On the relative belief transform

Fabio Cuzzolin [a]

[a] *Department of Computing and Communication Technologies*
Oxford Brookes University
Wheatley campus, Oxford OX33 1HX, United Kingdom

Abstract

In this paper we discuss the semantics and properties of the relative belief transform, a probability transformation of belief functions closely related to the classical plausibility transform. We discuss its rationale in both the probability-bound and Shafer's interpretations of belief functions. Even though the resulting probability (as it is the case for the plausibility transform) is not consistent with the original belief function, an interesting rationale in terms of optimal strategies in a non-cooperative game can be given in the probability-bound interpretation to both relative belief and plausibility of singletons. On the other hand, we prove that relative belief commutes with Dempster's orthogonal sum, meets a number of properties which are the duals of those met by the relative plausibility of singletons, and commutes with convex closure in a similar way to Dempster's rule. This supports the argument that relative plausibility and belief transform are indeed naturally associated with the D-S framework, and highlights a classification of probability transformations in two families, according to the operator they relate to. Finally, we point out that relative belief is only a member of a class of "relative mass" mappings, which can be interpreted as low-cost proxies for both plausibility and pignistic transforms.

Key words: *Theory of evidence, probability transformation, relative plausibility and belief of singletons, duality, Dempster's combination, commutativity.*

---

1 Introduction

The theory of evidence [38] extends classical probability theory through the notion of *belief function*, a mathematical entity which independently assigns probability values to *sets* of possibilities rather than single events. A belief

Email address: `fabio.cuzzolin@brookes.ac.uk` (Fabio Cuzzolin).
URL: `http://cms.brookes.ac.uk/staff/FabioCuzzolin/` (Fabio Cuzzolin).

function $b : 2^\Theta \rightarrow [0,1]$ on a finite set or *frame* Θ has the form $b(A) = \sum_{B \subseteq A} m_b(B)$, where the function $m_b : 2^\Theta \rightarrow [0,1]$ (called *basic probability assignment* or *basic belief assignment* b.b.a.) is both non-negative $m_b(A) \geq 0$ $\forall A \subseteq \Theta$ and normalized $\sum_{A \subseteq \Theta} m_b(A) = 1$. Events A associated with non-zero basic probabilities $m_b(A) \neq 0$ are called *focal elements*. A basic probability assignment m_b can be uniquely recovered from a belief function b by Moebius transform: $m_b(A) = \sum_{B \subseteq A} (-1)^{|A-B|} b(B)$. Special belief functions assigning non-zero masses to singletons only ($m_b(A) = 0$ whenever $|A| > 1$, $A \subseteq \Theta$) are called *Bayesian* belief functions, and are in 1-1 correspondence with probability distributions on Θ. Different operators have been proposed for the combination of two or more belief functions, starting from the orthogonal sum originally formulated by A. Dempster [19,18].
Belief functions possess a number of alternative semantics in terms of multi-valued mappings [40], random sets [35,29], inner measures [36,26], transferable beliefs [45] or hints [32]. One such interpretation is based on the fact that a belief function corresponds to a set of upper and lower bounds to the values of probability measures on Θ, which in turn determine a convex set $\mathcal{P}[b]$ of such probabilities (often called *consistent* with b):

$$\mathcal{P}[b] \doteq \Big\{ p \in \mathcal{P} : b(A) \leq p(A) \leq pl_b(A) \; \forall A \subseteq \Theta \Big\}, \qquad (1)$$

where the plausibility function $pl_b : 2^\Theta \rightarrow [0,1]$, $pl_b(A) = 1 - b(A^c) = \sum_{B \cap A \neq \emptyset} m_b(B)$ carries the same evidence as b. Such interpretation has been criticized as incompatible with Dempster's rule of combination [52].
On the other hand, in the original model in which belief functions are induced by multi-valued mappings of probability distributions, Dempster's conditioning can be judged inappropriate from a Bayesian point of view. In order to deal with such criticism, in his "Transferable Belief Model" (TBM) [42,45] Smets abandons all notions of multivalued mapping to define belief directly in terms of basis belief assignments ("credal" level).

1.1 Probability transformation of belief functions

The relation between belief and probability, in particular, has been an important subject of study in the theory of evidence. Given a frame of discernment Θ, let us denote by $\mathcal{B}$ the set of all belief functions on Θ, and by $\mathcal{P}$ the set of all probability distributions on Θ. According to [17], we call a *probability transform* of belief functions an operator $pt : \mathcal{B} \rightarrow \mathcal{P}$, $b \mapsto pt[b]$ mapping belief measures onto probability distributions, such that $b(x) \leq pt[b](x) \leq pl_b(x) = 1 - b(\{x\}^c)$. Note that such definition requires the probability which results from the transform to be compatible with the upper and lower bounds the original belief function b enforces *on the singletons only*, and not on all the focal sets as in Equation (1). This is a minimal,

sensible constraint which does not require probability transforms to adhere to the upper-lower probability semantic of belief functions.
A number of papers have been published on the issue of probability transform [53,33,1,55,22,23,28]. Many of these proposals seek efficient implementations of the rule of combination. Tessem [49], for instance, incorporates only the highest-valued focal elements in his m_{klx} approximation. A similar approach inspires the summarization technique formulated by Lowrance *et al.* [34].
A different, decision based approach to probability transformation is at the foundation of the TBM, where decisions are made via the *pignistic probability*

$$BetP[b](x) = \sum_{A \supseteq \{x\}} \frac{m_b(A)}{|A|}, \tag{2}$$

generated by what he calls the *pignistic transform*: $BetP : \mathcal{B} \rightarrow \mathcal{P}$, $b \mapsto BetP[b]$. Initially justified by the Principle of Insufficient Reason, the pignistic probability is the result of a redistribution process in which the mass of each focal element A is re-assigned to all its elements $x \in A$ on an equal basis, and is perfectly compatible with the upper-lower probability semantics of belief functions, as it is the center of mass of the polytope (1) of consistent probabilities [4].
Other proposals have been recently brought forward by Dezert et al. [24], Burger [3], Sudano [48] and others, based on redistribution processes similar to that of the pignistic transform. New Bayesian approximations of belief functions, such as the orthogonal projection of a belief function b onto the probability simplex:

$$\pi[b](x) = \sum_{A \supseteq \{x\}} m_b(A) \Big(\frac{1 + |A^c| 2^{1-|A|}}{n} \Big) + \sum_{A \not\supseteq \{x\}} m_b(A) \Big(\frac{1 - |A| 2^{1-|A|}}{n} \Big) \tag{3}$$

have been derived from purely geometric considerations [9] in the context of the geometric approach to the theory of evidence [11].

1.2 Relative plausibility and belief transforms

Originally developed by Voorbraak [51] as a probabilistic approximation intended to limit the computational cost of operating with belief functions in the Dempster-Shafer framework, the *plausibility transform* [5] has later been supported by Cobb and Shenoy in virtue of its commutativity properties with respect to Dempster's sum. Even though initially defined in terms of commonality values, the plausibility transform $\tilde{pl} : \mathcal{B} \rightarrow \mathcal{P}$, $b \mapsto \tilde{pl}[b]$ maps each belief function b onto the probability distribution $\tilde{pl}[b] = \tilde{pl}_b$ obtained by normalizing the plausibility values $pl_b(x)$ [1] of the element of Θ:

$$\tilde{pl}_b(x) = \frac{pl_b(x)}{\sum_{y \in \Theta} pl_b(y)}. \tag{4}$$

We call the output $\tilde{pl}_b$ (4) of the plausibility transform *relative plausibility of singletons*. Voorbraak proved that the latter is a perfect representative of b when combined with other probabilities $p \in \mathcal{P}$ through Dempster's rule $\oplus$:

$$\tilde{pl}_b \oplus p = b \oplus p \quad \forall p \in \mathcal{P}. \tag{5}$$

Dually, a *relative belief transform* $\tilde{b} : \mathcal{B} \rightarrow \mathcal{P}$, $b \mapsto \tilde{b}[b]$ mapping each belief function to the corresponding *relative belief of singletons* $\tilde{b}[b] = \tilde{b}$ [10,13,27,17]

$$\tilde{b}(x) = \frac{b(x)}{\sum_{y \in \Theta} b(y)} \tag{6}$$

can be defined. Unlike the relative plausibility of singletons, however, $\tilde{b}[b]$ exists iff b assigns some mass to singleton focal sets:

$$\sum_{x \in \Theta} m_b(x) \neq 0. \tag{7}$$

The notion of relative belief transform (under the name of "normalized belief of singletons") has first been proposed by Daniel [17]. Some preliminary analyses of the relative belief transform and its close relationship with the (relative) plausibility transform have been presented in [10,13].
Whatever the rationale for proposing a probability transformation of belief functions (decision making, as in the pignistic transform, or computational burden in many other cases), there are many ways of investigating its design: geometrical properties, principle of insufficient reason, commutativity properties, etcetera. A detailed discussion of the geometrical properties of $\tilde{b}$ and $\tilde{pl}$, for instance, has been given in [14]. Here we focus on commutativity properties, making a distinction between transforms which commute with respect to affine combination of belief functions (pignistic and orthogonal probabilities), and those which commute with respect to Dempster's rule (and therefore are consistent with the original Dempster-Shafer framework): relative plausibility and belief transform. This outlines a classification of probability transformations into two classes, according to the operator they commute with.

[1] With a harmless abuse of notation we denote the values of mass, belief and plausibility functions on a singleton x by $m_b(x)$, $b(x)$ and $pl_b(x)$ rather than $m_b(\{x\})$, $b(\{x\})$ and $pl_b(\{x\})$.

1.3 Paper contribution and outline

As belief functions have different, rather conflicting interpretations, in Section 2 we discuss the semantics of relative belief and plausibility in both the probability-bound and Shafer's interpretations of the theory. Within the probability-bound interpretation (Section 2.1), as neither transforms are associated with a valid redistribution of the mass of the focal elements to the singletons, it is easy to prove that they are not consistent with the original belief function. However, an interesting betting semantic for such transforms in this interpretation can be provided in an adversarial game theory scenario [50] in which an opponent is free to pick any probability function in the set determined by a belief function, and the decision maker's goal is to maximize their minimal reward (or minimize their maximal loss). In Shafer's formulation of the theory of evidence as an evidence combination process, arguments similar to those formulated for the plausibility transform can be resorted to in the case of the relative belief transform (Section 2.2).
Indeed, as we argue here, the relative plausibility and belief transforms are closely related probability transformations (Section 3). Not only the latter can be seen as the relative plausibility of singletons of the associated plausibility function (Section 3.2), but both transforms meet a number of dual properties with respect to Dempster's rule of combination (Section 3.3). In particular, while $\tilde{pl}_b$ commutes with Dempster's sum *of belief functions*, $\tilde{b}$ commutes with the orthogonal sum of *plausibility functions*. Similarly, while $\tilde{pl}_b$ perfectly represents the belief function b when combined with any probability distribution (5), $\tilde{b}$ perfectly represents *the associated plausibility function* pl_b when combined with a probability through the natural extension of Dempster's sum (Section 3.4). Such a duality is illustrated in the following table:

$$\begin{array}{ccc} b & \leftrightarrow & pl_b \\ \tilde{pl}_b & \leftrightarrow & \tilde{b} \\ b \oplus p = \tilde{pl}_b \oplus p \ \ \forall p & \leftrightarrow & pl_b \oplus p = \tilde{b} \oplus p \ \ \forall p \\ \tilde{pl}_b[b_1 \oplus b_2] = \tilde{pl}_b[b_1] \oplus \tilde{pl}_b[b_2] & \leftrightarrow & \tilde{b}[pl_{b_1} \oplus pl_{b_2}] = \tilde{b}[pl_{b_1}] \oplus \tilde{b}[pl_{b_2}]. \end{array}$$

This outlines the classification of probability transformations into two major classes: those commuting with affine combination versus those commuting with Dempster's rule (Section 3.5). The behavior of the plausibility transform w.r.t. affine combination, closely mimicking that of Dempster's rule itself, confirms that this second family is indeed more consistent with the original Dempster-Shafer framework.
The symmetry/duality between (relative) plausibility and belief is broken, however, as the existence of the relative belief of singletons is subject to a strong condition (7), stressing the issue of its applicability (Section 4). Even

though this situation is “singular” (in the sense that it excludes most belief and probability measures, Section 4.1), in practice the situation in which the mass of all singletons is nil is common. In Section 4.2, however, we point out that relative belief is only a member of a class of *relative mass* transformations, which can be interpreted as low-cost proxies for both plausibility and pignistic transforms (4.3). We discuss their applicability as approximate transformations in two significant scenarios.

2 Rationale of relative belief and plausibility of singletons in two interpretations of the theory of evidence

The original semantic of belief functions derives from Dempster's analysis of the effect of multi-valued mappings $\Gamma : \Omega \to 2^{\Theta}$, $x \in \Omega \mapsto \Gamma(x) \subseteq \Theta$ on evidence available in the form of a probability distribution on a “top” domain Ω on a “bottom” decision set Θ. As such, belief values are probabilities of events implying other events. In some of his papers [20], however, Dempster himself claimed that the mass $m_b(A)$ associated with a non-singleton event $A \subseteq \Theta$ could be understood as a “floating probability mass” which could not be attached to any particular singleton event $x \in A$ because of the lack of precision of the (multi-valued) operator that quantifies our knowledge via the mass function. This has originated a popular but controversial interpretation of belief functions as coherent sets of probabilities determined by sets of lower and upper bounds to their probability values.
As Shafer admits, there is a sense in which a single belief function can indeed be interpreted as a consistent system of probability bounds. However, the issue with the probability-bound interpretation of belief functions becomes evident when considering two or more belief functions addressing the same question but representing conflicting items of evidence, i.e., when Dempster's rule is applied to aggregate evidence. In [38,39], Shafer disavowed any probability-bound interpretation, a position later seconded by Dempster [21].
We will come back to this point in Section 2.2, in which we will link the relative belief transform to Cobb and Shenoy's arguments [5] in favor of the plausibility transform as a link between Shafer's theory of evidence (endowed with Dempster's rule) and Bayesian reasoning. To corroborate this argument, in Section 2.1 we show that both plausibility and relative belief transforms (unlike Smets' pignistic transform) are not consistent with a probability-bound interpretation of belief functions. Even in this scenario, however, a rationale for such transformations can be given via a utility theoretical argument, as in the case of the pignistic probability.

In their static, probability-bound interpretation, belief functions $b : 2^\Theta \rightarrow [0,1]$ determine each a convex set $\mathcal{P}[b]$ of "consistent" probability distributions (1). It can be proven that a probability distribution on Θ is consistent with b in the above way iff it is the result of a *redistribution process*, in which the mass of each focal element is shared between its elements in an arbitrary proportion [12]. One such probability is central in Smets' Transferable Belief Model, in which decisions are made at the *pignistic* level by applying the pignistic transform to convert the available belief function into a probability distribution. The nature of the mapping was originally based on the Principle of Insufficient Reason (PIR) proposed by Bernoulli, Laplace, and Keynes, which states that "if there is no known reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability".
A direct consequence of the PIR [2] in the probability-bound interpretation of belief functions is that, when considering how to redistribute the mass of an event A, it is wise to assume equiprobability amongst its singletons. This yields exactly the pignistic transform (2).

It is easy to prove that relative belief and plausibility of singletons are not the result of such a redistribution process, and therefore are not consistent with the original belief function in the sense defined above. Indeed, the relative plausibility of singletons (4) is the result of a process in which:

- for each singleton $x \in \Theta$ a redistribution process (there could be more than one) is selected in which the mass of all the events containing it is reassigned to x, yielding $\{pl_b(x), x \in \Theta\}$;
- however, as different redistribution processes are supposed to hold for different singletons (many of which belong to the same higher-size focal elements), this scenario is not compatible with the existence of a single redistribution of mass to the singletons, as the mass of the same higher cardinality event is assigned to different singletons;
- the obtained plausibility values $pl_b(x)$, $x \in \Theta$ are nevertheless normalized to yield a formally admissible probability distribution.

Similarly, for the relative belief of singletons (6):

- for each singleton $x \in \Theta$ a redistribution process is selected in which only the mass of $\{x\}$ itself is re-assigned to x, yielding $\{b(x) = m_b(x), x \in \Theta\}$;
- once again this scenario does not correspond to a single valid redistribution

[2] Later on, however, Smets [43] advocated that the PIR could not justify by itself the uniqueness of the pignistic transform, and proposed a justification based on a number of axioms.

process, as the mass of all higher-size focal elements is not assigned to any singletons;

- the obtained values $b(x)$, $x \in \Theta$ are nevertheless normalized to produce a valid probability.

The fact that both such probability transformations derive from assuming at the same time a number of incompatible redistribution processes is reflected by the fact that the resulting probability distributions are not guaranteed to belong to the set of probabilities (1) consistent with b.

Theorem 1 *The relative belief of singletons is not always consistent.*

Proof. We just need a simple counterexample. Consider a belief function $b : 2^\Theta \to [0,1]$ on $\Theta = \{x_1, x_2, ..., x_n\}$, $k_{m_b} \doteq \sum_{x \in \Theta} m_b(x)$ the total mass it assigns to singletons, with b.b.a. $m_b(x_i) = k_{m_b}/n$ for all i, $m_b(\{x_1, x_2\}) = 1 - k_{m_b}$. Then:

$$\begin{aligned} b(\{x_1, x_2\}) &= 2 \cdot \frac{k_{m_b}}{n} + 1 - k_{m_b} = 1 - k_{m_b}\Big(\frac{n-2}{n}\Big), \\ \tilde{b}(x_1) = \tilde{b}(x_2) = \frac{1}{n} \Rightarrow \tilde{b}(\{x_1, x_2\}) &= \frac{2}{n}. \end{aligned}$$

For $\tilde{b}$ to be consistent with b (Equation (1)) it is necessary that $\tilde{b}(\{x_1, x_2\}) \geq b(\{x_1, x_2\})$, in other words:

$$\frac{2}{n} \geq 1 - k_{m_b} \frac{n-2}{n} \equiv k_{m_b} \geq 1,$$

i.e., $k_{m_b} = 1$. If $k_{m_b} < 1$ (b is not a probability) its relative belief of singletons is not consistent. □

A similar counterexample can be found for $\tilde{pl}_b$.

Theorem 2 *The relative plausibility of singletons is not always consistent.*

Proof. Let us pick for sake of simplicity a frame of discernment with just three elements: $\Theta = \{x_1, x_2, x_3\}$, and the following b.b.a.:

$$m_b(\{x_i\}^c) = \frac{k}{3} \; \forall i = 1, 2, 3, \quad m_b(\{x_1, x_2\}^c) = m_b(\{x_3\}) = 1 - k.$$

In this case, the plausibility of $\{x_1, x_2\}$ is obviously: $pl_b(\{x_1, x_2\}) = 1 - (1 - k) = k$, while the plausibilities of the singletons are: $pl_b(x_1) = pl_b(x_2) = 2/3k$, $pl_b(x_3) = 1 - 1/3k$. Therefore $\sum_{x \in \Theta} pl_b(x) = 1 + k$ and the relative plausibility values are: $\tilde{pl}_b(x_1) = \tilde{pl}_b(x_2) = \frac{2/3k}{1+k}$, $\tilde{pl}_b(x_3) = \frac{1-1/3k}{1+k}$.
For $\tilde{pl}_b$ to be consistent with b we need:

$$\tilde{pl}_b(\{x_1, x_2\}) = \tilde{pl}_b(x_1) + \tilde{pl}_b(x_2) = \frac{4}{3} k \frac{1}{1+k} \leq pl_b(\{x_1, x_2\}) = k,$$

which happens if and only if $k \geq 1/3$. Therefore, for $k < 1/3$ $\tilde{pl}_b \notin \mathcal{P}[b]$. □

As an additional example, consider a belief function on $\Theta = \{x_1, x_2, ..., x_n\}$ with two focal elements:

$$m_b(x_1) = 0.01, \quad m_b(\{x_2, ..., x_n\}) = 0.99. \tag{8}$$

This can be interpreted as the following real-world situation. A number of people $x_2, ..., x_n$ have no money of their own but they are all candidates to inherit the wealth of a very rich relative. Person x_1 is not, but has some little money of their own. Note that it is not correct to interpret $x_2, ..., x_n$ as assured, joint owners of a certain wealth (say, shares of the same company), as (8) is indeed consistent (in the probability-bound interpretation) with a distribution which assigns probability 0.99 to a single person of the group $x_2, ..., x_n$.
The relative belief of singletons associated with (8) is the distribution with $\tilde{b}(x_1) = 1$, $\tilde{b}(x_i) = 0$ for $i = 2, ..., n$. Clearly this is not a good representative of the set of probabilities consistent with the above belief function, as it does not contemplate at all the chance all the heirs $x_2, ..., x_n$ have to gain a remarkable amount of money. Indeed, according to Theorem 1, $\tilde{b}$ in this example is not at all consistent with (8).

What $\tilde{b}$ and $\tilde{pl}_b$ do is to set respectively a lower and an upper bound to the probability values for each element $x \in \Theta$ of the frame under the constraint represented by the belief function b, as in Dempster's original interpretation. However, even though in the probability-bound interpretation the two transforms do not appear as valid approximations of belief functions, an interesting interpretation for them can be provided in a game/utility theory context [50,46,30]. The argument we propose here recalls somehow the betting rationale for the pignistic transform in the TBM [54], where a MAP decision is taken at the pignistic level as follows:

$$x_* = \arg\max_{x \in \Theta} \Big\{ BetP[b]\,(x) \Big\}.$$

In expected utility theory [50], a decision maker can choose between a number of "lotteries" (probability distributions) L_i in order to maximize their expected return or utility calculated as

$$E(L_i) = \sum_{x \in \Theta} u(x) \cdot p_i(x),$$

where u is a utility function $u : \Theta \to \mathbb{R}^+$ which measures the relative satisfaction (for us) of the different outcomes $x \in \Theta$ of the lottery, and $p_i(x)$ is the probability of x under lottery L_i.

Consider instead the following game theory scenario, inspired by Strat's expected utility approach to decision making with belief functions [47,37].

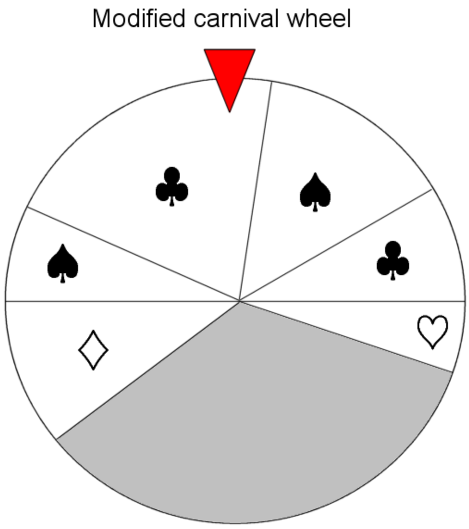

Fig. 1. The modified carnival wheel, in which part of the spinning wheel is cloaked.

In a country fair, people are asked to bet on one of the possible outcomes of a spinning carnival wheel. Suppose the outcomes are $\{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$, and that they each have the same utility (return) to the player. This is equivalent to a lottery (probability distribution), in which each outcome has a probability proportional to the area of the corresponding sectors on the wheel. However, the fair manager decides to make the game more interesting by covering part of the wheel. Players are still asked to bet on a single outcome, knowing that the manager is allowed to rearrange the hidden sector of the wheel as he pleases (see Figure 1). Clearly, this situation can be described as a belief function, in particular one in which the fraction of area associated with the hidden sector is assigned as mass to the whole decision space $\{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$. If additional (partial) information is provided, for instance that $\diamondsuit$ cannot appear in the hidden sector, different belief functions must be chosen instead.

Regardless the particular belief function b (set of probabilities) at hand, the rule allowing the manager to pick an arbitrary distribution of outcomes in the hidden section mathematically translates into allowing him/her to choose *any* probability distribution $p \in \mathcal{P}[b]$ consistent with b in order to damage the player. Supposing the aim of the player is to maximize their minimal chance of winning the bet, which outcome (singleton) should they pick?
In the probability-bound interpretation, the belief value of each singleton $x \in \Theta$ measures the minimal support x can receive from a distribution of the family $\mathcal{P}[b]$ associated with the belief function b:

$$b(x) = \min_{p \in \mathcal{P}[b]} p(x).$$

Hence $x_{maximin} \doteq \arg\max_{x \in \Theta} b(x)$ is the outcome which maximizes such minimal support. In the example of Figure 1, as $\clubsuit$ is the outcome which occupies the largest share of the visible part of the wheel, the safest bet (the one which guarantees the maximal chance in the worst case) is indeed ♣. Formally, ♣ is the singleton with the largest belief value. Now, if we normalize to compute the relative belief of singletons this outcome is obviously conserved:

$$x_{maximin} = \arg\max_{x\in\Theta} \tilde{b}(x) = \arg\max_{x\in\Theta} \min_{p\in\mathcal{P}[b]} p(x).$$

In conclusion, if the utility function is constant (i.e., if no element of Θ can be preferred over the others), $x_{maximin}$ (the peak(s) of the relative belief of singletons) represents the best possible defensive strategy aimed at maximizing the minimal utility of the possible outcomes.

Dually, $pl_b(x)$ measures the maximal possible support to x by a distribution consistent with b, so that

$$x_{minimax} = \arg\min_{x\in\Theta} \tilde{pl}_b(x) = \arg\min_{x\in\Theta} \max_{p\in\mathcal{P}[b]} p(x)$$

is the outcome which minimizes the maximal possible support.
Suppose for sake of simplicity that the loss function $l : \Theta \rightarrow \mathbb{R}^+$ which measures the relative dissatisfaction of the outcomes is constant, and that in the same game theory setup our opponent is (again) free to pick a consistent probability distribution $p \in \mathcal{P}[b]$. Then the element with minimal relative plausibility is the best possible defensive strategy aimed at minimizing the maximum possible loss.
Note that when the utility function is not constant the above minimax and maximin problems naturally generalize as:

$$x_{maximin} = \arg\max_{x\in\Theta} \tilde{b}(x)u(x), \qquad x_{minimax} = \arg\min_{x\in\Theta} \tilde{pl}_b(x)l(x).$$

While in classical utility theory the decision maker has to select the best "lottery" (probability distribution) in order to maximize the expected utility, here the "lottery" is chosen by his/her opponent (given the available partial evidence), and the decision maker is left with betting on the safest strategy (element of Θ). Relative belief and plausibility of singletons play a crucial role in determining the safest betting strategy in an adversarial scenario in which the decision maker has to minimize their maximal loss/maximize their minimal return.

2.2 Semantics within Shafer's interpretation

Shafer has strongly argued against a probability-bound interpretation of belief functions. When these are not taken in isolation but as pieces of evidence to combine, such an interpretation forces us to consider only groups of belief functions whose degrees of belief, when interpreted as probability bounds, can be satisfied simultaneously (in other words, when their sets of consistent probabilities have non-empty intersection). In Shafer's (and Shenoy's) view, though, when belief functions are combined via Dempster's rule this is irrelevant, even though consistent probabilities that simultaneously bound all the belief functions being combined as well as the resulting belief function do exist when no renormalization is required in their Dempster's combination. Consequently, citing Shafer, authors who support a probability-bound interpretation of belief functions are uncomfortable with renormalization [56].

In this context, Cobb and Shenoy [5] have argued in favor of the plausibility transform as a link between Shafer's theory of evidence (endowed with Dempster's rule) and Bayesian reasoning. Besides some general arguments supporting probability transformations of belief functions in general, their points more specifically about the plausibility transform can be summarized as follows:

- a probability transformation consistent with Dempster's rule can improve our understanding of the theory of evidence by providing probabilistic semantics for belief functions, i.e., "meanings" of basic probability assignments in the context of betting for hypotheses in the frame Θ;
- in opposition to some literature on belief functions suggesting that the theory of evidence is more expressive than probability theory (since the probability model obtained by using the pignistic transformation leads to non-intuitive results [2]), they show that by using the plausibility transformation method the original belief function model and the corresponding probability model yield the same qualitative results;
- a probability transformation consistent with Dempster's rule allows to build probabilistic models by converting/transforming belief function models obtained by using the belief function semantics of distinct evidence [41].

Mathematically, they proved [7] that the plausibility transform commutes with Dempster's rule, and meets a number of additional properties which they claim "allow an integration of Bayesian and D-S reasoning that takes advantage of the efficiency in computation and decision-making provided by Bayesian calculus while retaining the flexibility in modeling evidence that underlies D-S reasoning".

In this paper we prove that a similar set of (dual) properties hold for the relative belief transform, associating relative belief and relative plausibility transforms in a family of probability transformations strongly related to Shafer's interpretation of the theory of evidence via Dempster's rule.

3 Duality

Relative belief and plausibility of singletons are, as we show here, linked by a form of duality, as $\tilde{b}$ can be interpreted as the relative plausibility of singletons *of the plausibility function* pl_b associated with b. Furthermore, $\tilde{b}$ and $\tilde{pl}_b$ share a close relationship with Dempster's evidence combination rule $\oplus$, as they meet a set of dual properties with respect to $\oplus$. This suggests a classification of all the probability transformations of belief functions in terms of the operator they relate to.

3.1 Relative plausibility, Dempster's rule, and pseudo belief functions

Definition 1 *The* orthogonal sum *or* Dempster's sum *of two belief functions* $b_1, b_2 : 2^\Theta \rightarrow [0,1]$ *is a new belief function* $b_1 \oplus b_2 : 2^\Theta \rightarrow [0,1]$ *with b.b.a.:*

$$m_{b_1 \oplus b_2}(A) = \frac{\sum_{B \cap C = A} m_{b_1}(B)\, m_{b_2}(C)}{\sum_{B \cap C \neq \emptyset} m_{b_1}(B)\, m_{b_2}(C)}, \tag{9}$$

where m_{b_i} *denotes the b.b.a. associated with* b_i.

We denote by $k(b_1, b_2)$ the denominator of Equation (9).
Cobb and Shenoy [7] proved that the relative plausibility function $\tilde{pl}_b$ commutes with Dempster's rule, and meets a number of additional properties [3].

Proposition 1 *(1) If* $b = b_1 \oplus \cdots \oplus b_m$ *then* $\tilde{pl}_b = \tilde{pl}_{b_1} \oplus \cdots \oplus \tilde{pl}_{b_m}$*: Dempster's sum and relative plausibility commute.*
(2) If m_b *is idempotent with respect to Dempster's rule, i.e.* $m_b \oplus m_b = m_b$*, then* $\tilde{pl}_b$ *is idempotent with respect to Bayes' rule.*
(3) Let us define the limit of a belief function b *as*

$$b^\infty \doteq \lim_{n \rightarrow \infty} b^n \doteq \lim_{n \rightarrow \infty} \Big(b \oplus \cdots \oplus b \Big) \quad (n \text{ times}); \tag{10}$$

if $\exists x \in \Theta$ *such that* $pl_b(x) > pl_b(y)$ *for all* $y \neq x, y \in \Theta$*, then* $\tilde{pl}_{b^\infty}(x) = 1$*,* $\tilde{pl}_{b^\infty}(y) = 0$ *for all* $y \neq x$*.*
(4) If $\exists A \subseteq \Theta$ $(|A| = k)$ *such that* $pl_b(x) = pl_b(y)$ *for all* $x, y \in A$ *and* $pl_b(x) > pl_b(z)$ *for all* $x \in A, z \in A^c$*, then* $\tilde{pl}_{b^\infty}(x) = \tilde{pl}_{b^\infty}(y) = 1/k$ *for all* $x, y \in A$*, while* $\tilde{pl}_{b^\infty}(z) = 0$ *for all* $z \in A^c$*.*

On his side, Voorbraak has shown [51] that:

Proposition 2 *The relative plausibility of singletons* $\tilde{pl}_b$ *is a perfect representative of* b *in the probability space when combined through Dempster's rule:* $b \oplus p = \tilde{pl}_b \oplus p$*,* $\forall p \in \mathcal{P}$*.*

The relative belief of singletons meets analogous dual properties. Their study requires first to extend our analysis to a more general class of objects. Sum

[3] The original statements from [6] have been reformulated according to the notation of this paper.

functions of the form $\varsigma(A) = \sum_{B\subseteq A} m_\varsigma(B)$ whose Moebius transform m_ς meets the normalization axiom, $\varsigma(\Theta) = \sum_{\emptyset\subsetneq A\subseteq\Theta} m_\varsigma(A) = 1$, but is not necessarily non-negative, are called *pseudo belief functions* [44].
Plausibility functions are pseudo belief functions too, as they meet the normalization constraint $pl_b(\Theta) = 1$ for all b. Their Moebius transform [8]

$$\mu_b(A) \doteq \sum_{B\subseteq A} (-1)^{|A\setminus B|} pl_b(B) = (-1)^{|A|+1} \sum_{B\supseteq A} m_b(B), \;\; A \neq \emptyset \qquad (11)$$

is called *basic plausibility assignment* ($\mu_b(\emptyset) = 0$).
Both belief and plausibility functions can be represented as vectors of a Cartesian space $\mathcal{B}$ called *belief space* [11]. In that space they can be written as affine combinations of the *categorical* belief functions b_A (such that $m_{b_A}(A) = 1$, $m_{b_A}(B) = 0 \; \forall B \neq A$), with coefficients given by their b.b.a. or basic plausibility assignment, respectively:

$$b = \sum_{\emptyset\neq A\subseteq\Theta} m_b(A) b_A, \quad pl_b = \sum_{\emptyset\neq A\subseteq\Theta} \mu_b(A) b_A. \qquad (12)$$

3.2 A (broken) symmetry

A direct consequence of the duality between belief and plausibility measures is the existence of a striking symmetry between (relative) plausibility and belief transform. A formal proof of this symmetry is based on the following interesting property of the basic plausibility assignment μ_b [15].

Lemma 1 $\sum_{A\supseteq\{x\}} \mu_b(A) = m_b(x)$.

Theorem 3 *Given a pair of belief/plausibility functions $b, pl_b : 2^\Theta \to [0,1]$, the relative belief transform of the belief function b coincides with the plausibility transform of the associated plausibility function pl_b (interpreted as a pseudo belief function):*

$$\tilde{b}[b] = \tilde{pl}[pl_b].$$

Proof. Each pseudo belief function admits a (pseudo) plausibility function, as in the case of standard belief functions, which can be computed as $pl_\varsigma(A) = \sum_{B\cap A\neq\emptyset} m_\varsigma(B)$.
For the class of pseudo belief functions ς which correspond to the plausibility of some belief function b ($\varsigma = pl_b$ for some $b \in \mathcal{B}$), the pseudo plausibility function is $pl_{pl_b}(A) = \sum_{B\cap A\neq\emptyset} \mu_b(B)$, as μ_b (11) is the Moebius inverse of pl_b. When applied to the elements $x \in \Theta$ of the common frame of b, pl_b this yields $pl_{pl_b}(x) = \sum_{B\ni x} \mu_b(B) = m_b(x)$ by Lemma 1, which implies

$$\tilde{pl}[pl_b](x) = \frac{pl_{pl_b}(x)}{\sum_{y\in\Theta} pl_{pl_b}(y)} = \frac{m_b(x)}{\sum_{y\in\Theta} m_b(y)} = \tilde{b}[b]. \qquad \square$$

The symmetry between relative plausibility and belief of singletons is broken by the fact that the latter is not defined for belief functions with no singleton focal sets. Since $\tilde{b}$ is itself an instance of relative plausibility (of a plausibility function pl_b), and $\tilde{pl}_b$ always exists, this seems to contradict Theorem 3.
This superficial paradox finds an explanation in the combinatorial nature of belief, plausibility, and commonality functions. As we proved in [15], while belief measures are sum functions of the form $b(A) = \sum_{B \subset A} m_b(B)$ whose Moebius transform m_b is both normalized and non-negative, plausibility measures are sum functions whose Moebius transform μ_b is not necessarily non-negative (while commonality functions are not even normalized). Hence, the quantity

$$\sum_x pl_{pl_b}(x) = \sum_x \sum_{A \supseteq \{x\}} \mu_b(A) = \sum_{A \supseteq \Theta} \mu_b(A)|A|$$

can be equal to zero, in which case $\tilde{pl}_{pl_b} = \tilde{b}$ does not exist.

3.3 Dual properties of the relative belief operator

The duality between $\tilde{b}$ and $\tilde{pl}_b$ (albeit imperfect to some extent) extends to the transformations' behavior with respect to Dempster's rule of combination (9). We first need to note that the orthogonal sum can be naturally extended to a pair ς_1, ς_2 of pseudo belief functions too [16], by applying (9) to their Moebius inverses m_{ς_1}, m_{ς_2}.

Proposition 3 *When applied to a pair of pseudo belief functions ς_1, ς_2, Dempster's rule defined as in Equation (9) yields again a pseudo belief function.*

We still denote the orthogonal sum of two pseudo belief functions ς_1, ς_2 by $\varsigma_1 \oplus \varsigma_2$. As plausibility functions are pseudo belief functions, Dempster's rule can then be formally applied to them as well. We can then prove a dual commutativity result for relative beliefs. To this purpose, it is convenient to introduce a dual form of the relative belief operator, mapping a plausibility function to the corresponding relative belief of singletons: $\tilde{b} : \mathcal{PL} \to \mathcal{P}$, $pl_b \mapsto \tilde{b}[pl_b]$, where $\mathcal{PL}$ is the space of all plausibility functions, and

$$\tilde{b}[pl_b](x) \doteq \frac{m_b(x)}{\sum_{y \in \Theta} m_b(y)} \quad \forall x \in \Theta \tag{13}$$

is defined as usual for belief functions b such that $\sum_y m_b(y) \neq 0$.
Indeed, as b and pl_b are in 1-1 correspondence, we can indifferently define an operator mapping a *belief function* b to its relative belief $\tilde{b}$, *or* mapping the unique plausibility function pl_b associated with b to $\tilde{b}$. The following commutativity theorem follows, as the dual of point 1) in Proposition 1.

Theorem 4 *The relative belief operator commutes with respect to Dempster's combination* of plausibility functions*:* $\tilde{b}[pl_1 \oplus pl_2] = \tilde{b}[pl_1] \oplus \tilde{b}[pl_2]$.

Proof. The basic plausibility assignment of $pl_1 \oplus pl_2$ is, according to (9):

$$\mu_{pl_1 \oplus pl_2}(A) = \frac{1}{k(pl_1, pl_2)} \sum_{X \cap Y = A} \mu_1(X)\mu_2(Y).$$

Therefore, according to Lemma 1, the corresponding relative belief of singletons $\tilde{b}[pl_1 \oplus pl_2](x)$ (13) is proportional to:

$$\begin{aligned} m_{pl_1 \oplus pl_2}(x) &= \sum_{A \supseteq \{x\}} \mu_{pl_1 \oplus pl_2}(A) \\ &= \frac{\sum_{A \supseteq \{x\}} \sum_{X \cap Y = A} \mu_1(X)\mu_2(Y)}{k(pl_1, pl_2)} = \frac{\sum_{X \cap Y \supseteq \{x\}} \mu_1(X)\mu_2(Y)}{k(pl_1, pl_2)}, \end{aligned} \tag{14}$$

where $m_{pl_1 \oplus pl_2}(x)$ denotes the b.b.a. of the (pseudo) belief function which corresponds to the plausibility function $pl_1 \oplus pl_2$. On the other hand, as $\sum_{X \supseteq \{x\}} \mu_b(X) = m_b(x)$:

$$\tilde{b}[pl_1](x) \propto m_1(x) = \sum_{X \supseteq \{x\}} \mu_1(X), \;\; \tilde{b}[pl_2](x) \propto m_2(x) = \sum_{X \supseteq \{x\}} \mu_2(X).$$

Their Dempster's combination is therefore:

$$(\tilde{b}[pl_1] \oplus \tilde{b}[pl_2])(x) \propto \Big(\sum_{X \supseteq \{x\}} \mu_1(X) \Big) \Big(\sum_{Y \supseteq \{x\}} \mu_2(Y) \Big) = \sum_{X \cap Y \supseteq \{x\}} \mu_1(X)\mu_2(Y),$$

and by normalizing we get (14). □
Theorem 4 implies that

$$\tilde{b}[(pl_b)^n] = (\tilde{b}[pl_b])^n. \tag{15}$$

As an immediate consequence, an idempotence property which is the dual of point 2) of Proposition 1 holds for the relative belief of singletons.

Corollary 1 *If pl_b is idempotent with respect to Dempster's rule, i.e. $pl_b \oplus pl_b = pl_b$, then $\tilde{b}[pl_b]$ is itself idempotent: $\tilde{b}[pl_b] \oplus \tilde{b}[pl_b] = \tilde{b}[pl_b]$.*

Proof. By Theorem 4 $\tilde{b}[pl_b] \oplus \tilde{b}[pl_b] = \tilde{b}[pl_b \oplus pl_b]$, and if $pl_b \oplus pl_b = pl_b$ the thesis immediately follows. □
The dual results of the remaining two statements of Proposition 1 can be proven in a similar fashion.

Theorem 5 *If $\exists x \in \Theta$ such that $b(x) > b(y)$ $\forall y \neq x, y \in \Theta$, then*

$$\tilde{b}[pl_b^\infty](x) = 1, \quad \tilde{b}[pl_b^\infty](y) = 0 \;\forall y \neq x.$$

Proof. Taking the limit on both sides of Equation (15) we get

$$\tilde{b}[pl_b^\infty] = (\tilde{b}[pl_b])^\infty. \tag{16}$$

Let us consider the quantity $(\tilde{b}[pl_b])^\infty = \lim_{n\to\infty}(\tilde{b}[pl_b])^n$ on the right hand side. Since $(\tilde{b}[pl_b])^n(x) = K(b(x))^n$ (where K is a constant independent from x), and x is the unique most believed state, it follows that

$$(\tilde{b}[pl_b])^\infty(x) = 1, \quad (\tilde{b}[pl_b])^\infty(y) = 0 \;\; \forall y \neq x. \tag{17}$$

Hence by (16) $\tilde{b}[pl_b^\infty](x) = 1$, and $\tilde{b}[pl_b^\infty](y) = 0$ for all $y \neq x$. $\square$

A similar proof can be given for the following generalization of Theorem 5.

Corollary 2 *If $\exists A \subseteq \Theta$ ($|A| = k$) s.t. $b(x) = b(y)$ $\forall x, y \in A$, $b(x) > b(z)$ $\forall x \in A, z \in A^c$, then*

$$\tilde{b}[pl_b^\infty](x) = \tilde{b}[pl_b^\infty](y) = 1/k \; \forall x, y \in A, \quad \tilde{b}[pl_b^\infty](z) = 0 \; \forall z \in A^c.$$

It is crucial to point out that commutativity (Theorem 4) and idempotence (Corollary 1) hold for combinations *of plausibility functions*, and not of belief functions. Consider as an example the belief function b on the frame of size four $\Theta = \{x, y, z, w\}$ determined by the following basic probability assignment:

$$m_b(\{x, y\}) = 0.4, \quad m_b(\{y, z\}) = 0.4, \quad m_b(w) = 0.2. \tag{18}$$

Its basic plausibility assignment is, according to (11), given by:

$$\begin{array}{c} \mu_b(x) = 0.4, \; \mu_b(y) = 0.8, \; \mu_b(z) = 0.4, \\ \mu_b(w) = 0.2, \; \mu_b(\{x, y\}) = -0.4, \; \mu_b(\{y, z\}) = -0.4. \end{array} \tag{19}$$

To check the validity of Theorems 4 and 5 let us analyze the two series $(\tilde{b}[pl_b])^n$ and $\tilde{b}[(pl_b)^n]$. By applying Dempster's rule to the basic plausibility assignment (19) ($pl_b^2 = pl_b \oplus pl_b$) we get a new basic plausibility assignment μ_b^2 with values $\mu_b^2(x) = 4/7$, $\mu_b^2(y) = 8/7$, $\mu_b^2(z) = 4/7$, $\mu_b^2(w) = -1/7$, $\mu_b^2(\{x, y\}) = -4/7$, $\mu_b^2(\{y, z\}) = -4/7$ (see Figure 2). To compute the corresponding relative belief of singletons $\tilde{b}[pl_b^2]$ we first need to get the plausibility values:

$$\begin{array}{c} pl_b^2(\{x, y, z\}) = \mu_b^2(x) + \mu_b^2(y) + \mu_b^2(z) + \mu_b^2(\{x, y\}) + \mu_b^2(\{y, z\}) = 8/7, \\ pl_b^2(\{x, y, w\}) = 1, \quad pl_b^2(\{x, z, w\}) = 1, \quad pl_b^2(\{y, z, w\}) = 1, \end{array}$$

which imply (by definition $pl_b(A) \doteq 1 - b(A^c)$): $b^2(w) = -1/7$, $b^2(z) = b^2(y) = b^2(x) = 0$, i.e., $\tilde{b}[pl_b^2] = [0, 0, 0, 1]'$ (representing probability distributions as vectors of the form $[p(x), p(y), p(z), p(w)]'$).

{y,z}		{y}	{z}		{y}	{y,z}
{x,y}	{x}	{y}			{x,y}	{y}
{w}				{w}		
{z}			{z}			{z}
{y}		{y}			{y}	{y}
{x}	{x}				{x}	
	{x}	{y}	{z}	{w}	{x,y}	{y,z}

Fig. 2. Intersection of focal elements in Dempster's combination of the basic plausibility assignment (19) with itself. Non-zero mass events for each addendum $\mu_1 = \mu_2 = \mu_b$ correspond to rows/columns of the table, each entry of the table hosting the related intersection.

Theorem 4 is confirmed as, by (18) (being $\{w\}$ the only singleton with non-zero mass), $\tilde{b} = [0,0,0,1]'$ so that $\tilde{b} \oplus \tilde{b} = [0,0,0,1]'$ and $\tilde{b}[.]$ commutes with $pl_b \oplus$. By combining pl_b^2 with pl_b one more time we get the basic plausibility assignment:

$$\mu_b^3(x) = 16/31, \quad \mu_b^3(y) = 32/31, \quad \mu_b^3(z) = 16/31, \quad \mu_b^3(w) = -1/31,$$
$$\mu_b^3(\{x,y\}) = -16/31, \quad \mu_b^3(\{y,z\}) = -16/31,$$

which corresponds to $pl_b^3(\{x,y,z\}) = 32/31$, $pl_b^3(\{x,y,w\}) = 1$, $pl_b^3(\{x,z,w\}) = 1$, $pl_b^3(\{y,z,w\}) = 1$, i.e.: $b^3(w) = -1/31$, $b^3(z) = b^3(y) = b^3(x) = 0$, and $\tilde{b}[pl_b^3] = [0,0,0,1]'$, which again is equal to $\tilde{b} \oplus \tilde{b} \oplus \tilde{b}$ as Theorem 4 guarantees. The series of basic plausibility assignments $(\mu_b)^n$ clearly converges to:

$$\mu_b^n(x) \to 1/2^+, \quad \mu_b^3(y) \to 1^+, \quad \mu_b^3(z) \to 1/2^+, \quad \mu_b^3(w) \to 0^-,$$
$$\mu_b^3(\{x,y\}) \to -1/2^-, \quad \mu_b^3(\{y,z\}) \to -1/2^-,$$

associated with the following plausibility values: $\lim_{n\to\infty} pl_b^n(\{x,y,z\}) = 1^+$, $pl_b^n(\{x,y,w\}) = pl_b^n(\{x,z,w\}) = pl_b^n(\{y,z,w\}) = 1 \ \forall n \geq 1$, which in turn correspond to the following values of belief of singletons: $\lim_{n\to\infty} b^n(w) = 0^-$, $b^n(z) = b^n(y) = b^n(x) = 0 \ \forall n \geq 1$. Therefore:

$$\begin{aligned} \lim_{n\to\infty} \tilde{b}[pl_b^\infty](w) &= \lim_{n\to\infty} \tfrac{b^n(w)}{b^n(w)} = 1, \\ \lim_{n\to\infty} \tilde{b}[pl_b^\infty](x) &= \lim_{n\to\infty} \tilde{b}[pl_b^\infty](y) = \lim_{n\to\infty} \tilde{b}[pl_b^\infty](z) \\ &= \lim_{n\to\infty} \tfrac{0}{b^n(w)} = \lim_{n\to\infty} 0 = 0, \end{aligned}$$

in perfect agreement with Theorem 5.

A dual of the representation theorem (Proposition 2) for the relative belief transform can also be proven, once we recall a useful result on Dempster's sum of affine combinations [16].

Proposition 4 *The orthogonal sum $b \oplus \sum_i \alpha_i b_i$, $\sum_i \alpha_i = 1$ of a belief function b with any*[4] *affine combination of belief functions is itself an affine combination of the partial sums $b \oplus b_i$, namely:*

$$b \oplus \sum_i \alpha_i b_i = \sum_i \gamma_i (b \oplus b_i), \tag{20}$$

where $\gamma_i = \frac{\alpha_i k(b,b_i)}{\sum_j \alpha_j k(b,b_j)}$ and $k(b, b_i)$ is the normalization factor of the partial Dempster's sum $b \oplus b_i$.

Again, the duality between $\tilde{b}$ and $\tilde{pl}_b$ implies that the relative belief of singletons represents the associated *plausibility* function pl_b, and *not* the corresponding belief function b: $\tilde{b} \oplus p \neq b \oplus p$.

Theorem 6 *The relative belief of singletons $\tilde{b}$ represents perfectly the corresponding plausibility function pl_b when combined with a probability via (extended) Dempster's rule: $\tilde{b} \oplus p = pl_b \oplus p$ for each Bayesian belief function $p \in \mathcal{P}$.*

Proof. In virtue of Equation (12) we can express a plausibility function as an affine combination of all the categorical belief functions b_A. We can then apply the commutativity property (20), obtaining:

$$pl_b \oplus p = \sum_{A \subseteq \Theta} \nu(A) p \oplus b_A \tag{21}$$

where $\nu(A) = \frac{\mu_b(A) k(p,b_A)}{\sum_{B \subseteq \Theta} \mu_b(B) k(p,b_B)}$ and $p \oplus b_A = \frac{\sum_{x \in A} p(x) b_x}{k(p,b_A)}$, with $k(p, b_A) = \sum_{x \in A} p(x)$. Once replaced these expressions in (21) we get: $pl_b \oplus p =$

$$= \frac{\sum_{A \subseteq \Theta} \mu_b(A) \Big(\sum_{x \in A} p(x) b_x \Big)}{\sum_{B \subseteq \Theta} \mu_b(B) \Big(\sum_{y \in B} p(y) \Big)} = \frac{\sum_{x \in \Theta} p(x) \Big(\sum_{A \supseteq \{x\}} \mu_b(A) \Big) b_x}{\sum_{y \in \Theta} p(y) \Big(\sum_{B \supseteq \{y\}} \mu_b(B) \Big)} = \frac{\sum_{x \in \Theta} p(x) m_b(x) b_x}{\sum_{y \in \Theta} p(y) m_b(y)},$$

once again by Lemma 1. But this is exactly $\tilde{b} \oplus p$, as a direct application of Dempster's rule (9) shows. □

[4] In fact the collection $\{b_i\}$ is required to include *at least* a belief function which is combinable with b, [16].

Theorem 6 can be obtained from Proposition 2 by replacing b with pl_b and $\tilde{pl}_b$ with $\tilde{b}$ in virtue of their duality. It is natural to suppose other properties of upper probabilities could in the future be found by analogous transformations of known propositions on lower probabilities, as a useful mathematical characterization of the relation between them.

Once again, the representation theorem 6 is about combinations of *plausibility functions* (as pseudo belief functions) and *not* combinations of proper belief functions. Going back to the previous example, the combination $b \oplus b$ of b with itself has b.b.a.:

$$
\begin{aligned}
m_{b\oplus b}(\{x,y\}) &= \frac{m_b(\{x,y\}) \cdot m_b(\{x,y\})}{k(b,b)} = \frac{0.16}{0.68} = 0.235,\\
m_{b\oplus b}(\{y,z\}) &= \frac{m_b(\{y,z\}) \cdot m_b(\{y,z\})}{k(b,b)} = \frac{0.16}{0.68} = 0.235,\\
m_{b\oplus b}(w) &= \frac{m_b(w) \cdot m_b(w)}{k(b,b)} = \frac{0.04}{0.68} = 0.058,\\
m_{b\oplus b}(y) &= \frac{m_b(\{x,y\}) \cdot m_b(\{y,z\}) + m_b(\{y,z\}) \cdot m_b(\{x,y\})}{k(b,b)} = 0.47,
\end{aligned}
$$

which obviously yields $\widetilde{b \oplus b} = \left[0, \frac{0.47}{0.528}, 0, \frac{0.058}{0.528}\right]' \neq \tilde{b} \oplus \tilde{b} = [0,0,0,1]'$.
The basic reason for that is that the plausibility function of a sum of two belief functions is *not* the sum of the associated plausibilities: $[pl_{b_1} \oplus pl_{b_2}] \neq pl_{b_1 \oplus b_2}$.

3.5 Two families of probability transforms

The following table summarizes the duality results we just presented:

$$
\begin{array}{ccc}
b & \leftrightarrow & pl_b \\
\tilde{pl}_b & \leftrightarrow & \tilde{b} \\
b \oplus p = \tilde{pl}_b \oplus p \ \ \forall p \in \mathcal{P} & \leftrightarrow & pl_b \oplus p = \tilde{b} \oplus p \ \ \forall p \in \mathcal{P} \\
\tilde{pl}_b[b_1 \oplus b_2] = \tilde{pl}_b[b_1] \oplus \tilde{pl}_b[b_2] & \leftrightarrow & \tilde{b}[pl_{b_1} \oplus pl_{b_2}] = \tilde{b}[pl_{b_1}] \oplus \tilde{b}[pl_{b_2}] \\
b \oplus b = b \vdash \tilde{pl}[b] \oplus \tilde{pl}[b] = \tilde{pl}[b] & \leftrightarrow & pl_b \oplus pl_b = pl_b \vdash \tilde{b}[pl_b] \oplus \tilde{b}[pl_b] = \tilde{b}[pl_b].
\end{array}
$$

Note that, just as Voorbraak's and Cobb's results are not valid for all pseudo belief functions but only for proper belief functions, the above dual results do not hold for all pseudo belief functions either, but only for those pseudo belief functions which are plausibility functions. These results bring about a subdivision of all probability transformations in two families, related to Dempster's sum and affine combination respectively. Once we recall that [9]

Proposition 5 *Both pignistic $BetP[b]$ and orthogonal $\pi[b]$ transform (3) com-*

mute with respect to affine combination. Whenever $\sum_i \alpha_i = 1$ *we have that:*

$$\pi\Big[\sum_i \alpha_i b_i\Big] = \sum_i \alpha_i \pi[b_i], \;\; BetP\Big[\sum_i \alpha_i b_i\Big] = \sum_i \alpha_i BetP[b_i].$$

we realize that we can in fact distinguish two families of probability transformations, determined by their behavior with respect to two operators acting on belief functions: affine combination (in the space of belief functions) and Dempster's rule [38,19,18].

The notion that there exist two distinct families of probability transformations, each determined by the operator they commute with, was already implicitly present in the literature. Smets' linearity axiom [45], which lays at the foundation of the pignistic transform, obviously corresponds (even though expressed in a somewhat different language) to commutativity with affine combination of belief functions. To address the criticism such axiom was subject to, Smets introduced later its formal justification based on an expected utility argument in the presence of conditional evidence [43].
On the other hand, Cobb and Shenoy defended the commutativity with respect of Dempster's rule, on the basis that the Dempster-Shafer theory of evidence is a coherent framework of which Dempster's rule is an integral part, and that a Dempster-compatible transformation can provide a useful probabilistic semantic for belief functions.

Incidentally, there seems to be a flaw in Smets' argument that the pignistic transform is uniquely determined as the probability transformation which commutes with affine combination: in [9] we indeed proved that the orthogonal transform (3) also enjoys the same property.
Analogously, we showed here that the plausibility transform is not unique as a probability transformation which commutes with $\oplus$ (even though, in this latter case, the transformation is applied to different objects).

We add a further element to this debate here, by proving that the plausibility transform, even though it does not obviously commute with affine combination, does commute with the *convex closure* (22) of belief functions in the belief space $\mathcal{B}$:

$$Cl(b_1, ..., b_k) = \Big\{ b \in \mathcal{B} : b = \alpha_1 b_1 + \cdots + \alpha_k b_k, \sum_i \alpha_i = 1, \; \alpha_i \geq 0 \; \forall i \Big\}. \quad (22)$$

The two facts are not in contradiction: as a matter of fact, the behavior of the plausibility transform in this respect reflects a similar behavior by Dempster's rule (proved in [16]), supporting the argument that the plausibility transform is indeed naturally associated with the D-S framework.
Let us first understand its relation with affine combination.

Lemma 2 *For all $\alpha \in \mathbb{R}$ we have that:*

$$\tilde{pl}\Big[\alpha b_1 + (1-\alpha) b_2\Big] = \beta_1 \tilde{pl}[b_1] + \beta_2 \tilde{pl}[b_2],$$

where

$$\beta_1 = \frac{\alpha k_{pl_1}}{\alpha k_{pl_1} + (1-\alpha) k_{pl_2}}, \quad \beta_2 = \frac{\alpha k_{pl_2}}{\alpha k_{pl_1} + (1-\alpha) k_{pl_2}}.$$

Proof. By definition, the plausibility values of the affine combination $\alpha b_1 + (1-\alpha) b_2$ are $pl[\alpha b_1 + (1-\alpha) b_2](x) =$

$$\begin{aligned}
&= \sum_{A \supseteq \{x\}} m_{\alpha b_1 + (1-\alpha) b_2}(A) = \sum_{A \supseteq \{x\}} \Big[\alpha m_1(A) + (1-\alpha) m_2(A)\Big] \\
&= \alpha \sum_{A \supseteq \{x\}} m_1(A) + (1-\alpha) \sum_{A \supseteq \{x\}} m_2(A) = \alpha pl_1(x) + (1-\alpha) pl_2(x).
\end{aligned}$$

Hence, after denoting by $k_{pl_i} = \sum_{y \in \Theta} pl_i(y)$ the total plausibility of the singletons w.r.t. b_i, the values of the relative plausibility of singletons can be computed as: $\tilde{pl}[\alpha b_1 + (1-\alpha) b_2](x) =$

$$\begin{aligned}
&= \frac{\alpha pl_1(x) + (1-\alpha) pl_2(x)}{\sum_{y \in \Theta} [\alpha pl_1(y) + (1-\alpha) pl_2(y)]} = \frac{\alpha pl_1(x) + (1-\alpha) pl_2(x)}{\alpha k_{pl_1} + (1-\alpha) k_{pl_2}} \\
&= \frac{\alpha pl_1(x)}{\alpha k_{pl_1} + (1-\alpha) k_{pl_2}} + \frac{(1-\alpha) pl_2(x)}{\alpha k_{pl_1} + (1-\alpha) k_{pl_2}} \\
&= \frac{\alpha k_{pl_1}}{\alpha k_{pl_1} + (1-\alpha) k_{pl_2}} \tilde{pl}_1(x) + \frac{(1-\alpha) k_{pl_2}}{\alpha k_{pl_1} + (1-\alpha) k_{pl_2}} \tilde{pl}_2(x).
\end{aligned}$$

$= \beta_1 \tilde{pl}_1(x) + \beta_2 \tilde{pl}_2(x)$. □

Theorem 7 *The relative plausibility operator commutes with convex closure in the belief space: whenever $b_1, ..., b_m \in \mathcal{B}$ are belief functions defined on the same frame, $\tilde{pl}[Cl(b_1, ..., b_m)] = Cl(\tilde{pl}[b_1], ..., \tilde{pl}[b_m])$.*

Proof. The proof follows the structure of that of Theorem 3 and Corollary 3 in [16], on the commutativity of Dempster's rule and convex closure.
Formally, we need to prove that:

(1) whenever $b = \sum_k \alpha_k b_k$, $\alpha_k \geq 0$, $\sum_k \alpha_k = 1$, we have that $\tilde{pl}[b] = \sum_k \beta_k \tilde{pl}[b_k]$ for some convex coefficients β_k;

(2) whenever $p \in Cl(\tilde{pl}[b_k], k)$ (i.e., $p = \sum_k \beta_k \tilde{pl}[b_k]$ with $\beta_k \geq 0$, $\sum_k \beta_k = 1$), there exists a set of convex coefficients $\alpha_k \geq 0$, $\sum_k \alpha_k = 1$ such that $p = \tilde{pl}[\sum_k \alpha_k b_k]$.

Point (1) follows directly from Lemma 2. Proving (2), instead, is equivalent to proving that there exist $\alpha_k \geq 0$, $\sum_k \alpha_k = 1$ such that:

$$\beta_k = \frac{\alpha_k k_{pl_k}}{\sum_j \alpha_j k_{pl_j}} \quad \forall k = 1, ..., m, \tag{23}$$

which is equivalent to:

$$\alpha_k = \frac{\beta_k}{k_{pl_k}} \cdot \sum_j \alpha_j k_{pl_j} \propto \frac{\beta_k}{k_{pl_k}} \quad \forall k = 1, ..., m,$$

as $\sum_j \alpha_j k_{pl_j}$ does not depend on k. If we pick $\alpha_k = \frac{\beta_k}{k_{pl_k}}$ the system (23) is met: by further normalization we obtain as desired. □

It is left to future work to complete this analysis, and check whether other transforms commute with either affine combination or Dempster's rule, therefore enriching our understanding of the two families of transformations.

4 Generalizations of the relative belief operator

A serious issue with the relative belief of singletons is its applicability. In opposition to relative plausibility, $\tilde{b}$ does not exist for a large class of belief functions (those which assign no mass to singletons). Even though this singular case involves only a small fraction of all belief measures (Section 4.1), this issue arises in many practical cases, for instance when using fuzzy membership functions to model the evidence.

4.1 Zero mass to singletons as a singular case

In the binary case $\Theta = \{x, y\}$, according to (7) the only belief function which does not admit a relative belief of singletons is the vacuous one b_Θ: $m_{b_\Theta}(\Theta) = 1$. Its b.b.a. is $m_{b_\Theta}(x) = m_{b_\Theta}(y) = 0$ so that $\sum_x m_{b_\Theta}(x) = 0$ and $\tilde{b}_\Theta$ does not exist. Symmetrically, the pseudo belief function $\varsigma = pl_{b_\Theta}$ (for which $pl_{b_\Theta}(x) = pl_{b_\Theta}(y) = 1$) is such that $pl_{pl_{b_\Theta}} = b_\Theta$, so that $\tilde{pl}_{pl_{b_\Theta}}$ does not exist. Figure 3-left illustrates the location of $\tilde{b}$ in the simple binary case (in which each pseudo belief function can be represented as a vector of $\mathbb{R}^2$), and those of the dual singular points b_Θ, $\varsigma = pl_{b_\Theta}$.
As illustrated by the binary case, the set of belief functions for which $\tilde{b}$ does not exist is a lower-dimensional fraction of the set $\mathcal{B}$ of all belief functions. To prove this, let us compute the region spanned by the most common probability transformations: the plausibility and the pignistic transforms.
Theorem 7 proves that the plausibility transform commutes with convex closure (22). As (by Proposition 4, [9]) the pignistic transform (2) commutes with affine combination, we have that $BetP$ also commutes with Cl:

$$BetP[Cl(b_1, ..., b_m)] = Cl(BetP[b_i], i = 1, ..., m).$$

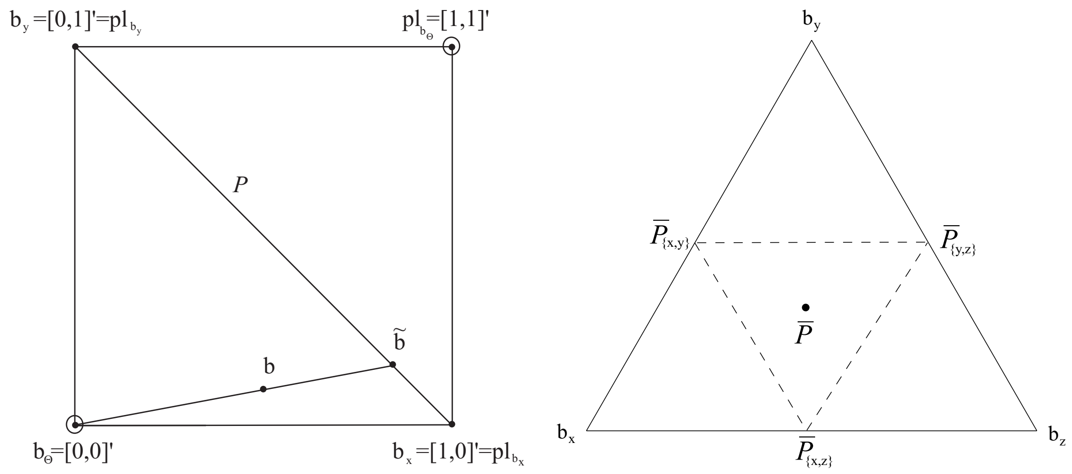

Fig. 3. Left: belief functions $b = [m_b(x), m_b(y)]'$ and plausibility functions $pl_b = [pl_b(x) = 1 - m_b(y), pl_b(y) = 1 - m_b(x)]'$ on $\Theta = \{x, y\}$ can be represented as points of $\mathbb{R}^2$ [11]. The locations of $\tilde{b} = [\frac{m_b(x)}{m_b(x)+m_b(y)}, \frac{m_b(y)}{m_b(x)+m_b(y)}]'$ and the singular points $b_\Theta = [0,0]'$ and $pl_{b_\Theta} = [1,1]'$ are shown. Right: For the class of belief functions $\{b : \sum_x m_b(x) = 0\}$, pignistic function and relative plausibility are allowed to span only a proper subset of the probability simplex (delimited by dashed lines in the ternary case $\Theta = \{x, y, z\}$).

In the case of both transformations, therefore, to determine the image of any convex set $Cl(b_1, ..., b_m)$ of belief functions it is sufficient to compute the transformations of its vertices. The space of all belief functions $\mathcal{B} \doteq \{b : 2^\Theta \rightarrow [0,1]\}$, in particular, is the convex closure of all the categorical belief functions b_A: $\mathcal{B} = Cl(b_A, A \subseteq \Theta)$ [11].
The image of a categorical belief function b_A (a vertex of $\mathcal{B}$) under either plausibility or pignistic transform is:

$$\tilde{pl}_{b_A}(x) = \frac{\sum_{B \supseteq \{x\}} m_{b_A}(B)}{\sum_{B \supseteq \{x\}} m_{b_A}(B)|B|} = \begin{cases} \frac{1}{|A|} & x \in A \\ 0 & else \end{cases} \doteq \overline{\mathcal{P}}_A = \sum_{B \supseteq \{x\}} \frac{m_{b_A}(B)}{|B|} =$$

$= BetP[b_A](x)$, so that $BetP[\mathcal{B}] = Cl(BetP[b_A], A \subseteq \Theta) = Cl(\overline{\mathcal{P}}_A, A \subseteq \Theta) = \mathcal{P} = \tilde{pl}[\mathcal{B}]$. The outputs of both pignistic and relative plausibility transforms span *the whole* probability simplex $\mathcal{P}$.
Consider, however, the set of (singular) belief functions which assign zero mass to singletons. They live in $Cl(b_A, |A| > 1)$ as, according to Equation (12) they have the form $b = \sum_{|A|>1} m_b(A) b_A$, with $m_b(A) \geq 0$, $\sum_{|A|>1} m_b(A) = 1$. The region of $\mathcal{P}$ spanned by their probability transforms is therefore:

$$\begin{aligned} \tilde{pl}[Cl(b_A, |A| > 1)] &= Cl(\tilde{pl}_{b_A}, |A| > 1) = Cl(\overline{\mathcal{P}}_A, |A| > 1) \\ &= Cl(BetP[b_A], |A| > 1) = BetP[Cl(b_A, |A| > 1)]. \end{aligned}$$

The result is illustrated by Figure 3-right in the ternary case $\Theta = \{x, y, z\}$. If (7) is not met, both probability transformations span only a limited region

$$Cl(\overline{\mathcal{P}}_{\{x,y\}}, \overline{\mathcal{P}}_{\{x,z\}}, \overline{\mathcal{P}}_{\{y,z\}}, \overline{\mathcal{P}}_{\Theta}) = Cl(\overline{\mathcal{P}}_{\{x,y\}}, \overline{\mathcal{P}}_{\{x,z\}}, \overline{\mathcal{P}}_{\{y,z\}})$$

of the probability simplex (the triangle delimited by dashed lines in Figure 3-right).

4.2 The family of relative mass probability transformations

One may argue that even though the "singular" case concerns only a small fraction of all belief and probability measures, in many practical application there is a bias towards some particular models which are the most exposed to the problem.
For example, uncertainty is commonly represented using a fuzzy membership function [31]. If the membership function has only a finite number of values, then it is equivalent to a belief function whose focal sets are linearly ordered under set inclusion $A_1 \subset \cdots \subset A_n = \Theta$, $|A_i| = i$, or *consonant belief function* [38,25]. In that case, at most one focal element A_1 is a singleton. So, the vast majority of the useful information in the b.b.a. is contained in the non-singleton focal elements.

Relative belief is in fact only one element of an entire family of probability transformations. Indeed, $\tilde{b}$ can be thought of as the transform which, given a belief function b:

(1) retains the focal elements of size 1 only, yielding an unnormalized belief function;
(2) computes (indifferently) the latter's relative plausibility/pignistic transformation:

$$\tilde{b}(x) = \frac{\sum_{A \supseteq x, |A|=1} m_b(A)}{\sum_y \sum_{A \supseteq x, |A|=1} m_b(A)} = \frac{m_b(x)}{k_{m_b}} = \frac{\sum_{A \supseteq x, |A|=1} \frac{m_b(A)}{|A|}}{\sum_y \sum_{A \supseteq x, |A|=1} \frac{m_b(A)}{|A|}}.$$

Following this scheme, a family of natural generalizations of the relative belief transform is obtained by, given an arbitrary belief function b:

(1) retaining the focal elements of size s only;
(2) computing either the resulting relative plausibility ...
(3) ... or the associated pignistic transformation.

Now, both alternatives 2) or 3) *yield the same probability distribution*. Indeed, the application of the relative plausibility transform yields:

$$p(x) = \frac{\sum\limits_{A\supseteq\{x\}:|A|=s} m_b(A)}{\sum\limits_{y\in\Theta}\sum\limits_{A\supseteq\{y\}:|A|=s} m_b(A)} = \frac{\sum\limits_{A\supseteq\{x\}:|A|=s} m_b(A)}{\sum\limits_{A\subseteq\Theta:|A|=s} m_b(A)|A|} = \frac{\sum\limits_{A\supseteq\{x\}:|A|=s} m_b(A)}{s\sum\limits_{A\subseteq\Theta:|A|=s} m_b(A)},$$

while applying the pignistic transform yields:

$$p(x) = \frac{\sum\limits_{A\supseteq\{x\}:|A|=s} \dfrac{m_b(A)}{|A|}}{\sum\limits_{y\in\Theta}\sum\limits_{A\supseteq\{y\}:|A|=s} \dfrac{m_b(A)}{|A|}} = \frac{s\sum\limits_{A\supseteq\{x\}:|A|=s} m_b(A)}{s\sum\limits_{y\in\Theta}\sum\limits_{A\supseteq\{y\}:|A|=s} m_b(A)}, \tag{24}$$

i.e., the same result. The following natural extension of the relative belief operator is therefore well defined.

Definition 2 *Given a belief function $b : 2^\Theta \rightarrow [0,1]$ with b.b.a. m_b, we call* relative mass transformation *of b of level s the transform $\tilde{M}_s[b]$ which maps b to the probability distribution (24).*

We denote by $\tilde{m}_s$ the output of the relative mass transform of level s.

4.3 Classical probability transformations as convex combinations of relative mass transformations

It is easy too see that both relative plausibility of singletons and pignistic probability are convex combinations of all the (n) relative mass probabilities $\{\tilde{m}_s, s = 1,...,n\}$. Namely, let us denote by $k_{b,s} = \sum_{A\subseteq\Theta:|A|=s} m_b(A)$ the total mass of focal elements of size s, and by $pl_b(x;s) = \sum_{A\supseteq\{x\}:|A|=s} m_b(A)$ the contribution to the plausibility of x of the same size-s focal elements.
Immediately, $\sum_y pl_b(y) =$

$$= \sum_y \sum_{A\supseteq\{y\}} m_b(A) = \sum_{A\subseteq\Theta} m_b(A)|A| = \sum_{r=1}^n r\Big(\sum_{A\subseteq\Theta,|A|=r} m_b(A)\Big) = \sum_{r=1}^n r k_{b,r}.$$

Therefore we obtain for the relative plausibility of singletons the following convex decomposition into relative mass probabilities $\tilde{m}_s$:

$$\begin{aligned}\tilde{pl}_b(x) &= \frac{pl_b(x)}{\sum_y pl_b(y)} = \frac{\sum_s pl_b(x;s)}{\sum_r r k_{b,r}} = \sum_s \frac{pl_b(x;s)}{\sum_r r k_{b,r}} = \sum_s \frac{pl_b(x;s)}{s k_{b,s}}\frac{s k_{b,s}}{\sum_r r k_{b,r}} \\ &= \sum_s \alpha_s \tilde{m}_s(x),\end{aligned} \tag{25}$$

as $\tilde{m}_s(x) = \frac{pl_b(x;s)}{sk_{b,s}}$, whose coefficients

$$\alpha_s = \frac{sk_{b,s}}{\sum_r rk_{b,r}} \propto sk_{b,s} = \sum_y pl_b(y;s)$$

measure for each level s the total plausibility contribution of the focal elements of size s. In the case of the pignistic probability we get:

$$\begin{array}{l} BetP[b](x) = \displaystyle\sum_{A \supseteq \{x\}} \frac{m_b(A)}{|A|} = \sum_s \sum_{A \supseteq \{x\}, |A|=s} \frac{m_b(A)}{s} = \sum_s \frac{1}{s} \sum_{A \supseteq \{x\}, |A|=s} m_b(A) \\ \qquad = \displaystyle\sum_s \frac{1}{s} pl_b(x;s) = \sum_s k_{b,s} \frac{pl_b(x;s)}{sk_{b,s}} = \sum_s k_{b,s} \tilde{m}_s(x), \end{array} \tag{26}$$

where the coefficients $\beta_s = k_{b,s}$ measure for each level s the mass contribution of the focal elements of size s.

4.4 Relative mass transforms as low-cost proxies

Accordingly, the relative mass probabilities can be seen as basic components of both the pignistic and the plausibility transform, associated with the evidence carried by focal elements of a specific size.
As such transforms can be computed just by considering size-s focal elements, they can also be thought of as low-cost proxies for both relative plausibility and pignistic probability, since only the $\binom{n}{s}$ size-s focal elements (instead of the initial 2^n) have to be stored, while all the others can be dropped without further processing.
We can think of two natural criteria for such an approximation of $\tilde{pl}$, $BetP$ via the relative mass transforms:

- (C1) we retain the component s whose coefficient α_s/β_s is the largest in the convex decomposition (25)/(26);
- (C2) we retain the component associated with the *minimal size* focal elements.

Clearly, the relative belief transformation coincides with this second approximation if $\sum_x m_b(x) \neq 0$. When the mass of singletons is nil, instead, the second criterion delivers a natural extension of the relative belief operator:

$$\tilde{b}^{ext}(x) \doteq \frac{\sum_{A \supseteq \{x\}:|A|=min} m_b(A)}{|A|_{min} \sum_{A \subseteq \Theta:|A|=min} m_b(A)}. \tag{27}$$

The two approximation criteria favor different aspects of the original belief function. (C1) focuses (in two different ways) on the strength of the evidence carried by focal elements of equal size. Note that the optimal (C1) approximations of plausibility and pignistic transform are in principle distinct:

$$\hat{s}[\tilde{pl}] = \arg\max_s s k_{b,s}, \quad \hat{s}[BetP] = \arg\max_s k_{b,s}.$$

The best approximation of the pignistic probability is not necessarily the best approximation of the relative plausibility of singletons. Criterion (C2) favors instead the *precision* of the pieces of evidence that make up the belief function b. Let us compare these two approaches in two simple scenarios.

While (C1) is (at least superficially) a sensible, rational principle (the selected proxy must be the greatest contributor to the actual classical probability transformation), (C2) seems harder to justify. Why should one retain only the smallest focal elements, regardless their mass?
The attractive feature of the relative belief of singletons, among (C2) approximations, is its simplicity: the original mass is directly re-distributed onto the singletons. What about the "extended" operator (27)?

Consider a scenario in which we want to approximate the plausibility/pignistic transform of a belief function $b : 2^\Theta \rightarrow [0,1]$, with b.b.a. $m_b(A) = m_b(B) = \epsilon$, $|A| = |B| = 2$, and $m_b(\Theta) = 1 - 2\epsilon \gg m_b(A)$ (Figure 4-left). Its relative plausibility of singletons is given by:

$$\begin{array}{ll} \tilde{pl}_b(x) \propto m_b(A) + m_b(\Theta), & \tilde{pl}_b(y) \propto m_b(A) + m_b(B) + m_b(\Theta), \\ \tilde{pl}_b(z) \propto m_b(B) + m_b(\Theta), & \tilde{pl}_b(w) \propto m_b(\Theta) \ \ \forall w \neq x, y, z. \end{array}$$

Its pignistic probability reads instead as:

$$\begin{array}{ll} BetP(x) = \frac{m_b(A)}{2} + \frac{m_b(\Theta)}{n}, & BetP(y) = \frac{m_b(A)+m_b(B)}{2} + \frac{m_b(\Theta)}{n}, \\ BetP(z) = \frac{m_b(B)}{2} + \frac{m_b(\Theta)}{n}, & BetP(w) = \frac{m_b(\Theta)}{n} \ \ \forall w \neq x, y, z. \end{array}$$

Both transformations have a profile similar to that of Figure 4-right (when assuming $m_b(A) > m_b(B)$). Now, according to criterion (C1), the best approximation (among all relative mass transformations) of both $\tilde{pl}_b$ and $BetP[b]$ is given by selecting the focal elements of size n, i.e., Θ, as the greatest contributor to both the convex sums (25) and (26).
However, it is easy to see that this yields as an approximation the average probability $p(w) = 1/n \ \forall w \in \Theta$, which carries no information at all. In particular, the fact that the available evidence supports to a limited extent the singletons x, y and z is completed discarded, and no decision is possible.
If, on the other hand, we operate according to the criterion (C2), we end up selecting the size-2 focal elements A and B. The resulting approximation is

$$\tilde{m}_2(x) \propto m_b(A), \quad \tilde{m}_2(y) \propto m_b(A) + m_b(B), \quad \tilde{m}_2(z) \propto m_b(B),$$

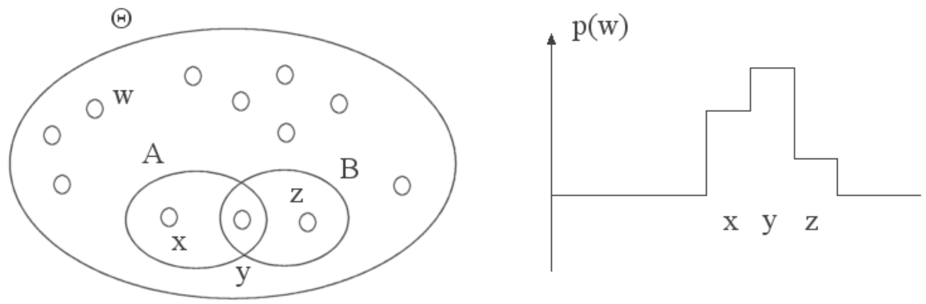

Fig. 4. Left: the original belief function in the first scenario discussed in the text. Right: corresponding profile of both relative plausibility of singletons and pignistic probability.

$\tilde{m}_2(w) = 0 \; \forall w \neq x, y, z$. This has the same profile as that of $\tilde{pl}_b$ or $BetP[b]$ (Figure 4-right): the decision made accordingly corresponds to that made based on $\tilde{pl}_b$ or $BetP[b]$.
We can conclude that, at least in some situations, $\tilde{m}_2 = \tilde{b}^{ext}$ is the best approximation of both plausibility and pignistic transforms in a decision-making sense: we end up making the same decision, at a much lower (in general) computation cost.

Consider however a second scenario, in which a belief function has only two focal elements A and B, with $|A| > |B|$ and $m_b(A) \gg m_b(B)$ (Figure 5-left). Both relative plausibility and pignistic probability have the following values:

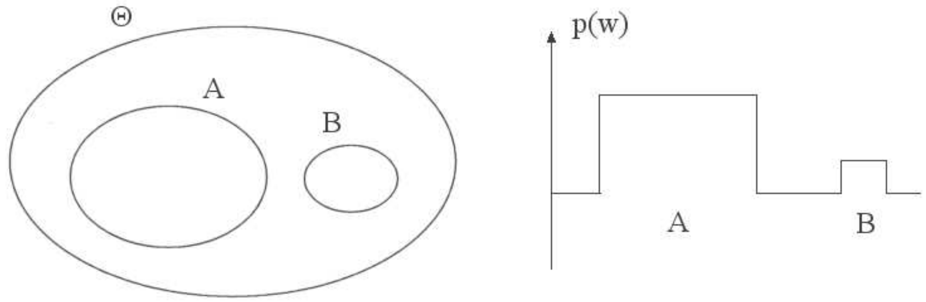

Fig. 5. Left: the belief function of the second scenario. Right: corresponding profile of both relative plausibility of singletons and pignistic probability.

$$\tilde{pl}_b(w) = BetP(w) \propto m_b(A) \; w \in A, \quad \tilde{pl}_b(w) = BetP(w) \propto m_b(B) \; w \in B,$$

and correspond to the profile of Figure 5-right. In this second case, (C1) and (C2) generate the uniform probability on the elements of A (as $m_b(A) \gg m_b(B)$) and the uniform probability on the elements of B (as $|B| < |A|$), respectively. Therefore, in this scenario it is (C1) that yields the best approximation of both plausibility and pignistic transforms in a decision-making perspective.

The second scenario corresponds to a situation in which the evidence is highly conflicting. In such a case we are given two opposite decision alternatives, and it is quite difficult to say which one makes more sense. Should we privilege precision or evidence support?
Some insight on this issue comes from recalling that higher-size focal elements are expression of "epistemic" uncertainty (in Smets' terminology), as they come from missing data/lack of information on the problem at hand. Besides, by their own nature they allow less resolution for decision making (in the second scenario above, if we believe to the result of (C1) we are left uncertain on whether to pick one of $|A|$ outcomes, while if we believe in (C2) the uncertainty is restricted to $|B|$ outcomes). In conclusion, it is not irrational, in case of conflicting evidence, to favor precision over evidence support. This amounts to choosing the approximation criterion (C2), which ultimately supports the case for the relative belief operator and its natural extension (27).

5 Conclusions

In this paper we discussed the rationale of the relative belief transform in both the probability-bound and Dempster-Shafer interpretations of belief functions. Even though neither the relative belief of singletons nor the relative plausibility of singletons are consistent with the original belief function, an interesting rationale in terms of optimal strategies in a non-cooperative game can be attached to such mappings when one assumes a belief function is a set of probability distributions. We proved that relative belief commutes with Dempster's orthogonal sum, meets a number of properties which are the duals of those met by the relative plausibility of singletons, and commutes with convex closure in a similar way as Dempster's rule does, supporting the argument that relative plausibility and belief transform are indeed naturally associated with the D-S framework, and highlighting a classification of probability transformations into two families. To address the issue of its limited applicability, we pointed out that relative belief is just a member of a class of relative mass transformations, which can be interpreted as low-cost proxies for both plausibility and pignistic transforms.

References

[1] M. Bauer, *Approximation algorithms and decision making in the Dempster-Shafer theory of evidence–an empirical study*, International Journal of Approximate Reasoning **17** (1997), no. 2–3, 217–237.

[2] P.L. Bogler, *Shafer-Dempster reasoning with applications to multisensor target*

identification systems, IEEE Transactions on Systems, Man and Cybernetics **17** (1987), no. 6, 968–977.

[3] T. Burger, *Defining new approximations of belief functions by means of Dempster's combination*, Proceedings of the First Workshop on the theory of Belief Functions (BELIEF'10), 2010.

[4] A. Chateauneuf and J.Y. Jaffray, *Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion*, Mathematical Social Sciences **17** (1989), 263–283.

[5] B.R. Cobb and P.P. Shenoy, *A comparison of Bayesian and belief function reasoning*, Information Systems Frontiers **5** (2003), no. 4, 345–358.

[6] ______, *On the plausibility transformation method for translating belief function models to probability models*, International Journal of Approximate Reasoning **41** (2006), no. 3, 314–330.

[7] ______, *A comparison of methods for transforming belief function models to probability models*, Proceedings of ECSQARU'2003, Aalborg, Denmark, July 2003, pp. 255–266.

[8] F. Cuzzolin, *Geometry of upper probabilities*, Proceedings of ISIPTA'03, Lugano, Switzerland, 2003.

[9] ______, *Two new Bayesian approximations of belief functions based on convex geometry*, IEEE Transactions on Systems, Man, and Cybernetics - Part B **37** (2007), no. 4, 993–1008.

[10] ______, *Dual properties of the relative belief of singletons*, Proceedings of the Tenth Pacific Rim Conference on Artificial Intelligence (PRICAI'08), Hanoi, Vietnam, December 2008.

[11] ______, *A geometric approach to the theory of evidence*, IEEE Transactions on Systems, Man, and Cybernetics - Part C **38** (2008), no. 4, 522–534.

[12] ______, *On the structure of simplex of inner Bayesian approximations*, Proceedings of the European Conference on Logics in Artificial Intelligence (JELIA'08), Dresden, Germany, 2008.

[13] ______, *Semantics of the relative belief of singletons*, Workshop on Uncertainty and Logic, Kanazawa, Japan, March 2008.

[14] ______, *The geometry of consonant belief functions: simplicial complexes of necessity measures*, Fuzzy Sets and Systems **161** (2010), no. 10, 1459–1479.

[15] ______, *Three alternative combinatorial formulations of the theory of evidence*, Intelligent Decision Analysis **14** (2010), no. 4, 439–464.

[16] ______, *Geometry of Dempster's rule of combination*, IEEE Transactions on Systems, Man and Cybernetics part B **34** (April 2004), no. 2, 961–977.

[17] M. Daniel, *On transformations of belief functions to probabilities*, International Journal of Intelligent Systems **21** (2006), no. 3, 261–282.

[18] A.P. Dempster, *A generalization of Bayesian inference*, Journal of the Royal Statistical Society, Series B **30** (1968), 205–247.

[19] ———, *Upper and lower probabilities generated by a random closed interval*, Annals of Mathematical Statistics **39** (1968), 957–966.

[20] ———, *A generalization of Bayesian inference*, Classic works of the Dempster-Shafer theory of belief functions, 2008, pp. 73–104.

[21] ———, *Lindley's paradox: Comment*, Journal of the American Statistical Association **77** (1982), no. 378, 339–341.

[22] T. Denoeux, *Inner and outer approximation of belief structures using a hierarchical clustering approach*, Int. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **9** (2001), no. 4, 437–460.

[23] T. Denoeux and A. Ben Yaghlane, *Approximating the combination of belief functions using the Fast Moebius Transform in a coarsened frame*, International Journal of Approximate Reasoning **31** (2002), no. 1-2, 77–101.

[24] J. Dezert and F. Smarandache, *A new probabilistic transformation of belief mass assignment*, 2007.

[25] D. Dubois and H. Prade, *Possibility theory*, Plenum Press, New York, 1988.

[26] R. Fagin and J.Y. Halpern, *Uncertainty, belief and probability*, Proceedings of IJCAI'89, 1989, pp. 1161–1167.

[27] R. Haenni, *Aggregating referee scores: an algebraic approach*, COMSOC'08, 2nd International Workshop on Computational Social Choice (U. Endriss and W. Goldberg, eds.), 2008, pp. 277–288.

[28] R. Haenni and N. Lehmann, *Resource bounded and anytime approximation of belief function computations*, International Journal of Approximate Reasoning **31** (2002), no. 1-2, 103–154.

[29] H.T. Hestir, H.T. Nguyen, and G.S. Rogers, *A random set formalism for evidential reasoning*, Conditional Logic in Expert Systems, North Holland, 1991, pp. 309–344.

[30] J.Y. Jaffray and P.P. Wakker, *Decision making with belief functions: compatibility and incompatibility with the sure-thing principle*, Journal of Risk and Uncertainty **8** (1994), 255–271.

[31] G.J. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*, Prentice Hall PTR, Upper Saddle River, NJ, 1995.

[32] J. Kohlas and P.-A. Monney, *A mathematical theory of hints. an approach to Dempster-Shafer theory of evidence*, Lecture Notes in Economics and Mathematical Systems, vol. 425, Springer-Verlag, 1995.

[33] I. Kramosil, *Approximations of believeability functions under incomplete identification of sets of compatible states*, Kybernetika **31** (1995), no. 5, 425–450.

[34] J.D. Lowrance, T.D. Garvey, and T.M. Strat, *A framework for evidential-reasoning systems*, Proceedings of the National Conference on Artificial Intelligence, 1986, pp. 896–903.

[35] H.T. Nguyen, *On random sets and belief functions*, J. Mathematical Analysis and Applications **65** (1978), 531–542.

[36] E.H. Ruspini, *Epistemic logics, probability and the calculus of evidence*, Proceedings of IJCAI'87, 1987, pp. 924–931.

[37] J. Schubert, *On 'rho' in a decision-theoretic apparatus of Dempster-Shafer theory*, International Journal of Approximate Reasoning **13** (1995), 185–200.

[38] G. Shafer, *A mathematical theory of evidence*, Princeton Univ. Press, 1976.

[39] ———, *Constructive probability*, Synthese **48** (1981), 309–370.

[40] ———, *Perspectives on the theory and practice of belief functions*, International Journal of Approximate Reasoning **4** (1990), 323–362.

[41] P.P. Shenoy, *No double counting semantics for conditional independence*, Working Paper No. 307. School of Business, University of Kansas, 2005.

[42] Ph. Smets, *Belief functions versus probability functions*, Uncertainty and Intelligent Systems (Saitta L. Bouchon B. and Yager R., eds.), Springer Verlag, Berlin, 1988, pp. 17–24.

[43] ———, *Decision making in the TBM: the necessity of the pignistic transformation*, International Journal of Approximate Reasoning **38** (2005), no. 2, 133–147.

[44] ———, *The canonical decomposition of a weighted belief*, Proceedings of IJCAI'95, Montréal, Canada, 1995, pp. 1896–1901.

[45] Ph. Smets and R. Kennes, *The transferable belief model*, Artificial Intelligence **66** (1994), no. 2, 191–234.

[46] T.M. Strat, *Making decisions with belief functions*, Proceedings of UAI'89, 1989, pp. 351–360.

[47] ———, *Decision analysis using belief functions*, International Journal of Approximate Reasoning **4** (1990), 391–417.

[48] J.J. Sudano, *Equivalence between belief theories and nave Bayesian fusion for systems with independent evidential data*, Proceedings of ISIF'03, 2003.

[49] B. Tessem, *Approximations for efficient computation in the theory of evidence*, Artificial Intelligence **61** (1993), no. 2, 315–329.

[50] J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*, Princeton University Press, 1944.

[51] F. Voorbraak, *A computationally efficient approximation of Dempster-Shafer theory*, International Journal on Man-Machine Studies **30** (1989), 525–536.

[52] P. Walley, *Belief function representations of statistical evidence*, Annals of Statistics **15** (1987), 1439–1465.

[53] T. Weiler, *Approximation of belief functions*, IJUFKS **11** (2003), no. 6, 749–777.

[54] H. Xu, Y.-T. Hsia, and Ph. Smets, *The transferable belief model for decision making in the valuation-based systems*, IEEE Transactions on Systems, Man, and Cybernetics part A **26** (1996), 698–707.

[55] A. Ben Yaghlane, T. Denoeux, and K. Mellouli, *Coarsening approximations of belief functions*, Proceedings of ECSQARU (S. Benferhat and P. Besnard, eds.), 2001, pp. 362–373.

[56] L.A. Zadeh, *A simple view of the Dempster-Shafer theory of evidence and its implications for the rule of combination*, AI Magazine **7** (1986), no. 2, 85–90.