Belief Function Independence: I. The Marginal Case.

Boutheina Ben Yaghlane *
AMID, Université de Tunis, Tunisia

Philippe Smets
IRIDIA, Université Libre de Bruxelles, Belgium

Khaled Mellouli
AMID, Université de Tunis, Tunisia

May 29, 2001

Abstract

In this paper, we study the notion of marginal independence between two sets of variables when uncertainty is expressed by belief functions as understood in the context of the transferable belief model. We define the concepts of non-interactivity and irrelevance, that are not equivalent. Doxastic independence for belief functions is defined as irrelevance and irrelevance preservation under Dempster's rule of combination. We prove that doxastic independence and non-interactivity are equivalent.

Keywords. Belief functions, Transferable Belief Model, Non-interactivity, Irrelevance, Doxastic Independence.

1 Introduction

An important requirement, for uncertainty reasoning system management, is to specify the conditions under which one item of information is considered independent from another, given what we know, and to represent knowledge in structures that display these conditions. In the probabilistic framework, these conditions are identified with the notion of *independence*, also called *irrelevance* or *informational irrelevance* [16].

*Address correspondence to Boutheina Ben Yaghlane, AMID, Université de Tunis, IHEC-Carthage Présidence 2016 Tunis, Tunisia. E-mail: boutheina.yaghlane@ihec.rnu.tn and psmets@ulb.ac.be

The notion of informational irrelevance has been extensively studied in probability theory [7], [8], [15], [16], where it is identified with independence. The concept of independence has also been studied in other non-probabilistic frameworks such that Spohn's theory of ordinal conditional functions [22], Zadeh's possibility theory [1], [4], [9], [11], [24], [28], upper and lower probabilities theory [3], [5], [6], [25], and in an abstract framework that unifies different uncertainty calculi called valuation-based system [19]. However, the concept of independence has not been widely treated in belief functions theory.

The aim of this paper is to investigate some ways to define independence relationships between variables when uncertainty is expressed under the form of belief functions. Some other researches studying this topic are [3], [17], and [23]. We focus on belief functions in the context of the transferable belief model [21], [20], a model to represent quantified beliefs where the beliefs are represented by belief functions. In order to avoid possible confusion, we insist on the fact that the TBM is unrelated to lower probability models. We concentrate on the intuitive meaning of each definition of independence and we discuss the possible links between them. In this paper, we consider only the marginal case, leaving the conditional case for the second part.

In addition to the obvious theoretical reasons for the study of independence, there are also practical interests. Indeed, thanks to independence, many computational tasks can be simplified. Complex evidential problem can be modularized into simpler components in such a way that we only treat the pieces of information having relevance to the question we are interested in.

Furthermore, the practical importance of independence is captured in three processes supported by expert systems: elicitation, inference, and explanation [10]. For instance in the phase of eliciting probabilistic models from human experts, qualitative independencies among variables can often be easily and confidently asserted whereas numerical assessments can be very imprecise.

The main question is the definition of independence. In the literature there are two main approaches to define independence:

- **Irrelevance approach**: Two variables are said to be *independent* if no piece of information that can be learned about one of them can change our state of knowledge about the other. This form of independence is called *irrelevance*. It can be expressed by experts.

- **Decomposition approach**: Two variables are said to be *independent* if the global information about the two variables can be expressed as a combination of two pieces of knowledge, one for each variable. This form of independence is called here *non-interactivity*. It permits efficient computation by local computations without loosing any information.

In the case of probability theory, both approaches are equivalent, so the distinction is not essential. However, these approaches do not have identical meaning in belief function theory.

The rest of this paper is organized as follows. In section 2, we present some useful definitions and notations needed for belief function context. Next, in section 3, we present the definitions of irrelevance, non-interactivity and independence in probability theory. To extend these definitions to belief function case, we recall, in section 4, two definitions proposed by Shafer [17] called cognitive and evidential independence. After this, we define the concepts of marginal non-interactivity (section 5), irrelevance (section 6) and independence (section 7) for belief functions, making clear the links between them. Finally, in section 8, we make some concluding remarks.

2 Belief Function Theory

2.1 The Transferable Belief Model

The theory of belief functions, also known as Dempster-Shafer theory and theory of evidence, aims to model someone's degree of belief. It is regarded as a generalization of the Bayesian approach. Since this theory was developed by Shafer [17], many interpretations have been proposed. Among them, we can distinguish:

- **a lower probability model** where beliefs are represented by families of probability functions and the belief functions are the lower envelop of these families. This model is considered as a special case of imprecise probabilities [26].
- **Dempster's model** derived from probability theory where a probability space is mapped by a one-to-many mapping on another space. It is essentially at the core of the hint theory [13].
- **the transferable belief model (TBM)** where beliefs are represented by belief functions [21] [20]. This model is unrelated to probability models whereas the other two are generalization of them.

In this paper, we are only concerned with the TBM, so we will use the concepts as defined in it.

Definition 1 *Let Ω be a finite set of elements, called the frame of discernment. The mapping $bel : 2^\Omega \rightarrow [0,1]$ is an (unnormalized) belief function if and only if there exists a basic belief assignment (bba) $m : 2^\Omega \rightarrow [0,1]$ such that:*

$$(i) \quad \sum_{A \subseteq \Omega} m(A) = 1,$$

$$(ii) \quad bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B),$$

$$(iii) \quad bel(\emptyset) = 0.$$

The value $m(A)$ represents the degree of belief that is exactly committed to A. Due to the lack of information, $m(A)$ cannot support any more specific event. The value $bel(A)$ quantifies the strength of the belief that the event A occurs. A subset A such that $m(A) > 0$ is called a *focal element* of *bel*. *bel* is *vacuous* if the only focal element is Ω. In the TBM, we accept that $m(\emptyset)$ can be positive. It reflects either the fact that the actual value of Ω might not be in Ω (open-world assumptions, [21]) or the conflict that underlies the pieces of evidence used to build m.

Given a belief function *bel*, we can define a *plausibility function pl*: $2^\Omega \to [0,1]$ and a *commonality function q*: $2^\Omega \to [0,1]$ as follows: for $A \subseteq \Omega$,

$$pl(A) = bel(\Omega) - bel(\overline{A}) = \sum_{B \cap A \neq \emptyset} m(B),$$
$$pl(\emptyset) = 0,$$
$$q(A) = \sum_{A \subseteq B \subseteq \Omega} m(B).$$

The value $pl(A)$ quantifies the maximum amount of potential specific support that could be given to A. The commonality function $q(A)$ represents a measure of uncertainty in the context where A holds and it is useful for simplifying some computations. It is proved that m, bel, pl and q are in one-to-one correspondence with each other [17].

2.2 Notations and Properties of Belief Functions

In this section, we present some notations and properties necessary when belief functions are used.

2.2.1 Variables.

Let U $= \{X, Y, Z, ...\}$ be a set of finite variables, $\Theta_X = \{x_1, ..., x_n\}$ be the domain relative to the variable X (with a finite cardinality n), and x represents any instance of X. For simplicity sake, we denote Θ_X by X, Θ_Y by Y... Let Ω be a frame of discernment [17] (or universe of discourse). It is the Cartesian product of the domains of the variables in U. For example, $X \times Y$ represents the product space of variables X and Y, and when there is no ambiguity, it is simply denoted by XY. The elements of X (Y ...) are represented by indexed variables like x_i (y_j ...) whereas x (y ...) denote subsets of X (Y ...).

For $x \subseteq X$ and $y \subseteq Y$, (x, y) is defined by $(x, y) = \{(x_i, y_j) : x_i \in x, y_j \in y\}$, and similarly for (x, y, z) ...

By construction, Ω is a common refinement of X, Y, Z... ([17], page 121). The variables X, Y, Z.... are themselves independent coarsenings of Ω ([17], page 127) where, for instance when $\Omega = XYZ$, the "independence" means that:

$$(x_i, Y, Z) \cap (X, y_j, Z) \cap (X, Y, z_k) \neq \emptyset, \; \forall x_i \in X, y_j \in Y, z_k \in Z.$$

While studying marginal independence, all we need is two disjoint subsets of variables in U. We can as well redefine two new variables X and Y which domains are the Cartesian product of those variables considered in each set, respectively. This simplifies the notation as we may consider only two variables X and Y, and $\Omega = XY$.

We also often use the set $Rect_{XY}$ which is the set of subsets of XY that can be represented as (x, y) for some $x \subseteq X$ and some $y \subseteq Y$.

2.2.2 Belief functions.

Let BK denotes the background knowledge that holds and that underlies the beliefs. In BK, we find the classical conditioning events. We introduce the following notations and their related properties:

- $bel^{\Omega\downarrow X}$ is the marginal of bel^{Ω} on X. The Ω superscript will not be mentioned when there is no risk of confusion. In particular, we have:

$$bel^{XY\downarrow X}(x) = bel^{XY}(x, Y),$$

$$pl^{XY\downarrow X}(x) = pl^{XY}(x, Y).$$

- $bel^{\Omega}[BK]$ denotes the belief function on Ω when BK holds. It can be seen as a vector in a $2^{|\Omega|}$ dimensional space. Classically, it was denoted as $bel^{\Omega}(. \mid BK)$, but the bracket notation turns out to be more convenient.

- $bel^{\Omega}[BK](A)$ denotes the value of $bel^{\Omega}[BK]$ at $A \subseteq \Omega$. When BK is the proposition that states that the actual value of Ω belongs to $B \subseteq \Omega$, its value is given by:

$$\begin{aligned} bel^{\Omega}[B](A) &= bel^{\Omega}(A \cup \overline{B}) - bel^{\Omega}(\overline{B}) \\ pl^{\Omega}[B](A) &= pl^{\Omega}(A \cap B) \\ q^{\Omega}[B](A) &= q^{\Omega}(A) \text{ if } A \subseteq B, \\ &= 0 \text{ otherwise.} \end{aligned}$$

  These are the so called Dempster's rule of conditioning (except for the normalization factor).

- The $\oplus$ symbol represents Dempster's rule of combination in its normalized form and $\circledcirc$ represents the conjunctive combination, i.e., the same operation as Dempster's rule of combination except the normalization (the division by $1 - m(\emptyset)$) is not performed. The conjunctive combination rule can be written equivalently as:

$$m_{1\circledcirc 2}(A) = m_1 \circledcirc m_2(A) = \sum_{B,C\subseteq\Omega, B\cap C=A} m_1(B)m_2(C)$$

$$q_{1\circledcirc 2}(A) = q_1(A)q_2(A).$$

- Note that conditioning and marginalization do not commute, so the order of the symbols is important. $bel^{XY}[y]^{\downarrow X}$ is the belief function obtained by conditioning bel^{XY} on y and the result is then marginalized on X.

In order to distinguish between normalized belief functions as those defined by Shafer and the unnormalized ones used in the transferable belief model, we use the following convention. Normalized functions have their first letter in upper-cases, whereas the unnormalized ones have theirs in lower cases. So M, Bel, Pl and Q denote the normalized forms, whereas m, bel, pl and q denote the unnormalized forms. To get the normalized forms, one just divide the unnormalized ones by the factor $1 - m(\emptyset)$ (putting $M(\emptyset) = 0$) or identically by $bel(\Omega)$ or $pl(\Omega)$.

Notations. The following notations are used repeatedly:
N1. $f^{XY}(A, B) = f^{XY}(\{(x, y) : x \in A, y \in B\}) \quad \forall A \subseteq X, B \subseteq Y$ and $f \in \{m, bel, pl, q\}$.

A particular Ω is often used. Its elements are defined as follows:
N2. Let $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2\}$. We define $\Omega = X \times Y = \{w_1, w_2, w_3, w_4\}$ with:

$$\begin{aligned} w_1 &= (x_1, y_1) \\ w_2 &= (x_1, y_2) \\ w_3 &= (x_2, y_1) \\ w_4 &= (x_2, y_2) \end{aligned}$$

We next prove two useful lemmas and then we present several classical properties of belief functions theory for the reader's convenience.

Lemma 1 *For any plausibility function pl^{XY} defined on XY, we have*

$$pl^{XY}[y]^{\downarrow X}(x) = pl^{XY}(x, y), \ \forall x \subseteq X, \forall y \subseteq Y$$

Proof. Consider $m^{XY}(w)$ for $w \subseteq XY$.

Case 1: $w \cap y = \emptyset$ then $m^{XY}(w)$ will be included in $m^{XY}[y]^{\downarrow X}(\emptyset)$. Thus it will neither be included in $pl^{XY}[y]^{\downarrow X}(x)$ nor in $pl^{XY}(x, y)$. So it is absent from both terms of the lemma.

Case 2: $w \cap y \neq \emptyset$ then either $w \cap y \cap x = \emptyset$ and $m^{XY}(w)$ is not included in the two pl of the lemma, or $w \cap y \cap x \neq \emptyset$ and $m^{XY}(w)$ is included in both pl of the lemma.

This holds for any $w \subseteq XY$ and any $y \subseteq Y$. Therefore, both pl of the lemma contain the same bbm and they are thus equal. □

Lemma 2 *For any plausibility function pl^{XY} defined on XY, we have*

$$pl^{XY\downarrow X}(x) = pl^{XY}(x, Y)$$

Proof. Consider $m^{XY}(w)$ for $w \subseteq XY$. Either $w \cap x = \emptyset$ in which case $m^{XY}(w)$ is neither included in $pl^{XY}(x, Y)$ nor in $pl^{XY\downarrow Y}(x)$, or $w \cap x \neq \emptyset$ in which case it is included in both pl of the lemma. So, both pl of the lemma contain the same bbm and they are thus equal. □

Some useful properties. We list some relations that will be used in the future and are classical in belief function theory.

P1. $q^{\Omega}(A) = pl^{\Omega}(A)$ when $|A| = 1, A \subseteq \Omega$.

P2. $m^{\Omega}(\emptyset) = \sum_{A \subseteq \Omega} (-1)^{|A|} q^{\Omega}(A)$.

P3. $q^{\Omega}(A) = \sum_{B \subseteq A} (-1)^{|B|+1} pl^{\Omega}(B),\ \forall A, \emptyset \neq A \subseteq \Omega\ ,\ q^{\Omega}(\emptyset) = 1$.

P4. $q^{\Omega}(A) \geq q^{\Omega}(A \cup B),\ \forall A, B \subseteq \Omega$.

3 Independence in Probability Theory

First, we recall the definition of probabilistic independence. We say that two random variables X and Y are (*marginally*) *independent* with respect to a distribution P on the space XY, denoted by $X \perp\!\!\!\perp_P Y$, if and only if one of the following conditions is satisfied: for all $x \subseteq X$, all $y \subseteq Y$,

- $P^{XY}(x, y) = P^{XY\downarrow X}(x) P^{XY\downarrow Y}(y)$
  where $P^{XY\downarrow X}$ and $P^{XY\downarrow Y}$ are the marginal probabilities of P^{XY} on X and Y, respectively.
- $P^{XY}[y]^{\downarrow X}(x) = P^{XY\downarrow X}(x)$
  where $P^{XY}[y]^{\downarrow X}$ is the conditional probability on X given y.

In fact, thanks to the additivity of probability measures, it is sufficient that the property holds for all $x_i \in X, y_j \in Y$.

Remark. Our notation is more cumbersome than the usual one (i.e. such as in [8], [16]), but it helps when belief functions are involved as it avoids confusion.

The first definition of independence is presented in terms of the *factorization* of the joint probability distribution through its marginal distributions on X and Y, respectively (a mathematical property). It is also called *separability* or *non-interactivity* [28] (see section 5). We will use the last name.

The second definition can be interpreted in terms of *irrelevance* of information, it means that any information about Y is irrelevant to the uncertainty about X (an epistemic property).

In probability context, irrelevance and independence imply each other, hence the distinction is usually left aside.

4 Cognitive and Evidential Independence

To extend these definitions of independence to the case of belief functions, Shafer ([17], page 147 *et seq.*) proposes two definitions of independence. After recalling these definitions, we introduce our definitions of (marginal) non-interactivity, irrelevance and doxastic independence for variables and we study the links between them. Shafer's definitions are based on normalized belief functions.

4.1 Cognitive Independence: Weak Independence

Following Shafer [17], two variables are "*cognitively independent*" with respect to a belief function if new evidence that bears on only one of them does not change the degree of belief for propositions discerned by the other. This notion of "cognitive independence" is also called *weak independence* by Kong [14]. The formal definition of "cognitive independence" is the following:

Definition 2 *([17], page 149): The variables X and Y are "cognitively independent" with respect to M^{XY} if and only if: for all $x \subseteq X$, all $y \subseteq Y$,*

$$Pl^{XY}(x, y) = Pl^{XY\downarrow X}(x)\ Pl^{XY\downarrow Y}(y) \tag{1}$$

4.2 Evidential Independence: Strong Independence

Shafer [17] proposed another notion of independence called "*evidential independence*": two variables are "*evidentially independent*" if their joint belief function is represented by the combination of their marginals using Dempster's rule of combination. This notion of "evidential independence" is also called *strong independence* by Kong [14]. The formal definition of "evidential independence" is as follows:

Definition 3 *([17], page 147): The variables X and Y are "evidentially independent" with respect to M^{XY} if and only if: for all $x \subseteq X$, all $y \subseteq Y$,*

$$Pl^{XY}(x, y) = Pl^{XY\downarrow X}(x)\ Pl^{XY\downarrow Y}(y) \tag{2}$$

$$Bel^{XY}(x, y) = Bel^{XY\downarrow X}(x)\ Bel^{XY\downarrow Y}(y) \tag{3}$$

"Cognitive independence" is a weaker condition than "evidential independence": if two variables are "evidentially independent" with respect to a belief function, then they will be "cognitively independent" with respect to it. Indeed, "evidential independence" requires constraints on Bel and on Pl whereas "cognitive independence" requires only constraints on Pl.

Based on the definition of "evidential independence", let us state the following theorems:

Theorem 1 *The variables X and Y are "evidentially independent" with respect to M^{XY} if and only if:*

$$\begin{aligned} M^{XY}(w) &= M^{XY\downarrow X}(x)\ M^{XY\downarrow Y}(y),\ \textit{if } w=(x,y) \\ &= 0,\ \textit{otherwise.} \end{aligned} \tag{4}$$

where x is the projection of w on X, and y is the projection of w on Y.

Proof. See Shafer [17] page 149 and page 166 . □

This theorem just states that the focal elements of M^{XY}, i.e., those subsets w of XY where $M^{XY}(w) > 0$, belong to $Rect_{XY}$.

Theorem 2 *The variables X and Y are "evidentially independent" with respect to M^{XY} if and only if:*

$$Q^{XY}(w) = Q^{XY\downarrow X}(x)\ Q^{XY\downarrow Y}(y),\ \forall w \subseteq XY \tag{5}$$

where x is the projection of w on X, and y is the projection of w on Y.

Proof. Let $\omega \subseteq XY$ with $x(\omega)$ and $y(\omega)$ its projections on X and Y, respectively. Relation (5) can be rewritten, for all $\omega \subseteq XY$, as:

$$\begin{aligned} Q^{XY}(\omega) &= Q^{XY\downarrow X}(x(\omega))\ Q^{XY\downarrow Y}(y(\omega)) \\ \sum_{\omega':\ \omega\subseteq\omega'\subseteq XY} M^{XY}(\omega') &= \sum_{x\subseteq X: x(\omega)\subseteq x} M^{XY\downarrow X}(x) \sum_{y\subseteq Y: y(\omega)\subseteq y} M^{XY\downarrow Y}(y) \\ &= \sum_{x\subseteq X, y\subseteq Y: (x(\omega),y(\omega))\subseteq(x,y)} M^{XY\downarrow X}(x)\ M^{XY\downarrow Y}(y) \end{aligned}$$

The sum of all masses $M^{XY\downarrow X}(x)\ M^{XY\downarrow Y}(y)$ over all $x \subseteq X$ and $y \subseteq Y$ is 1 as it corresponds to $Q^{XY}(\emptyset)$ which is always 1. As far as the relation between Q and M is one-to-one, the only non zero masses $M^{XY}(\omega)$ are those that admit the representation as $M^{XY\downarrow X}(x)\ M^{XY\downarrow Y}(y)$, i.e., all focal elements of M^{XY} belong to $Rect_{XY}$. So that relation (5) holds is equivalent to the fact that relation (4) holds, and thus it is equivalent to X and Y being evidentially independent. □

Remarks.

- "Cognitive independence" may hold whereas "evidential independence" fails, so "cognitive independence" does not imply theorem 1. In addition, neither (2) nor (3) implies the other. This is shown in the example 7.3 of Shafer ([17], page 151) and in [2].
- All definitions by Shafer and Kong assume normalized belief functions ($m(\emptyset) = 0$).

- Shafer [17] does not explain the signification of the two adjectives: cognitive and evidential.
- Shafer speaks of "independence" but we will prefer the expression 'non-interactivity' as this definition is essentially a mathematical one and we keep the word 'independence' for the common sense property (see section 5).
- One could wonder why a definition based on only

  $$Bel^{XY}(x,y) = Bel^{XY\downarrow X}(x)\ Bel^{XY\downarrow Y}(y)$$

  was not proposed?. It is probably because it is useless.

5 Belief Function Non-Interactivity

In this section, we propose the definition of decompositional independence for belief functions. In possibility theory, there is an analogous definition introduced by Zadeh [28] where the decompositional independence between two variables is represented by the **non-interactivity** relation. We use this last terminology. The non-interactivity is a mathematical property useful for computations considerations when propagating beliefs in evidential networks [27].

Intuitively, the *non-interactivity* of two variables X and Y with respect to m^{XY} means that the joint mass can be reconstructed from its marginals. The purpose is that for any functions $f \in \{m, bel, pl, q\}$, we have that f^{XY} is some function of f^X and f^Y. As far as once it is true for m, it is true for all of them, we propose the following definition of non-interactivity.

Definition 4 ***Non-interactivity.*** *Given two variables X and Y, and $m = m^{XY}$ on XY. X and Y are non-interactive with respect to m, denoted by $X \perp_m Y$, if and only if:*

$$pl^{XY}(X,Y)\ m^{XY} = m^{XY\downarrow X} \circledcirc m^{XY\downarrow Y} \tag{6}$$

The scalar $pl^{XY}(X,Y)$ is introduced because we tolerate unnormalized belief functions. We could (almost) identically propose the definition as:

$$M^{XY} = M^{XY\downarrow X} \oplus M^{XY\downarrow Y}$$

which is equal to:

$$\frac{m^{XY}}{pl^{XY}(X,Y)} = \frac{m^{XY\downarrow X}}{pl^{XY\downarrow X}(X)} \oplus \frac{m^{XY\downarrow Y}}{pl^{XY\downarrow Y}(Y)}$$

These definitions are all equal once $pl^{XY}(X,Y) > 0$. The 'almost' qualification covers the highly degenerated case where $pl^{XY}(X,Y) = 0$, relation (6) being still valid, whereas the others become undefined.

One might also ask why we use the $\circledcirc$ operator in definition 4. Its origin comes from the concept of specialization detailed in Klawonn and Smets [12]. The $\circledcirc$ is the only specialization that is associative and covers conditioning. Details can be found in [12].

Theorem 3 *Let m^{XY} be a bba on XY and $X \perp_{m^{XY}} Y$. Then the focal elements of m^{XY} belong to $Rect_{XY}$.*

Proof. Notice that a non-normalized belief function on Ω can be normalized by just dividing it by $pl(\Omega)$. So if we divide both terms of relation (6) by $(pl^{XY})^2$, we get:

$$M^{XY} = M^{XY\downarrow X} \oplus M^{XY\downarrow Y},$$

where M denotes normalized bba. In that case non-interactivity and Shafer's evidential independence definitions are equivalent, therefore, by theorem 1, the focal elements of M^{XY} belong to $Rect_{XY}$. As far as m^{XY} is proportional to M^{XY}, its focal elements are those of M^{XY} and the empty set which belongs also to $Rect_{XY}$. □

¿From theorem 1, non-interactivity and Shafer's evidential independence definitions are equivalent when we consider normalized belief functions.

Thus relations (2) and (3) can be directly generalized into:

$$pl^{XY}(X,Y)\ pl^{XY}(x,y) = pl^{XY\downarrow X}(x)\ pl^{XY\downarrow Y}(y) \tag{7}$$
$$pl^{XY}(X,Y)\ pl^{XY}(x,y) = pl^{XY}(x,Y)\ pl^{XY}(X,y) \tag{8}$$
$$pl^{XY}(X,Y)\ bel^{XY}(x,y) = bel^{XY\downarrow X}(x)\ bel^{XY\downarrow Y}(y) \tag{9}$$

where relation (8) is obtained from relation (7) and lemma 2.

Non-interactive belief functions can easily be build by using any pair of belief functions, one being defined on X, the other on Y, provided they give the same bba to the empty set.

Theorem 4 *Let m^X and m^Y be two bba defined on X and Y where X and Y are independent coarsenings of Ω with $m^X(\emptyset) = m^Y(\emptyset) = \alpha$ with $\alpha \in [0,1]$. Then the bba m^{XY} defined on XY by:*

$$(1-\alpha)\ m^{XY} = m^X \circledcirc m^Y$$

with $m^{XY}(\emptyset) = 1$ if $\alpha = 1$, satisfies non-interactivity: $X \perp_{m^{XY}} Y$.

Proof. The proof consists in showing that m^X and m^Y are the marginals of m^{XY} and that $pl^{XY}(X,Y) = 1-\alpha$.

If $\alpha = 1$, we have $m^X(\emptyset) = m^Y(\emptyset) = m^{XY}(\emptyset) = 1$, so both m^X and m^Y are the marginals of m^{XY} on X and Y, respectively.

If $\alpha > 0$, then $m^{XY} = \frac{1}{1-\alpha}\, m^X \circledcirc m^Y$. Marginalization of m^{XY} on X satisfies axiom M1 of Shenoy ([19], page 209), so $m^{XY\downarrow X} = \frac{1}{1-\alpha} m^X \circledcirc m^{Y\downarrow X}$. $m^{Y\downarrow X}$ is a bba with $m^{Y\downarrow X}(\emptyset) = \alpha$ and $m^{Y\downarrow X}(X) = 1-\alpha$. So the $\circledcirc$ results only in a multiplication of every term by $m^{Y\downarrow X}(X)$, i.e., by $1-\alpha$. Therefore $m^{XY\downarrow X} = m^X$. By symmetry the same holds for $m^{XY\downarrow Y} = m^Y$. Finally, using lemma 2, we get $pl^{XY}(X,Y) = pl^{XY\downarrow X}(X) = pl^X(X) = 1-\alpha$, hence the theorem is proved.□

6 Belief Function Irrelevance

In probability theory, the notion of independence can be defined in term of **irrelevance**. This kind of independence is based on *conditioning*. The intuitive meaning of irrelevance is that knowing the value y_j of Y does not affect the beliefs on X. In belief functions theory, we propose the following definition of irrelevance:

Definition 5 ***Irrelevance.*** *Given two variables X and Y, and $m = m^{XY}$ on XY, Y is irrelevant to X with respect to m, denoted by $IR_m(X,Y)$, if and only if:*
$\forall y \subseteq Y$ such that $pl^{XY}(X,y) > 0$

$$m^{XY}[y]^{\downarrow X}(x) \quad \propto \quad m^{XY\downarrow X}(x),\ \forall x \subseteq X, x \neq \emptyset \tag{10}$$

and $\forall y \subseteq Y$ such that $pl^{XY}(X,y) = 0$

$$m^{XY}[y]^{\downarrow X}(x) = 0,\ \forall x \subseteq X, x \neq \emptyset,\ \textit{and } m^{XY}[y]^{\downarrow X}(\emptyset) = 1.$$

In relation (10), we need $\propto$ because in the TBM context we do not normalize when applying Dempster's rule of conditioning. Under normalization, proportionality becomes equality.

Theorem 5 *Given two variables X and Y, and $m = m^{XY}$ on XY, $IR_m(X,Y)$ if and only if $pl^{XY}[y]^{\downarrow X} = \alpha_y\ pl^{XY\downarrow X}$, $\forall y \subseteq Y$, where*

$$\alpha_y = \frac{pl^{XY}(X,y)}{pl^{XY}(X,Y)}.$$

Proof. We prove that the proportionality between the plausibility functions is equivalent to the proportionality between the bba's.

Let $\mathbf{P}$ be the $2^{|\Omega|} \times 2^{|\Omega|}$ matrix operator that transforms a bba defined on Ω into its related plausibility function. The elements $P_{A,B}$ of $\mathbf{P}$ for $A \subseteq \Omega, B \subseteq \Omega$ are given by:

$$\begin{aligned} P_{A,B} &= 1 \text{ if and only if } A \cap B \neq \emptyset \\ &= 0 \text{ otherwise.} \end{aligned}$$

Then for any bba on Ω, $pl^\Omega = \mathbf{P} m^\Omega$ where pl^Ω and m^Ω denote here the vectors corresponding to the plausibility function and the bba, respectively.

Let $\Omega = X$. The multiplication by the matrix $\mathbf{P}$ of the two bba encountered in $IR_m(X,Y)$ definition (relation 10) gives:

$$\mathbf{P}m^{XY}[y]^{\downarrow X} = \alpha_y \mathbf{P}m^{XY\downarrow X}$$

where α_y is independent of $x \subseteq X$. We have $\mathbf{P}m^{XY}[y]^{\downarrow X} = pl^{XY}[y]^{\downarrow X}$ and $\mathbf{P}m^{XY\downarrow X} = pl^{XY\downarrow X}$. The non proportionality of $m^{XY}[y]^{\downarrow X}(\emptyset)$ has no impact on the lemma as $P_{\emptyset,C} = P_{C,\emptyset} = 0$, $\forall C \subseteq X$, what just translates the fact that $pl(\emptyset) = 0$. Thus $pl^{XY}[y]^{\downarrow X} = \alpha_y pl^{XY\downarrow X}$ for any $y \subseteq Y$.

In particular, $pl^{XY}[y]^{\downarrow X}(X) = \alpha_y pl^{XY\downarrow X}(X)$. Applying lemmas 1 and 2, we get

$$\alpha_y = \frac{pl^{XY}(X,y)}{pl^{XY}(X,Y)}$$

□

Based on the definition of irrelevance, we can deduce the following consequences.

Theorem 6 *Given two variables X and Y, and $m = m^{XY}$ on XY, the following assertions are equivalent:*

1. $IR_m(X,Y)$

2. $$pl^{XY}[y']^{\downarrow X} = \beta \; pl^{XY}[y'']^{\downarrow X} \qquad (11)$$

where $\beta = \dfrac{pl^{XY}(X,y')}{pl^{XY}(X,y'')}$, ($\beta$ *independent of* x)

3. $$pl^{XY}(x,y) = \frac{pl^{XY}(x,Y)\; pl^{XY}(X,y)}{pl^{XY}(X,Y)} \qquad (12)$$

Furthermore,

$$IR_m(X,Y) = IR_m(Y,X) \qquad (13)$$

Proof.

1. We prove that **1** and **2** are equivalent. By theorem 5 and considering y' and y'' subsets of Y, we get

$$pl^{XY\downarrow X} = \frac{1}{\alpha_{y'}}\; pl^{XY}[y']^{\downarrow X} = \frac{1}{\alpha_{y''}}\; pl^{XY}[y'']^{\downarrow X}$$

$$pl^{XY}[y']^{\downarrow X} = \frac{pl^{XY}(X,y')}{pl^{XY}(X,y'')}\; pl^{XY}[y'']^{\downarrow X}$$

2. Using lemmas 1, 2, theorem 5 and relation (11), relation (12) can be rewritten, for any $x \subseteq X$, as

$$pl^{XY}(x,y) = pl^{XY}[y]^{\downarrow X}(x) = \frac{pl^{XY}(X,y)}{pl^{XY}(X,Y)}\; pl^{XY\downarrow X}(x) =$$

$$\frac{pl^{XY}(X,y)}{pl^{XY}(X,Y)}\, pl^{XY}(x,Y)$$

Hence, (12) is proved.

3. Relation (13) results directly from the symmetry of relation (12). □

The third item of theorem 6 implies that IR is equivalent to Shafer's cognitive independence when belief functions are normalized.

Consider the particular belief function on XY that allocates a bba 1 to some (x,y) for $x \subseteq X$ and $y \subseteq Y$. We show it satisfies the irrelevance property.

Lemma 3 *Let m be a bba defined on XY so that $m(x,y) = 1$ for some $x \subseteq X, y \subseteq Y$. Then $IR_m(X,Y)$.*

Proof. After conditioning m on $y' \subseteq Y$, we have: $m[y']^{\downarrow X}(x) = 1$ for any y' such that $y' \cap y \neq \emptyset$, and $m[y']^{\downarrow X}(\emptyset) = 1$ for any y' such that $y' \cap y = \emptyset$. Thus we have $IR_m(X,Y)$. □

In the following example, we show that irrelevance does not imply non-interactivity between variables.

Example. Let the sets X and Y be as defined by **N2**. Table 1 presents a very symmetrical bba m^{XY} on XY that satisfies the irrelevance constraints but not the non-interactivity ones. The marginals are: $m^{XY\downarrow X}(x_1) = m^{XY\downarrow X}(x_2) = .3$, $m^{XY\downarrow X}(X) = .4$ and $m^{XY\downarrow Y}(y_1) = m^{XY\downarrow Y}(y_2) = .3$, $m^{XY\downarrow Y}(Y) = .4$. Irrelevance is satisfied as can be controlled by computing the plausibilities that enter in the third item of theorem 6. Nevertheless non-interactivity is not satisfied as $m^{XY}(w_1) = .15$ and $m^{XY\downarrow X}(x_1)\, m^{XY\downarrow Y}(y_1)/pl^{XY}(X,Y) = .3 \times .3/1. = .09$, hence violating relation (6). □

for ω such that:	$m^{XY}(w)$	$pl^{XY}(w)$
$w = \emptyset$	.00	.00
$\lvert w\rvert = 1$	.15	.49
$w \in \{(x_1,Y),(x_2,Y),(X,y_1),(X,y_2)\}$	.00	.70
$w \in \{(w_1,w_4),(w_2,w_3)\}$	.04	.66
$\lvert w\rvert = 3$	.02	.85
$w = XY$	.24	1.00

Table 1: For each subset of XY, listed in column 1, columns 2 and 3 present the value of m^{XY} and of its related pl^{XY}.

We feel the next property should also be satisfied by irrelevance. Let A_1 and A_2 denote two agents whose beliefs are considered. Suppose the first agent A_1 states that Y is irrelevant to X and produces his beliefs on XY and the second agent A_2 states that Y is irrelevant to X and produces his own beliefs on XY. Then the two agent beliefs on XY are combined by Dempster's rule of combination. We want that Y is still irrelevant to X under this combined belief function on XY.

This idea can be formalized by the next property, called *Irrelevance Preservation under Dempster' rule of combination*, and denoted by $IRP\circledcirc$.

Definition 6 ***Irrelevance Preservation under Dempster's rule of combination.*** *If $IR_{m_1}(X,Y)$ and $IR_{m_2}(X,Y)$ then $IR_{m_1 \circledcirc m_2}(X,Y)$.*

We are going to show the major result of this paper, that is that IR & $IRP\circledcirc$ is equivalent to non-interactivity. We first prove it for the case where $|X| = |Y| = 2$ (section 6.1) and then for general X and Y (section 6.2).

Remark. This property is not described in probability theory as the concept of combination and the $\circledcirc$ operation are hardly considered.

6.1 The 2×2 Case

We consider the case where $|X| = |Y| = 2$ and prove that IR & $IRP\circledcirc$ implies non-interactivity. The proof is based on an analysis of the relation between $m(\emptyset)$ and the commonality function. Relation **P2** gives:

$$m(\emptyset) = 1 - \sum_{i=1}^{4} q(w_i) + q(x_1, Y) + q(x_2, Y) + q(X, y_1) + q(X, y_2) \\ + q(\{w_1, w_4\}) + q(\{w_2, w_3\}) - \sum_{|A|=3, A \subseteq XY} q(A) + q(X, Y).$$

We first analyze the term:

$$1 - m(\emptyset) - \sum_{i=1}^{4} q(w_i) + q(x_1, Y) + q(x_2, Y) + q(X, y_1) + q(X, y_2)$$

(lemma 4) then we show that $m(\{\omega_1, \omega_4\}) = m(\{\omega_2, \omega_3\}) = 0$ (lemma 5 and 6), and that $m(w) = 0$ for $|w| = 3$ (lemma 7), all these implying that the focal elements of m belong to the set of rectangles $Rect_{XY}$ defined on XY.

Lemma 4 *Let $\Omega = XY$ be defined by **N2**. Let $m = m^{XY}$ and $IR_m(X,Y)$ then:*

$$1 - m(\emptyset) - \sum_{i=1}^{4} q(w_i) + q(x_1, Y) + q(x_2, Y) + q(X, y_1) + q(X, y_2) \\ = \frac{m^{\downarrow X}(X) m^{\downarrow Y}(Y)}{pl(X, Y)} \quad (14)$$

Proof. All belief functions hereafter are initially defined on XY. For simplicity sake, we do not indicate the XY superscript.

By **P1**, $q(w_i) = pl(w_i)$ for i=1,2,3,4

By **P3**, $q(x_i, Y) = pl(x_i, y_1) + pl(x_i, y_2) - pl(x_i, Y)$ and, $q(X, y_i) = pl(x_1, y_i) + pl(x_2, y_i) - pl(X, y_i)$ for i=1,2.

Relation (14) becomes:

$$1 - m(\emptyset) + \sum_{i=1}^{4} pl(w_i) - pl(x_1, Y) - pl(x_2, Y) - pl(X, y_1) - pl(X, y_2) = \frac{m^{\downarrow X}(X) m^{\downarrow Y}(Y)}{pl(X, Y)} \quad (15)$$

For i=1,2, j=1,2, let

$\theta_i = m^{\downarrow X}(x_i) = m(x_i, y_1) + m(x_i, y_2) + m(x_i, Y)$

$\eta_j = m^{\downarrow Y}(y_j) = m(x_1, y_j) + m(x_2, y_j) + m(X, y_j)$

$\Theta = m^{\downarrow X}(X)$

$H = m^{\downarrow Y}(Y)$, and

$Z = m(\emptyset)$.

By $IR_m(X, Y)$, we have: $pl(x_i, y_j) = \frac{pl(x_i, Y)\ pl(X, y_j)}{pl(X, Y)} = \frac{(\theta_i + \Theta)\ (\eta_j + H)}{(1 - Z)}$, and: $1 - m(\emptyset) = pl(X, Y) = pl^{\downarrow X}(X) = pl^{\downarrow Y}(Y) = 1 - Z$.

Then the relation (15) becomes:

$1 - Z + (\theta_1 + \Theta)(\eta_1 + H)/(1 - Z) - (\theta_1 + \Theta)$

$+ \quad (\theta_1 + \Theta)(\eta_2 + H)/(1 - Z) - (\theta_2 + \Theta)$

$+ \quad (\theta_2 + \Theta)(\eta_1 + H)/(1 - Z) - (\eta_1 + H)$

$+ \quad (\theta_2 + \Theta)(\eta_2 + H)/(1 - Z) - (\eta_2 + H)$

$= (1 - Z)^{-1}[(1 - Z)^2 + (1 - Z + \Theta)(1 - Z + H) - (1 - Z)(2(1 - Z) + \Theta + H)]$

$= (1 - Z)^{-1}[2(1 - Z)^2 + (1 - Z)(\Theta + H) + \Theta H - 2(1 - Z)^2 - (1 - Z)(\Theta + H)]$

$= (1 - Z)^{-1}\Theta H$.

This last term is equal to $\frac{m^{\downarrow X}(X) m^{\downarrow Y}(Y)}{pl(X, Y)}$, hence the lemma is proved. □

Lemma 5 *Let $\Omega = XY$ be defined by **N2**. Let $m = m^{XY}$ and $IR_m(X, Y)$ then*

$$m(\{w_1, w_4\}) \leq m(X, Y) \text{ and } m(\{w_2, w_3\}) \leq m(X, Y).$$

Proof. We use the same notations as in the proof of lemma 4. By **P2**, we have: $0 = 1 - Z - \sum_{i=1}^{4} q(w_i) + q(x_1, Y) + q(x_2, Y) + q(X, y_1) + q(X, y_2) + q(\{w_1, w_4\}) + q(\{w_2, w_3\}) - \sum_{|A|=3, A \subseteq XY} q(A) + q(X, Y)$.

With Lemma 4, this becomes:
$0 = \frac{H\Theta}{1 - Z} + q(\{w_1, w_4\}) + q(\{w_2, w_3\}) - \sum_{|A|=3, A \subseteq XY} q(A) + q(X, Y)$.

By transforming q into m, we get:
$0 = \frac{H\Theta}{1-Z} + m(\{w_1, w_4\}) + m(\{w_2, w_3\}) + \sum_{|A|=3, A\subseteq XY} m(A) + 2m(X,Y) - \sum_{|A|=3, A\subseteq XY} m(A) - 4m(X,Y) + m(X,Y)$

Thus,
$0 = \frac{H\Theta}{1-Z} + m(\{w_1, w_4\}) + m(\{w_2, w_3\}) - m(X,Y)$
or
$m(\{w_1, w_4\}) + m(\{w_2, w_3\}) + \frac{H\Theta}{1-Z} = m(X,Y)$

All terms being non negative, we have:
$m(\{w_1, w_4\}) \leq m(X,Y)$ and $m(\{w_2, w_3\}) \leq m(X,Y)$ □

Lemma 6 *Let* $\Omega = XY$ *be defined by* ***N2***. *Let* $m = m^{XY}$ *and* $IR_m(X,Y)$. *Then* $IRP\circledcirc$ *implies:* $m(\{w_1, w_4\}) = m(\{w_2, w_3\}) = 0$

Proof. By the previous lemmas 4 and 5, we have the next inequality:
$m(\{w_1, w_4\}) + m(\{w_1, w_2, w_4\}) + m(\{w_1, w_3, w_4\}) + m(X,Y)$
$\leq m(X,Y) + m(\{w_1, w_2, w_4\}) + m(\{w_1, w_3, w_4\}) + m(X,Y)$
Thus,
$q(\{w_1, w_4\}) \leq q(\{w_1, w_2, w_4\}) + q(\{w_1, w_3, w_4\})$.

Let $\hat{q} = max(q(\{w_1, w_2, w_4\}), q(\{w_1, w_3, w_4\}))$
Using **P4**, we have $\hat{q} \leq q(\{w_1, w_4\}) \leq 2\hat{q}$.
This inequality results from $IR_m(X,Y)$ and must still be satisfied by $\circledcirc_{i=1}^{n} m$ in order to satisfy the property of irrelevance preservation under Dempster's rule of combination (IRP$\circledcirc$).
So, $\hat{q}^n \leq q^n(\{w_1, w_4\}) \leq 2\hat{q}^n$ for all $n > 0$.
Thus, $\hat{q} \leq q(\{w_1, w_4\}) \leq 2^{\frac{1}{n}}\hat{q}$ for all $n > 0$.
Therefore $q(\{w_1, w_4\}) = \hat{q}$ as $lim_{n\to\infty} 2^{\frac{1}{n}} = 1$.
This equality means that $m(\{w_1, w_4\}) + m(\{w_1, w_2, w_4\}) + m(\{w_1, w_3, w_4\}) + m(X,Y) = max(m(\{w_1, w_2, w_4\}), m(\{w_1, w_3, w_4\})) + m(X,Y)$
Thus, $m(\{w_1, w_4\}) = 0$
The proof for $m(\{w_2, w_3\})$ follows the same steps. □

Lemma 7 *Let* $\Omega = XY$ *be defined by* ***N2***. *Let* $m = m^{XY}$ *and* $IR_m(X,Y)$. *Then* $IRP\circledcirc$ *implies:* $m(A) = 0$, $\forall A$ *with* $|A| = 3$.

Proof. Suppose $m(\{\omega_1, \omega_2, \omega_3\}) = x > 0$. By symmetry, we build m' on XY by permuting w_1 with w_4. In that case, $m(\{\omega_2, \omega_3, \omega_4\}) = x$, and we have $IR_{m'}(X,Y)$.

Consider now $m\circledcirc m'$. The product $m(\{\omega_1, \omega_2, \omega_3\})m'(\{\omega_2, \omega_3, \omega_4\}) = x^2 > 0$ is allocated to $m\circledcirc m'(\{\omega_2, \omega_3\})$. By lemma 5, this implies that we do not have $IR_{m\circledcirc m'}(X,Y)$ what in contrary to what $IRP\circledcirc$ requires. Thus $IRP\circledcirc$ implies $m(\{\omega_1, \omega_2, \omega_3\}) = 0$.

By symmetry the same holds for any $A \subseteq XY$ such that $| A |= 3$. □

6.2 The General Case

We have thus proved that IR & $IRP\circledcirc$ implies that all focal elements of XY belong to $Rect_{XY}$ for the case defined by **N2**. We now prove that the same holds in general.

Theorem 7 *Let $\Omega = XY$. Let $m = m^{XY}$ and $IR_m(X, Y)$. Then IRP$\circledcirc$ implies that $m(A) = 0$, whenever A does not belong to $Rect_{XY}$.*

Proof. Suppose $A \subseteq XY$ does not belong to $Rect_{XY}$, thus does not admit a representation as (x, y), for some $x \subseteq X, y \subseteq Y$ and $m(A) > 0$. Then there exist a subset B of A such that $B = \{(x_1, y_1), (x_2, y_2)\}$ or $B = \{(x_1, y_1), (x_2, y_2), (x_1, y_2)\}$ for some $x_1, x_2 \in X, y_1, y_2 \in Y, x_1 \neq x_2, y_1 \neq y_2$.

Let $C = \{(x_1, y_1), (x_2, y_2), (x_1, y_2), (x_2, y_1)\}$ and let m_C be the bba that induces the conditioning of m on C: so $m_C(C) = 1$.

By lemma 3, we have $IR_{m_C}(X, Y)$. Combining m with m_C leads to a bba on C, with all bba given to subsets of C. We are thus back to the conditions used in Lemmas 6 and 7. The bba $m(A)$ is transferred by conditioning on C to B, and thus we have a mass that should be null in order to satisfy $IR_{m\circledcirc m_C}(X, Y)$ & $IRP\circledcirc$. Hence $m(A)$ must be null.

Therefore the bba $m(D)$ can be positive only if $D \in Rect_{XY}$. □

We can now prove that $IR_m(X, Y)$ & $IRP\circledcirc$ implies $X \perp_m Y$.

Theorem 8 *Let $\Omega = XY$ and $m = m^{XY}$. If $IR_m(X, Y)$ and if for all m' defined on XY such that $IR_{m'}(X, Y)$, we have $IR_{m\circledcirc m'}(X, Y)$, then $X \perp_m Y$.*

Proof. Suppose that both m^{XY} and m' are normalized. It is sufficient then to prove that Shafer's evidential independence holds as this property is equivalent to non-interactivity (see section 5). We have already obtained that irrelevance implies Shafer's cognitive independence (12), so we must only prove that the product relation (3) for *bel* holds as well.

Let $A^{\downarrow X}$ denote the projection of $A \subseteq XY$ on X. We have for $x \neq \emptyset$:

$$\begin{aligned} m^{XY}[y]^{\downarrow X}(x) &= \sum_{A^{\downarrow X}=x} m^{XY}[y](A) \\ &= \sum_{A^{\downarrow X}=x, B\subseteq(X,\overline{y})} m^{XY}(A \cup B) \end{aligned}$$

In theorem 7, we have proved that $m(D) > 0$ only if D admits a representation as $D = (x, y)$, for some $x \subseteq X, y \subseteq Y$. Let R* denote that property.

By theorem 5 and with $pl^{XY}(X, Y) = 1$, $IR_m(X, Y)$ implies $\forall x \subseteq X, x \neq \emptyset$:

$$pl^{XY}[y]^{\downarrow X}(x) = pl^{XY}(X, y)\ pl^{XY\downarrow X}(x).$$

Applying on both side the transformation between plausibility functions and bba, one gets:

$$m^{XY}[y]^{\downarrow X}(x) = pl^{XY}(X, y)\ m^{XY\downarrow X}(x).$$

By the R* property, it becomes:

$$m^{XY}[y]^{\downarrow X}(x) = pl^{XY}(X,y) \sum_{y' \subseteq Y} m^{XY}(x,y').$$

Computing then the *bel* functions by summing over all $x' \subseteq x, x' \neq \emptyset$, we get:

$$bel^{XY}[y]^{\downarrow X}(x) = pl^{XY}(X,y)\ bel^{XY}(x,Y).$$

By the marginalization and conditioning definitions and R*, we get:

$$\begin{aligned}
bel^{XY}[y]^{\downarrow X}(x) &= bel^{XY}[y](x,y) \\
&= \sum_{\emptyset \neq x' \subseteq x, \emptyset \neq y' \subseteq y} m^{XY}[y](x',y') && \text{using R*} \\
&= \sum_{A \subseteq (X,\overline{y})} \sum_{\emptyset \neq x' \subseteq x, \emptyset \neq y' \subseteq y} m^{XY}((x',y') \cup A) \\
&= \sum_{y'' \subseteq \overline{y}} \sum_{\emptyset \neq x' \subseteq x, \emptyset \neq y' \subseteq y} m^{XY}((x',(y' \cup y'') && \text{using R*} \\
&= \sum_{\emptyset \neq x' \subseteq x, y' \cap y \neq \emptyset} m^{XY}(x',y') \\
&= \sum_{\emptyset \neq x' \subseteq x, \emptyset \neq y' \subseteq Y} m^{XY}(x',y') - \sum_{\emptyset \neq x' \subseteq x, \emptyset \neq y' \subseteq \overline{y}} m^{XY}(x',y') \\
&= bel^{XY}(x,Y) - bel^{XY}(x,\overline{y}) && \text{using R*}
\end{aligned}$$

Therefore we have:

$$\begin{aligned}
pl^{XY}(X,y)\ bel^{XY}(x,Y) &= bel^{XY}(x,Y) - bel^{XY}(x,\overline{y}) \\
(1 - pl^{XY}(X,y))\ bel^{XY}(x,Y) &= bel^{XY}(x,\overline{y}) \\
bel^{XY}(X,\overline{y})\ bel^{XY}(x,Y) &= bel^{XY}(x,\overline{y})
\end{aligned}$$

This last relation being true for any x and y, we thus have proved that the product rule characterizing *bel* in Shafer's definition of evidential independence holds.

If the bba are unnormalized, the proof proceeds similarly. It consists in building the normalized bba m_N^{XY} and m'_N, proving as just done, that Shafer's definition of evidential independence holds, and then multiplying all terms by $(pl^{XY}(X,Y))^2$. □

Theorem 8 means that when Y is irrelevant to X with respect to m and this irrelevance is preserved under Dempster' rule of combination with any other m^* such that Y is also irrelevant to X with respect to m^*, then X and Y are non-interactive with respect to m. We also show the reverse.

Theorem 9 $X \perp_m Y \Rightarrow IR_m(X, Y)$.

Proof. Immediate as $X \perp_m Y$ is equivalent to Shafer's evidential independence (5), up to the normalization factor. So the product rule for the *pl* function (2) is satisfied, in this case irrelevance is also satisfied (12). The normalization factor is handled by the proportionality factor introduced in relations (6) and (10). □

Theorem 10 $X \perp_m Y \Rightarrow IRP\circledcirc$.

Proof. Suppose two bba m_1^{XY} and m_2^{XY} on XY such that $X \perp_{m_1^{XY}} Y$ and $X \perp_{m_2^{XY}} Y$. By theorem 2, the non-interactivity implies that $\forall w \subseteq XY$:

$$\begin{aligned} q_1^{XY}(w) &= q_1^{XY\downarrow X}(x)\; q_1^{XY\downarrow Y}(y), \\ q_2^{XY}(w) &= q_2^{XY\downarrow X}(x)\; q_2^{XY\downarrow Y}(y), \end{aligned}$$

where $x = \omega^{\downarrow X}$ and $y = \omega^{\downarrow Y}$, the projections of ω on X and Y, respectively.

Applying conhunctive combination, we get:

$$\begin{aligned} (q_1^{XY} \circledcirc q_2^{XY})(w) &= q_1^{XY}(\omega)\; q_2^{XY}(\omega), \\ &= q_1^{XY\downarrow X}(x)\; q_1^{XY\downarrow Y}(y)\; q_2^{XY\downarrow X}(x)\; q_2^{XY\downarrow Y}(y) \\ &= (q_1^{XY\downarrow X} \circledcirc q_2^{XY\downarrow X})(x)\; (q_1^{XY\downarrow Y} \circledcirc q_2^{XY\downarrow Y})(y). \end{aligned}$$

Therefore, X and Y are also non-interactive with respect to $m_1^{XY} \circledcirc m_2^{XY}$, in which case we have $IR_{m_1^{XY} \circledcirc m_2^{XY}}(X, Y)$ by theorem 9. So irrelevance is preserved under Dempster's rule of combination. □

7 Doxastic Independence

The most obvious difference between probabilistic independence and belief function independence is that irrelevance and independence have not identical meaning in the belief function framework. This distinction is not commonly considered in probabilistic framework where authors like Pearl [16] and Dawid [8] use the words irrelevance and independence interchangeably.

In order to enhance this distinction, we use the expression **doxastic independence** for belief function independence. In Greek, 'doxein' means 'to believe'. The formal definition of doxastic independence is as follows:

Definition 7 ***Doxastic Independence.*** *Given two variables X and Y, and a bba m on XY. Variables X and Y are doxastically independent with respect to m, denoted by* $X \perp\!\!\!\perp_m Y$*, if and only if m satisfies:*

- $IR_m(X, Y)$
- $\forall m_0$ *on* $XY : IR_{m_0}(X, Y) \Rightarrow IR_{m \circledcirc m_0}(X, Y)$

The intuitive meaning of this definition is that two variables are considered as doxastically independent only when they are irrelevant and this irrelevance is preserved under Dempster's rule of combination.

Theorem 11 *Doxastic independence preservation under* $\circledcirc$.
If $X \perp\!\!\!\perp_{m_1} Y$ *and* $X \perp\!\!\!\perp_{m_2} Y$ *then* $X \perp\!\!\!\perp_{m_1 \circledcirc m_2} Y$

Proof. Let BIR be the set of bba m such that $IR_m(X,Y)$. Both m_1 and m_2 belong to BIR.

According to the definition of doxastic independence of X and Y, with respect to m_1 and m_2, respectively, we have:

$$IR_{m_1}(X,Y) \quad \text{and} \quad \forall m \in BIR,\ IR_{m_1 \circledcirc m}(X,Y), \tag{16}$$
$$IR_{m_2}(X,Y) \quad \text{and} \quad \forall m \in BIR,\ IR_{m_2 \circledcirc m}(X,Y). \tag{17}$$

As $m_2 \in BIR$, we replace m by m_2 in (16), we obtain thus $IR_{m_1 \circledcirc m_2}(X,Y)$.

For what concerns the IRP$\circledcirc$, we must prove that for all $m \in BIR$,

$$IR_{m_1 \circledcirc m_2 \circledcirc m}(X,Y).$$

This is true if $m_2 \circledcirc m \in BIR$ for all $m \in BIR$. By (17) we know that for any $m \in BIR$, $IR_{m_2 \circledcirc m}(X,Y)$, therefore $m_2 \circledcirc m \in BIR$. □

The link between doxastic independence and non-interactivity is given by the next theorem:

Theorem 12 *Given two variables X and Y, and a bba m on XY. The variables X and Y are doxastically independent with respect to m if and only if X and Y are non-interactive with respect to m:*

$$X \perp\!\!\!\perp_m Y \Leftrightarrow X \perp_m Y$$

Proof. Proof immediate from theorems 8, 9 and 10. □

8 Conclusion

In this paper, we have studied different concepts of marginal independence for belief functions. Of special interest for us is to clarify the relationships between the concepts of non-interactivity, irrelevance and doxastic independence when uncertainty is expressed under the form of belief functions as in TBM.

Non-interactivity is defined by the 'mathematical' property that the joint belief function can be described by its marginals. Irrelevance is defined by the 'common sense' property that the result of conditioning the joint belief function on one variable and marginalizing it on the other variable produce a belief function that is the same whatever the conditioning event. Finally doxastic independence is defined by a particular form of irrelevance, the one preserved under Dempster's rule of combination. Irrelevance alone does not imply non-interactivity. The major theorem proves that doxastic independence is equivalent to non-interactivity, thus equating the 'common sense' definition with the 'mathematical' definition.

These concepts of marginal independence for belief functions can be extended to conditional case. The study of *conditional independence* in the framework of belief functions theory is similar to the marginal case and will be presented in a forthcoming paper.

Acknowledgments

The authors are grateful to Foued Ben Abdelaziz, Salem Benferhat, Fabio Cozman, Phil Dawid, Michel Grabish, Romano Scozzafava, Prakash Shenoy, Gleen Shafer and Milan Studeny for their interest on this work. Many thanks to the referees for their helpful comments and suggestions.

References

[1] S. Benferhat, D. Dubois and H. Prade, *Expressing Independence in Possibilistic Framework and its application to default reasoning*, European Conference on Artificial Intelligence (ECAI'94), 150-153, 1994.

[2] B. Ben Yaghlane, Ph. Smets and K. Mellouli, *Independence Concepts for Belief Functions*, Proceedings of 8^{th} International Conference on Information Processing and Management of Uncertainty in Knowledge Based Systems IPMU'2000, Madrid, vol.1, 357-364, 2000.

[3] L.M. Campos and J.F. Huete, *Independence Concepts in Upper and Lower Probabilities*, Uncertainty in Intelligent Systems, Bouchon-Meunier et al. (Eds), 85-96, 1993.

[4] L.M. Campos, Huete J.F. and S. Moral, *Possibilistic Independence*, Third European Congress on Intelligent Techniques and Soft Computing EUFIT'95, Germany, vol. 1, 69-73, 1995.

[5] I. Couso, S. Moral and P. Walley, *Examples of Independence for Imprecise Probabilities*, 1st International Symposium on Imprecise Probabilities and Their Applications ISIPTA'99, Ghent, 121-130, 1999.

[6] F.G. Cozman, *Irrelevance and Independence Axioms in Quasi-Bayesian Theory*, European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty ECSQARU'99, London, Lecture Notes in AI 1638, A.Hunter and S. Parsons (Eds), Springer-Verlag, 128-136, 1999.

[7] A.P. Dawid, *Conditional Independence in Statistical Theory*, Journal of the Royal Statistical Society, Series B, Vol. 41, 1-31, 1979.

[8] A.P. Dawid, *Conditional Independence*, In Encyclopedia of Statistical Science (Update) Volume 3, Wiley, New York, 1999.

[9] P. Fonck, *Conditional Independence in Possibility Theory*, Uncertainty in Artificial Intelligence, 221-226, 1994.

[10] D. Geiger, T. Verma and J. Pearl, *Identifying Independence in Bayesian Networks*, Networks 20, 507-533, 1990.

[11] E. Hisdal, *Conditional Possibilities Independence and Non interaction*, Fuzzy sets end Systems, Vol. 1, 1978.

[12] F. Klawonn and P. Smets, *The dynammic of belief in the transferable belief model and specialization- generalization matrices*, in Dubois, D, Wellman, M. P., D'Ambrosio, B. and Smets, P. (Eds.), Uncertainty in Artificial Intelligence 92, Morgan Kaufman, San Mateo, Ca, 130-137, 1992.

[13] J. Kohlas, J. and P. A. Monney, *A Mathematical Theory of Hints: An Approach to Dempster-Shafer Theory of Evidence*, Lecture Notes in Economics and Mathematical Systems No. 425, Springer-Verlag, 1995.

[14] C.T.A. Kong, *A Belief Function Generalization of Gibbs Ensemble*, Joint Technical Report, S-122 Harvard University and N239 University of Chicago, Departments of Statistics, 1988.

[15] S. L. Lauritzen, A.P. Dawid, B.N. Larsen and H.G. Leimer, *Independence Properties of Directed Markov Fields*, Networks, 20(5), 491-505, 1990.

[16] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmman, Los Altos, CA, 1988.

[17] G. Shafer, *Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.

[18] G. Shafer, P.P. Shenoy, and K. Mellouli, *Propagating Belief Functions in Qualitative Markov Trees*, International Journal of Approximate Reasoning, 1(4), 349- 400, 1987.

[19] P.P. Shenoy, *Conditional Independence in Valuation-Based Systems*, International Journal of Approximate Reasoning, 10, 203- 234, 1994.

[20] Smets P. *The transferable belief model for quantified belief representation.*, Handbooks for Defeasible Reasoning and Uncertainty Management Systems, Volume 1: The Representation of Uncertainty. Gabbay D. and Smets P. (Eds.), Kluwer, Boston. 267-301, 1998.

[21] P. Smets and R. Kennes, *The Transferable Belief Model*, Artificial Intelligence, 66, 191-234, 1994.

[22] W. Spohn, *Ordinal Conditional Functions: A Dynamic Theory of Epistemic States*, Causation in decision, belief changes and statistics, W. Harper and B. Skyrms (Eds), 105-134, 1988.

[23] M. Studeny, *Formal Properties of Conditional Independence in Different Calculi of Artificial Intelligence*, ECSQARU'93, K. Clarke K., R. Kruse and S. Moral (Eds), Springer-Verlag, 341-348, 1993.

[24] J. Vejnarova, *Conditional Independence Relations in Possibility Theory*, 1st International Symposium on Imprecise Probabilities and Their Applications ISIPTA'99, Ghent, 343-351, 1999.

[25] P. Vicig, *Epistemic Independence for Imprecise Probabilities*, International Journal of Approximate Reasoning, 24, 235-250, 2000.

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

[27] H. Xu, *An efficient implementation of the belief function propagation*, in Bonissone, P. P. and Henrion, M. and Kanal, L. N. and Lemmer, J. F. (Eds.) Uncertainty in Artificial Intelligence 6, North Holland, Amsterdam, The Netherlands, 425-432, 1991.

[28] L.A. Zadeh, *Fuzzy Sets as a Basis for Theory of Possibility*, Fuzzy Sets and Systems, 1:3-28, 1978.