

Plato’s Cave in the Dempster-Shafer Land: the Link between Pignistic and Plausibility Transformations

Chunlai Zhou^{1,2}, Biao Qin¹, Xiaoyong Du¹

¹DEKE(MOE), Computer Science Dept., Renmin University of China

²State Key Lab of Computer Science, Institute of Software, Chinese Academy of Sciences
{czhou,qinbiao,duyong}@ruc.edu.cn

Abstract

In reasoning under uncertainty in AI, there are (at least) two useful and different ways of understanding beliefs: the first is as *absolute belief* or degree of belief in propositions and the second is as *belief update* or measure of change in belief. Pignistic and plausibility transformations are two well-known probability transformations that map belief functions to probability functions in the Dempster-Shafer theory of evidence. In this paper, we establish the link between pignistic and plausibility transformations by devising a belief-update framework for belief functions where plausibility transformation works on belief update while pignistic transformation operates on absolute belief. In this framework, we define a *new* belief-update operator connecting the two transformations, and interpret the framework in a belief-function model of parametric statistical inference. As a metaphor, these two transformations projecting the belief-update framework for belief functions to that for probabilities are likened to the fire projecting reality into shadows on the wall in Plato’s cave.

1 Introduction

The Greek philosopher Plato presented *the metaphor of cave* to compare “the effect of education and the lack of it on our nature.” A gathering of prisoners have lived chained to the wall of a cave all of their lives. They face a blank wall, watch shadows projected onto the wall from things passing in front of a fire behind them and mistakenly interpret the shadows as reality. One man breaks free from his chains and runs out of the cave. He sees the real world and learns that it is far beyond the shadows he had been seeing. He then goes back to tell his fellow prisoners how reality is mapped into shadows.

In this paper we employ this metaphor to describe relationships between the belief-update frameworks for belief and probability functions¹. Dempster-Shafer theory of evidence (or simply DS theory) [Shafer, 1976] generalizes the

¹Here we adapt the metaphor from Judea Pearl who used it to explain a similar relationship of causal reasoning to probabilistic inference.

Bayesian theory of subjective probability. This formalism represents a degree of belief (also referred to as a *mass*) as a belief function rather than a Bayesian probability distribution. A belief function assigns its masses to *all* non-empty subsets while a probability function assigns non-zero masses *only* to singletons. So, a belief function can be more expressive than a probability function. Belief functions may account for partially specified knowledge states, for example total ignorance, which can’t be represented by a probability measure. Moreover, Dempster-Shafer theory allows one to combine evidence from two different sources according to Dempster’s rule, and arrive at a measure of belief (represented by a belief function) that takes into account all available evidence. Dempster’s conditioning is a special case of this rule when one source is to establish the conditioning event with certainty.

In this paper we study probability transformations which map belief functions into probability functions. Probability transformations are classified into two major types [Cuzzolin, 2012]: those commuting with Dempster’s rule of combination and those commuting with affine combination. The first type seeks efficient probabilistic approximation intended to reduce the computational cost operating with belief functions due to the high computational complexity of Dempster’s rule. So, probability transformations in this family are consistent with the Dempster-Shafer semantics and commute with Dempster’s rule. Plausibility transformation is a representative of this family [Voorbraak, 1989; Cobb and Shenoy, 2006]. The second type of probability transformations is a decision-based approach which redistributes the mass of each focal element in a belief function to all its elements according to some prescribed mechanism. Such mass allocation scheme usually makes use of the linearity property or commutes with affine combination [Smets, 2005]. The most famous one among this family is Smets’ pignistic transformation [Smets and Kennes, 1994].

Our main contribution establishes the link between plausibility and pignistic transformations by setting up a framework for belief update in Dempster-Shafer theory (Theorem 3.1). There are two important components in a general framework of belief updates in the process of reasoning under uncertainty [Heckerman, 1986]: absolute belief and belief update. *Absolute belief* measures quantities which reflect the absolute degree to which propositions are believed while *belief update*

refers to the degree to which beliefs in propositions change as evidence about them becomes known, and functions as a mapping from absolute beliefs to absolute beliefs. The well-known Bayes theorem exactly characterizes the belief-update framework for probability functions. In this paper, we devise a belief-update paradigm for belief functions. We follow Smets' approach by representing the evidence in an observation with the combination of conditional beliefs conditioned on the observation [Smets, 1993]. The main difficulty in constructing the paradigm is defining a belief-update operator combining a prior belief with the evidence. We propose a *new* operator which treats the prior as absolute belief and the evidence as belief-update (Eq. (7)). Moreover, we *axiomatically justify* this operator with three natural properties. Under this paradigm, we show that plausibility transformation performs on beliefs as belief update while pignistic transformation works on beliefs as absolute beliefs. These two transformations project the belief-update framework for belief functions to that for probability functions, which can be described by the analogy of Plato's cave where the fire projects the reality to shadows on the wall. In this sense, our established framework gives a *new* generalized Bayesian Theorem for belief functions.

2 Dempster-Shafer Theory

Let Ω be a frame of discernment and $\mathcal{A} = 2^\Omega$ be the Boolean algebra of propositions. $|A|$ denotes the cardinality of a subset A . A *mass assignment* (or *mass function*) over Ω is a mapping $m : \mathcal{A} \rightarrow [0, 1]$ satisfying $\sum_{A \in \mathcal{A}} m(A) = 1$. A mass function m is called *normal* if $m(\emptyset) = 0$. Without further notice, all mass functions in this paper are assumed to be normal. A *belief function* is a function $bel : \mathcal{A} \rightarrow [0, 1]$ satisfying the conditions: $bel(\emptyset) = 0$, $bel(\Omega) = 1$ and $bel(\bigcup_{i=1}^n A_i) \geq \sum_{\emptyset \neq I \subseteq \{1, \dots, n\}} (-1)^{|I|+1} bel(\bigcap_{i \in I} A_i)$ where $A_i \in \mathcal{A}$ for all $i \in \{1, \dots, n\}$. A mapping $f : \mathcal{A} \rightarrow [0, 1]$ is a belief function if and only if its Möbius transform is a mass assignment [Shafer, 1976]. In other words, if $m : \mathcal{A} \rightarrow [0, 1]$ is a mass assignment, then it determines a belief function $bel : \mathcal{A} \rightarrow [0, 1]$ as follows: $bel(A) = \sum_{B \subseteq A} m(B)$ for all $A \in \mathcal{A}$. Moreover, given a belief function bel , we can obtain its corresponding mass function m as follows: $m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} bel(B)$ for all $A \in \mathcal{A}$. Intuitively, for a subset event A , $m(A)$ measures the belief that an agent commits *exactly* to A , not the total belief $bel(A)$ that an agent commits to A . Elements A with non-zero masses are called *focal elements*. The belief function bel is called *Bayesian* if $m(A) = 0$ for all non-singletons A . It is called *categorical* if it has only one focal set. And it is called *vacuous* if Ω is the only focal element and its mass function is denoted by m_V . The vacuous belief represents the state of total ignorance. The corresponding *plausibility function* $pl_m : 2^\Omega \rightarrow [0, 1]$ and *commonality function* $Q_m : 2^\Omega \rightarrow [0, 1]$ are defined respectively by $pl_m(A) = \sum_{E \cap A \neq \emptyset} m(E)$ and $Q_m(A) = \sum_{E \supseteq A} m(E)$ for all $A \subseteq \Omega$. Whenever the context is clear, we drop the subscript m . In this paper, If m_1 and m_2 are two mass functions on Ω induced by two *independent* evidential sources, the combined mass function is calculated according to *Demp-*

ster's rule of combination: for any $C \subseteq \Omega$,

$$(m_1 \oplus m_2)(C) = \frac{\sum_{A \cap B = C} m_1(A) m_2(B)}{\sum_{A \cap B \neq \emptyset} m_1(A) m_2(B)} \quad (1)$$

When an event E is observed, then the conditional mass function of m is obtained according to *Dempster conditioning*: for any $C \subseteq \Omega$, $m(C | E) = \frac{\sum_{B \cap E = C} m(B)}{pl(E)}$. Suppose that $\Omega' \supseteq \Omega$ and m is a mass function over Ω . The mass function m can be identified with a mass function $\vec{m}_{\Omega'}$ over the bigger frame Ω' : for any $E' \subseteq \Omega'$, $\vec{m}_{\Omega'}(E') = m(E)$ if $E' = E \cup (\Omega' \setminus \Omega)$ and $\vec{m}_{\Omega'}(E') = 0$ otherwise. Such defined $\vec{m}_{\Omega'}$ is called the *conditional embedding* of m to Ω' . When the context is clear, we usually drop the subscript Ω' . It is easy to see that conditional embedding is inverse to the above Dempster conditioning.

From now on, we fix a frame of discernment Ω . Let \mathcal{P} and \mathcal{M} denote the sets of probability and mass functions on Ω , respectively. The *Bayesian parallel combination* $\odot : \mathcal{P} \times \mathcal{P} \rightarrow \mathcal{P}$ is a binary operator on probabilities defined as follows: for any two probability functions $p_1, p_2 \in \mathcal{P}$,

$$(p_1 \odot p_2)(\omega) := \frac{p_1(\omega) p_2(\omega)}{\sum_{\omega \in \Omega} p_1(\omega) p_2(\omega)} \text{ for all } \omega \in \Omega. \quad (2)$$

When both m_1 and m_2 are probability functions, then the Dempster combination (Eq. (1)) degenerates into the Bayesian parallel combination. A function τ from \mathcal{M} to \mathcal{P} is called a *probability transformation*. A probability transformation τ is called *commuting with affine combination* if for any mass functions m_1, \dots, m_n , $\tau(\sum_{i=1}^n a_i m_i) = \sum_{i=1}^n a_i \tau(m_i)$ for all nonnegative a_1, \dots, a_n such that $\sum_{i=1}^n a_i = 1$. It is called *commuting with Dempster's rule* if, for any two mass functions m_1 and m_2 , $\tau(m_1 \oplus m_2) = \tau(m_1) \odot \tau(m_2)$.

There are many probability transformation methods used in the literature for transforming belief functions to probability functions. The two most prominent are Smets' pignistic transformation and plausibility transformation. Pignistic transformation is decision-based and it is at the foundation of Smets' theory of transferable belief models [Smets and Kennes, 1994]. A transferable belief model is a two-level mental model: the credal level where beliefs are represented by a mass function, and the pignistic level where decisions are made according to a probability distribution. It is a probability distribution *derived from* the mass function through *pignistic transformation* $Betp : \mathcal{M} \rightarrow \mathcal{P}$. The pignistic transformation for the above mass function m is given by

$$Betp(m)(\omega) := \sum_{\omega \in B \subseteq \Omega} \frac{1}{|B|} m(B) \text{ for any } \omega \in \Omega. \quad (3)$$

Note that $Betp(m)$ is a probability function on Ω and is called a *pignistic probability function*. The *pignistic probability function depends on not only the belief m but also the structure of the frame*. An alternative probability transformation called *plausibility transformation* was originally proposed by Voorbraak [Voorbraak, 1989] for efficient probabilistic approximation in the Dempster-Shafer framework. The plausibility transformation $Pl_P : \mathcal{M} \rightarrow \mathcal{P}$ is given by: for any $\omega \in \Omega$,

$$Pl_P(m)(\omega) := \frac{pl(\omega)}{\sum_{\omega \in \Omega} pl(\omega)} \quad (4)$$

where pl is the corresponding plausibility function of m . Plausibility transformation Pl_P perfectly preserves the updating power of a mass function m as belief update according to the Dempster combination [Voorbraak, 1989] in the sense that, for any probability function pr , $pr \oplus m = pr \oplus Pl_P(m)$. This transformation has been later supported by Cobb and Shenoy [Cobb and Shenoy, 2006] in virtue of its commuting with Dempster's rule of combination: for any two mass functions m_1 and m_2 ,

$$Pl_P(m_1 \oplus m_2) = Pl_P(m_1) \odot Pl_P(m_2) \quad (5)$$

3 Plato's Cave in the Dempster-Shafer Land

In this section, we first present Heckerman's general framework for belief updates, and then devise a framework for belief functions by defining a new operator combining a prior belief with the evidence in an observation. This operator is axiomatically justified by three natural properties. Within this framework, we establish the connections between plausibility and pignistic transformations. In the last part, we interpret this framework in a belief-function model of parametric statistical inference.

3.1 A General Framework for Belief Updates

The problem of reasoning under uncertainty can be decomposed into three distinct components: problem formulation, belief assignment, and belief entailment [Heckerman, 1986]. *Problem formulation* refers to the task of specifying the propositions or events of interest as well as possible outcomes of each proposition and relations among propositions. The specification in this phase is usually treated as *generic knowledge*, i.e., it pertains to a class of situations as a whole. The *assignment of belief* refers to the process of constructing and assigning beliefs to propositions of interests. Finally, belief entailment focuses on the change of beliefs in propositions when evidence about them becomes known. *Absolute beliefs* indicate both quantities which reflect the absolute degree to which propositions are believed in the second phase and the derived beliefs in the third while *belief updates* are quantities which reflect the degree to which beliefs in propositions change in the third.

The basic assumption of the updating procedure says that a belief update $Up(bel, Ev)$ with the evidence Ev , in conjunction with the prior belief bel , is sufficient to determine the posterior belief bel' . Both prior and posterior beliefs represent absolute beliefs. Cox [Cox, 1946] (later updated by Halpern [Halpern, 1999]) has presented a small number of intuitive properties for a measure of absolute belief and proved that any measure that satisfies these properties must be some monotonic transformation of a probability function. A detailed Cox-style justification for quantifying beliefs by belief functions is given in [Smets, 1997].

Another important property concerns the combination of belief updates representing different and independent sources of evidence. The prior belief bel is updated by the first evidence Ev_1 to the posterior bel_1 . This belief update is denoted as $Up(bel, Ev_1)$. Then, the posterior bel_1 is updated in the

conjunction with the second evidence Ev_2 to a new posterior bel_2 . The *composition property* of the belief-update paradigm tells that the *sequential updates* of bel with two pieces of evidence Ev_1 and Ev_2 should be the same as the single update by the *combination* of evidences Ev_1 and Ev_2 . A formal characterization of the belief update paradigm in a probabilistic context is essentially the well-known Bayes Theorem and has been provided in [Heckerman, 1986] where the belief update for a piece of evidence is represented by some monotonic transformation of a likelihood ratio.

3.2 The Framework for Belief Functions

For a binary operator $\otimes : \mathcal{M} \times \mathcal{M} \rightarrow \mathcal{M}$, we are mainly interested in the following three properties:

- (Composition) For any $m, m_{up}^{(1)}$ and $m_{up}^{(2)} \in \mathcal{M}$,

$$(m \otimes m_{up}^{(1)}) \otimes m_{up}^{(2)} = m \otimes (m_{up}^{(1)} \oplus m_{up}^{(2)}) \quad (6)$$

- (Plato's projection) For any $m, m_{up}, m' \in \mathcal{M}$,

$$\begin{array}{ccc} m & \otimes & m_{up} & = & m' \\ \downarrow \text{Betp} & & \downarrow \text{Pl}_P & & \downarrow \text{Betp} \\ pr & \odot & pr_{up} & = & pr' \end{array}$$

- (Proportionality) For any $A, B \subseteq \Omega, m_{up}^1, m_{up}^2 \in \mathcal{M}$,

$$\frac{(m_V \otimes m_{up}^1)(A)}{(m_V \otimes m_{up}^1)(B)} : \frac{m_{up}^1(A)}{m_{up}^1(B)} = \frac{(m_V \otimes m_{up}^2)(A)}{(m_V \otimes m_{up}^2)(B)} : \frac{m_{up}^2(A)}{m_{up}^2(B)}$$

where m_V is the vacuous mass function.

The composition property is simply the belief-function version of the composition property in the general belief-update paradigm. Two distinct pieces of evidence are combined according to Dempster's rule. In the above diagram about Plato's projection property, the first layer is the belief-update paradigm for belief functions and the second layer for probability functions. In the metaphor of Plato's cave, the dynamics of the reality to those of shadows is an analogy of the belief-update framework for belief functions to that for probability functions. The fire represents those transformations. For this metaphor, we call this important property Plato's projection property. The last proportionality property essentially tells us that, if the prior belief is vacuous, then the posterior is proportional to the *relative* strength of the evidence.

In this section, we propose a *new* operator $\circ : \mathcal{M} \times \mathcal{M} \rightarrow \mathcal{M}$ which is *axiomatically justified* by the above three natural properties. For any mass functions m and m_{up} ,

$$(m \circ m_{up})(E) := \frac{\sum_{E=A \cap B} m(A) \frac{|E|}{|A|} m_{up}(B)}{K_\circ} \quad (7)$$

where $K_\circ := \sum_{A \cap B \neq \emptyset} m(A) \frac{|A \cap B|}{|A|} m_{up}(B) \neq 0$. It is not well-defined when $K_\circ = 0$. Note that the operator \circ may be obtained from the well-known Dempster combination \oplus just by replacing $m(A)$ with $m(A) \frac{|E|}{|A|}$. The factor $\frac{|E|}{|A|}$ is used to measure the proportion in A of the intersection subset $E = (A \cap B)$ which is updated by $m_{up}(B)$ and hence quantifies how the focal elements interact as in [Zhang, 1994].

However, these two combinations are essentially different. The operator \circ is not commutative while the Dempster combination \oplus is. Dempster's combination \oplus relies on a basic assumption that the combined mass functions from different pieces of evidence are considered to be on a par, i.e., play the same role [Ma *et al.*, 2011]. The noncommutative property of the operator \circ reflects our motivation that m and m_{up} should be treated differently: m as absolute belief and m_{up} as belief update.

Theorem 3.1 *The operator \circ is the unique binary operator $\otimes : \mathcal{M} \times \mathcal{M} \rightarrow \mathcal{M}$ that satisfies composition, Plato's projection and proportionality properties.*

Proof. First we show the necessary part. For any $E \subseteq \Omega$,

$$\begin{aligned} & [(m \circ m_{up}^{(1)}) \circ m_{up}^{(2)}](E) \\ &= \frac{\sum_{E=B \cap A_2} \frac{|E|}{|B|} \left[\frac{\sum_{B=A \cap A_1} \frac{|B|}{|A|} m(A) m_{up}^{(1)}(A_1)}{K_1} \right] m_{up}^{(2)}(A_2)}{K_2} \\ &= \frac{\sum_{E=A \cap A_1 \cap A_2} \frac{|E|}{|A|} m(A) m_{up}^{(1)}(A_1) m_{up}^{(2)}(A_2)}{K_1} \end{aligned} \quad (8)$$

where K_1 is some normalization factor. On the other hand,

$$\begin{aligned} & [m \circ (m_{up}^{(1)} \oplus m_{up}^{(2)})](E) \\ &= \frac{\sum_{E=A \cap B} \frac{|E|}{|A|} m(A) \left[\frac{\sum_{B=A_1 \cap A_2} \frac{m_{up}^{(1)}(A_1) m_{up}^{(2)}(A_2)}{K_{\oplus}} \right]}{K_{12}} \\ &= \frac{\sum_{E=A \cap A_1 \cap A_2} \frac{|E|}{|A|} m(A) m_{up}^{(1)}(A_1) m_{up}^{(2)}(A_2)}{K_2} \end{aligned} \quad (9)$$

where K_2 is the corresponding normalization factor. Note that both normalization factors are constant and independent of all other parameters. Since the numerators in these two expressions Eqs.(8) and (9) are the same for all E , these expressions are equal to each other: $(m \circ m_{up}^{(1)}) \circ m_{up}^{(2)} = m \circ (m_{up}^{(1)} \oplus m_{up}^{(2)})$. Next we show that the update operator \circ meets the requirement in Plato's projection property. For any $w \in \Omega$,

$$\begin{aligned} \text{Betp}(m')(w) &= \sum_{w \in C} \frac{m'(C)}{|C|} \\ &= \frac{1}{K_{\circ}} \sum_{w \in A} \frac{m(A)}{|A|} \sum_{w \in B} m_{up}(B) \\ &= \text{Betp}(m)(w) \text{Pl}_P(m_{up})(w) / K \end{aligned}$$

for some constants K_{\circ} and K . It is straightforward to check that \circ also satisfies the proportionality property.

Now we prove the sufficient part. Assume that \otimes satisfies the above three properties. First we show a claim:

CLAIM 1 For any $E \subseteq \Omega$, $(m_V \otimes m_{up})(E) = \frac{|E| m_{up}(E)}{K_V}$ where K_V is the normalization factor.

From the proportionality property, we define $k(A)c(m_{up}) := \frac{(m_V \otimes m_{up})(A)}{m_{up}(A)}$ where $k(A)$ and $c(m_{up})$ depends only on A

and m_{up} , respectively. Let m' denote $m_V \otimes m_{up}$. It follows that $\text{Betp}(m')(w) = \sum_{w \in A} \frac{k(A)c(m_{up})}{|A|} m_{up}(A)$. On the other hand, $\text{Pl}_P(m_{up})(w) = \frac{\sum_{w \in A} m_{up}(A)}{K_{Pl}}$ where K_{Pl} is the corresponding normalization factor. From Plato's projection property, we have that $\text{Betp}(m')(w) = \text{Pl}_P(m_{up})(w)$ since the pignistic transformation of the vacuous mass function m_V is the uniform probability distribution. In other words, $\sum_{w \in A} \frac{k(A)c(m_{up})}{|A|} m_{up}(A) = \frac{\sum_{w \in A} m_{up}(A)}{K_{Pl}}$. Since this equality holds for all $m_{up}(A)$, we can conclude that the corresponding coefficients of $m_{up}(A)$ should be equal to each other, namely $\frac{k(A)c(m_{up})}{|A|} = \frac{1}{K_{Pl}}$. Also $K_{Pl} = K_V$. So we have shown the claim.

Now we continue with the sufficiency part. From the claim, we know that, for any mass function m , there is a *unique* mass function $m_{\bar{V}}$ such that $m_V \otimes m_{\bar{V}} = m$: for all $E \neq \emptyset$, $m_{\bar{V}}(E) := m(E) \frac{K_V}{|E|}$. For any $m_{up} \in \mathcal{M}$,

$$\begin{aligned} & (m \otimes m_{up})(E) \\ &= [(m_V \otimes m_{\bar{V}}) \otimes m_{up}](E) \quad (\text{the Claim}) \\ &= [m_V \otimes (m_{\bar{V}} \oplus m_{up})](E) \quad (\text{Composition Property}) \\ &= \frac{|E| (m_{\bar{V}} \oplus m_{up})(E)}{K'} \quad (\text{the Claim}) \\ &= K_V \frac{\sum_{A \cap B = E} \frac{m_{\bar{V}}(A) |A| |E|}{K_V} m_{up}(B)}{K_{\oplus} K'} \\ &= \frac{\sum_{A \cap B = E} m(A) \frac{|E|}{|A|} m_{up}(B)}{K} \quad (\text{the Claim}) \\ &= (m \circ m_{up})(E) \end{aligned}$$

where K' , K_{\oplus} and K are the corresponding normalization factors. Now we have shown the sufficiency part and hence the whole theorem. QED

In our theory, vacuous belief as belief-update plays no role in the combination rules because it maps absolute beliefs (priors) to absolute beliefs (posteriors) while as absolute belief it translates belief-updates (evidence) to absolute beliefs (posteriors). Claim 1 characterizes exactly this translation with vacuous priors. It also says that absolute belief consists of two parts: one is the strength of the supporting evidence and the other the nonspecificity (measured by the cardinality). Our characterization of the operator \circ is totally different from those for Dempster's combination \oplus [Dubois and Prade, 1986] and for Smets' conjunctive combination [Klammann and Smets, 1992]. With the operator \circ defined in Eq. (7), we have established a belief-update framework for belief functions.

It is easy to see that the belief-update operator \circ is consistent with Bayes' rule in the following sense [Shafer, 1982]: if the prior m is a probability distribution, then $m \circ m_{up} = m \oplus m_{up}$ for any mass function m_{up} . If an event C is observed and is represented by the categorical mass function m_C with C being its only focal set, we treat m_C as belief update. Within our framework, we *derive* the following conditioning rule:

$$m(E|C) := (m \circ m_C)(E) = \frac{\sum_{E=C \cap A} \frac{|A \cap C|}{|A|} m(A)}{K} \quad (10)$$

where $K = \sum_{A \cap C \neq \emptyset} \frac{|A \cap C|}{|A|} m(A)$. Our conditioning rule is consistent with pignistic transformation: for any E , $Betp(m(\cdot|C))(E) = Betp(m)(E|C)$ where $\cdot|$ is the Bayesian conditioning.

By assimilating the cardinality in absolute belief m , we define the average commonality function $\bar{Q}_m : 2^\Omega \rightarrow [0, 1]$ as $\bar{Q}_m(A) := \sum_{E \supseteq A} \frac{m(E)}{|E|}$ for all $A \subseteq \Omega$. From the definition of the update operator \circ in Eq. (7), we obtain the following form of the generalized Bayesian theorem for belief functions:

Theorem 3.2 For any $A, B \subseteq \Omega$, and for the absolute belief m and belief-update function m_{up} in Eq. (7), it holds that

$$\frac{\bar{Q}_m(A)}{\bar{Q}_m(B)} \cdot \frac{Q_{m_{up}}(A)}{Q_{m_{up}}(B)} = \frac{\bar{Q}_{m \circ m_{up}}(A)}{\bar{Q}_{m \circ m_{up}}(B)} \quad (11)$$

where $Q_{m_{up}}$ is the commonality function for m_{up} .

From next subsection about statistical model, we will see that, among the above three quotients, $\frac{Q_{m_{up}}(A)}{Q_{m_{up}}(B)}$ is the counterpart of likelihood-ratio in the standard Bayesian theorem, and $\frac{\bar{Q}_m(A)}{\bar{Q}_m(B)}$ and $\frac{\bar{Q}_{m \circ m_{up}}(A)}{\bar{Q}_{m \circ m_{up}}(B)}$ are the counterparts of odds there.

3.3 Belief-function Model of Statistical Evidence

On one hand, belief functions as absolute beliefs are intended to quantify someone's degree of belief and model subjective, personal beliefs, i.e., what the Bayesians claim to be their domain of applications [Shafer, 1976; Smets, 1993]. For example, the vacuous belief is used to model the cognitive state of total ignorance. In Smets' transferable belief models, when a decision must be made, pignistic transformation maps beliefs at the credal level to a probability measure at the pignistic one. On the other hand, viewing belief function as a representation of belief-update essentially amounts to viewing it as a likelihood function representing statistical evidence [Shafer, 1982; Halpern and Fagin, 1992; Walley, 1987; Smets, 1993]. First we recall the standard framework of likelihood representation of statistical evidence in a Bayesian setting [Shafer, 1976; Halpern and Fagin, 1992]. Suppose that Ω consists of *basic hypotheses* $\omega_1, \dots, \omega_n$ and we have another set \mathcal{O} consisting of *basic observations* Ob_1, \dots, Ob_m . Intuitively, we are considering a situation where exactly one of these hypotheses holds, and we are testing which one it is. The basic observations are the data given to us by our tests. We assume that for each basic hypothesis ω_i , we have a probability measure pr_i on $2^\mathcal{O}$. Intuitively, $pr_i(Ob)$ is the belief of observing $Ob (\subseteq \mathcal{O})$ given that the hypothesis w_i holds. The probability $pr_i(Ob)$ is also viewed as the *likelihood* of ω_i given the observation Ob . We would expect that observing Ob would provide more support to ω_i than ω_j if $pr_i(Ob) > pr_j(Ob)$. The corresponding *likelihood function* \mathcal{L}_{Ob} on Ω is defined as $\mathcal{L}_{Ob}(\omega_i) := pr_i(Ob)$. Those likelihoods $\mathcal{L}_{Ob}(\omega_i)$ characterize the support for each hypothesis ω_i by the evidence in Ob . According to the well-known Bayes' theorem, the posterior belief pr' in ω is obtained by normalizing the multiplication of the prior belief

$pr(\omega)$ with the likelihood $\mathcal{L}_{Ob}(\omega)$. In the *Bayesian* setting, we are mainly concerned about the ratios of these likelihoods. So we choose to work on the *probabilized likelihood function* $l_{Ob} : \Omega \rightarrow [0, 1]$ which is defined as: for any $\omega \in \Omega$, $l_{Ob}(\omega) = \frac{\mathcal{L}_{Ob}(\omega)}{\sum_{\omega'} \mathcal{L}_{Ob}(\omega')}$. It is easy to see that l_{Ob} is a probability function and $pr' = pr \circ l_{Ob}$. If Ob^1 and Ob^2 are two *independent* observations, the probabilized likelihood function for the joint observation (Ob^1, Ob^2) is the Bayesian parallel combination of l_{Ob^1} and l_{Ob^2} where l_{Ob^1} and l_{Ob^2} are the probabilized likelihood functions of Ob^1 and Ob^2 , respectively. In other words, $l_{(Ob^1, Ob^2)} = l_{Ob^1} \circ l_{Ob^2}$.

The evidence in Ob can also be represented by a belief function [Smets, 1993]. The parametric model $\{pr_i : i = 1, \dots, n\}$ can be described by the following belief function over $\Omega \times \mathcal{O}$: $bel := \vec{pr}_1 \oplus \dots \oplus \vec{pr}_n$ where $\vec{pr}_1, \dots, \vec{pr}_n$ are the corresponding conditional embedding of pr_1, \dots, pr_n to $\Omega \times \mathcal{O}$, respectively. Let m denote its mass function. Note that bel is a belief function and not necessarily a probability distribution over $\Omega \times \mathcal{O}$. For simplicity, let bel_{Ob} denote $bel(\cdot|\Omega \times \mathcal{O}) \downarrow_\Omega$, which is the marginalization on Ω of the Dempster conditioning of bel with respect to $\Omega \times \mathcal{O}$ and hence is a belief function over Ω . Then the evidence in the observation Ob can be represented by the belief function bel_{Ob} in the sense that the above posterior probability function pr' can be obtained by the Dempster's combination of the prior pr with bel_{Ob} , namely $pr' = pr \oplus bel_{Ob} = pr \circ l_{Ob}$. It is easy to check that $Pl_P(m_{Ob}) = l_{Ob}$ where m_{Ob} is the corresponding mass function. For two *independent* observations Ob^1 and Ob^2 , the belief function representing the joint observation (Ob^1, Ob^2) is the Dempster's combination of those representing the two individual observations Ob^1 and Ob^2 . That is to say, $bel_{(Ob^1, Ob^2)} = bel_{Ob^1} \oplus bel_{Ob^2}$. It is easy to check that $Pl_P(m_{Ob^1} \oplus m_{Ob^2}) = l_{Ob^1} \circ l_{Ob^2}$ where m_{Ob^1} and m_{Ob^2} are the corresponding mass functions. So plausibility transformation connects these two equivalent representations of evidence in observations: one is the probabilized likelihood function l_{Ob} and the other is the mass function m_{Ob} .

More generally, the same representational approach applies to the evidence on observations in a parametric model $\{bel_i : 1 \leq i \leq n\}$ where bel_i is a conditional belief function. Similarly, this general parametric model $\{bel_i : 1 \leq i \leq n\}$ can be described by a belief function Bel over $\Omega \times \mathcal{O}$: $Bel := \vec{bel}_1 \oplus \dots \oplus \vec{bel}_n$ where $\vec{bel}_1, \dots, \vec{bel}_n$ are their corresponding conditional embeddings. The evidence in an observation Ob is represented by $(Bel(\cdot|\Omega \times \mathcal{O})) \downarrow_\Omega$. For simplicity, we denote this restriction over Ω by Bel_{Ob} . For two *independent* observations Ob^1 and Ob^2 , the belief function $Bel_{(Ob^1, Ob^2)}$ representing the joint observation (Ob^1, Ob^2) is the Dempster's combination of the belief functions Bel_{Ob^1} and Bel_{Ob^2} representing the two individual observations Ob^1 and Ob^2 . Let m_0 denote a prior as absolute belief and m_{Ob} the mass function representing evidence in Ob in the parametric model $\{bel_i : 1 \leq i \leq n\}$. We obtain the posterior m' by combining m_0 and m_{Ob} with the update operator \circ , i.e., $m' = m_0 \circ m_{Ob}$. The Plato's projection property in Theorem 3.1 tells us that this equality induces the well-known Bayes' theorem under the two probability transformations.

Example 3.3 Here we adapt a simple example from [Shafer, 1982]. Consider the case that $\Omega = \{\theta, \bar{\theta}\}$ and $\mathcal{O} = \{x, \bar{x}\}$. Assume that $pr_{\theta}(x) = \frac{1}{2}$ and $pr_{\bar{\theta}}(x) = \frac{1}{2}$. The mass function $m_{\Omega \times \mathcal{O}}$ representing the parametric model $\{pr_{\theta}, pr_{\bar{\theta}}\}$ is obtained by combining the conditional embeddings of pr_{θ} and $pr_{\bar{\theta}}$ according to Dempster’s rule. Its focal elements are $\{(\theta, x), (\bar{\theta}, x)\}, \{(\theta, x), (\bar{\theta}, \bar{x})\}, \{(\theta, \bar{x}), (\bar{\theta}, x)\}$ and $\{(\theta, \bar{x}), (\bar{\theta}, \bar{x})\}$. All of their masses are the same and are equal to $\frac{1}{4}$. The mass function m_x representing the evidence in the observation x can be obtained by conditioning $m_{\Omega \times \mathcal{O}}$ on x (actually on $\Omega \times \{x\}$) and marginalizing on Ω : $m_x(\{\theta\}) = m_x(\{\bar{\theta}\}) = m_x(\{\theta, \bar{\theta}\}) = \frac{1}{3}$. Assume that the prior is the vacuous m_V . Our difference from those in the literature [Shafer, 1982; Smets, 1993; Halpern and Fagin, 1992; Denoeux, 2014] lies in the rule combining the prior m_V with the mass function representing the evidence in the observation x ; in this paper, we combine them according to the newly proposed update operator \circ while, in the previous approaches, they are combined according to Dempster’s (or conjunctive) rule of combination. In the following table, one can find a comparison of our result with those using Dempster’s combination:

	$m_V \oplus m_x$	$m_V \circ m_x$
$\{\theta\}$	$\frac{1}{3}$	$\frac{1}{4}$
$\{\bar{\theta}\}$	$\frac{1}{3}$	$\frac{1}{4}$
$\{\theta, \bar{\theta}\}$	$\frac{1}{3}$	$\frac{1}{2}$

In the parametric model $\{pr_i : 1 \leq i \leq n\}$ where pr_i are conditional probabilities, if the prior is a probability function, then we may employ Walley’s general framework [Walley, 1987] to show that all belief-function representations of evidence, including our belief-update framework, produce the same posterior as what is obtained in the Bayesian setting. However, the above example shows that, when the prior is a belief function and not necessarily a probability function, the posterior derived in our framework is not a probability function and hence may not be obtained by the well-known Bayes’ theorem. So the generalization of the belief-update framework from probability functions to belief functions does provide some extra power over the Bayesian approach.

4 Related Works and Conclusion

In the literature, there are several proposals for belief function representations of statistical evidence. [Shafer, 1976] and [Denoeux, 2014] considered consonant belief functions to capture the evidence in an observation Ob while [Halpern and Fagin, 1992] provided a probabilistic representation, which is closely related to plausibility transformation. All of them *axiomatically* justified their representations by assuming the *strong* likelihood principle. In this paper, we follow Smets’ approach [Smets, 1993] by assuming the *weak* likelihood principle to represent the evidence with a belief function which is not necessarily a consonant belief or probability function. Walley set up a more general framework for belief-function representations and questioned the use of Dempster’s rule for combining the prior with evidence as in this paper [Walley, 1987].

Halpern and Fagin distinguished two different views of belief functions: belief as generalized probability and belief as

evidence, which is similar to our distinction between absolute belief and belief update [Halpern and Fagin, 1992]. They also considered the link between these two views. They interpreted generalized probabilities as envelopes, and evidence as a mapping from probability functions to probability functions. But the problem with this interpretation is the case when the prior *bel* is a belief function which is not necessarily a probability function. Let pr_{Ev} be a probability function representing the evidence. The lower envelope of the Dempster’s combination of the evidence pr_{Ev} with probability functions dominating the prior *bel* is not necessarily a belief function. [Yen, 1986] and [Zhou and Feng, 2015] also distinguished two different treatments of beliefs in the belief-update framework by [Heckerman, 1986] as in this paper. [Heckerman, 1985] made a similar distinction for certainty factors. But their interpretations of belief update are totally different from our likelihood-based semantics: they presented the framework in random set theory or probability theory. Our derived conditioning rule (Eq. (10)) fits well with the conditioning approach in [Dubois and Denoeux, 2012; Zhou and Feng, 2015] proposing a tradeoff between prediction and revision. The distinction between absolute belief and belief update is also present in the classical framework of probability distributions [Pearl, 1988]. Theorem 3.2 is the generalization of the likelihood-odds form of Bayesian theorem there.

One may refer to [Cuzzolin, 2007; 2012; 2015] for a detailed account of many different probability transformations in the literature. Cuzzolin classified them into two major classes: those commuting with affine combination and those commuting with Dempster’s rule [Cuzzolin, 2012]. Pignistic and plausibility transformations are the two most well-known among these two classes. Pignistic transform was first proposed by [Dubois and Prade, 1982] (not with this name, though, as the name is due to Smets). Plausibility transform was first studied for transforming possibility into probability in [Dubois and Prade, 1980]. Plausibility transformation for belief functions was proposed in [Voorbraak, 1989] and supported by [Cobb and Shenoy, 2006]. [Wilson, 1993] considered pignistic transformation in decision-making. But none of them has studied the connections between these two transformations. The results in this paper can be easily generalized to connect those probability transformations commuting with affine combinations and with Dempster’s rule. It may be interesting to investigate the connection of the framework in this paper with the more general imprecise-probability representation of statistical evidence [Walley, 1990].

Acknowledgments

The first author is partly supported by Key project for basic research from the Ministry of Science and Technology of China (Grant No. 2012CB316205) and NSF of China (Grant No. 61370053). The second author is partly supported by NSF of China (Grant No. 61472425). The third author is partly supported by Key project for basic research from the Ministry of Science and Technology of China (Grant No. 2012CB316205) and NSF of China (Grant No. 61432006).

References

- [Cobb and Shenoy, 2006] Barry Cobb and Prakash Shenoy. On the plausibility transformation method for translating belief function models to probability models. *Int. J. Approx. Reasoning*, 41(3):314–330, 2006.
- [Cox, 1946] Richard Cox. Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1):1–13, 1946.
- [Cuzzolin, 2007] Fabio Cuzzolin. Two new bayesian approximations of belief functions based on convex geometry. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 37(4):993–1008, 2007.
- [Cuzzolin, 2012] Fabio Cuzzolin. On the relative belief transform. *Int. J. Approx. Reasoning*, 53(5):786–804, 2012.
- [Cuzzolin, 2015] Fabio Cuzzolin. *Geometry of Uncertainty*. Springer, 2015. to appear.
- [Denoeux, 2014] Thierry Denoex. Likelihood-based belief function: Justification and some extensions to low-quality data. *Int. J. Approx. Reasoning*, 55(7):1535–1547, 2014.
- [Dubois and Denoex, 2012] Didier Dubois and Thierry Denoex. Conditioning in dempster-shafer theory: Prediction vs. revision. In *Belief Functions*, pages 385–392, 2012.
- [Dubois and Prade, 1980] Didier Dubois and Henri Prade. *Fuzzy sets and systems: theory and applications*, volume 144 of *Mathematics in science and engineering*. Acad. Press, 1980.
- [Dubois and Prade, 1982] Didier Dubois and Henri Prade. On several representations of an uncertain body of evidence. In E. Sanchez M.M. Gupta, editor, *Fuzzy Information and Decision Processes*, pages 167–181, 1982.
- [Dubois and Prade, 1986] Didier Dubois and Henri Prade. On the unicity of dempster rule of combination. *Int. J. Intell. Syst.*, 1(2):133–142, 1986.
- [Halpern and Fagin, 1992] Joseph Halpern and Ron Fagin. Two views of belief: Belief as generalized probability and belief as evidence. *Artif. Intell.*, 54(2):275–317, 1992.
- [Halpern, 1999] Joseph Halpern. A counterexample to theorems of cox and fine. *J. Artif. Intell. Res.*, 10:67–85, 1999.
- [Heckerman, 1985] David Heckerman. Probabilistic interpretation for mycin’s certainty factors. In *UAI ’85*, pages 167–196, 1985.
- [Heckerman, 1986] David Heckerman. An axiomatic framework for belief updates. In *UAI ’86*, pages 11–22, 1986.
- [Klawonn and Smets, 1992] Frank Klawonn and Philippe Smets. The dynamic of belief in the transferable belief model and specialization-generalization matrices. In *UAI ’92, Stanford, CA, USA, July 17-19, 1992*, pages 130–137, 1992.
- [Ma et al., 2011] Jianbing Ma, Weiru Liu, Didier Dubois, and Henri Prade. Bridging jeffrey’s rule, agm revision and dempster conditioning in the theory of evidence. *International J. on AI Tools*, 20 (4):691–720, 2011.
- [Pearl, 1988] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems - Networks of Plausible Inference*. Morgan Kaufmann series in representation and reasoning. Morgan Kaufmann, 1988.
- [Shafer, 1976] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, N.J., 1976.
- [Shafer, 1982] Glenn Shafer. Belief function and parametric models (with discussion). *J. Roy. Statist. Soc. Ser. B*, 23:322–352, 1982.
- [Smets and Kennes, 1994] Philippe Smets and Robert Kennes. The transferable belief model. *Artif. Intell.*, 66(2):191–234, 1994.
- [Smets, 1993] Philippe Smets. Belief functions: The disjunctive rule of combination and the generalized bayesian theorem. *Int. J. Approx. Reasoning*, 9(1):1–35, 1993.
- [Smets, 1997] Philippe Smets. The normative representation of quantified beliefs by belief functions. *Artif. Intell.*, 92(1-2):229–242, 1997.
- [Smets, 2005] Philippe Smets. Decision making in the tbm: the necessity of the pignistic transformation. *Int. J. Approx. Reasoning*, 38(2):133–147, 2005.
- [Voorbraak, 1989] Frans Voorbraak. A computationally efficient approximation of dempster-shafer theory. *International Journal of Man-Machine Studies*, 30(5):525–536, 1989.
- [Walley, 1987] Peter Walley. Belief function representations of statistical evidence. *The Annals of Statistics*, 15(4):1439–1465, 1987.
- [Walley, 1990] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1990.
- [Wilson, 1993] Nic Wilson. Decision-making with belief functions and pignistic probabilities. In *ECSQARU’93*, pages 364–371, 1993.
- [Yen, 1986] John Yen. A reasoning model based on an extended dempster-shafer theory. In *AAAI’86*, pages 125–131, 1986.
- [Zhang, 1994] Lianwen Zhang. Representation, independence, and combination of evidence in the dempster-shafer theory. In J. Kacprzyk R. R. Yager and M. Fedrizzi., editors, *Advances in Fuzzy Set Theory and Applications*, pages 51–69. John Wiley and Sons, Inc., New York, 1994.
- [Zhou and Feng, 2015] Chunlai Zhou and Yuan Feng. Extending transferable belief models with probilistic priors. In *UAI*, pages 962–971, 2015.