = w, Y = y, Z = z] = \underline{E}[g(X)|Z = z],$$

whenever $(Y, Z) = (y, z)$ is possible, where the minimum is taken over all values W such that $(W, Y, Z) = (w, y, z)$ is possible. Using this result and applying another general coherence relationship, $\underline{E}[g(X)|Z = z] \geq \underline{E}[\underline{E}[g(X)|Y, Z = z]|Z = z] \geq \underline{E}[\underline{E}[g(X)|Z = z]|Z = z] = \underline{E}[g(X)|Z = z]$. Because the first and last expressions are identical, there must be equality in each step, so $\underline{E}[\underline{E}[g(X)|Y, Z = z]|Z = z] = \underline{E}[g(X)|Z = z]$. Using the positivity assumption and $\underline{E}[g(X)|Y = y, Z = z] \geq \underline{E}[g(X) \mid Z = z]$ whenever $(Y, Z) = (y, z)$ is possible, this implies that $\underline{E}[g(X)|Y = y, Z = z] = \underline{E}[g(X)|Z = z]$ whenever $(Y, Z) = (y, z)$ is possible. Thus $(Y \text{ IR } X|Z)$.

DWU: Suppose that $(X, Y, Z) = (x, y, z)$ is possible. Then $(X, Z) = (x, z)$ and $(Y, Z) = (y, z)$ are both possible, so the positivity assumption and irrelevance hypothesis imply that $\underline{P}[Y = y|X = x, Z = z] > 0$. Let h be any function of W. By the generalized Bayes rule (2), $\underline{E}[h(W)|Y = y, X = x, Z = z]$ is the unique solution λ of the equation

$$\underline{E}\big[(h(W) - \lambda)I_{\{y\}}(Y)|X = x, Z = z\big] = 0.$$

Because $\underline{E}\big[(h(W) - \lambda)I_{\{y\}}(Y)|X = x, Z = z\big] = \underline{E}\big[(h(W) - \lambda)I_{\{y\}}(Y)|Z = z\big]$ by hypothesis, $\underline{E}[h(W)|Y = y, X = x, Z = z]$ is the unique solution λ of

$$\underline{E}\big[(h(W) - \lambda)I_{\{y\}}(Y)|Z = z\big] = 0,$$

and it is therefore equal to $\underline{E}[h(W)|Y = y, Z = z]$ because the generalized Bayes rule has a unique solution. This establishes that $(X \text{ IR } W\,|\,(Y,Z))$.

RWU: Suppose that (w, y, z) is a possible value of (W, Y, Z). Let g be any function of X. As $((W,Y) \text{ IR } X\,|\,Z)$ by hypothesis, $\underline{E}[g(X)\,|\,W = w, Y = y, Z = z] = \underline{E}[g(X)\,|\,Z = z]$. Using the positivity assumption and applying property RD gives $(Y \text{ IR } X\,|\,Z)$, so that $\underline{E}[g(X)\,|\,Y = y, Z = z] = \underline{E}[g(X)\,|\,Z = z]$, and then $\underline{E}[g(X)\,|\,W = w, Y = y, Z = z] = \underline{E}[g(X)\,|\,Y = y, Z = z]$. This establishes that $(W \text{ IR } X\,|\,(Y, Z))$. □

Other general properties of epistemic irrelevance can be derived from the properties in the preceding theorem. Some properties are listed in the following corollary.

Corollary 1. Epistemic irrelevance satisfies, for all random variables W, X, Y, Z:

1. If W is a function of Z then $(X \text{ IR } Y \mid Z) \Rightarrow (X \text{ IR } (W, Y) \mid Z)$.
   Under the positivity assumption (7), epistemic irrelevance also satisfies:
2. $(W \text{ IR } X \mid Z)$ & $(Y \text{ IR } X \mid (W, Z)) \Rightarrow (Y \text{ IR } X \mid Z)$.
3. $(W \text{ IR } X \mid Z)$ & $(Y \text{ IR } X \mid (W, Z)) \Rightarrow (W \text{ IR } X \mid (Y, Z))$.

Proof.

1. Suppose that $(X, Z) = (x, z)$ is possible. Given a function $h(W, Y)$, define the function $h'(y) = h(w(z), y)$. Then $\underline{E}[h(W,Y)|X = x, Z = z] = \underline{E}[h(w(z), Y)|X = x, Z = z] = \underline{E}[h'(Y)|X = x, Z = z] = \underline{E}[h'(Y)|Z = z] = \underline{E}[h(W,Y)|Z = z]$. This shows that $(X \text{ IR } (W,Y)|Z)$.

2&3. The assumptions imply that $((W,Y) \text{ IR } X|Z)$ by RC. Using the positivity assumption, this implies that $(Y \text{ IR } X|Z)$ by RD, and it also implies that $(W \text{ IR } X|(Y,Z))$ by RWU. □

As a special case of property 2, consider the case where Z is omitted and Y provides partial information about W alone, in the sense that $(Y \text{ IR } X|W)$. Suppose also that $\underline{P}(Y = y) > 0$ for all possible Y. Then property 2 implies that if $(W \text{ IR } X)$ then $(Y \text{ IR } X)$. In other words, if W is irrelevant to X then learning partial information about the value of W does not change the uncertainty concerning X. This property seems necessary for 'irrelevance' of W to X in the intuitive sense. Campos and Moral gave an example which, they claimed, showed that learning partial information about W could change uncertainty about X [10]. But their example relies, in an essential way, on conditioning on events of probability zero. The later Example 1 shows that properties 2 and 3 of the Corollary can fail when the positivity assumption is violated.

Examples will be given later to show that epistemic irrelevance can violate the direct contraction and intersection properties. Although epistemic irrelevance does not always satisfy direct contraction, it does so whenever particular probability distributions are precisely determined.[11]

Theorem 2. Suppose that all the credal sets $\mathcal{M}(W \mid Y, Z)$ are *precise*, i.e., for all possible values of (y, z), the credal set $\mathcal{M}(W \mid Y = y, Z = z)$ contains only a single

[11] The key to this result is that, because of the precision assumption in the Theorem, the credal set $\mathcal{M}(W, Y|Z)$ is uniquely determined by $\mathcal{M}(Y|Z)$ and $\mathcal{M}(W|Y, Z)$. A similar argument shows that, if all the credal sets $\mathcal{M}(X|Y, Z)$ are precise, then epistemic irrelevance satisfies the reverse intersection property.

probability distribution. Then epistemic irrelevance satisfies the direct contraction property: $(X \,\text{IR}\, Y|Z)\ \&(X \,\text{IR}\, W|(Y,Z)) \Rightarrow (X \,\text{IR}\, (W,Y)|Z)$.

Proof. Suppose that all the assumptions are satisfied and (x, z) is a possible value of (X, Z). Let h be any function of (W, Y). By the precision assumption and the second irrelevance assumption, the prevision $E[h(W, Y)|X = x, Y = y, Z = z] = E[h(W,Y)|Y = y, Z = z]$ is precisely determined whenever $(X, Y, Z) = (x, y, z)$ is possible. With all terms conditioned on $(X, Z) = (x, z)$, it is necessary for coherence of the conditional lower previsions that [32, Property 6.7.3],

$$\begin{aligned}\underline{E}[h(W,Y)|X=x,Z=z] &= \underline{E}[E[h(W,Y)|X=x,Y,Z=z]|X=x,Z=z]\\ &= \underline{E}[E[h(W,Y)|Y,Z=z]|X=x,Z=z]\\ &= \underline{E}[E[h(W,Y)|Y,Z=z]|Z=z]\\ &= \underline{E}[h(W,Y)|Z=z],\end{aligned}$$

where the assumption that $(X \,\text{IR}\, Y|Z)$ is used to eliminate X from the nested prevision $\underline{E}[E[h(W,Y)|Y, Z = z]|X = x, Z = z]$. This establishes that $(X \,\text{IR}\, (W,Y)\,|\,Z)$. □

5. Epistemic independence

Epistemic independence of two variables requires that uncertainty about each of the variables be unchanged by information about the other variable. Consequently, we say that two variables X and Y are *epistemically independent* conditional on Z, denoted by $(X \,\text{IN}\, Y\,|\,Z)$, when $(X \,\text{IR}\, Y\,|\,Z)$ and $(Y \,\text{IR}\, X\,|\,Z)$, i.e., each variable is epistemically irrelevant to the other variable. In terms of lower previsions, $(X \,\text{IN}\, Y\,|\,Z)$ is equivalent to: $\underline{E}[f(X)\,|\,Y = y, Z = z] = \underline{E}[f(X)\,|\,Z = z]$ for all functions $f(X)$ and all possible values of $(y,\ z)$, and $\underline{E}[g(Y)\,|\,X = x, Z = z] = \underline{E}[g(Y)|Z = z]$ for all functions $g(Y)$ and all possible values of $(x,\ z)$ [2, 31, 32].[12] The definition of epistemic independence can again be expressed in terms of equivalence of credal sets.

Unlike epistemic irrelevance, epistemic independence is a symmetric relation. In the special case of a precise probability model, epistemic independence agrees with the second definition of stochastic independence that was given in section 3.

Because epistemic independence is the symmetric version of epistemic irrelevance, its properties follow directly from the properties of irrelevance that were

[12] Again, it would be reasonable to strengthen this definition of epistemic independence by requiring also that X and Y be logically independent conditional on Z. The relationship between epistemic independence and logical independence is discussed in [32, section 9.1].

listed in Theorem 1. The next theorem summarises the valid graphoid properties of epistemic independence.

Theorem 3. Epistemic independence satisfies, for all random variables W, X, Y, Z:

Symmetry: $(X \text{ IN } Y \mid Z) \Rightarrow (Y \text{ IN } X \mid Z)$

Redundancy: $(X \text{ IN } Y \mid X)$.

Under the positivity assumption (7), epistemic independence also satisfies:

Decomposition: $(X \text{ IN } (W, Y) \mid Z) \Rightarrow (X \text{ IN } Y \mid Z)$

Weak union: $(X \text{ IN } (W, Y) \mid Z) \Rightarrow (X \text{ IN } W \mid (Y, Z))$.

6. Examples of relevance and irrelevance

The main purpose of the examples in this section is to show that there are simple models in which epistemic irrelevance violates the symmetry, direct contraction and intersection axioms, and epistemic independence violates contraction and intersection, and that this behaviour accords with informal notions of irrelevance. The examples show that, except for the properties listed in Theorem 1, *all* versions of the graphoid axioms (not just the 'direct' and 'reverse' versions) can be violated by epistemic irrelevance. The most important conclusions from the examples are:

- The symmetry axiom for epistemic irrelevance fails in Examples 1, 2, 3 and 4.
- When the positivity condition (7) fails, the reverse decomposition and reverse weak union axioms for epistemic irrelevance can fail, as shown in Example 1.
- The direct contraction axiom for epistemic irrelevance fails in Examples 1, 2, and 3, and the contraction axiom for epistemic independence fails in Example 1.
- The direct and reverse intersection axioms for epistemic irrelevance fail in Examples 4 and 5, and the intersection axiom for epistemic independence fails in Example 5.

Example 1. The first example shows that irrelevance can be asymmetric, and that both irrelevance and epistemic independence can violate the contraction axiom. The example involves one toss of a fair coin and one drawing of a ball from an urn of unknown composition.

Let Y denote the outcome of the coin toss, with possible values h (heads) and t (tails). Let W denote the colour of the ball drawn from the urn, with possible values r (red) and b (blue). There are two urns from which the ball can be drawn: urn R contains two red balls and one blue ball; urn B contains one red ball and two blue balls. The probability distributions associated with the two urns, P_R and P_B, are determined by $P_R(r) = 2/3$ or $P_B(r) = 1/3$.

Suppose that there is a *selection rule* that determines which urn the ball will be drawn from, as a function of the outcome of the coin toss. Using R and B to refer to the urn that is selected, there are four possible selection rules:

	$Y = h$	$Y = t$
$S = S_1$	R	R
$S = S_2$	B	B
$S = S_3$	R	B
$S = S_4$	B	R

Let X be a variable that gives partial information about the selection rule: $X = s$ if the *same* urn is used whatever the outcome of the coin toss, and $X = d$ if a *different* urn is used in the two possible cases (heads or tails). Initially we are completely ignorant about the selection rule, and therefore we are completely ignorant about the value of X.

In this example, it is clear that W is irrelevant to X, i.e., learning the colour of the ball that was drawn (W) tells us nothing about whether a different urn would have been chosen if the coin had landed differently. But it is also clear that X is relevant to W. Initially we are completely ignorant about which urn will be selected, so we know only that the chance of drawing a red ball (i.e., $W = r$) lies between 1/3 and 2/3. But if we learn that $X = d$, i.e., that the selected urn differs according to whether the coin lands heads or tails, then we know that the chance of drawing a red ball is precisely (1/2)(1/3) + (1/2)(2/3) = 1/2. Learning that $X = d$ changes our uncertainty about W, so that X is relevant to W. Epistemic irrelevance therefore violates the symmetry property: $(W \,\text{IR}\, X) \Rightarrow (X \,\text{IR}\, W)$.

It is also clear that X is irrelevant to Y, i.e., learning partial information about the selection rule, or indeed the complete rule, does not change our uncertainty about whether the coin will land heads or tails. Also X is irrelevant to W given Y, i.e., whatever the outcome of the coin toss, learning whether or not the selected urn depends on the outcome does not give us any useful information about which urn will be selected. But we have already established that X is relevant to W. Since X is therefore relevant to (W, Y) by the direct decomposition property, epistemic irrelevance violates the direct contraction axiom: $(X \,\text{IR}\, Y)$ & $(X \,\text{IR}\, W \mid Y) \Rightarrow (X \,\text{IR}\, (W, Y))$.

Furthermore, learning the outcome of the coin toss (Y) does not change uncertainty about X, and, given the outcome of the coin toss, learning the colour of the ball that was drawn (W) gives no information about X. Thus Y is irrelevant to X, and W is irrelevant to X given Y. Combining these with previously established relations, X is epistemically independent of Y, and X is independent of W given Y. Epistemic independence therefore violates the contraction property: $(X \,\text{IN}\, Y)$ & $(X \,\text{IN}\, W|Y) \Rightarrow (X \,\text{IN}\, (W, Y))$.

These conclusions can be confirmed by defining the appropriate credal sets. The four possible selection rules produce four possible joint probability distributions for (W, Y), which are denoted by Q_1, Q_2, Q_3, Q_4. These four distributions are the extreme points of $\mathcal{M}(W, Y)$. They are determined, for $w = r$ or $w = b$, by: $Q_1(w, h) = Q_1(w, t) = P_R(w)/2$ (corresponding to S_1); $Q_2(w, h) = Q_2(w, t) = P_B(w)/2$ (corresponding to S_2); $Q_3(w, h) = P_R(w)/2$ and $Q_3(w, t) = P_B(w)/2$ (corresponding to S_3); $Q_4(w, h) = P_B(w)/2$ and $Q_4(w, t) = P_R(w)/2$ (corresponding to S_4).

Using this notation, $\mathcal{M}(W, Y|X = s) = \{Q_1, Q_2\}$, $\mathcal{M}(W, Y|X = d) = \{Q_3, Q_4\}$, $\mathcal{M}(W|\, X, Y) = \mathcal{M}(W|Y) = \{P_R, P_B\} \cong \mathcal{M}(W)$, $\mathcal{M}(Y|X) = \mathcal{M}(Y) = \{U\}$ where U is the probability distribution that assigns probability 1/2 each to heads and tails, and $\mathcal{M}(X|W, Y) = \mathcal{M}(X|W) = \mathcal{M}(X|Y) = \mathcal{M}(X) = \{(0, 1), (1, 0)\}$. By comparing the appropriate credal sets, we see that $(X \,\mathrm{IR}\, Y), (X \,\mathrm{IR}\, W|Y), (Y \,\mathrm{IR}\, X), (W \,\mathrm{IR}\, X)$ and $(W \,\mathrm{IR}\, X|Y)$. Hence $(X \,\mathrm{IN}\, Y)$ and $(X \,\mathrm{IN}\, W|Y)$. But X is relevant to W since, for example, the upper and lower probabilities of drawing a blue ball are $\overline{P}(b) = 2/3$ and $\underline{P}(b) = 1/3$ before observing the value of X, but they change to $\overline{P}(b|X = d) = \underline{P}(b|X = d) = 1/2$ after learning that $X = d$.

Both the reverse decomposition and reverse weak union properties of Theorem 1 are violated in this example, since (X, Y) is irrelevant to W, but X is relevant to W, and Y is relevant to W given X. Similarly, properties 2 and 3 of Corollary 1 are also violated here. This shows that these properties can fail when the positivity assumption (7) is violated, as it is in this example: because we are completely ignorant about X, each possible value of X has lower probability zero.

The irrelevance and independence properties that are satisfied in this example are:

$$(X \,\mathrm{IN}\, Y), (Y \,\mathrm{IR}\, W), (W \,\mathrm{IR}\, X), ((X, Y) \,\mathrm{IR}\, W),$$
$$((W, Y) \,\mathrm{IR}\, X), (X \,\mathrm{IN}\, W|Y), (Y \,\mathrm{IR}\, X|W).$$

All the other irrelevance properties of this type (involving W, X, Y) are violated. The following versions of the contraction axiom are violated:

$$(X \,\mathrm{IR}\, Y) \;\&\; (W \,\mathrm{IR}\, X \,|\, Y) \Rightarrow (X \,\mathrm{IR}\, (W, Y)); \quad (Y \,\mathrm{IR}\, X) \quad (X \,\mathrm{IR}\, W|Y) \Rightarrow (X \,\mathrm{IR}\, (W, Y));$$
$$(Y \,\mathrm{IR}\, X) \;\&\; (W \,\mathrm{IR}\, X \,|\, Y) \Rightarrow (X \,\mathrm{IR}\, (W, Y)).$$

Example 2. To show that the violation of symmetry and direct contraction in Example 1 can occur when all possible configurations of the variables (W, X, Y) have positive lower probability, consider the following modification. The problem is exactly as in Example 1 except that now we know that the value of X is determined by tossing a second fair coin, unrelated to the first coin toss: if the second coin lands heads then $X = s$, and otherwise $X = d$. That is, Y and X are determined independently by tossing two fair coins. If $X = s$ then the urn is selected by either selection rule S_1 or S_2, and we are completely ignorant about which rule will be used. If $X = d$ then the urn is selected by either S_3 or S_4, and again we are completely ignorant about which rule. Finally, a ball is drawn from the selected urn to determine W.

In this modified problem, both X and Y have precise marginal probability distributions, since each value is determined by tossing a fair coin. The only imprecision is in the probability distribution of W conditional on (X, Y).

Let T denote the overall rule that is used to choose an urn. There are four possible rules T, where S_1, S_2, S_3 and S_4 are the selection rules used in Example 1:

	$X = s$	$X = d$
$T_{1,3}$	S_1	S_3
$T_{1,4}$	S_1	S_4
$T_{2,3}$	S_2	S_3
$T_{2,4}$	S_2	S_4

Define the probability distributions Q_i for (W, Y) as in Example 1. The joint credal set $\mathcal{M}(W, X, Y)$ has four extreme points (one for each rule $T_{i,j}$), which correspond to the pairs (Q_i, Q_j) with $i = 1$, 2 and $j = 3$, 4. For each such extreme point, the marginal distribution of X is uniform on $\{s,d\}$, the distribution of (W, Y) conditional on $X = s$ is Q_i, and the distribution of (W, Y) conditional on $X = d$ is Q_j. Since each Q_i assigns positive probability to every possible pair (w, y), it follows that all eight possible outcomes (w, x, y) now have positive lower probability.

When the value of X becomes known, this modified example is identical to Example 1. Consequently the credal sets $\mathcal{M}(Y|X), \mathcal{M}(W|X,Y), \mathcal{M}(W,Y|X = s)$ and $\mathcal{M}(W,Y|X = d)$ are exactly as in Example 1. We find that $\mathcal{M}(Y) = \mathcal{M}(Y|X)$ and $\mathcal{M}(W|Y) \cong \mathcal{M}(W|X,Y)$. It follows that, as in Example 1, irrelevance violates the direct contraction axiom.

However, other credal sets differ from those in Example 1. We find that $\mathcal{M}(X|Y) = \mathcal{M}(X) = \{U\} \ncong \{P_R, P_B\} = \mathcal{M}(X|W,Y)$, so that W is relevant to X given Y. Because X and W are no longer independent given Y, epistemic independence satisfies the contraction axiom in this example, unlike Example 1. Because $(X \,\text{IR}\, W|Y)$ but not $(W \,\text{IR}\, X|Y)$, epistemic irrelevance again violates the symmetry axiom.

The marginal upper and lower probabilities of drawing a blue ball are now $\overline{P}(b) = 7/12$ and $\underline{P}(b) = 5/12$, whereas the conditional upper and lower probabilities are still $\overline{P}(b|X = s) = 2/3$, $\underline{P}(b|X = s) = 1/3$, and $\overline{P}(b|X = d) = \underline{P}(b|X = d) = 1/2$. Because $\mathcal{M}(W)$ differs from Example 1, (X, Y) is now relevant to W and properties RD and RWU of Theorem 1 are satisfied.

Although this example involves only a small modification to Example 1, the modification destroys most of the irrelevance properties that were satisfied in Example 1. The only irrelevance properties involving W, X, Y that are satisfied in the modified problem are

$$(X \,\text{IN}\, Y), \quad \text{and} \quad (X \,\text{IR}\, W|Y).$$

The following versions of the contraction property are violated:

$$(X \,\text{IR}\, Y) \;\&\; (X \,\text{IR}\, W \,|\, Y) \Rightarrow ((W,Y) \,\text{IR}\, X); \quad (Y \,\text{IR} X) \;\&\; (X \,\text{IR}\, W \,|\, Y) \Rightarrow ((W,Y) \text{IR}\, X).$$

Example 3. Consider another modification of Example 1. Suppose that a second fair coin is tossed in a way that is unrelated to the first coin toss. Let V denote the outcome of the second toss. The introduction of V does not alter the irrelevance relations discussed in Example 1, but now the following versions of the decomposition, weak union and contraction axioms are violated: $(W \text{ IR } (X,V)) \Rightarrow (W \text{ IR } (W)$; $((X,V) \text{ IR } W) \Rightarrow (W \text{ IR } X)$; $(W \text{ IR } (X,V)) \Rightarrow (X \text{ IR } W|V)$; $((X,V) \text{ IR } W) \Rightarrow (W \text{ IR } X|V)$; and $(W \text{ IR } X)$ & $(V \text{ IR } W \mid X) \Rightarrow ((X,V) \text{ IR } W)$.

Example 4. In the preceding examples, epistemic irrelevance satisfies the direct intersection and reverse intersection properties. This example shows that these, and all other, versions of the intersection property can fail in simple cases.

Again W, X, Y are binary variables. Let W and Y denote the outcomes of two stochastically independent tosses of a fair coin, each with possible values h and t. Let X denote the colour of a ball that is drawn from an urn, with possible values r (red) and b (blue). The urn contains one red ball, one blue ball, and one other ball whose colour depends on the outcomes of the coin tosses. The third ball is selected to be red if a particular coin lands in a particular way, and to be blue if this coin lands in the other way, but we are completely ignorant about which coin determines the colour and about which outcome is associated with which colour.

There are four possible selection rules:

	$W = h$	$W = t$
$S = S_1$	R	B
$S = S_2$	B	R

	$Y = h$	$Y = t$
$S = S_3$	R	B
$S = S_4$	B	R

The joint credal set $\mathcal{M}(W,Y,X)$ contains four probability distributions, one corresponding to each selection rule, which are defined as follows. Let P_R and P_B denote the two possible probability distributions for X, determined by $P_R(r) = 2/3$ or $P_B(r) = 1/3$. Each probability distribution in $\mathcal{M}(W,Y,X)$ has a uniform distribution for (W, Y), since these are the outcomes of independent tosses of a fair coin, and the probability distribution of X conditional on $(W = w, Y = y)$ in each case is: (1) P_R if $W = h$, P_B if $W = t$; (2) P_B if $W = h$, P_R if $W = t$; (3) P_R if $Y = h$, P_B if $Y = t$; (4) P_B if $Y = h$, P_R if $Y = t$. It follows that W, X, Y are logically independent and all eight possible outcomes (w, y, x) have positive lower probability.

It can be verified that $\mathcal{M}(X|W,Y) = \{P_R, P_B\} \cong \mathcal{M}(X|W) = \mathcal{M}(X|Y)$. Consequently $(W \text{ IR } X|Y)$ and $(Y \text{ IR } X|W)$. But $\mathcal{M}(X)$ contains only the uniform distribution on $\{r, b\}$, and therefore (W, Y) is relevant to X. It follows that epistemic irrelevance violates the reverse intersection property: if W and Y are logically independent, $(W \text{ IR } X|Y)$ & $(Y \text{ IR } X|W) \Rightarrow ((W,Y) \text{ IR } X)$.[13]

[13] Despite this example, irrelevance must 'almost' satisfy reverse intersection: if W and Y are logically independent, $(W \text{ IR } X|Y)$ and $(Y \text{ IR } X|W)$, then, for every function $f(X)$, $\underline{E}[f(X)|W = w, Y = y]$ must be constant (independent of w and y), and the constant value must be less than or equal to $\underline{E}[f(X)]$.

Also $\mathcal{M}(Y|W, X = \mathcal{M}(Y|X) \cong \{P_R, P_B\}$, so that $(W \text{ IR } Y|X)$. But since $\mathcal{M}(X|\ W) \ncong \mathcal{M}(X)$, W is relevant to X, and by the direct decomposition property of Theorem 1, W is relevant to (X, Y). It follows that epistemic irrelevance violates the direct intersection property: if W and Y are logically independent, $(W \text{ IR } Y|X)$ & $(W \text{ IR } X|Y) \nRightarrow (W \text{ IR } (X, Y))$.

This failure of direct and reverse intersection again accords with intuition. First consider reverse intersection. Before we learn the outcomes of the coin tosses, there is probability 1/2 that the third ball in the urn will be red, whatever the selection rule, because of the randomness in each coin toss. So, initially, there is precise probability 1/2 of drawing a red ball from the urn: $\underline{P}(X = r) = \overline{P}(X = r) = 1/2$. But after we learn the outcome of one or both coin tosses, some of the randomness disappears, and the four selection rules assign different probabilities, ranging from 1/3 to 2/3, to drawing a red ball: $\underline{P}(X = r|W = w) = \underline{P}(X = r|W = w, Y = y) = 1/3$ and $\overline{P}(X = r|W = w) = \overline{P}(X = r|W = w, Y = y) = 2/3$ for all possible values h, t of w and Y.[14] Learning the outcome of one or both coin tosses (W, Y) changes our uncertainty about the outcome of the drawing (X). That is, (W, Y) is relevant to X. But after learning the outcome of one coin toss, we are completely ignorant about the colour of the third ball in the urn, and learning the outcome of the second coin toss does not change this state of ignorance. Thus W is irrelevant to X given Y, and Y is irrelevant to X given W.

The argument for violation of direct intersection is similar. The preceding argument shows that W is irrelevant to X given Y, and also that W is relevant to X which implies that W is relevant to (X, Y). Finally, W is irrelevant to Y given X because, given the colour of the ball that was drawn (X), learning the outcome of one coin toss tells us nothing about the outcome of the other coin toss.

The direct contraction axiom is also violated here: W is irrelevant to Y because the two coin tosses are stochastically independent, W is irrelevant to X given Y by the previous argument, but W is relevant to (X, Y).

The following irrelevance and independence properties are satisfied in this example:

$$(W \text{ IN } Y), \quad (W \text{ IN } Y|X), \quad (Y \text{ IR } X|W) \quad \text{and} \quad (W \text{ IR } X|Y).$$

All the other irrelevance properties involving W, X, Y are violated. From this we can see that all eight possible versions of the intersection property fail in this example. The symmetry property also fails because $(Y \text{ IR } X|W)$ but not $(X \text{ IR } Y|W)$.

Example 5. In all the preceding examples, epistemic independence satisfies the intersection axiom. This final example shows that intersection can fail for epistemic

[14] This behaviour, in which conditioning on a variable W makes probabilities less precise whatever value W may take, is called *dilation* in [25]. The preceding footnote shows that dilation occurs whenever the reverse intersection property is violated, since then conditioning on (W, Y) must make probabilities concerning X less precise. Dilation also occurs in Example 5.

independence. The example involves three tosses of a fair coin with unknown interaction between the tosses. Each of the variables W, X, Y represents the outcome of a different coin toss, with possible values h (heads) and t (tails). The variables W, X, Y are logically independent. The joint credal set $\mathcal{M}(W, X, Y)$ contains all probability distributions for (W, X, Y) under which the marginal probability distribution for each variable is uniform on $\{h, t\}$. Let (S, U, V) be any permutation of (W, X, Y). The joint credal set is invariant under such a permutation. By definition of the joint credal set, $\mathcal{M}(S)$ contains just the uniform distribution on $\{h, t\}$.

One possible type of interaction between the tosses is that toss U is fair, but the outcomes of tosses S and V are always identical to the outcome of U. This would produce a probability distribution with $P(S = h|U = h) = 1$. Another possibility is that the outcomes of S and V are always different from the outcome of U, which would produce $P(S = h|U = h) = 0$. Hence $\overline{P}(S = h|U = h) = 1, \underline{P}(S = h|U = h) = 0$, and the conditional credal set $\mathcal{M}(S|U)$ is *vacuous*, i.e., it contains all the probability distributions on $\{h, t\}$. A similar argument shows that $\mathcal{M}(S|U, V)$ is also vacuous. Since $\mathcal{M}(S|U, V) = \mathcal{M}(S|U)$, we have $(V \,\text{IR}\, S|U)$ for each permutation (S, U, V) of (W, X, Y). Hence $(X \,\text{IN}\, W|Y)$ and $(X \,\text{IN}\, Y|W)$.

But X is not independent of (W, Y), since $\mathcal{M}(X)$ is precise whereas $\mathcal{M}(X|W, Y)$ is vacuous, i.e., $\overline{P}(X = h) = \underline{P}(X = h) = 1/2$ but $\overline{P}(X = h|W = w, Y = y) = 1$ and $\underline{P}(X = h|W = w, Y = y) = 0$ for all possible values of w and y. Initially we know that the probability of heads on toss X is precisely 1/2, but, because we know nothing about the interaction between tosses, learning the outcome of one or both of the other tosses produces complete ignorance about the probability of heads on toss X. (This is another example of dilation.) It follows that epistemic independence violates the intersection axiom.

The six extreme points of $\mathcal{M}(W, X, Y)$ are listed in the Appendix. The only irrelevance relations that hold are $(V \,\text{IR}\, S\,|\,U)$ for each permutation (S, U, V) of (W, X, Y).

Under this model, $\underline{P}(W = w, Y = y) = 0$ for all possible values of w and y. To show that the intersection axiom can fail even when all possible configurations of the variables (w, x, y) have positive lower probability, we can modify the example by putting an upper bound on the degree of interaction between tosses. To be specific, suppose that we reduce the joint credal set $\mathcal{M}(W, X, Y)$ by imposing the extra constraint: $P(S = s|U = u, V = v) \geq 1/3$ whenever (S, U, V) is a permutation of (W, X, Y) and s, u, v take values in $\{h, t\}$.

The following argument shows that, in this modified example, $\mathcal{M}(S|U, V) = \mathcal{M}(S|U)$. It suffices to show that $\underline{P}(S = s|U = u, V = v) = \underline{P}(S = s|U = u) = 1/3$ for all possible values of s, u, v, since S has only two possible values and therefore the lower probabilities determine the corresponding credal sets. By definition of the model, $\underline{P}(S = s|U = u, V = v) \geq 1/3$. Using a general coherence relationship, $\underline{P}(S = s|U = u) \geq \underline{E}[\underline{P}(S = s|U = u, V)|U = u] \geq \underline{E}[1/3|U = u] = 1/3$. To show that $\underline{P}(S = s|U = u, V = v) \leq 1/3$ and $\underline{P}(S = s|U = u) \leq 1/3$, it suffices to find a probability distribution P^* in $\mathcal{M}(S, U, V)$ that satisfies $P^*(S = s|U = u, V = v) = P^*(S = s|U = u) = 1/3$. Define P^* so that V is stochastically independent of (S, U), the marginal distribution of V is uniform on $\{h, t\}$, and the distribution of (S, U) is: $P^*(s, u) = P^*(s', u') = 1/6$ and $P^*(s, u') = P^*(s', u) = 1/3$, where (s, s') and (u, u') are permutations of (h, t). It can be verified that P^* has the required properties.

Hence, in the modified example, $(X \text{ IN } W|Y)$ and $(X \text{ IN } Y|W)$. But again (W, Y) is relevant to X, since $\overline{P}(X = h) = \underline{P}(X = h) = 1/2$ but $\underline{P}(X = h|W = w, Y = y) = 1/3$ and $\overline{P}(X = h|W = w, Y = y) = 1 - \underline{P}(X = t|W = w, Y = y) = 2/3$, for all possible values of w and y. Again the intersection property fails for epistemic independence. The epistemic irrelevance relation in this case violates all eight possible versions of the intersection axiom, including direct intersection and reverse intersection.

The 16 extreme points of $\mathcal{M}(W, X, Y)$ in this modified example are listed in the Appendix. Again, the only irrelevance relations that hold are $(V \text{ IN } S|U)$ for each permutation (S, U, V) of (W, X, Y).

7. Conclusion

The mathematical properties of Walley's concepts of epistemic irrelevance and epistemic independence depart in several ways from the properties of stochastic independence. The most basic difference is that irrelevance and independence are not identical in Walley's framework. This distinction is not commonly made in formal studies. For example, both Dawid [7] and Pearl [22], in defending the graphoid axioms, use the terms 'irrelevance' and 'independence' almost interchangeably. Dawid introduces the graphoid axioms, excluding the intersection axiom, by saying that "we can rephrase these as assertions of irrelevance" and he suggests that we can consider them "as a reasonable set of axioms for a general concept of irrelevance" [7].

But irrelevance and independence do not have the same meaning in ordinary language. The concept of irrelevance is asymmetric: X may be irrelevant to Y without Y being irrelevant to X. On the other hand, symmetry seems to be essential to any concept of independence, since two variables are independent only when they are irrelevant to each other.

Such considerations of symmetry are crucial to appreciate the results in this paper. For example, consider the contraction property. The failure of direct contraction may seem strange at first, but all that is necessary about contraction is contained in the irrelevance property of reverse contraction. In fact, Dawid and Pearl, in discussing the contraction axiom, defend only reverse contraction. To quote Pearl [22, p. 85]:

> The contraction axiom states that if we judge W irrelevant to X after learning some irrelevant information Y, then W must have been irrelevant before we learned Y.

This property of reverse contraction is indeed a desirable property of irrelevance, but the quite different property of direct contraction seems necessary only if a preconceived notion of symmetry is combined with reverse contraction.

The examples in section 6 show that epistemic irrelevance can violate the symmetry, direct contraction and intersection axioms, that epistemic independence can violate the contraction and intersection axioms and that these violations are consistent with our intuitive understanding of irrelevance. This suggests that structures that are more general than graphoids, and especially ones that weaken the symmetry, contraction and intersection axioms, deserve further attention. The properties of epistemic irrelevance that are listed in Theorem 1 could be taken as basic axioms for a more general concept of irrelevance. It seems unlikely that the properties listed in Theorem 1 completely characterize epistemic irrelevance, especially as there is no finite system of axioms that characterizes stochastic independence [27]. It may be necessary to add further properties to those listed in Theorem 1, in order to produce a mathematical structure that is interesting and useful.

Acknowledgements

Fabio Cozman's interest in this topic arose from questions posed by Milan Studený. Thanks also to David Avis for making the software *lrs* available (at http://cgm.cs.mcgill.ca/~avis/C/lrs.html). Fabio Cozman was partially supported by CNPq through grant 300183/98-4. Peter Walley was supported by FAPESP during part of this project.

Appendix

Extreme points for Example 5

The extreme probability distributions in Example 5 were computed using the software *lrs*. In the first version of the example, $\mathcal{M}(W, X, Y)$ consists of all probability distributions P such that $P(S = h) = 1/2$ when S is either W, X or Y. The extreme points of $\mathcal{M}(W, X, Y)$ are listed in table 1. The modified version of the example adds the constraints: $P(S = s|U = u, V = v) \geq 1/3$ whenever (S, U, V) is a permutation of (W, X, Y)

Table 1
The six extreme points of $\mathcal{M}(W, X, Y)$ for the first version of Example 5.

$P(W, X, Y)$	$P(hhh)$	$P(hht)$	$P(hth)$	$P(htt)$	$P(thh)$	$P(tht)$	$P(tth)$	$P(ttt)$
P_1	0	0	0	1/2	1/2	0	0	0
P_2	0	0	1/2	0	0	1/2	0	0
P_3	0	1/2	0	0	0	0	1/2	0
P_4	1/2	0	0	0	0	0	0	1/2
P_5	0	1/4	1/4	0	1/4	0	0	1/4
P_6	1/4	0	0	1/4	0	1/4	1/4	0

Table 2
The 16 extreme points of $\mathcal{M}(W, X, Y)$ for the second version of Example 5.

$P(W, X, Y)$	$P(hhh)$	$P(hht)$	$P(hth)$	$P(htt)$	$P(thh)$	$P(tht)$	$P(tth)$	$P(ttt)$
P_1	1/12	1/12	1/6	1/6	1/6	1/6	1/12	1/12
P_2	1/12	1/6	1/12	1/6	1/6	1/12	1/6	1/12
P_3	1/12	1/6	1/6	1/12	1/12	1/6	1/6	1/12
P_4	1/12	1/6	1/6	1/12	1/6	1/12	1/12	1/6
P_5	1/6	1/12	1/12	1/6	1/12	1/6	1/6	1/12
P_6	1/6	1/12	1/12	1/6	1/6	1/12	1/12	1/6
P_7	1/6	1/12	1/6	1/12	1/12	1/6	1/12	1/6
P_8	1/6	1/6	1/12	1/12	1/12	1/12	1/6	1/6
P_9	1/14	1/7	1/7	1/7	1/7	1/7	1/7	1/14
P_{10}	1/7	1/14	1/7	1/7	1/7	1/7	1/14	1/7
P_{11}	1/7	1/7	1/14	1/7	1/7	1/14	1/7	1/7
P_{12}	1/7	1/7	1/7	1/14	1/14	1/7	1/7	1/7
P_{13}	1/10	1/10	1/10	1/5	1/5	1/10	1/10	1/10
P_{14}	1/10	1/10	1/5	1/10	1/10	1/5	1/10	1/10
P_{15}	1/10	1/5	1/10	1/10	1/10	1/10	1/5	1/10
P_{16}	1/5	1/10	1/10	1/10	1/10	1/10	1/10	1/5

and s, u, v take values in $\{h, t\}$. The extreme points of $\mathcal{M}(W, X, Y)$ in this case are listed in table 2.

References

1. J.-M. Bernard, T. Seidenfeld and M. Zaffalon, eds., in: *Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications* (Carleton Scientific, Lugano, Switzerland, 2003).
2. I. Couso, S. Moral and P. Walley, A survey of concepts of independence for imprecise probabilities, Risk, Decision and Policy 5 (2000) 165–181.
3. F.G. Cozman, Irrelevance and independence relations in quasi-Bayesian networks, in: *XIV Conference on Uncertainty in Artificial Intelligence*, eds. G. Cooper and S. Moral (Morgan Kaufmann, San Francisco, 1998) pp. 89–96.
4. F.G. Cozman, Irrelevance and independence axioms in quasi-Bayesian theory, in: *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, eds. A. Hunter and S. Parsons (Springer, Berlin Heidelberg New York, 1999) pp. 128–136.
5. F.G. Cozman, Credal networks, Artificial Intelligence 120 (2000) 199–233.
6. F.G. Cozman, Algorithms for conditioning on events of zero lower probability, in: *Fifteenth International Florida Artificial Intelligence Research Society Conference*, eds. S. Haller and G. Simmons (AAAI, Menlo Park, California, 2002) pp. 248–252.
7. A.P. Dawid, Conditional independence, in: *Encyclopedia of Statistical Sciences, Update Volume 2*, eds. S. Kotz, C.B. Read and D.L. Banks (Wiley, New York, 1999) pp. 146–153.
8. A.P. Dawid, Separoids: A mathematical framework for conditional independence and irrelevance, Annals of Mathematics and Artificial Intelligence 32 (2001) 335–372.
9. C.P. de Campos and F.G. Cozman, Belief updating and learning in semi-qualitative probabilistic networks, in: *Conference on Uncertainty in Artificial Intelligence*, eds. F. Bacchus and T. Jaakkola (AUAI, Edinburgh, 2005).

10. L. de Campos and S. Moral, Independence concepts for convex sets of probabilities, in: *XI Conference on Uncertainty in Artificial Intelligence*, eds. P. Besnard and S. Hanks (Morgan Kaufmann, San Francisco, 1995) pp. 108–115.
11. G. de Cooman, F.G. Cozman, S. Moral and P. Walley, eds., in: *Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications* (Imprecise Probabilities Project, Ghent, Belgium, 1999).
12. G. de Cooman, T.L. Fine and T. Seidenfeld, eds., in: *Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications* (Shaker, Maastricht, The Netherlands, 2001).
13. D. Galles and J. Pearl, Axioms of causal relevance, Artificial Intelligence 97 (1997) 9–43.
14. D. Geiger, T. Verma and J. Pearl, Identifying independence in Bayesian networks, Networks 20 (1990) 507–534.
15. J.Y. Halpern, *Reasoning About Uncertainty* (MIT, Cambridge, Massachusetts, 2003).
16. P. Hansen and B. Jaumard, Probabilistic satisfiability. Technical Report, Les Cahiers du GERAD, École Polytechique de Montréal, G-96-31, 1996.
17. I. Levi, *The Enterprise of Knowledge* (MIT, Cambridge, Massachusetts, 1980).
18. F. Matús and M. Studený, eds., in: *Workshop on Conditional Independence Structures and Graphical Models* (Fields Institute for Research in Mathematical Sciences, Toronto, 1999).
19. S. Moral, Epistemic irrelevance on sets of desirable gambles, in: *Second International Symposium on Imprecise Probabilities and Their Applications*, eds. G. de Cooman, T.L. Fine and T. Seidenfeld (Shaker, Maastricht, The Netherlands, 2001) pp. 247–254.
20. N.J. Nilsson, Probabilistic logic, Artificial Intelligence 28 (1986) 71–87.
21. S. Parsons. *Qualitative Approaches to Reasoning Under Uncertainty* (MIT, Cambridge, Massachusetts, 2001).
22. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, San Mateo, California, 1988).
23. J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, Cambridge, UK, 2000).
24. J. Pearl, On two pseudo-paradoxes in Bayesian analysis, Annals of Mathematics and Artificial Intelligence 32 (2001) 171–177.
25. T. Seidenfeld and L. Wasserman, Dilation for sets of probabilities, Annals of Statistics 21 (1993) 1139–1154.
26. P. Spirtes, C. Glymour and R. Scheines, *Causation, Prediction, and Search*, 2nd ed. (MIT, Cambridge, Massachusetts, 2003).
27. M. Studený, Conditional independence relations have no finite complete characterizations, in: *Transactions of the Eleventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, eds. S. Kubik and J.A. Visek (Kluwer, Dordrecht, 1992) pp. 377–396.
28. M. Studený, Semigraphoids and structures of probabilistic conditional independence, Annals of Mathematics and Artificial Intelligence 21 (1997) 71–98.
29. B. Vantaggi, Graphical models for conditional independence structures, in: *Second International Symposium on Imprecise Probabilities and Their Applications*, eds. G. de Cooman, T.L. Fine and T. Seidenfeld (Shaker, Maastricht, The Netherlands, 2001) pp. 332–341.
30. J. Vejnarová. Conditional independence relations in possibility theory, in: *First International Symposium on Imprecise Probabilities and Their Applications*, eds. G. de Cooman, F.G. Cozman, S. Moral and P. Walley (Imprecise Probabilities Project, Ghent, Belgium, 1999) pp. 343–351.
31. P. Vicig, Epistemic independence for imprecise probabilities, International Journal of Approximate Reasoning 24 (2000) 235–250.
32. P. Walley, *Statistical Reasoning with Imprecise Probabilities* (Chapman and Hall, London, 1991).
33. P. Walley, Measures of uncertainty in expert systems, Artificial Intelligence 83 (1996) 1–58.

34. P. Walley and T.L. Fine, Towards a frequentist theory of upper and lower probability, Annals of Statistics 10 (1982) 741–761.
35. 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 (2004) 119–151.
36. M.P. Wellman, Fundamental concepts of qualitative probabilistic networks, Artificial Intelligence 44 (1990) 257–303.