Total Belief Theorem and Generalized Bayes' Theorem

Jean Dezert
The French Aerospace Lab
Palaiseau, France.
jean.dezert@onera.fr

Albena Tchamova
Inst. of I&C Tech., BAS
Sofia, Bulgaria.
tchamova@bas.bg

Deqiang Han
Inst. of Integrated Automation
Xi'an Jiaotong University, China.
deqhan@gmail.com

***Abstract*—This paper presents two new theoretical contributions for reasoning under uncertainty: 1) the Total Belief Theorem (TBT) which is a direct generalization of the Total Probability Theorem, and 2) the Generalized Bayes' Theorem drawn from TBT. A constructive justification of Fagin-Halpern belief conditioning formulas proposed in the nineties is also given. We also show how our new approach and formulas work through simple illustrative examples.**

Keywords: Total Belief Theorem (TBT), Generalized Bayes' Theorem (GBT), belief functions.

I. Introduction

This paper presents new theoretical results for reasoning under uncertainty with belief functions (BF) introduced by Shafer in [1] in Dempster-Shafer Theory (DST). The first important result is the Total Belief Theorem (TBT) which is a generalization of the Total Probability Theorem (TPT) for the belief functions framework. From TBT, one can provide a solid justification of Fagin-Halpern (FH) belief conditioning formulas [3]–[5] which are generalizations of the classical conditional probability formulas. These theoretical results allow us to establish rigorously the Generalized Bayes' Theorem (GBT). The belief conditioning problem is challenging, not new, and one of the two main methods usually adopted by users working with BF is : 1) Shafer's belief conditioning method based on Dempster's rule of combination [1], or 2) the belief conditioning method consistent with imprecise probability calculus bounds [2], [6], [7] based on the lower and upper probability interpretation of belief functions popularized by Fagin and Halpern [3]. In this paper we focus on the second approach of belief conditioning because Dempster's rule of combination presents serious problems as reported in [8]–[16]. Smets did also attempt to generalize Bayes' Theorem (BT) and did propose his own GBT [17] on the basis of conditional embedding, conjunctive merging and Shafer's conditioning. Unfortunately, Smets' approach remains doubtful as reported in [18]. Our new GBT establishment is obtained by a direct constructive manner from TBT. It does not need extra assumptions nor some underlying ad-hoc construction principles. Also, we prove that our TBT and GBT presented in this work are fully consistent with classical TPT and BT as soon as the belief functions are Bayesian.

This paper starts with a brief review of very basics of Probability Theory, including the Total Probability Theorem (TPT) and Bayes' Theorem (BT) in Section II because this helps to have a better understanding of the generalizations we propose. A brief review of belief functions is given in Section III, followed by classical Shafer's and Fagin-Halpern's belief conditioning methods respectively in Sections IV and V. In Section VI, we present the decomposition of the set of focal elements of any basic belief assignment (BBA) that allows us to establish formally the TBT and its generalization on Cartesian product space. The Section VII presents and justifies the new belief conditioning formulas drawn from TBT which are fully consistent with Fagin-Halpern conditioning formulas. This section also presents the generalization of Bayes' theorem in the framework of belief functions. We illustrate our new theoretical results with a quite simple GBT example in Section VIII to show how to make derivations of GBT and to prove that Shafer's conditioning results are inconsistent with GBT. Section IX concludes this paper.

II. Total Probability Theorem & Bayes' formula

A. *Total Probability Theorem*

In probability theory, the elements θ_i of the space Θ are experimental outcomes. The subsets of Θ are called events and the event $\{\theta_i\}$ consisting of the single element θ_i is an elementary event. The space Θ is called the *sure event* and the empty set $\emptyset$ is the *impossible event*. We assign to each event A a number $P(A)$ in $[0,1]$, called the probability of A, which satisfies the three Kolmogorov's conditions: 1) $P(\emptyset)=0$; 2) $P(\Theta)=1$; and 3) if $A\cap B=\{\emptyset\}$, then $P(A\cup B)=P(A)+P(B)$. These conditions are the axioms of the theory of probability [20]. The fundamental Theorem of the probability theory is the Total Probability Theorem (TPT), also called a the law of total probability, see [20] which can be stated as follows.

Total Probability Theorem (TPT): Consider an event B and any partition[1] $\{A_1, A_2, \ldots, A_k\}$ of the space Θ. Then

$$P(B)=P(B\cap A_1)+P(B\cap A_2)+\ldots+P(B\cap A_k) \quad (1)$$

B. *Conditional probability and Bayes' formula*

Starting from TPT formula (1) and assuming $P(B)>0$, we get for any $i\in\{1,\ldots,k\}$ after dividing each side of (1) by $P(B)$ and

[1]A partition of Θ is a collection of exclusive subsets of Θ whose union equals Θ.

rearranging terms the equality

$$\frac{P(A_i \cap B)}{P(B)} = 1 - \sum_{\substack{j=1,\dots,k \\ j\neq i}} \frac{P(A_j \cap B)}{P(B)} = 1 - \frac{P(\bar{A}_i \cap B)}{P(B)} \quad (2)$$

which allows us to define the conditional probability $P(A_i|B)$ by[2]

$$P(A_i|B) \triangleq P(A_i \cap B)/P(B) \quad (3)$$

Similarly, by considering an event A_i of Θ and the partition $\{B, \bar{B}\}$ of Θ, the TPT formula $P(A_i) = P(A_i \cap B) + P(A_i \cap \bar{B})$ applies, and by dividing it by $P(A_i)$ (assuming $P(A_i) > 0$), one gets

$$\frac{P(A_i \cap B)}{P(A_i)} = 1 - \frac{P(A_i \cap \bar{B})}{P(A_i)} \quad (4)$$

which allows to define the conditional probability $P(B|A_i)$ by

$$P(B|A_i) \triangleq P(A_i \cap B)/P(A_i) \quad (5)$$

From (3) and (5), one deduces the equality

$$P(A_i \cap B) = P(A_i|B)P(B) = P(B|A_i)P(A_i) \quad (6)$$

From equality (6) and assuming $P(B) > 0$ and $P(A_i) > 0$, we get

$$P(A_i|B) = P(B|A_i)P(A_i)/P(B) \quad (7)$$

$$P(B|A_i) = P(A_i|B)P(B)/P(A_i) \quad (8)$$

Using (1) and noting that $P(A_i \cap B) = P(B|A_i)P(A_i)$, we get

$$P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i) \quad (9)$$

Substituting (9) in (7), we obtain Bayes' Theorem (BT) formula stated mathematically as the following equation

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)} \quad (10)$$

One can verify that the conditional probability defined by (3) satisfies the three axioms of the Theory of Probability [20].

Previously, A_i and B were events (subsets) of the same space Θ. If $A_i \subseteq \Theta_1$ and $B \subseteq \Theta_2$ with $\Theta_1 \neq \Theta_2$, which corresponds to a so-called combined experiment [20], similar conditioning formulas can also be established by working in the Cartesian product space $\Theta \triangleq \Theta_1 \times \Theta_2$ whose elementary elements are all the ordered pairs (x_p, y_q) with $x_p \in \Theta_1$ and $y_q \in \Theta_2$. The two experiments are viewed as a single combined one whose outcomes are pairs (x_p, y_q). In this space $\Theta = \Theta_1 \times \Theta_2$, x_p is not an elementary element but a subset of n elements of Θ, i.e. $\{x_p\} = \{(x_p, y_1), \dots, (x_p, y_n)\}$. Similarly, y_q is not an elementary element but a subset of m elements of Θ, i.e. $\{y_q\} = \{(x_1, y_q), \dots, (x_m, y_q)\}$. If $A_i \subseteq \Theta_1$ and $B \subseteq \Theta_2$, then $A_i \times B = \{(x_p, y_q)|x_p \in A; y_q \in B\} \subseteq \Theta$. If one forms $A_i \times \Theta_2$ and $\Theta_1 \times B$ one sees that $A_i \times B = (A_i \times \Theta_2) \cap (\Theta_1 \times B) = (\Theta_1 \times B) \cap (A_i \times \Theta_2)$. Because the event $A_i \times \Theta_2$ occurs in the combined experiment if the event A_i of the experiment 1 occurs no matter what the outcome of experiment 2 is, one has $P(A_i \times \Theta_2) = P_1(A_i)$ where $P_1(A_i)$ is the probability of event A_i in the experiment 1. Similarly, the event $\Theta_1 \times B$ occurs if B occurs in experiment 2 no matter what the outcome of experiment 1 is, so that $P(\Theta_1 \times B) = P_2(B)$ where $P_2(B)$ is the probability of event B in the experiment 2. Considering a partition $\{A_1, A_2, \dots, A_k\}$ of Θ_1 and a subset (event) $B \subseteq \Theta_2$, and based on set theory and property of Cartesian product, one can establish also TPT formula

$$P(\Theta_1 \times B) = \sum_{i=1,\dots,k} P((\Theta_1 \times B) \cap (A_i \times \Theta_2))$$

[2]the notation $\triangleq$ means *equal by definition*

and Bayes' formula

$$P(A_i \times \Theta_2|\Theta_1 \times B) = \frac{P(\Theta_1 \times B|A_i \times \Theta_2)P(A_i \times \Theta_2)}{\sum_{i=1}^{k} P(\Theta_1 \times B|A_i \times \Theta_2)P(A_i \times \Theta_2)}$$

That is why, for notation convenience (and notation abuse), we can just use classical formulas even when working with different sets of experimental outcomes Θ_1 and Θ_2. One just has to keep in mind that in this case A_i must be understood as $A_i \times \Theta_2$ and B as $\Theta_1 \times B$.

III. Basics of belief functions

Based on Dempster's works [2], [19], Shafer did introduce Belief Functions (BF) to model the epistemic uncertainty[3] and to reason under uncertainty [1]. We consider a finite discrete frame of discernement (FoD) $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$, with $n > 1$, and where all exhaustive and exclusive elements of Θ represent the set of the potential solutions of the problem under concern. The set of all subsets of Θ is the power-set of Θ denoted by 2^Θ. The number of elements (i.e. the cardinality) of 2^Θ is $2^{|\Theta|}$. A basic belief assignment (BBA) associated with a given source of evidence is defined as the mapping $m(\cdot) : 2^\Theta \to [0, 1]$ satisfying the conditions $m(\emptyset) = 0$ and $\sum_{A\in 2^\Theta} m(A) = 1$. The quantity $m(A)$ is the mass of belief of subset A committed by the source of evidence (SoE). A focal element X of a BBA $m(\cdot)$ is an element of 2^Θ such that $m(X) > 0$. Note that the empty set $\emptyset$ is not a focal element of a BBA because $m(\emptyset) = 0$ (closed-world assumption of Shafer's model for the FoD). The set of all focal elements of $m(\cdot)$ is denoted

$$\mathcal{F}_\Theta(m) \triangleq \{X \subseteq \Theta|m(X) > 0\} = \{X \in 2^\Theta|m(X) > 0\} \quad (11)$$

The set of focal elements of $m(\cdot)$ included in $A \subseteq \Theta$ is denoted

$$\mathcal{F}_A(m) \triangleq \{X \in \mathcal{F}_\Theta(m)|X \cap A = X\} \quad (12)$$

Note that if $A \subseteq B \subseteq \Theta$, then $\mathcal{F}_A(m) \subseteq \mathcal{F}_B(m)$. Also, $\forall A, B \subseteq \Theta$ one has $\mathcal{F}_{A\cap B}(m) = \mathcal{F}_A(m) \cap \mathcal{F}_B(m)$, but $\mathcal{F}_{A\cup B}(m) \neq \mathcal{F}_A(m) \cup \mathcal{F}_B(m)$ in general. The set $\mathcal{F}_\Theta(m)$ can always be partitioned as $\{\mathcal{F}_A(m), \mathcal{F}_{\bar{A}}(m), \mathcal{F}_{A^*}(m)\}$ where[4]

$$\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_A(m) - \mathcal{F}_{\bar{A}}(m) \quad (13)$$

$$= \{X \in \mathcal{F}_\Theta(m)|X \cap A \neq \emptyset \text{ and } X \cap \bar{A} \neq \emptyset\} \quad (14)$$

represents the set of focal elements of $m(\cdot)$ which are not subsets of A and not subsets of $\bar{A} \triangleq \Theta - \{A\} = \{X|X \in \Theta \text{ and } X \notin A\}$, where $\bar{A}$ is the complement of A in Θ and the minus symbol denotes the set difference operator.

Belief and plausibility functions are defined by[5]

$$Bel(A) = \sum_{\substack{X\in 2^\Theta \\ X\subseteq A}} m(X) = \sum_{\substack{X\in \mathcal{F}_\Theta(m) \\ X\subseteq A}} m(X) = \sum_{X\in \mathcal{F}_A(m)} m(X) \quad (15)$$

$$Pl(A) = \sum_{\substack{X\in 2^\Theta \\ X\cap A\neq\emptyset}} m(X) = \sum_{\substack{X\in \mathcal{F}_\Theta(m) \\ X\cap A\neq\emptyset}} m(X) = 1 - \text{Bel}(\bar{A}). \quad (16)$$

The width $U(A^*) = Pl(A) - Bel(A)$ of the belief interval $[Bel(A), Pl(A)]$ is called the *uncertainty on* A committed by the SoE. It represents the imprecision on the (subjective) probability of A granted by the SoE which provides the BBA $m(\cdot)$. The uncertainty $U(A^*)$ can also be expressed directly as

$$U(A^*) = \sum_{X\in \mathcal{F}_{A^*}(m)} m(X) \quad (17)$$

[3]Also called sometimes the cognitive uncertainty by some authors.

[4]For notation convenience, we use A^* to denote focal elements of $m(\cdot)$ which are not in A, nor in $\bar{A}$.

[5]By convention, a *sum of non existing terms* (if it occurs in formulas depending on the given BBA) is always set to zero.

It is worth noting that $U(\bar{A}^*) = Pl(\bar{A}) - Bel(\bar{A}) = (1 - Bel(A)) - (1 - Pl(A)) = Pl(A) - Bel(A) = U(A^*)$, or equivalently

$$U(\bar{A}^*) = \sum_{X \in \mathcal{F}_{\bar{A}^*}(m)} m(X) \quad (18)$$

where $\mathcal{F}_{\bar{A}^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{\bar{A}}(m) - \mathcal{F}_A(m) = \mathcal{F}_{A^*}(m)$.

When all elements of $\mathcal{F}_\Theta(m)$ are only singletons, $m(\cdot)$ is called a *Bayesian BBA* [1] and its corresponding $Bel(\cdot)$ and $Pl(\cdot)$ functions are homogeneous to a same (subjective) probability measure $P(\cdot)$. In this case $\mathcal{F}_{A^*}(m) = \mathcal{F}_{\bar{A}^*}(m) = \emptyset$. According to Shafer's Theorem 1 below, see [1] page 39 with its proof on page 51, the belief functions can be characterized without referencing to a BBA.

Theorem 1: If Θ is a FoD, then a function $Bel : 2^\Theta \mapsto [0, 1]$ is a belief function if and only if it satisfies the following conditions:

- B1) Belief in impossible event is zero, that is $Bel(\emptyset) = 0$.
- B2) Belief in the certain event is one, that is $Bel(\Theta) = 1$.
- B3) For every positive integer n and every collection $A_1, \ldots, A_n$ of subsets of Θ

$$Bel(A_1 \cup \ldots \cup A_n) \geq \sum_{\substack{I \subset \{1,\ldots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} Bel(\bigcap_{i \in I} A_i) \quad (19)$$

Quantities $m(\cdot)$ and $Bel(\cdot)$ are one-to-one, and for any $A \subseteq \Theta$ the BBA $m(\cdot)$ is obtained from $Bel(\cdot)$ by Möbius inverse formula (see [1], p.39)

$$m(A) = \sum_{B \subseteq A \subseteq \Theta} (-1)^{|A-B|} Bel(B) \quad (20)$$

Shafer [1] did propose to combine $s \geq 2$ distinct sources of evidence represented by BBAs $m_1(.), \ldots, m_s(.)$ over the same FoD with Dempster's rule (i.e. the normalized conjunctive rule). However Dempster's rule has been strongly disputed from both theoretical and practical standpoints as reported in [16], [21], [22]. In particular, the high (or even very low) conflict level between the sources can be totally ignored by Dempster's rule which is a very serious problem [15]. Also, Shafer's conditioning (based on Dempster's rule) is inconsistent with the probabilistic conditioning (see next section).

IV. Shafer's conditioning

A. Shafer's conditioning formulas

Shafer's conditioning formulas are established in Theorem 3.6 p. 66 of [1] from Dempster's rule of combination of the original BBA $m(\cdot)$ with the BBA $m_B(B) = 1$ focused on B. We review them for convenience. For $A, B \subseteq \Theta$ with $Pl(B) > 0$, $Bel(A|B)$ and $Pl(A|B)$ are given by

$$Bel(A|B) = (Bel(A \cup \bar{B}) - Bel(\bar{B}))/(1 - Bel(\bar{B})) \quad (21)$$

$$Pl(A|B) = Pl(A \cap B)/Pl(B) \quad (22)$$

The expression (21) of $Bel(A|B)$ is equivalent to

$$Bel(A|B) = (Pl(B) - Pl(B \cap \bar{A}))/Pl(B) \quad (23)$$

because one has always (from definition of belief functions) $Pl(B) = 1 - Bel(\bar{B})$ and the numerator of (21) can be written as

$$Bel(A \cup \bar{B}) - Bel(\bar{B}) = Pl(B) - Pl(B \cap \bar{A})$$

If $A = \emptyset$, $Bel(\emptyset|B) = Pl(\emptyset|B) = 0$, and if $A = \Theta$, $Bel(\Theta|B) = Pl(\Theta|B) = 1$. Also, if $B = \Theta$, $Bel(A|\Theta) = Bel(A)$ and $Pl(A|\Theta) = Pl(A)$. Note that if $B = A$ in (22)–(23), we get $Bel(A|A) = Pl(A|A) = 1$ which fits with the common sense.

In reversing the roles played by A and B and switching the notations in previous expressions, the following formulas also hold (assuming $Pl(A) > 0$)

$$Bel(B|A) = (Pl(A) - Pl(A \cap \bar{B}))/Pl(A) \quad (24)$$

$$Pl(B|A) = Pl(B \cap A)/Pl(A) \quad (25)$$

From (22) and (25), one deduces that

$$Pl(A \cap B) = Pl(A|B)Pl(B) = Pl(B|A)Pl(A)$$

Hence, the following formula applies for conditional plausibilities when $Pl(B) > 0$

$$Pl(A|B) = Pl(B|A)Pl(A)/Pl(B) \quad (26)$$

Shafer's formula (25) is similar to conditional probabilities (3) when replacing plausibility by probability. So, at first glance it seems appealing. In the sequel we show why this is not the case.

B. Drawback of Shafer's conditioning

The main drawback of Shafer's conditioning is that the bounds of belief interval $[Bel(A|B), Pl(A|B)]$ obtained by (21)-(22) are in general incompatible with lower and upper bounds of the conditional probability $P(A|B)$. This problem makes Shafer's conditioning based on Dempster's rule very disputable and cast doubts on pertinence (validity) of Shafer's conditioning results when used in applications. This serious problem has already been reported and addressed by several authors [3], [6], [7], [11] with some examples. To easily show this incompatibility of Shafer's conditioning with probability calculus we present briefly the famous Ellsberg urn example [23].

Example 1 (Ellsberg urn): We consider an urn with red (R) balls, black (B) and yellow (Y) balls. One knows that 1/3 of balls are red balls and 2/3 or balls are black and yellow balls. So the a priori information about the chance to pick a ball in the urn can be represented by a (parametric) probability mass function $P(\cdot)$

$$P(R) = 1/3 \qquad P(B) = 2/3 - x \qquad P(Y) = x$$

where x is an unknown number/parameter in $[0, 2/3]$. Therefore, $P(B)$ and $P(Y)$ are unknown but their bounds are known. In fact, this problem can be seen as a problem of imprecise probabilities where $P(R) + P(B) + P(Y) = 1$ with

$$P(R) \in [1/3, 1/3] \qquad P(B) \in [0, 2/3] \qquad P(Y) \in [0, 2/3]$$

Now let's suppose that someone picks a ball at random in the urn and tell us that the color of the ball is not black, i.e. the event $\bar{B} = R \cup Y$ has occurred. How do we must revise (update) our prior probabilities with this new information? The correct answer to this question is obtained by computing the conditional probabilities $P(R|\bar{B})$, $P(B|\bar{B})$ and $P(Y|\bar{B})$ and by analyzing their bounds. This is done using the fact that $P(\bar{B}) = P(R \cup Y) = P(R) + P(Y) - P(R \cap Y) = P(R) + P(Y) = (1/3) + x$. Indeed, $P(R \cap Y) = 0$ because the events R and Y are mutually exclusive. So, we get

$$P(R|\bar{B}) = \frac{P(R \cap (R \cup Y))}{P(R \cup Y)} = \frac{P(R)}{(1/3) + x} = \frac{1/3}{(1/3) + x}$$

$$P(B|\bar{B}) = \frac{P(B \cap (R \cup Y))}{P(R \cup Y)} = \frac{P(\emptyset)}{(1/3) + x} = \frac{0}{(1/3) + x}$$

$$P(Y|\bar{B}) = \frac{P(Y \cap (R \cup Y))}{P(R \cup Y)} = \frac{P(Y)}{(1/3) + x} = \frac{x}{(1/3) + x}$$

If $x = 0$, then $P(R|\bar{B}) = 1$ and $P(Y|\bar{B}) = 0$. If $x = 2/3$, then $P(R|\bar{B}) = 1/3$ and $P(Y|\bar{B}) = 2/3$. Therefore after conditioning we get

$$P(R|\bar{B}) \in [1/3, 1] \qquad P(B|\bar{B}) \in [0, 0] \qquad P(Y|\bar{B}) \in [0, 2/3]$$

Let's examine what we get with Shafer's conditioning. The problem is modeled using the *a priori BBA* $m(\cdot)$ defined on the FoD $\Theta = \{R, B, Y\}$ with $m(R) = 1/3$ and $m(B \cup Y) = 2/3$ which gives the belief intervals $[Bel(R), Pl(R)] = [1/3, 1/3]$, $[Bel(B), Pl(B)] = [0, 2/3]$ and $[Bel(Y), Pl(Y)] = [0, 2/3]$. With Shafer's conditioning formulas and noting that $Pl(R) = 1/3$, $Pl(B) = 2/3$, $Pl(Y) = 2/3$, and $Pl(R \cup Y) = 1$, we get

$$Bel(R|\bar{B}) = \frac{Pl(R\cup Y) - Pl((R\cup Y)\cap(B\cup Y))}{Pl(R\cup Y)} = \frac{1-Pl(Y)}{1} = 1/3$$
$$Bel(B|\bar{B}) = \frac{Pl(R\cup Y) - Pl((R\cup Y)\cap(R\cup Y))}{Pl(R\cup Y)} = \frac{1-Pl(R\cup Y)}{1} = 0$$
$$Bel(Y|\bar{B}) = \frac{Pl(R\cup Y) - Pl((R\cup Y)\cap(R\cup B))}{Pl(R\cup Y)} = \frac{1-Pl(R)}{1} = 2/3$$
$$Pl(R|\bar{B}) = \frac{Pl(R\cap(R\cup Y))}{Pl(R\cup Y)} = \frac{Pl(R)}{Pl(R\cup Y)} = 1/3$$
$$Pl(B|\bar{B}) = \frac{Pl(B\cap(R\cup Y))}{Pl(R\cup Y)} = \frac{Pl(\emptyset)}{1} = 0$$
$$Pl(Y|\bar{B}) = \frac{Pl(Y\cap(R\cup Y))}{Pl(R\cup Y)} = \frac{Pl(Y)}{Pl(R\cup Y)} = 2/3$$

Hence with Shafer's conditioning we get results incompatible with the real bounds of conditional probabilities because

$$[Bel(R|\bar{B}), Pl(R|\bar{B})] = [1/3, 1/3] \neq [1/3, 1]$$
$$[Bel(B|\bar{B}), Pl(B|\bar{B})] = [0, 0]$$
$$[Bel(Y|\bar{B}), Pl(Y|\bar{B})] = [2/3, 2/3] \neq [0, 2/3]$$

V. Fagin-Halpern conditioning

Fagin and Halpern (FH) proposed in [3], [4] to define the conditional belief as the lower envelope (i.e. the infimum) of a family of conditional probability functions to make belief conditioning consistent with imprecise conditional probability calculus.

A. Fagin-Halpern conditioning formulas

Assuming $Bel(B) > 0$, Fagin and Halpern proposed the following conditional formulas (FH formulas for short)

$$Bel(A|B) = Bel(A\cap B)/(Bel(A\cap B) + Pl(\bar{A}\cap B)) \quad (27)$$

$$Pl(A|B) = Pl(A\cap B)/(Pl(A\cap B) + Bel(\bar{A}\cap B)) \quad (28)$$

They prove in [3] that $Bel(A|B)$ given by (27) satisfies the three conditions of Theorem 1 and so FH belief conditioning is an appealing solution for BF conditioning. However, it is quite obscure how Fagin and Halpern did obtain (construct) FH formulas. A justification has been given by Sundberg and Wagner in [7] (p. 268) but it is not very easy to follow. In this paper, we justify clearly and directly the establishment of FH formulas from the simple and direct consequence of the Total Belief Theorem (TBT).

Similarly, by switching notations and assuming $Bel(A) > 0$, the previous FH formulas can be rewritten as

$$Bel(B|A) = Bel(A\cap B)/(Bel(A\cap B) + Pl(\bar{B}\cap A)) \quad (29)$$

$$Pl(B|A) = Pl(A\cap B)/(Pl(A\cap B) + Bel(\bar{B}\cap A)) \quad (30)$$

As we see, FH formulas are also consistent with Bayes' formula when the underlying BBA $m(\cdot)$ is Bayesian. Indeed if $m(\cdot)$ is Bayesian, then $Pl(A\cap B) = Bel(A\cap B) = P(A\cap B)$, $Pl(\bar{A}\cap B) = Bel(\bar{A}\cap B) = P(\bar{A}\cap B)$ and $Pl(\bar{B}\cap A) = Bel(\bar{B}\cap A) = P(\bar{B}\cap A)$ and FH formulas become equivalent to

$$Bel(A|B) = Pl(A|B) = P(A\cap B)/(P(A\cap B) + P(\bar{A}\cap B)) \quad (31)$$

Thanks to TPT formula (1), the denominator involved in these formula is $P(A\cap B) + P(\bar{A}\cap B) = P(B)$, therefore

$$Bel(A|B) = Pl(A|B) = P(A\cap B)/P(B) = P(A|B) \quad (32)$$

Similarly, one can also easily verify that

$$Bel(B|A) = Pl(B|A) = P(A\cap B)/P(A) = P(B|A) \quad (33)$$

B. Advantage of Fagin-Halpern conditioning

The advantage of FH conditioning is its complete compatibility with the conditional probability calculus [7], [25]. We show what provides FH conditioning in the previous Ellsberg urn example.

Ellsberg urn example revisited: Applying FH conditioning formulas with the conditioning event $\bar{B} = R \cup Y$ we obtain

$$Bel(R|\bar{B}) = \frac{Bel(R\cap(R\cup Y))}{Bel(R\cap(R\cup Y)) + Pl((B\cup Y)\cap(R\cup Y))} = \frac{Bel(R)}{Bel(R) + Pl(Y)} = \frac{1/3}{(1/3)+(2/3)} = 1/3$$
$$Pl(R|\bar{B}) = \frac{Pl(R\cap(R\cup Y))}{Bel((B\cup Y)\cap(R\cup Y)) + Pl(R\cap(R\cup Y))} = \frac{Pl(R)}{Bel(Y) + Pl(R)} = \frac{1/3}{0+(1/3)} = 1$$

Similarly, we can verify that $Bel(B|\bar{B}) = 0$, $Pl(B|\bar{B}) = 0$, $Bel(Y|\bar{B}) = 0$ and $Pl(Y|\bar{B}) = 2/3$. Therefore with these conditioning formulas, we get the correct bounds of the imprecise conditional probabilities

$$[Bel(R|\bar{B}), Pl(R|\bar{B})] = [1/3, 1]$$
$$[Bel(B|\bar{B}), Pl(B|\bar{B})] = [0, 0]$$
$$[Bel(Y|\bar{B}), Pl(Y|\bar{B})] = [0, 2/3]$$

One can also verify that $Bel(\emptyset|\bar{B}) = 0$, $Bel(R\cup B|\bar{B}) = 1/3$, $Bel(R\cup Y|\bar{B}) = 1$, $Bel(B\cup Y|\bar{B}) = 0$ and $Bel(R\cup B\cup Y|\bar{B}) = 1$. Applying Möbius inverse formula (20) with $Bel(\cdot|\bar{B})$, one gets the conditional BBA $m(R|\bar{B}) = 1/3$ and $m(R\cup Y|\bar{B}) = 2/3$, whereas with Shafer's conditioning one gets $m(R|\bar{B}) = 1/3$ and $m(Y|\bar{B}) = 2/3$. One sees that with Shafer's conditioning, because $(B\cup Y)\cap(R\cup Y) \neq \emptyset$ the mass $m(B\cup Y) = 2/3$ is entirely transferred (optimistically) to the most specific focal element Y included in $\bar{B} = R\cup Y$. With FH conditioning, the mass $m(B\cup Y) = 2/3$ is entirely transferred (pessimistically, or cautiously) to the least specific focal element $R\cup Y$ included in $\bar{B} = R\cup Y$.

VI. Total Belief Theorem (TBT)

In this section, we extend TPT theorem to BF and we establish the Total Belief Theorem (TBT) based on a decomposition of $\mathcal{F}_\Theta(m)$.

A. Decomposition of $\mathcal{F}_\Theta(m)$

Let us consider a FoD $\Theta = \{\theta_1, \ldots, \theta_{|\Theta|}\}$ with $|\Theta| > 1$ elements, and a BBA $m(\cdot)$ defined on 2^Θ with a given set of focal elements $\mathcal{F}_\Theta(m)$. Considering any partition $\{A_1, A_2, \ldots, A_k\}$ of the FoD Θ, then $\mathcal{F}_\Theta(m)$ can be obtained by the union of following subsets

$$\mathcal{F}_\Theta(m) = \mathcal{F}_{A_1}(m) \cup \ldots \cup \mathcal{F}_{A_k}(m) \cup \mathcal{F}_{A^*}(m) \quad (34)$$

where $\mathcal{F}_{A_i}(m)$ $(i = 1, \ldots, k)$ is the set of focal elements of $m(\cdot)$ included in A_i, and $\mathcal{F}_{A^*}(m)$ is the set of focal elements of $m(\cdot)$ which are not included in A_i, $i = 1, \ldots, k$. We use the notation A^* for representing the entity characterized by the focal set $\mathcal{F}_{A^*}(m)$ mathematically defined by

$$\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \ldots - \mathcal{F}_{A_k}(m) \quad (35)$$

The entity A^* has in general no explicit form and it is used only for notation convenience and conciseness. Because A_i for $i = 1, \ldots, k$ are mutually exclusive (disjoint), the sets $\mathcal{F}_{A_i}(m)$ are also mutually exclusive and therefore $\cap_{i=1,\ldots,k}(\mathcal{F}_\Theta(m) - \mathcal{F}_{A_i}(m)) = \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \ldots - \mathcal{F}_{A_k}(m)$ because all possible intersections of focal sets including $\mathcal{F}_{A_i}(m) \cap \mathcal{F}_{A_j}(m)$ for $j \neq i$ equal the empty set. Hence $\mathcal{F}_{A^*}(m)$ can also be expressed as

$$\mathcal{F}_{A^*}(m) = \cap_{i=1,\ldots,k}\bar{\mathcal{F}}_{A_i}(m) \quad (36)$$

where $\bar{\mathcal{F}}_{A_i}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_i}(m) = \mathcal{F}_{\bar{A}_i}(m) + \mathcal{F}_{A_i^*}(m)$ because when partitioning Θ as $\{A_i, \bar{A}_i\}$ one has $\mathcal{F}_{A_i^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_i}(m) - \mathcal{F}_{\bar{A}_i}(m)$.

Example 2: Consider $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ and a BBA $m(\cdot)$ defined on 2^Θ, with set of focal elements $\mathcal{F}_\Theta(m) = \{X_1, X_2, \ldots, X_8\}$ chosen as follows: $X_1 = \theta_1$, $X_2 = \theta_1 \cup \theta_2$, $X_3 = \theta_2 \cup \theta_3$, $X_4 = \theta_3 \cup \theta_4$, $X_5 = \theta_4$, $X_6 = \theta_4 \cup \theta_5$, $X_7 = \theta_1 \cup \theta_3 \cup \theta_5$, and $X_8 = \theta_5$. Consider also the partition $\{A_1, A_2, A_3\}$ of Θ with $A_1 = \{\theta_1, \theta_2\}$, $A_2 = \{\theta_3, \theta_4\}$ and $A_3 = \{\theta_5\}$. Therefore,

$$\begin{aligned}
\mathcal{F}_{A_1}(m) &= \{X_1, X_2\} = \{\theta_1, \theta_1 \cup \theta_2\}\\
\mathcal{F}_{A_2}(m) &= \{X_4, X_5\} = \{\theta_3 \cup \theta_4, \theta_4\}\\
\mathcal{F}_{A_3}(m) &= \{X_8\} = \{\theta_5\}\\
\mathcal{F}_{A^*}(m) &= \{X_1, \ldots, X_8\} - \{X_1, X_2\} - \{X_4, X_5\} - \{X_8\}\\
&= \{X_3, X_6, X_7\} = \{\theta_2 \cup \theta_3, \theta_4 \cup \theta_5, \theta_1 \cup \theta_3 \cup \theta_5\}\\
\bar{\mathcal{F}}_{A_1}(m) &= \mathcal{F}_\Theta(m) - \{X_1, X_2\} = \{X_3, X_4, X_5, X_6, X_7, X_8\}\\
\bar{\mathcal{F}}_{A_2}(m) &= \mathcal{F}_\Theta(m) - \{X_4, X_5\} = \{X_1, X_2, X_3, X_6, X_7, X_8\}\\
\bar{\mathcal{F}}_{A_3}(m) &= \mathcal{F}_\Theta(m) - \{X_8\} = \{X_1, X_2, X_3, X_4, X_5, X_6, X_7\}
\end{aligned}$$

Applying (36), one gets

$$\bar{\mathcal{F}}_{A_1}(m) \cap \bar{\mathcal{F}}_{A_2}(m) \cap \bar{\mathcal{F}}_{A_3}(m) = \{X_3, X_6, X_7\} = \mathcal{F}_{A^*}(m)$$

B. Total Belief Theorem (TBT)

Based on the previous decomposition of $\mathcal{F}_\Theta(m)$ according to any partition $\{A_1, \ldots, A_k\}$ of the FoD Θ, the following TBT holds.

Total Belief Theorem (TBT): Let's consider a FoD Θ with $|\Theta| \geq 2$ elements and a BBA $m(\cdot)$ defined on 2^Θ with the set of focal elements $\mathcal{F}_\Theta(m)$. For any chosen partition $\{A_1, \ldots, A_k\}$ of Θ and for any $B \subseteq \Theta$, one has

$$Bel(B) = \sum_{i=1,\ldots,k} Bel(A_i \cap B) + U(A^* \cap B) \tag{37}$$

where $\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \ldots - \mathcal{F}_{A_k}(m)$ and

$$U(A^* \cap B) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X). \tag{38}$$

Proof of TBT: See appendix.

A^* is a shorthand notation for the entity associated to the set of focal elements $\mathcal{F}_{A^*}(m)$ of the BBA $m(\cdot)$ involved in the summation (38) of $U(A^* \cap B)$. From (38), one sees that $U(A^* \cap B) \in [0, 1]$. If one applies TBT with $B = \Theta$, we get for any chosen partition $\{A_1, \ldots, A_k\}$ of Θ, $\sum_{i=1,\ldots,k} Bel(A_i) + U(A^*) = 1$ where $U(A^*) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m)} m(X)$. This equality corresponds to TPT if $U(A^*) = 0$ (i.e. there is no uncertainty on the value of probabilities of A_i, $i = 1, \ldots, k$). Note that if $B = \Theta$ and if the FoD Θ is simply partitioned as $\{A \triangleq A_1, \bar{A} \triangleq A_2\}$, then $U(A^* \cap B) = U(A^* \cap \Theta) = U(A^*) = Pl(A) - Bel(A) = Pl(\bar{A}) - Bel(\bar{A})$.

Corollary 1 of TBT: If $m(\cdot)$ is Bayesian, then TBT is consistent with the Total Probability Theorem (TPT) because $U(A^* \cap B) = 0$ and $Bel(\cdot)$ is homogeneous to a probability measure.

In expressing $Bel(\bar{B})$ with TBT and noting that $Pl(B) = 1 - Bel(\bar{B})$, one can also easily establish the following (not so elegant) Total Plausibility Theorem (TPlT).

Total Plausibility Theorem (TPlT): For any partition $\{A_1, \ldots, A_k\}$ of Θ and any $B \subseteq \Theta$, one has

$$Pl(B) = \sum_{i=1,\ldots,k} Pl(\bar{A}_i \cup B) + 1 - k - U(A^* \cap \bar{B}) \tag{39}$$

C. Example for TBT

Consider the FoD $\Theta = \{\theta_i, i = 1, \ldots, 7\}$ and $\mathcal{F}_\Theta(m) = \{X_1, X_2, \ldots, X_9\}$ of a BBA $m(\cdot)$ defined over 2^Θ as in Table I. Consider also the partition $\{A_1, A_2, A_3\}$ of Θ with $A_1 \triangleq \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_7$, $A_2 \triangleq \theta_2 \cup \theta_5$ and $A_3 \triangleq \theta_6$ and the subset $B = \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$ of Θ. The Table II summarizes the belief values of different subsets of Θ which are needed to apply TBT.

Focal element X	BBA $m(X)$
$X_1 = \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_7$	$m(X_1) = 0.01$
$X_2 = \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$	$m(X_2) = 0.02$
$X_3 = \theta_3 \cup \theta_5 \cup \theta_6$	$m(X_3) = 0.03$
$X_4 = \theta_4 \cup \theta_7$	$m(X_4) = 0.04$
$X_5 = \theta_2$	$m(X_5) = 0.20$
$X_6 = \theta_6 \cup \theta_7$	$m(X_6) = 0.30$
$X_7 = \theta_2 \cup \theta_3 \cup \theta_7$	$m(X_7) = 0.20$
$X_8 = \theta_1 \cup \theta_4 \cup \theta_6$	$m(X_8) = 0.15$
$X_9 = \theta_6$	$m(X_9) = 0.05$

Table I
FOCAL ELEMENTS AND THEIR MASSES.

Subsets of Θ	$Bel(\cdot)$
$B = \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$	$Bel(B) = 0.39$
$A_1 = \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_7$	$Bel(A_1) = 0.04$
$A_2 = \theta_2 \cup \theta_5$	$Bel(A_2) = 0.20$
$A_3 = \theta_6$	$Bel(A_3) = 0.05$
$A_1 \cap B = \theta_4 \cup \theta_7$	$Bel(A_1 \cap B) = 0.04$
$A_2 \cap B = \theta_5$	$Bel(A_2 \cap B) = 0$
$A_3 \cap B = \theta_6$	$Bel(A_3 \cap B) = 0.05$

Table II
BELIEF VALUES USED FOR THE DERIVATIONS.

In this example, one has

$$\begin{aligned}
\mathcal{F}_B(m) &= \{X_4, X_6, X_9\} \text{ and } \mathcal{F}_{\bar{B}}(m) = \{X_5\}\\
\mathcal{F}_{A_1}(m) &= \{X_4\} \text{ and } \mathcal{F}_{\bar{A}_1}(m) = \{X_5, X_9\}\\
\mathcal{F}_{A_2}(m) &= \{X_5\} \text{ and } \mathcal{F}_{\bar{A}_2}(m) = \{X_4, X_6, X_8, X_9\}\\
\mathcal{F}_{A_3}(m) &= \{X_9\} \text{ and } \mathcal{F}_{\bar{A}_3}(m) = \{X_1, X_2, X_4, X_5, X_7\}\\
\mathcal{F}_{A^*}(m) &= \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \mathcal{F}_{A_2}(m) - \mathcal{F}_{A_3}(m)\\
&= \{X_1, X_2, X_3, X_6, X_7, X_8\}
\end{aligned}$$

Therefore, one has

$$U(A^* \cap B) = \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X) = m(X_6) = 0.30$$

In applying TBT formula (37), one can easily verify

$$\begin{aligned}
Bel(B) &= \sum_{i=1,\ldots,3} Bel(B \cap A_i) + U(A^* \cap B)\\
&= 0.04 + 0 + 0.05 + 0.30 = 0.39
\end{aligned}$$

D. Generalization of TBT

As explained in Section II-B, we have to work in Cartesian product space $\Theta = \Theta_1 \times \Theta_2$ if the partition $\{A_1, \ldots, A_k\}$ is related to a given FoD Θ_1 and B is a subset of an other FoD Θ_2. Because $\{A_1, \ldots, A_k\}$ is a partition of Θ_1, then $\{A_1 \times \Theta_2, \ldots, A_k \times \Theta_2\}$ defines a partition of $\Theta = \Theta_1 \times \Theta_2$ and because $\Theta_1 \times B = \bigcup_{i=1,\ldots,k}((\Theta_1 \times B) \cap (A_i \times \Theta_2))$, one can always apply TBT in the Cartesian space Θ. More precisely, one has

$$Bel(\Theta_1 \times B) = \sum_{i=1,\ldots,k} Bel(A_i \times B) + U(A^* \times B)) \tag{40}$$

and where $U(A^* \times B) \triangleq U((A^* \times \Theta_2) \cap (\Theta_1 \times B))$.

This formula can be used if and only if one knows the joint BBA $m(\cdot)$ (or equivalently the joint belief) defined over the powerset of the Cartesian space $\Theta = \Theta_1 \times \Theta_2$.

VII. Conditional belief functions and GBT

Before justifying FH conditioning from TBT and presenting the Generalized Bayes' Theorem for BF, we establish a useful lemma.

Lemma 1: Consider a FoD Θ with a given BBA $m(\cdot)$ defined over Θ, for partition $\{A_i, \bar{A}_i\}$ of Θ and any $B \subseteq \Theta$, one always has

$$0 \leq U((\bar{A}_i \cap B)^*) - U(A^* \cap B) \leq 1 \tag{41}$$

where $U((\bar{A}_i \cap B)^*) = \sum_{X \in \mathcal{F}_{(\bar{A}_i \cap B)^*}(m)} m(X)$ and $U(A^* \cap B) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X)$.

Proof of Lemma 1: See appendix.

A. Conditional belief and plausibility

We consider a partition $\{A_i, \bar{A}_i\}$ of the FoD Θ and a subset B of Θ. Using TBT, one has

$$Bel(B) = Bel(A_i \cap B) + Bel(\bar{A}_i \cap B) + U(A^* \cap B) \tag{42}$$

Hence

$$Bel(B) - U(A^* \cap B) = Bel(A_i \cap B) + Bel(\bar{A}_i \cap B) \tag{43}$$

Moreover, since one has (by definition)

$$U((\bar{A}_i \cap B)^*) = Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B) \tag{44}$$

from the equality (44), one gets

$$Bel(\bar{A}_i \cap B) = Pl(\bar{A}_i \cap B) - U((\bar{A}_i \cap B)^*) \tag{45}$$

Putting the expression of $Bel(\bar{A}_i \cap B)$ above into (43) and rearranging terms, one gets

$$Bel(B) + \Delta(U) = Bel(A_i \cap B) + Pl(\bar{A}_i \cap B) \tag{46}$$

where $\Delta(U) \triangleq U((\bar{A}_i \cap B)^*) - U(A^* \cap B)$, and $\Delta(U) \in [0, 1]$ because of Lemma 1.

Assuming $Bel(B) > 0$, and dividing left and right sides of the equality (46) by $Bel(B) + \Delta(U)$, one gets

$$1 = \frac{Bel(A_i \cap B)}{Bel(B) + \Delta(U)} + \frac{Pl(\bar{A}_i \cap B)}{Bel(B) + \Delta(U)} \tag{47}$$

Hence, the equality (47) suggests to define the conditional belief $Bel(A_i|B)$ and $Pl(\bar{A}_i|B)$ as follows

$$Bel(A_i|B) \triangleq Bel(A_i \cap B)/(Bel(B) + \Delta(U)) \tag{48}$$

$$Pl(\bar{A}_i|B) \triangleq Pl(\bar{A}_i \cap B)/(Bel(B) + \Delta(U)) \tag{49}$$

Using equality (46), the previous conditioning formulas can be rewritten more concisely as

$$Bel(A_i|B) = Bel(A_i \cap B)/(Bel(A_i \cap B) + Pl(\bar{A}_i \cap B)) \tag{50}$$

$$Pl(\bar{A}_i|B) = Pl(\bar{A}_i \cap B)/(Bel(A_i \cap B) + Pl(\bar{A}_i \cap B)) \tag{51}$$

Replacing $\bar{A}_i$ by A_i in notations of formulas (49)–(51) we get[6] the following expressions for conditional plausibility $Pl(A_i|B)$

$$Pl(A_i|B) \triangleq \frac{Pl(A_i \cap B)}{Bel(B) + U((A_i \cap B)^*) - U(A^* \cap B)} \tag{52}$$

$$Pl(A_i|B) = \frac{Pl(A_i \cap B)}{Bel(\bar{A}_i \cap B) + Pl(A_i \cap B)} \tag{53}$$

Formulas (50) and (53) coincide with FH formulas [4] originally proposed from a very good intuition. In this work, we derive them only from TBT by a direct constructive manner. Note that $Bel(A_i|B)$ given in (48) satisfies $Bel(\emptyset|B) = 0$, $Bel(\Theta|B) = 1$, and $Bel(A_i|B) \in [0, 1]$ conditions. To prove that $Bel(A_i|B)$ defined by (50) is a belief function one must also prove that it is an n-monotone ($n \geq 2$) Choquet's capacity [24] on the finite set Θ, or equivalently that the condition B3 of Theorem 1 holds for $Bel(\cdot|B)$. The proof of B3 is difficult, but three different proofs have been already given by Fagin and Halpern [3], Jaffray [6], and Sundberg and Wagner [7], the latter one being the clearest of fashion.

[6] It is worth to note that one has always $U(A^* \cap B) = \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X) = U(\bar{A}^* \cap B)$ because $\mathcal{F}_{A^*}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_{A_i}(m) - \mathcal{F}_{\bar{A}_i}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_{\bar{A}_i}(m) - \mathcal{F}_{A_i}(m) = \mathcal{F}_{\bar{A}^*}(m)$.

B. Generalization of Bayes' Theorem

Starting from (48) with $\Delta(U) \triangleq U((\bar{A}_i \cap B)^*) - U(A^* \cap B)$ and replacing $Bel(B)$ by the expression (37) of TBT, we get

$$Bel(A_i|B) = \frac{Bel(A_i \cap B)}{\sum_{i=1,\ldots,k} Bel(A_i \cap B) + U((\bar{A}_i \cap B)^*)} \tag{54}$$

Similarly, in assuming $Bel(A_i) > 0$, Fagin-Halpern expression of $Bel(B|A_i)$ given by

$$Bel(B|A_i) = \frac{Bel(B \cap A_i)}{Bel(B \cap A_i) + Pl(\bar{B} \cap A_i)} \tag{55}$$

is equivalent to the formula

$$Bel(B|A_i) = \frac{Bel(B \cap A_i)}{Bel(A_i) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i)} \tag{56}$$

where

$$U((\bar{B} \cap A_i)^*) \triangleq Pl(\bar{B} \cap A_i) - Bel(\bar{B} \cap A_i) \tag{57}$$

$$= \sum_{X \in \mathcal{F}_{(\bar{B} \cap A_i)^*}(m)} m(X) \tag{58}$$

with $\mathcal{F}_{(\bar{B} \cap A_i)^*}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_{\bar{B} \cap A_i}(m) - \mathcal{F}_{B \cup \bar{A}_i}(m)$, and where

$$U(B^* \cap A_i) \triangleq \sum_{X \in \mathcal{F}_{B^*}(m) | X \in \mathcal{F}_{A_i}(m)} m(X) \tag{59}$$

with $\mathcal{F}_{B^*}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_B(m) - \mathcal{F}_{\bar{B}}(m)$.

From (56), one obtains

$$Bel(A_i \cap B) = Bel(B|A_i)[Bel(A_i) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i)]$$

Replacing the above expression of $Bel(A_i \cap B)$ into the formula (54), we obtain the formula

$$Bel(A_i|B) = \frac{Bel(B|A_i) q(A_i, B)}{\sum_{i=1}^k Bel(B|A_i) q(A_i, B) + U((\bar{A}_i \cap B)^*)} \tag{60}$$

where the factor $q(A_i, B)$ introduced here for notation conciseness is defined by

$$q(A_i, B) \triangleq Bel(A_i) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i) \tag{61}$$

This allows to establish the Generalized Bayes' Theorem (GBT).

Generalized Bayes' Theorem (GBT): For any partition $\{A_1, \ldots, A_k\}$ of a FoD Θ, any belief function $Bel(\cdot) : 2^\Theta \mapsto [0, 1]$, and any subset B of Θ with $Bel(B) > 0$, then one has for $i \in \{1, \ldots, k\}$

$$Bel(A_i|B) = \frac{Bel(B|A_i) q(A_i, B)}{\sum_{i=1}^k Bel(B|A_i) q(A_i, B) + U((\bar{A}_i \cap B)^*)} \tag{62}$$

$$U((\bar{A}_i \cap B)^*) \triangleq Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B)$$

where $U((\bar{A}_i \cap B)^*) \triangleq \sum_{X \in \mathcal{F}_{(\bar{A}_i \cap B)^*}(m)} m(X) = Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B)$, and where the factor $q(A_i, B)$ is defined by (61).

Lemma 2: GBT reduces to BT if $Bel(\cdot)$ is a Bayesian BF.

Proof: See appendix.

Remark: When $A_i \subseteq \Theta_1$ and $B \subseteq \Theta_2$ with $\Theta_1 \neq \Theta_2$, we must work in the Cartesian product space $\Theta = \Theta_1 \times \Theta_2$ and the GBT formula is similar to (62) in replacing A_i by $A_i \times \Theta_2$, and B by $\Theta_1 \times B$. The application of GBT formula is not easy in general because it requires the knowledge of joint BBA $m(\cdot)$ defined over $2^{\Theta_1 \times \Theta_2}$ which is rarely known in practice. If the joint BBA $m(\cdot)$ can be expressed (or approximated) as a function of two marginal BBAs $m_1(\cdot)$ and $m_2(\cdot)$ (assumed to be known) defined respectively over Θ_1 and Θ_2, then GBT formula should become tractable.

VIII. Illustrative example of GBT

Consider $\Theta = \{\theta_i, i = 1, \ldots, 7\}$, $\mathcal{F}_\Theta(m) = \{X_1, X_2, \ldots, X_9\}$ and $m(\cdot)$ given in Table III.

Focal element X	BBA $m(X)$
$X_1 = \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_7$	$m(X_1) = 0.01$
$X_2 = \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$	$m(X_2) = 0.02$
$X_3 = \theta_3 \cup \theta_5 \cup \theta_6$	$m(X_3) = 0.03$
$X_4 = \theta_4 \cup \theta_7$	$m(X_4) = 0.04$
$X_5 = \theta_2$	$m(X_5) = 0.20$
$X_6 = \theta_6 \cup \theta_7$	$m(X_6) = 0.30$
$X_7 = \theta_2 \cup \theta_3 \cup \theta_7$	$m(X_7) = 0.20$
$X_8 = \theta_1 \cup \theta_4 \cup \theta_6$	$m(X_8) = 0.15$
$X_9 = \theta_6$	$m(X_9) = 0.05$

Table III
FOCAL ELEMENTS AND THEIR MASSES.

Consider the partition $\{A_1, A_2, A_3\}$ of Θ with $A_1 = \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_7$, $A_2 = \theta_2 \cup \theta_5$ and $A_3 = \theta_6$, and the subset $B = \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$ of Θ having belief $Bel(B) = m(X_4) + m(X_6) + m(X_9) = 0.39$. Table IV summarizes the BF values which are needed in the derivations.

Subsets X of Θ	$Bel(X)$	$Pl(X)$
$X = B = \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$	$Bel(B) = 0.39$	$Pl(B) = 0.80$
$X = A_1 = \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_7$	$Bel(A_1) = 0.04$	$Pl(A_1) = 0.75$
$X = A_2 = \theta_2 \cup \theta_5$	$Bel(A_2) = 0.20$	$Pl(A_2) = 0.46$
$X = A_3 = \theta_6$	$Bel(A_3) = 0.05$	$Pl(A_3) = 0.53$
$X = A_1 \cap B = \theta_4 \cup \theta_7$	$Bel(X) = 0.04$	$Pl(X) = 0.72$
$X = A_2 \cap B = \theta_5$	$Bel(X) = 0$	$Pl(X) = 0.04$
$X = A_3 \cap B = \theta_6$	$Bel(X) = 0.05$	$Pl(X) = 0.53$
$X = \bar{A}_1 \cap B = \theta_5 \cup \theta_6$	$Bel(X) = 0.05$	$Pl(X) = 0.54$
$X = \bar{A}_2 \cap B = \theta_4 \cup \theta_6 \cup \theta_7$	$Bel(X) = 0.39$	$Pl(X) = 0.80$
$X = \bar{A}_3 \cap B = \theta_4 \cup \theta_5 \cup \theta_7$	$Bel(X) = 0.04$	$Pl(X) = 0.75$
$X = A_1 \cap \bar{B} = \theta_1 \cup \theta_3$	$Bel(X) = 0$	$Pl(X) = 0.41$
$X = A_2 \cap \bar{B} = \theta_2$	$Bel(X) = 0.20$	$Pl(X) = 0.43$
$X = A_3 \cap \bar{B} = \emptyset$	$Bel(X) = 0$	$Pl(X) = 0$

Table IV
BELIEF AND PLAUSIBILITY VALUES USED FOR THE DERIVATIONS.

• **Results with Fagin-Halpern conditioning formulas**

Using (50) and (55) and the fact that $Pl(A_i|B) = 1 - Bel(\bar{A}_i|B)$ and $Pl(B|A_i) = 1 - Bel(\bar{B}|A_i)$, we get the values of Tables V–VI.

Subsets of Θ	$Bel(A_i\|B)$	$Pl(A_i\|B)$
A_1	$Bel(A_1\|B) \approx 0.0690$	$Pl(A_1\|B) \approx 0.9351$
A_2	$Bel(A_2\|B) = 0$	$Pl(A_2\|B) \approx 0.0930$
A_3	$Bel(A_3\|B) \approx 0.0625$	$Pl(A_3\|B) \approx 0.9298$

Table V
$Bel(A_i|B)$ AND $Pl(A_i|B)$ WITH FAGIN-HALPERN CONDITIONING.

Subsets of Θ	$Bel(B\|A_i)$	$Pl(B\|A_i)$
A_1	$Bel(B\|A_1) \approx 0.0889$	$Pl(B\|A_1) = 1$
A_2	$Bel(B\|A_2) = 0$	$Pl(B\|A_2) \approx 0.1667$
A_3	$Bel(B\|A_3) = 1$	$Pl(B\|A_3) = 1$

Table VI
$Bel(B|A_i)$ AND $Pl(B|A_i)$ WITH FAGIN-HALPERN CONDITIONING.

To verify GBT, one calculates $Bel(A_i)$, $U((\bar{B} \cap A_i)^*)$ and $U(B^* \cap A_i)$ for getting $q(A_i, B)$, and $U((\bar{A}_i \cap B)^*)$. These values are given in Table VII. $q(A_1, B) = 0.45$ is calculated by $q(A_1, B) \triangleq Bel(A_1) + U((\bar{B} \cap A_1)^*) - U(B^* \cap A_1) = 0.45$ because $Bel(A_1) = 0.04$, $U((\bar{B} \cap A_1)^*) = Pl(\bar{B} \cap A_1) - Bel(\bar{B} \cap A_1) = 0.41$ and $U(B^* \cap A_1) = \sum_{X \in \mathcal{F}_{A_1}(m) | X \in \mathcal{F}_{B^*}(m)} m(X) = 0$. $U((\bar{A}_1 \cap B)^*) = 0.49$ is calculated by $U((\bar{A}_1 \cap B)^*) = Pl(\bar{A}_1 \cap B) - Be(\bar{A}_1 \cap B) = 0.54 - 0.05 = 0.49$, and other values of Table VII are calculated similarly.

Subsets of Θ	$q(A_i, B)$	$U((\bar{A}_i \cap B)^*)$
A_1	0.45	0.49
A_2	0.43	0.41
A_3	0.05	0.71

Table VII
VALUES OF $q(A_i, B)$ AND $U((\bar{A}_i \cap B)^*)$ FOR GBT FORMULA.

One verifies that GBT formula (62) works because we retrieve correct values obtained with FH formula. Indeed, one has

$$\begin{aligned} Bel(A_1|B) &= \frac{Bel(B|A_1)q(A_1, B)}{\sum_{i=1}^{3} Bel(B|A_i)q(A_i, B) + U((\bar{A}_1 \cap B)^*)} \\ &\approx \frac{0.0889 \cdot 0.45}{(0.0889 \cdot 0.45) + (0 \cdot 0.43) + (1 \cdot 0.05) + 0.49} \\ &\approx 0.0690 \end{aligned}$$

Similarly, one can easily verify that one obtains $Bel(A_2|B) = 0$ and $Bel(A_3|B) \approx 0.0625$ with GBT.

• **Results with Shafer's conditioning formulas**

With formulas (22)–(23), we get the values of Tables VIII–IX.

Subsets of Θ	$Bel(A_i\|B)$	$Pl(A_i\|B)$
A_1	$Bel(A_1\|B) = 0.3250$	$Pl(A_1\|B) = 0.9000$
A_2	$Bel(A_2\|B) = 0$	$Pl(A_2\|B) = 0.0500$
A_3	$Bel(A_3\|B) = 0.0625$	$Pl(A_3\|B) = 0.6625$

Table VIII
$Bel(A_i|B)$ AND $Pl(A_i|B)$ WITH SHAFER'S CONDITIONING.

Subsets of Θ	$Bel(B\|A_i)$	$Pl(B\|A_i)$
A_1	$Bel(B\|A_1) \approx 0.4533$	$Pl(B\|A_1) \approx 0.9600$
A_2	$Bel(B\|A_2) \approx 0.0652$	$Pl(B\|A_2) \approx 0.0870$
A_3	$Bel(B\|A_3) = 1$	$Pl(B\|A_3) = 1$

Table IX
$Bel(B|A_i)$ AND $Pl(B|A_i)$ WITH SHAFER'S CONDITIONING.

One sees that the conditional values are not coherent since they do not verify GBT because we obtain in this example

$$\begin{aligned} Bel(A_1|B) &= 0.3250 \text{ (using (23))} \\ &\neq \frac{Bel(B|A_1)q(A_1, B)}{\sum_{i=1}^{3} Bel(B|A_i)q(A_i, B) + U((\bar{A}_1 \cap B)^*)} \\ &\approx \frac{0.4533 \cdot 0.45}{(0.4533 \cdot 0.45) + (0.0652 \cdot 0.43) + (1 \cdot 0.05) + 0.49} \\ &\approx 0.2642 \end{aligned}$$

Similarly, one can show that $Bel(A_2|B) = 0$ (using (23)) $\neq 0.0405$ (using GBT) and $Bel(A_3|B) = 0.0625$ (using (23)) $\neq 0.0504$ (using GBT). Hence, Ellsberg urn example and this example show clearly that Dempster's rule of combination used by Shafer to establish his belief conditioning formulas does not provide coherent and satisfactory results since they are inconsistent with lower and upper bounds of imprecise conditional probabilities and they do not satisfy GBT established directly by a constructive manner without ad-hoc assumption.

IX. Conclusion

This paper has presented new important results: the Total Belief Theorem (TBT), the justification of Fagin-Halpern conditioning from TBT, and the Generalized Bayes' Theorem (GBT). Our theoretical results allowed us to establish rigorously the Generalized Bayes' Theorem by a direct constructive manner from TBT. It does not need extra assumptions nor some underlying ad-hoc construction principles. Also, we prove that our TBT and GBT are fully consistent with classical TPT and Bayes Theorem as soon as the belief functions are Bayesian. That way this achievement could be an excellent ground for working in belief function framework. From Ellsberg's urn example and an illustrative example we have shown that Shafer's conditioning based on Dempster's rule provides results inconsistent with lower and upper bounds of imprecise conditional probabilities, and inconsistent with GBT. These new results should allow to reconcile practitioners of Bayesian reasoning with those of evidential reasoning.

Appendix

A. Proof of TBT

$$\begin{aligned}
Bel(B) &= \sum_{X\in\mathcal{F}_\Theta(m)|X\subseteq B} m(X)\\
&= \sum_{X\in\mathcal{F}_{A_1}(m)|X\in\mathcal{F}_B(m)} m(X) + \ldots\\
&\quad + \sum_{X\in\mathcal{F}_{A_k}(m)|X\in\mathcal{F}_B(m)} m(X)\\
&\quad + \sum_{X\in\mathcal{F}_{A^*}(m)|X\in\mathcal{F}_B(m)} m(X)\\
&= Bel(A_1\cap B) + \ldots + Bel(A_k\cap B)\\
&\quad + \sum_{X\in\mathcal{F}_{A^*}(m)|X\in\mathcal{F}_B(m)} m(X)\\
&= \sum_{i=1,\ldots,k} Bel(A_i\cap B) + U(A^*\cap B)
\end{aligned}$$

where $U(A^*\cap B) \triangleq \sum_{X\in\mathcal{F}_{A^*}(m)|X\in\mathcal{F}_B(m)} m(X)$.

B. Proof of Lemma 1

For notation convenience, we denote

$$\begin{aligned}
\Delta(U) &\triangleq U((\bar{A}_i\cap B)^*) - U(A^*\cap B)\\
&= [Pl(\bar{A}_i\cap B) - Bel(\bar{A}_i\cap B)]\\
&\quad - [Bel(A_i\cap B) + Bel(\bar{A}_i\cap B) - Bel(B)]\\
&= Pl(\bar{A}_i\cap B) - Bel(\bar{A}_i\cap B) + Bel(B)\\
&\quad - Bel(A_i\cap B) - Bel(\bar{A}_i\cap B)
\end{aligned}$$

To prove that $\Delta(U) \geq 0$, one needs to prove equivalently that $Pl(\bar{A}_i\cap B) - Bel(\bar{A}_i\cap B) + Bel(B) \geq Bel(A_i\cap B) + Bel(\bar{A}_i\cap B)$. Using TBT, one has $Bel(B) = Bel(A_i\cap B) + Bel(\bar{A}_i\cap B) + U(A^*\cap B)$, and replacing expression of $Bel(B)$ in the previous inequality, one must verify if the following equality is satisfied

$$\begin{aligned}
Pl(\bar{A}_i\cap B) - Bel(\bar{A}_i\cap B) + Bel(A_i\cap B) + Bel(\bar{A}_i\cap B) + U(A^*\cap B)\\
\geq Bel(A_i\cap B) + Bel(\bar{A}_i\cap B)
\end{aligned}$$

After simplification, we have to check if inequality below holds

$$Pl(\bar{A}_i\cap B) + U(A^*\cap B) \geq Bel(\bar{A}_i\cap B).$$

Because $Pl(\bar{A}_i\cap B) = Bel(\bar{A}_i\cap B) + U((\bar{A}_i\cap B)^*)$, one has to check if $Bel(\bar{A}_i\cap B) + U((\bar{A}_i\cap B)^*) + U(A^*\cap B) \geq Bel(\bar{A}_i\cap B)$. After simplification (omitting both $Bel(\bar{A}_i\cap B)$ in left and right side of the previous inequality), one just has to prove the inequality $U((\bar{A}_i\cap B)^*) + U(A^*\cap B) \geq 0$ in order to prove that $\Delta(U) \geq 0$. Because $U((\bar{A}_i\cap B)^*) \in [0,1]$ and $U(A^*\cap B) \in [0,1]$, the previous inequality always holds which proves that $U((\bar{A}_i\cap B)^*) - U(A^*\cap B) \geq 0$. Moreover because $U(A^*\cap B) \in [0,1]$, then $-U(A^*\cap B) \in [-1,0]$, and because $U((\bar{A}_i\cap B)^*) \in [0,1]$ one deduces that $\Delta(U) = U((\bar{A}_i\cap B)^*) - U(A^*\cap B) \leq 1$.

C. Proof of Lemma 2

If $Bel(\cdot) : 2^\Theta \mapsto [0,1]$ is a Bayesian belief function, then all focal elements of its corresponding BBA $m(\cdot)$ are singletons of 2^Θ. In this case $Bel(\cdot)$ and $Pl(\cdot)$ functions coincide and therefore one has $U((\bar{A}_i\cap B)^*) = Pl(\bar{A}_i\cap B) - Bel(\bar{A}_i\cap B) = 0$ and $U((\bar{B}\cap A_i)^*) = Pl(\bar{B}\cap A_i) - Bel(\bar{B}\cap A_i) = 0$. Any focal element (singleton) of $m(\cdot)$ is either a subset of B or a subset of $\bar{B}$ of the FoD Θ. Therefore, $\mathcal{F}_{B^*}(m) = \emptyset$, which implies $U(B^*\cap A_i) = 0$, so that $q(A_i,B) = Bel(A_i)$. The GBT formula (62) with in this case $q(A_i,B) = Bel(A_i)$ and $U((\bar{A}_i\cap B)^*) = 0$ reduces to the formula $Bel(A_i|B) = Bel(B|A_i)Bel(A_i)/\sum_{i=1}^k Bel(B|A_i)Bel(A_i)$. This coincides with formula (10) since $Bel(\cdot)$ (being a Bayesian belief function) is homogeneous to a probability measure $P(\cdot)$. This completes the proof that GBT formula is consistent with Bayes' Theorem formula when the Belief function is Bayesian.

References

[1] G. Shafer, A Mathematical Theory of Evidence, Princeton Press, 1976.
[2] A. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. of Math. Stat., (38):325-339, 1967.
[3] R. Fagin, J.Y. Halpern, A new approach to updating beliefs, UAI Conf. Proc., 317-325, 1991.
[4] J.Y. Halpern, R. Fagin, Two views of belief: belief as generalized probability and belief as evidence, Art. Intel., (54): 275-317, 1992.
[5] J.Y. Halpern, Reasoning about uncertainty, MIT Press, 2003.
[6] J.-Y. Jaffray, Bayesian updating and belief functions, IEEE Trans. on SMC, (22):1144-1152, 1992.
[7] C. Sunberg, C. Wagner, Generalized finite differences and Bayesian conditioning of Choquet capacities, Adv. in Appl. Math. (13), 1992.
[8] P. Diaconis, S.L. Zabell, Updating subjective probability, J. of Amer. Stat. Soc., (77):822-830, 1982.
[9] L. Zadeh, A Mathematical Theory of Evidence (book review), AI Magazine, (5):81-83, 1984.
[10] J. Lemmer, Confidence factors, empiricism and the Dempster-Shafer theory of evidence, Proc. of 1st UAI Conf., pp. 16–176, 1985.
[11] P.K. Black, Is Shafer general Bayes?, 3rd UAI Workshop, USA, 1987.
[12] P. Wang, A defect in Dempster-Shafer theory, UAI Conf. Proc., 1994.
[13] A. Gelman, The boxer, the wrestler, and the coin flip: a paradox of robust Bayesian inference and belief functions, Amer. Stat. (60):146-150, 2006.
[14] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan 1988.
[15] J. Dezert, P. Wang, A. Tchamova, On the validity of Dempster-Shafer theory, Fusion 2012 Proc., Singapore, July 9–12, 2012.
[16] J. Dezert, A. Tchamova, On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule, Int. J. of Intell. Syst., (29):223-252, 2014.
[17] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, IJAR, Vol. 9, No. 1, pp. 1–35, 1993.
[18] D. Dubois, T. Denoeux, Conditioning in Dempster-Shafer Theory: Prediction vs. Revision, Belief 2012 Conf. Proc., Compiègne, May 2012.
[19] A. Dempster, A generalization of Bayesian inference, J . of Royal Stat. Soc., (B30):205-247,1968.
[20] A. Papoulis, Probability, Random Variables, and Stochastic Processes, 2nd ed. New York: McGraw-Hill, pp. 37–38, 1984.
[21] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, Why Dempster's rule doesn't behave as Bayes rule with informative priors, INISTA 2013.
[22] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, Why Dempster's fusion rule is not a generalization of Bayes fusion rule, Fusion 2013 Proc.
[23] D. Ellsberg, Risk, ambiguity, and the Savage axioms, Quart. J. Econ. (75):643-669, 1961.
[24] G. Choquet, Theory of capacities, Ann. Inst. Fourier, (5):131-295, 1953.
[25] J.N Heendeni et al., A generalization of Bayesian inference in the Dempster-Shafer belief theoretic framework, Proc. of Fusion 2016.