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

Chunlai Zhou1,2, 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 transformations work 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 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

---

1Here we adapt the metaphor from Judea Pearl who used it to explain a similar relationship of causal reasoning to probabilistic inference.
referred 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 combing 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 \( \Omega \) be a frame of discernment and \( A = 2^\Omega \) be the Boolean algebra of propositions. \( |A| \) denotes the cardinality of a subset \( A \). A mass assignment (or mass function) over \( \Omega \) is a mapping \( m : A \rightarrow [0, 1] \) satisfying \( \sum_{A \in A} m(A) = 1 \). 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 \( bel : A \rightarrow [0, 1] \) satisfying the conditions: \( bel(\emptyset) = 0 \), \( bel(\Omega) = 1 \) and \( bel(\bigcup_{i=1}^n A_i) \geq \sum_{i \neq j \subseteq \{1, \ldots, n\}} (-1)^{|i|+1}bel(\cap_{i \in i} A_i) \) where \( A_i \in A \) for all \( i \in \{1, \ldots, n\} \). A mapping \( f : A \rightarrow [0, 1] \) is a belief function if and only if its Möbius transform is a mass assignment [Shafer, 1976]. In other words, if \( m : A \rightarrow [0, 1] \) is a mass assignment, then it determines a belief function \( bel : A \rightarrow [0, 1] \) as follows: \( bel(A) = \sum_{B \subseteq A} m(B) \) for all \( A \in A \). Moreover, given a belief function \( bel \), we can obtain its corresponding mass function \( m \) as follows: \( m(A) = \sum_{B \subseteq A} (-1)^{|A-B|}bel(B) \) for all \( A \in A \). 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 \( \Omega \) is the only focal element and its mass function is denoted by \( m_\emptyset \). 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 \( pl_m(A) = \sum_{E \supseteq A \emptyset} m(E) \) 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, \( m_1 \) and \( m_2 \) are two mass functions on \( \Omega \) induced by two independent evidential sources, the combined mass function is calculated according to Dempster’s rule of combination: for any \( C \subseteq \Omega \),

\[
(m_1 \oplus m_2)(C) = \frac{\sum_{A \cap B = C} m_1(A)m_2(B)}{\sum_{A \cap B \neq C} m_1(A)m_2(B)}
\]  

(1)

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

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

\[
(p_1 \ominus p_2)(\omega) := \frac{p_1(\omega)p_2(\omega)}{\sum_{\omega \in \Omega} p_1(\omega)p_2(\omega)} \text{ for all } \omega \in \Omega.
\]

(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 \( \tau \) from \( \mathcal{M} \) to \( \mathcal{P} \) is called a probability transformation. A probability transformation \( \tau \) is called commuting with affine combination if for any mass functions \( m_1, \ldots, m_n, \tau(\sum_{i=1}^n a_im_i) = \sum_{i=1}^n a_i\tau(m_i) \) for all nonnegative \( a_1, \ldots, a_n \) 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 \( m_2 \), \( \tau(m_1 \oplus m_2) \) = \( \tau(m_1) \ominus \tau(m_2) \).

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

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

(3)

Note that \( Betp(m) \) is a probability function on \( \Omega \) 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 \( PlP : \mathcal{M} \rightarrow \mathcal{P} \) is given by: for any \( \omega \in \Omega \),

\[
PlP(m)(\omega) := \frac{pl(\omega)}{\sum_{\omega \in \Omega} pl(\omega)}
\]

(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 \( m_2 \),

\[
PL_P(m_1 \oplus m_2) = PL_P(m_1) \odot PL_P(m_2)
\] (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 \( \oplus : \mathcal{M} \times \mathcal{M} \rightarrow \mathcal{M} \), we are mainly interested in the following three properties:

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

\[
(m \oplus m^{(1)}_{up}) \oplus m^{(2)}_{up} = m \oplus (m^{(1)}_{up} \oplus m^{(2)}_{up})
\] (6)

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

\[
\begin{align*}
pr & \quad \odot \quad pr_{up} \quad = \quad pr' \\
Betp & \quad \odot \quad PL_P \quad = \quad Betp'
\end{align*}
\]

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

\[
\frac{m_v \oplus m^1_{up}(A)}{m_v \oplus m^2_{up}(B)} = \frac{m_v \oplus m^1_{up}(A)}{m_v \oplus m^2_{up}(B)}
\]

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 paradigm for belief functions and the second layer 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 \( m_{up} \),

\[
(m \circ m_{up})(E) := \frac{\sum_{E=A \cap \bar{A}} m(A) |E| \cdot m_{up}(B)}{K_0}
\] (7)

where \( K_0 := \sum_{A \cap B \neq \emptyset} m(A) |A \cap B| \cdot m_{up}(B) \neq 0 \). It is not well-defined when \( K_0 = 0 \). Note that the operator \( \circ \) may be obtained from the well-known Dempster combination \( \oplus \) just by replacing \( m(A) \) with \( m(A) |E| \cdot |A| \). The factor \( |E| \cdot |A| \) 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 \( \oplus : \mathcal{M} \times \mathcal{M} \rightarrow \mathcal{M} \) that satisfies composition, Plato’s projection and proportionality properties.

**Proof.** First we show the necessary part. For any \( E \subseteq \Omega \),

\[
((m \circ m_{up}^{(1)}) \circ m_{up}^{(2)})(E) = \sum_{E = B \cap A_2} \frac{\sum_{B = A \cap A_1} |E| |A_1| m(A) m_{up}^{(1)}(A_1) m_{up}^{(2)}(A_2)}{K_2} \]

\[
= \sum_{E = A \cap A_1 \cap A_2} \frac{|E| |A_1| m(A) m_{up}^{(1)}(A_1) m_{up}^{(2)}(A_2)}{K_1}
\]

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

\[
\frac{m \circ (m_{up}^{(1)} \oplus m_{up}^{(2)})}{m_{up}^{(1)}}(E) = \sum_{E = A \cap B} \frac{|E| |A| m(A) (\sum_{B = A \cap A_1} |A_1| m_{up}^{(1)}(A_1) m_{up}^{(2)}(A_2))}{K_2}
\]

\[
= \sum_{E = A \cap A_1 \cap A_2} \frac{|E| |A_1| m(A) m_{up}^{(1)}(A_1) m_{up}^{(2)}(A_2)}{K_2}
\]

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 \( w \in \Omega \),

\[
Betp(m')(w) = \sum_{w \subseteq C} \frac{m'(C)}{|C|} = \frac{1}{K_\circ} \sum_{w \subseteq A} \frac{m(A)}{|A|} \sum_{w \subseteq B} m_{up}(B)
\]

\[
= Betp(m)(w) Pl_P(m_{up})(w)/K
\]

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 \( \circ \) satisfies the above three properties. First we show a claim:

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

From the proportionality property, we define \( k(A)c(m_{up}) := \frac{(m_V \oplus m_{up})(A)}{m_{up}(A)} \) where \( k(A) \) and \( c(m_{up}) \) depends only on \( A \) and \( m_{up} \), respectively. Let \( m' \) denote \( m_V \oplus m_{up} \). It follows that \( Betp(m')(w) = \sum_{w \subseteq A} \frac{k(A)c(m_{up})}{|A|} m_{up}(A) \). On the other hand, \( Pl_P(m_{up})(w) = \sum_{w \subseteq A} m_{up}(A) / K_{Pl} \) where \( K_{Pl} \) is the corresponding normalization factor. From Plato’s projection property, we have that \( Betp(m')(w) = Pl_P(m_{up})(w) \) since the pignistic transformation of the vacuous mass function \( m_V \) is the uniform probability distribution. In other words, \( \sum_{w \subseteq A} \frac{k(A)c(m_{up})}{|A|} m_{up}(A) = \sum_{w \subseteq A} m_{up}(A) / K_{Pl} \). 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 \( k(A)c(m_{up}) = \frac{1}{K_{Pl}} \). 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 \( m_\circ \) such that \( m_V \oplus m_\circ = m \) : for all \( E \neq \emptyset \), \( m_\circ(E) := \frac{m(E)}{K_V} \). For any \( m_{up} \in \mathcal{M} \),

\[
\frac{(m \circ m_{up})(E)}{m_{up}(E)} = \frac{(m_V \oplus m_{up})(E)}{m_{up}(E)} \quad \text{(the Claim)}
\]

\[
= \frac{[m_V \oplus (m \circ m_{up})](E)}{m_{up}(E)} \quad \text{(Composition Property)}
\]

\[
= \frac{E[(m_V \oplus m_{up})](E)}{K'} \quad \text{(the Claim)}
\]

\[
= K \sum_{A \subseteq B \subseteq E} \frac{|E| |A| m_{up}(B)}{K_{\circ} K} \quad \text{(the Claim)}
\]

\[
= (m \circ m_{up})(E)
\]

where \( K', K_{\circ} \) 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 [Klawonn 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:

\[
m[E|C] := (m \circ m_E)(E) = \sum_{E = C \cap A} \frac{|A| m(A)}{K} m_{up}(A)
\]

(10)
where \( K = \sum_{A \in C \neq \emptyset} \frac{|A \cap C|}{|A|} m(A) \). Our conditioning rule is consistent with pignistic transformation: for any \( E \), Betm(\( m(\cdot | C) \))|\( E \) = Betm(\( m(\cdot | C) \))|\( E \) where \( \cdot \) is the Bayesian conditioning.

By assimilating the cardinality in absolute belief \( m \), we define the average commonality function \( \hat{Q}_m : 2^\Omega \rightarrow [0, 1] \) as \( \hat{Q}_m(A) := \sum_{E \supseteq A} \frac{m(E)}{|E|} \) 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

\[
\frac{Q_m(A)}{Q_m(B)} \cdot \frac{Q_{m_{up}}(A)}{Q_{m_{up}}(B)} = \hat{Q}_{m_{up}}(A) \tag{11}
\]

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, \( \frac{Q_m(A)}{Q_m(B)} \) is the counterpart of likelihood-ratio in the standard Bayesian theorem, and \( \frac{Q_m(A)}{Q_m(B)} \cdot \frac{Q_{m_{up}}(A)}{Q_{m_{up}}(B)} \) 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 \( \Omega \) consists of basic hypotheses \( \omega_1, \cdots, \omega_n \) and we have another set \( O \) consisting of basic observations \( O_{b1}, \cdots, O_{bm} \). 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 \( \omega_i \), we have a probability measure \( pr_i \) on \( 2^\Omega \). Intuitively, \( pr_i(O) \) is the belief of observing \( O \in \Omega \) given that the hypothesis \( \omega_i \) holds. The probability \( pr_i(O) \) is also viewed as the likelihood of \( \omega_i \) given the observation \( O \). We would expect that observing \( O \) would provide more support to \( \omega_i \) than \( \omega_j \) if \( pr_i(O) > pr_j(O) \). The corresponding likelihood function \( L_{O_{b1}}(\omega_i) \) on \( \Omega \) is defined as \( L_{O_{b1}}(\omega_i) := pr_i(O) \). Those likelihoods \( L_{O_{b1}}(\omega_i) \) characterize the support for each hypothesis \( \omega_i \) by the evidence in \( O \). According to the well-known Bayes’ theorem, the posterior belief \( pr_i^{*} \) in \( \omega \) is obtained by normalizing the multiplication of the prior belief \( pr(\omega) \) with the likelihood \( L_{O_{b1}}(\omega) \). 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 \( \omega \in \Omega \),

\[
l_{ob}(\omega) = \frac{L_{O_{b1}O_{b2}}(\omega)}{\sum_{\omega'} L_{O_{b1}O_{b2}}(\omega')}. \tag{11}
\]

It is easy to see that \( l_{ob} \) is a probability function and \( pr_i^{*} = pr_i \odot l_{ob} \). If \( O_{b1} \) and \( O_{b2} \) are two independent observations, the probabilized likelihood function for the joint observation \((O_{b1}, O_{b2}) \) 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 \( O_{b1} \) and \( O_{b2} \), respectively. In other words, \( l_{(O_{b1},O_{b2})} = l_{ob_1} \odot l_{ob_2} \).

The evidence in \( O \) can also be represented by a belief function [Smets, 1993]. The parametric model \{\( pr_i : i = 1, \cdots, n \)\} can be described by the following belief function over \( \Omega \times O \): \( bel := \hat{p}_{O_1} \oplus \cdots \oplus \hat{p}_{O_n} \) where \( \hat{p}_{O_1}, \cdots, \hat{p}_{O_n} \) are the corresponding conditional embedding of \( pr_1, \cdots, pr_n \) to \( \Omega \times O \), respectively. Let \( m \) denote its mass function. Note that \( bel \) is a belief function and not necessarily a probability distribution over \( \Omega \times O \). For simplicity, let \( bel_{O} \) denote \( bel[: \Omega \times O] \downarrow_{\Omega} \), which is the marginalization on \( \Omega \) of the Dempster conditioning of \( bel \) with respect to \( \Omega \times O \) and hence is a belief function over \( \Omega \). Then the evidence in the observation \( O \) can be represented by the belief function \( bel_{O} \) in the sense that the above posterior probability function \( pr_i^{*} \) can be obtained by the Dempster’s combination of the prior \( pr \) with \( bel_{O} \), namely \( pr_i^{*} = pr \odot bel_{O} = pr \odot l_{ob} \). It is easy to check that \( Pl_{P}(m_{ob}) = l_{ob} \) where \( m_{ob} \) is the corresponding mass function. For two independent observations \( O_{b1} \) and \( O_{b2} \), the belief function representing the joint observation \((O_{b1}, O_{b2}) \) is the Dempster’s combination of those representing the two individual observations \( O_{b1} \) and \( O_{b2} \). That is to say, \( bel((O_{b1}, O_{b2})) = bel_{O_{b1}} \odot bel_{O_{b2}} \). It is easy to check that \( Pl_{P}(m_{ob_1} + m_{ob_2}) = l_{ob_1} \odot l_{ob_2} \) 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 \( \Omega \times O \):

\[
Bel := \hat{bel}_{1} \oplus \cdots \oplus \hat{bel}_{n} \text{ where } \hat{bel}_{1}, \cdots, \hat{bel}_{n} \text{ are their corresponding conditional embeddings. The evidence in an observation } O \text{ is represented by } (Bel[: \Omega \times O]) \downarrow_{\Omega}. \]

For two independent observations \( O_{b1} \) and \( O_{b2} \), the belief function \( Bel((O_{b1}, O_{b2})) \) representing the joint observation \((O_{b1}, O_{b2}) \) is the Dempster’s combination of the belief functions \( Bel_{O_{b1}} \) and \( Bel_{O_{b2}} \) representing the two individual observations \( O_{b1} \) and \( O_{b2} \). Let \( m_0 \) denote a prior as absolute belief and \( m_{ob} \) the mass function representing evidence in \( O \) 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 $pr_{\theta}$ and $pr_{\bar{\theta}}$ according to Dempster’s rule. Its focal elements are $\{(\theta, x), (\bar{\theta}, x), (\theta, \bar{x}), (\bar{\theta}, \bar{x})\}$ and $\{(\theta, \bar{x}), (\bar{\theta}, \bar{x})\}$. All of their masses are the same and are equal to $\frac{1}{2}$. 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 $\Omega$: $m_{x}(\{\theta\}) = m_{x}(\{\bar{\theta}\}) = m_{x}(\{\theta, \bar{\theta}\}) = \frac{1}{2}$. Assume that the prior is the vacuous mass function representing the evidence in the observation $x$. In this paper, we combine them according to the newly proposed update operator $\odot$ 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:

<table>
<thead>
<tr>
<th>${\theta}$</th>
<th>${\bar{\theta}}$</th>
<th>${\theta, \bar{\theta}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
</tbody>
</table>

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


