Accepted for publication in the International Journal of Approximate Reasoning.

A Decision Theory for Partially Consonant Belief Functions

Phan H. Giang*

George Mason University, 4400 University Dr., Fairfax, VA 22030, USA

Prakash P. Shenoy

University of Kansas, 1300 Sunnyside Ave., Lawrence, KS 66045, USA

Abstract

Partially consonant belief functions (pcb), studied by P. Walley, are the only class of Dempster-Shafer belief functions that are consistent with the likelihood principle of statistics. Structurally, the set of foci of a pcb is partitioned into non-overlapping groups and within each group, foci are nested. The pcb class includes both probability function and Zadeh's possibility function as special cases. This paper studies decision making under uncertainty described by pcb. We prove a representation theorem for preference relation over pcb lotteries to satisfy an axiomatic system that is similar in spirit to von Neumann and Morgenstern's axioms of the linear utility theory. The closed-form expression of utility of a pcb lottery is a combination of linear utility for probabilistic lottery and two-component (binary) utility for possibilistic lottery. In our model, the uncertainty information, risk attitude and ambiguity attitude are treated separately. A tractable technique to extract ambiguity attitude from a decision maker behavior is also discussed.

1. Introduction

In recent years, Dempster-Shafer (DS) belief function theory [3, 21, 23, 28] has drawn an increasing interest in the artificial intelligence and statistics

*Corresponding author

Email addresses: pgiang@gmu.edu (Phan H. Giang), pshenoy@ku.edu (Prakash P. Shenoy)

A preliminary version of this paper appeared in [9] community. The main appeal of the DS theory is its ability to faithfully express a wider class of uncertainty such as the notion of ambiguity that is not expressible by standard probability. Another advantage of DS theory, as its proponents argue, is the close link to evidence, which is the objective source of uncertainty.

The statistical inference problem is an important background for belief function theory. Dempster [3] and Shafer [22] have demonstrated how belief function theory generalizes Bayesian statistical inference. This generalization allows prior knowledge as well as conditional models to be described by belief functions rather than by probability functions.

The inclusion of Bayesian inference as a special case also gives rise to an opportunity for checking the validity of belief function theory against a number of fundamental principles on which statistics is founded. Walley [26] studies two functionals Q and R that represent, in terms of commonality functions, observational and prior evidences, respectively. He finds (see Theorem 3) that in order to be consistent with Bayes rule, observations must be represented by special belief functions. In this class, the set of foci are partitioned. Within each partition, foci are nested. Such belief functions are called *partially consonant* or pcb for short. The partially consonant class is rich enough to include as special cases probability functions (each singleton is a partition) and Zadeh's possibility functions (there is only one element in the partition).

Initially, *consonant* belief functions were used by Shafer [21] to represent statistical evidence. Later, Shafer [22] renounces the idea on the grounds that the set of consonant belief functions is not closed under Dempster's rule of combination. This property is desirable because from a statistical point of view, a series of independent observations can be viewed as a single (compound) observation. So either Dempster's rule is not suitable for combination of independent evidences or the consonant form is not appropriate for representation of evidence. Shafer gives up the latter and keeps the former. However, Walley [26], facing the same choices, comes to a different conclusion. Arguing that (1) the conditions by which Dempster's rule is consistent with Bayes' rule are too restrictive, and (2) Dempster's rule is not unique in satisfying a number of desirable properties for evidence combination, Walley concludes that Dempster's rule is neither suitable for combining independent observations nor for combining prior belief with observational evidence.

One still open problem is the use of belief functions for decision making. The main issue here is that a departure from probability also means the loss of

Bayesian decision theory which ranks alternatives by their expected utilities (EU). A number of proposals for decision making with belief functions have been proposed in literature. One basic idea is to find a transformation that converts a given belief function into a probability function and then use the probability function for decision making [23], [2]. von Neumann-Morgenstern (vNM) linear utility can be brought to use for belief functions, which could be viewed as lower probabilities [14], [16]. Another approach is to use techniques developed for more general uncertainty measure, e.g., lower prevision [27], capacity [19, 20], which includes belief functions as a special class. We will discuss these works in more details in Section 6.

In this paper, we propose a decision theory assuming that uncertainty is represented by a pcb. The paper is structured as follows. The derivation of pcb by Walley is reviewed in the next section. In section 4, after a brief review of the vNM axioms that lead to expected utility (EU) representation of probabilistic lotteries as well as the axioms that lead to binary qualitative utility (QU) representation of possibilistic lotteries, we introduce an axiom system for pcb lotteries and prove a representation theorem. We present one example in section 5. Section 6 discusses related literature. The final section has some concluding remarks.

For convenience, we list an incomplete inventory of notations used in the paper: Θ is used for parameter space; $\mathcal{U}$ - space of decision outcomes; w - an outcome; upper case letters at the beginning of alphabet A, B, C, E, F denote events or subsets of parameter space; upper case letters at the end X, Y, Z for variables; lower case letters x, y, z for their values; lower case letters f, g for acts which also are denoted as set with $\{,\}$ boundaries; L for lottery which also use square bracket boundaries $[,]$; P, p for probability; Π, π for possibility; Pl for plausibility; m for mass function; $\mathcal{L}$ for lottery set; Greek lower letters λ, ρ for the left and right components of a two-component utility; set of such elements is denoted by Ψ; $\succeq$ for preference relation; *Lik* for (normalized) likelihood; For the rest of this paper slash "/" does not denote arithmetic division, it is used to separate degree of uncertainty and associated outcome.

2. Partially Consonant Belief Functions

Walley [26] has derived pcb in the context of the statistical inference problem. The statistical inference problem is described by a triplet of sets $(\mathcal{X}, \Theta, \mathbf{P})$ where $\mathcal{X}$ is the sample space, Θ is the parameter space and $\mathbf{P}$ is the set of uncertainty measures on $\mathcal{X}$ indexed by parameter values in Θ. A statistical evidence/observation/data is a value $x \in \mathcal{X}$. The prior knowledge about parameters may or may not exist. The objective is to make inference about the unobserved parameter $\theta \in \Theta$ of the data generating process.

The Bayesian theory assumes that (1) $\mathbf{P} = \{P_\theta | \theta \in \Theta\}$ is the set of probability functions on $\mathcal{X}$ parameterized by elements of Θ; and (2) prior knowledge exists and is represented by a probability function on Θ. The observational evidence and prior knowledge are then combined by Bayes rule.

The likelihood principle (LP) of statistics states that information contained in an observation x is adequately captured by the likelihood function derived from it. The likelihood of a parameter θ given an observation x is the probability of observing x if θ is the true parameter i.e., $lik_x(\theta) = P_\theta(x)$. Moreover, proportional likelihood functions are equivalent (see, for example, [1] for a detailed discussion).

Dempster [3], and later Shafer [21, 22], arguing that prior knowledge and models can not always be adequately represented by probability, suggest a more general representation using belief functions. For the sake of self-containedness we repeat basic definitions and well known facts about DS theory. A *basic probability assignment* (bpa) or *probability mass* is a function

$$m : 2^\Theta \to [0, 1] \tag{1}$$

such that $m(\emptyset) = 0$ and $\sum_{A \subseteq \Theta} m(A) = 1$. The value $m(A)$ can be interpreted as the probability that a world in A will be the *true* world. A set with positive mass is called *focus*. From a mass function, a number of other functions can be defined

$$Bel(A) \stackrel{\text{def}}{=} \sum_{B \subseteq A} m(B) \tag{2}$$

$$Pl(A) \stackrel{\text{def}}{=} \sum_{B \cap A \neq \emptyset} m(B) \tag{3}$$

$$Q(A) \stackrel{\text{def}}{=} \sum_{A \subseteq B} m(B) \tag{4}$$

Bel is referred to as a *belief* function, *Pl* as a *plausibility* function and Q as a *commonality* function. It should be noted m, Bel, Pl, Q are different forms of the same belief function since starting from any of them, the other three can be completely recovered using Möbius transforms [21].

Belief function theory provides well-defined framework to discuss the notion of ambiguity.

Definition 1. *An event A is said to be* ambiguous *if there exists a focus that intersects with both a A and its complement $\bar{A}$.*

An ambiguous event A is characterized by a strict inequality $Pl(A)+Pl(\bar{A}) > 1$ (equivalently $Bel(A) + Bel(\bar{A}) < 1$). This is the case because (i) there exists a focus B whose strictly positive mass $m(B)$ is counted twice in both $Pl(A)$ and $Pl(\bar{A})$ and (ii) the mass of any other focus is counted at least in either $Pl(A)$ or $Pl(\bar{A})$. Conversely, an *unambiguous* event is characterized by equation $Pl(A)+Pl(\bar{A}) = 1$. If A is (un)ambiguous then so is $\bar{A}$. Intuitively, an unambiguous event and its complement separates the set of foci into two non-overlapping groups and the plausibility of an unambiguous event can be interpreted as probability.

The combination of independent belief functions is done via Dempster's rule. Suppose m_1, m_2 are two belief functions, their combination is another belief function denoted by $(m_1 \oplus m_2)$ defined as follows:

$$(m_1 \oplus m_2)(A) = k^{-1} \sum_{B_1 \cap B_2 = A} m_1(B_1).m_2(B_2) \tag{5}$$

where k is a normalization constant. Since $(1-k)$ can be viewed as the mass assigned to the empty set, it can be interpreted as the extent of inconsistency between m_1 and m_2.

In the special case when m_2 represents the observation B $(m_2(B) = 1)$, $m_1 \oplus m_2$ is called the conditional of m_1 given B $(m_1(\cdot|B))$. In terms of Pl, a conditional belief function assumes a familiar form

$$Pl(A|B) = \frac{Pl(A \cap B)}{Pl(B)} \tag{6}$$

In the statistical inference method argued by Shafer [22], prior knowledge, models and observation are represented in terms of belief functions. The inference is carried by combining these belief functions using Dempster's rule.

Since prior probability and conditional probabilities in Bayesian model are also belief functions, Walley asks under which conditions their combination by Dempster rule is consistent with Bayes rule. He chooses to work with commonality form for convenience. Specifically, Walley studies functional $\mathbf{Q}$ that translates likelihood function which summarizes the Bayesian model (X, Θ, P) and an observation x into a commonality function Q, and functional $\mathbf{R}$ that translates prior probability into a commonality function R. There are a number of desirable properties that $\mathbf{Q}, \mathbf{R}$ (two-place mappings) should satisfy. Some technical assumptions are made. Θ is finite $|\Theta| = N$. $\mathcal{S}$ is the set of likelihood vectors, $\mathcal{P}$ is the set of prior probability vectors. In the axioms listed below (according to Walley's original order), τ, σ stand for arbitrary likelihood vectors in $\mathcal{S}$, ρ for any prior probability in $\mathcal{P}$. Their components are denoted by subscripts. I_B is the characteristic function of subset $B \subseteq \Theta$.

- A1 $\mathbf{Q}(\cdot, \tau)$ is a commonality function over Θ.
- A2 $\mathbf{Q}(\cdot, \tau) \oplus \mathbf{Q}(\cdot, \sigma) = \mathbf{Q}(\cdot, \tau\sigma)$ if $\tau\sigma \in \mathcal{S}$
- A3 $\mathbf{R}(\cdot, \rho)$ is a commonality function over Θ.
- A4 If $\rho_j \tau_j > 0$ for some j, $\mathbf{R}(\cdot, \rho) \oplus \mathbf{Q}(\cdot, \tau) = \mathbf{R}(\cdot, \tau \circ \rho)$ where $\circ$ denotes Bayes' rule.
- A7 $\mathbf{Q}(\cdot, \tau) = \mathbf{Q}(\cdot, c\tau)$ for $0 < c < 1$.
- A8 $\mathbf{Q}(\cdot, \mathbf{1}) \oplus \mathbf{Q}(\cdot, \tau) = \mathbf{Q}(\cdot, \tau)$.
- A9 If $\tau \in \mathcal{S}$ and $\tau I_B \in \mathcal{S}$ then $\mathbf{Q}(A, \tau I_B) \propto \mathbf{Q}(A, \tau)$ when $A \subseteq B$ and $\mathbf{Q}(A, \tau I_B) = 0$ otherwise.

$A1$ and $A3$ require that any observational evidence (likelihood function) and prior probability can be converted into belief function form. $A2$ requires that two views on multiple independent observations as a compound evidence and as the combination of individual evidence are equivalent. $A4$ requires that belief function treatment is consistent with Bayesian treatment when applicable. $A7$ and $A8$ force $\mathbf{Q}$ under Dempster's rule to be consistent with the LP. $A9$ imposes consistency with Bayesian conditioning. Walley has the following theorem:

Theorem 1 (Walley 1987 [26]). *Assumptions A1, A3, A4, A7, A8 and A9 and a number of technical conditions hold if and only if there is some* $\lambda > 0$ *and some partition* $\{A_1, A_2, \ldots A_s\}$ *of* Θ *such that for all* $\rho \in \mathcal{P}$ *and* $\tau \in \mathcal{S}$

$$\boldsymbol{R}(\{\theta_i\}, \rho) = \rho_i^\lambda / \sum_{j=1}^{N} \rho_j^\lambda, \tag{7}$$

$$\boldsymbol{Q}(A,\tau) = k(\tau)\min\{\tau_i^{\lambda}|\theta_i \in A\} \textit{ when } A \in \bigcup_{1\le i\le s} 2^{A_i}, \tag{8}$$

$$\boldsymbol{Q}(\emptyset,\tau) = 1 \textit{ and } \boldsymbol{Q}(A,\tau) = 0 \textit{ otherwise} \tag{9}$$

where $k(\tau) = (\sum_{j=1}^{s}\max\{\tau_i^{\lambda}|\theta_i \in A_j\})^{-1}$.
In addition, assumption $A2$ *is satisfied only if* $s = N$.

It is an interesting result. It says that to be consistent with the likelihood principle and Bayesian updating when applicable, belief function representation must have a special form. It should be mentioned that λ in (8) can be interpreted as the scale parameter and can be used to manipulate the weight of evidence, if necessary. For convenience, we ignore the discounting evidence issue and assume hereafter that $\lambda = 1$.

Our attention is on belief functions that are obtained from statistical evidence i.e., the belief functions obtained from application of functional $\mathbf{Q}$ on likelihood vectors. The reason we ignore belief functions produced by the functional $\mathbf{R}$ is two fold. From a practical point of view, the prior is not always available. From a conceptual point of view, if the prior is available then, by Bayes theorem, (posterior) probability can be computed, therefore, Bayesian decision theory is applicable. In this case, a decision theory for belief functions becomes unnecessary.

It can be seen that (8, 9) of Theorem 1 imply a special arrangement of the foci of a belief function. First, (8) and (9) jointly say that the commonality of a set can be positive if and only if it is a subset of one of $A_1, A_2, \ldots A_s$ that form a partition of Θ. In terms of foci, because the commonality of a focus is positive, therefore it must be a subset of some A_i. Second, for any pair of foci B_1, B_2 that are subsets of A_i, one focus must be a subset of the other in order to satisfy (8). Let us consider the commonality values $\mathbf{Q}(B_1,\tau)$, $\mathbf{Q}(B_2,\tau)$ and $\mathbf{Q}(B_1 \cup B_2,\tau)$. By (8), $\mathbf{Q}(B_1,\tau) = \tau_{j_1}$ where $\tau_{j_1} = \min\{\tau_k|\theta_k \in B_1\}$, $\mathbf{Q}(B_2,\tau) = \tau_{j_2}$ where $\tau_{j_2} = \min\{\tau_l|\theta_l \in B_2\}$ and $\mathbf{Q}(B_1 \cup B_2,\tau) = \tau_j$ where $\tau_j = \min\{\tau_m|\theta_m \in B_1 \cup B_2\}$. Hence, either $\mathbf{Q}(B_1,\tau) = \mathbf{Q}(B_1 \cup B_2,\tau)$ or $\mathbf{Q}(B_2,\tau) = \mathbf{Q}(B_1 \cup B_2,\tau)$. From $\mathbf{Q}(B_1,\tau) = \mathbf{Q}(B_1 \cup B_2,\tau)$, it follows that $B_1 = B_1 \cup B_2$, equivalently $B_2 \subset B_1$. Suppose the contrary $B_1 \neq B_1 \cup B_2$. Since (i) any focus that is superset of $B_1 \cup B_2$ is also a superset of B_1 and (ii) focus B_1 is superset of itself but is not a superset of $B_1 \cup B_2$, the sum of masses of foci that go into $\mathbf{Q}(B_1,\tau)$ is strictly larger than the sum of masses that go into $\mathbf{Q}(B_1 \cup B_2,\tau)$ i.e., $\mathbf{Q}(B_1,\tau) > \mathbf{Q}(B_1 \cup B_2,\tau)$. This contradicts condition $\mathbf{Q}(B_1,\tau) = \mathbf{Q}(B_1 \cup B_2,\tau)$. A belief function that satisfies (8) and (9) is called a partially consonant belief function.

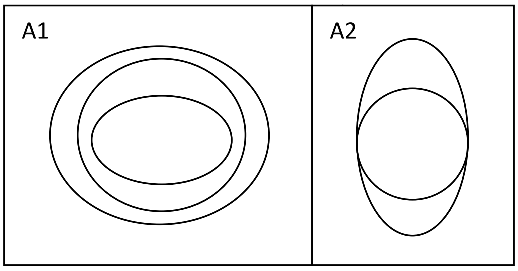

Figure 1: A partially consonant belief function: foci are depicted by ovals.

Walley's equations (8) and (9) characterize pcb in terms of commonality function. In the subsequent exposition, the mass and plausibility forms of belief functions are used instead of the commonality form. For reference convenience, we reformulate the definition of pcb in terms of mass function.

Definition 2. *A belief function defined on* Θ *is called* partially consonant *if its foci can be divided into groups such that* (*a*) *the foci of different groups do not intersect and* (*b*) *the foci of the same group are nested.*

The pcb concept includes as a special case the concept of *consonant* belief function considered in [21] in which foci are nested, one inside another. For a consonant belief function, its plausibility form Pl satisfies max-decomposition property $Pl(A \cup B) = \max(Pl(A), Pl(B))$ for $A, B \subseteq \Theta$. This property together with a normalization condition $Pl(\Theta) = 1$ ensure that the plausibility form of a consonant belief function satisfies characterizing axioms of the possibility measure which has origin in fuzzy set theory [29] and has been extensively studied by Dubois and Prade and others (see for example [5]). In this paper, however, we use the term *possibility function* according to the following definition.

Definition 3. *Suppose a belief function has nested foci* $B_1 \subset B_2 \subset \ldots B_n$*, then its plausibility function* Pl *is called* possibility function.

Note that the conditioning notion derived by Dempster's rule applied for consonant belief functions corresponds to the notion of *quantitative* conditioning often described in possibility theory literature [5].

Because the satisfaction of Dempster's rule requires evidence presented as probability ($s = N$), Walley suggests that Dempster's rule cannot be used to combine independent observations. However, all the axioms above (including

$A2$) are satisfied if Dempster's rule is replaced by another rule ($\otimes$) defined as follows.

$$(Q_1 \otimes Q_2)(A) \stackrel{\text{def}}{=} \begin{cases} k \min\{Q_1(\theta)Q_2(\theta) | \theta \in A\} & \text{if } Q_1(A).Q_2(A) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where k is a constant selected so that Q_{12} is a commonality function.

The significance of Walley's result is that (1) it points out the incompatibility between Dempster's rule and the likelihood principle in general[2], and (2) it isolates a subclass of belief functions representing likelihood information that is consistent with the likelihood principle and Bayesian inference (applicable when prior is available). The pcb class includes probability functions and possibility functions as special cases ($s = N$ and $s = 1$ respectively). In the intermediate case $1 < s < N$, pcb has a remarkable interpretation: it can be viewed as a model for several possibilistic variables conditioned on a probabilistic variable.

Given the importance of this result, it is somewhat puzzling that pcb has not received the attention it deserves. This situation could be explained by the fact that Walley himself seems to dismiss pcb usefulness on the ground that the sure-loss or "Dutch book" argument can still be made against the use of pcb in decision making. To reach this conclusion, Walley assumes that functions *Bel* and *Pl* of a pcb are interpreted as lower and upper betting rates. In the subsequent sections, we develop a decision theory with pcb which voids this Dutch book argument. Specifically, we argue that Walley's interpretation of plausibility as a upper betting rate is not justified for DS functions that come from statistical evidence. The difference between our treatment of belief functions and Walley's betting rate mirrors the difference between the notions of statistical likelihood and probability. It is obvious that given prior probability, statistical likelihood and probability are functionally dependent via Bayes formula. However, in the absence of prior, likelihood can not be treated as probability/betting rate.

[2]It is necessary to note that for *conditioning*, the most important special case of evidence combination (when new evidence is an observation B i.e., $Q_2(\theta) = 1$ if $\theta \in B$ and $Q_2(\theta) = 0$ otherwise), Dempster's rule and Walley's rule are equivalent ($Q_1 \oplus Q_2 = Q_1 \otimes Q_2$). For the rest of this work, we are interested in conditioning only and therefore Dempster's rule is completely adequate. In other words, results in this paper are not pre-conditioned on the rejection of Dempster's rule.

3. Decomposition of partially consonant belief function

In this section, we assume a pcb in the form of a plausibility function. First we prove some useful properties of pcb.

Definition 4. *Suppose* $\{A_i\}_{i=1}^s$ *is a partition of* Θ. *A pcb with foci* A_{ij} *with structure* $A_i = A_{i0} \supset A_{i1} \supset \dots A_{ik_i}$ *for* $1 \leq i \leq s$. $\mathcal{A}$ *denotes the algebra formed from collection* $\{A_1, A_2, \dots A_s\}$, *and* $\mathcal{A}_i$ *- the algebra formed from the elements of* A_i.

In this definition, we make a simplifying assumption each element of Θ must be in at least one focus.

Definition 5. (i) *An event* $B \subseteq \Theta$ *is called* possibilistic *wrt a plausibility function Pl if* $Pl(\cdot|B)$ *is a possibility function.*
(ii) *An event* $B \subseteq \Theta$ *is called* maximally possibilistic *if* B *is possibilistic and any strict superset of* B *is not a possibilistic event.*

Lemma 1. (a) *The plausibility function of a pcb in definition 4 has the form*

$$Pl(A) = \sum_{i=1}^{s} \max\{Pl(\theta) \mid \theta \in A_i \cap A\} \tag{11}$$

(b) *A conditional pcb is again a pcb.*
(c) *An event* B *is unambiguous iff* $B \in \mathcal{A}$.
(d) $B \subseteq A_i$ *for some* i *iff* B *is possibilistic.*

Proof: (a). Define $K(A) = \{A_{ij} | A_{ij} \cap A \neq \emptyset\}$ - the collection of foci that intersect with A. By definition

$$Pl(A) = \sum_{A_{ij} \in K(A)} m(A_{ij}) \tag{12}$$

$$Pl(A_i \cap A) = \sum_{A_{ij} \in K(A_i \cap A)} m(A_{ij}) \tag{13}$$

Since $\{A_i\}_{i=1}^s$ is a partition, $A = \cup_{i=1}^s A \cap A_i$. If $A_{ij} \cap A \neq \emptyset$ then $A_{ij} \cap (A_i \cap A) \neq \emptyset$ and $A_{ij} \cap (A_k \cap A) = \emptyset$ for $k \neq i$. In other words, if A_{ij} intersects with A then it intersects with $(A_i \cap A)$ only. It implies that $K(A) = \cup_{i=1}^s K(A \cap A_i)$ and $K(A \cap A_i) \cap K(A \cap A_k) = \emptyset$. From eqs. (12), (13) we have

$$Pl(A) = \sum_{i=1}^{s} Pl(A_i \cap A).$$

Now consider $Pl(A_i \cap A)$. Since the foci inside A_i are nested, for each A there is an innermost focus $A_{i\ell}$ for some ℓ it intersects with. So $Pl(A_i \cap A) = \sum_{j=0}^{\ell} m(A_{ij})$. For $\theta \in A \cap A_{i\ell}$, clearly $Pl(\theta) = \sum_{j=0}^{\ell} m(A_{ij})$. So, $Pl(A_i \cap A) = Pl(\theta)$. Since $\forall \theta \in A_i \cap A$, $Pl(A_i \cap A) \geq Pl(\theta)$, combine both facts we have

$$Pl(A_i \cap A) = \max\{Pl(\theta) | \theta \in A_i \cap A\}.$$

(b). By Dempster rule in eq. (5), conditioning on B creates new foci which are $A_{ij} \cap B$. Clearly, if A_{ij} have pcb structure then so do the conditional foci.
(c). Suppose B is unambiguous. This means there is no focus A_{ij} such that $A_{ij} \cap B \neq \emptyset$ and $A_{ij} \cap \bar{B} \neq \emptyset$. Equivalently, for any focus A_{ij} either $A_{ij} \cap B = \emptyset$ or $A_{ij} \cap \bar{B} = \emptyset$. This means $A_{ij} \subseteq \bar{B}$ or $A_{ij} \subseteq B$. Take $j = 0$, for all i either $A_i = A_{i0} \subseteq \bar{B}$ or $A_i = A_{i0} \subseteq B$. Since $\{A_i\}_{i=1}^{s}$ is a partition of Θ, $B = \cup_{i=1}^{s}(B \cap A_i) = \cup\{A_i | A_i \subseteq B\}$. If B is a union of some A_i then no focus can intersect with both B and $\bar{B}$ and hence it is unambiguous.
(d). If $Pl(\cdot|B)$ is a possibility function, then the foci of this conditional pcb are nested. The conditional foci are $A_{ij} \cap B$. Suppose $B \cap A_i \neq \emptyset$ and $B \cap A_k \neq \emptyset$ for $i \neq k$. For $j = 0$, since $A_{i0} = A_i$ and $\{A_i\}$ is a partition of Θ, $(B \cap A_i) \cap (B \cap A_k) = \emptyset$. This contradicts the supposition, therefore $B \subseteq A_i$ for some i. This completes the proof. ■

From part (c), it follows that $\mathcal{A}$ is the algebra of unambiguous events characterized by pcb i.e., restricted to $\mathcal{A}$, Pl is a probability function. From part (d) it is clear that A_i are maximally possibilistic.

Theorem 2. *For the plausibility function Pl given in (11) and* $1 \leq i \leq s$

$$P(A) \stackrel{\text{def}}{=} Pl(A) \text{ for } A \in \mathcal{A} \text{ is a probability function on } \mathcal{A} \quad (14)$$

$$\Pi_i(B) \stackrel{\text{def}}{=} Pl(B|A_i) \text{ for } B \in \mathcal{A}_i \text{ is a possibility function on } \mathcal{A}_i \quad (15)$$

Conversely, if a plausibility function Pl that satisfies (14) and (15) i.e., $Pl(\cdot)$ *is a probability function on* $\mathcal{A}$ *and* $Pl(\cdot|A_i)$ *are possibility functions on* $\mathcal{A}_i$ *then it is a pcb and satisfies (11).*

Proof: $(\Rightarrow)$ Given the plausibility function in (11), we need to show that functions P and Π_i defined by (14) and (15) are probability function on $\mathcal{A}$ and possibility functions on $\mathcal{A}_i$. Obviously, by (14) $P(A) \geq 0$ for any $A \in \mathcal{A}$ and $P(\Theta) = 1$. For P to be a probability, we only need to show that

$P(A \cup A') = P(A) + P(A')$ for $A, A' \in \mathcal{A}$ such that $A \cap A' = \emptyset$. By (11)

$$
\begin{aligned}
P(A \cup A') &= \sum_{i=1}^{s} \max\{Pl(\theta) \mid \theta \in A_i \cap (A \cup A')\} && (16)\\
P(A) &= \sum_{i=1}^{s} \max\{Pl(\theta) \mid \theta \in A_i \cap A\} && (17)\\
P(A') &= \sum_{i=1}^{s} \max\{Pl(\theta) \mid \theta \in A_i \cap A'\} && (18)
\end{aligned}
$$

Because A and A' are disjoint, each A_i cannot belong to both of them. If $\theta \in A_i \cap (A \cup A')$ then either $\theta \in A_i \cap A$ (counted in (17)) or $\theta \in A_i \cap A'$ (counted in (18)) but not both. Therefore, $P(A \cup A') = P(A) + P(A')$.

Next, we need to show that Π_i defined in (15) is a possibility function. It is enough to show that $\Pi_i(A_i) = 1$ and $\Pi_i(B \cup B') = \max(\Pi_i(B), \Pi_i(B'))$ for $B, B' \in \mathcal{A}_i$. By (6)

$$
\begin{aligned}
\Pi_i(A_i) &= \frac{Pl(A_i \cap A_i)}{Pl(A_i)} = 1 && (19)\\
\Pi_i(B \cup B') &= \frac{Pl(A_i \cap (B \cup B'))}{Pl(A_i)} = \frac{Pl(B \cup B')}{Pl(A_i)} && (20)
\end{aligned}
$$

The condition $B, B' \in \mathcal{A}_i$ means $B, B' \subseteq A_i$, therefore by (11)

$$
\begin{aligned}
Pl(B \cup B') &= \max\{Pl(\theta) \mid \theta \in (B \cup B')\} && (21)\\
Pl(B) &= \max\{Pl(\theta) \mid \theta \in B\} && (22)\\
Pl(B') &= \max\{Pl(\theta) \mid \theta \in B'\} && (23)
\end{aligned}
$$

This means $\Pi_i(B \cup B') = \max(\Pi_i(B), \Pi_i(B'))$.

($\Leftarrow$) Suppose $Pl(\cdot)$ is a probability function (denoted by P) on $\mathcal{A}$ and s conditionals $Pl(\cdot|A_i)$ are possibility functions (denoted by Π_i) on $\mathcal{A}_i$, we will show that Pl is a pcb and satisfies (11).

We use two well known facts: (1) the foci of a probability function are singletons in its frame and (2) the foci of a possibility function are nested[3]. From the fact that $\cup_{i=1}^{s} A_i = \Theta$ and the sum $\sum_{i=1}^{s} Pl(A_i) = 1$ we conclude

[3]Specifically, for possibility function Π_i one can stratify it domain A_i into strata $A_{i1}, A_{i2}, \ldots$ in such a way that (1) within each stratum the possibility of any element

that the foci of Pl are constrained within each A_i. Because if there were a focus that intersected with at least two A_i, A_j then the mass of that focus would be counted twice in $Pl(A_i)$ and $Pl(A_j)$ then the sum would be strictly more than 1. Therefore, conditioning on A_i does not change the foci within that set. It follows that the foci of conditional Π_i are also the foci of original Pl. Thus, Pl is a pcb and by lemma 1 satisfies (11).■

Using variables to represent Θ, the decomposition looks even more appealing. One has a probabilistic variable X whose domain has size s i.e. $x_1, x_2, \ldots, x_s$. Conditional on each value x_i one has a possibilistic variable Y_i whose domain has size m_i i.e., $y_{i1}, y_{i2}, \ldots y_{im_i}$. Then each element θ of Θ is characterized by a pair of values $(x_i y_{ik})$ of X and Y. For each i, partition A_i is the set $\{(x_i y_{ik}) | 1 < k < m_i\}$.

Usually, decomposition is not a information-preserving operation i.e., the original whole is not recoverable from the decomposed parts. However, the decomposition of a pcb into a probability function and s possibility functions is information-preserving.

Example 1 Suppose there are two variables X, Y whose domains are $\{x_1, x_2\}$ and $\{y_1, y_2, y_3\}$. A pcb and its decomposition are given as follows.

focus	$\{x_1y_1, x_1y_2, x_1y_3\}$	$\{x_1y_1, x_1y_2\}$	$\{x_1y_1\}$	$\{x_2y_1, x_2y_2\}$	$\{x_2y_1\}$
mass	0.3	0.2	0.1	0.1	0.3

$\Pi(Y\|X)$	y_1	y_2	y_3	$P(X)$
x_1	1.0	0.83	0.5	0.6
x_2	1.0	0.25	0.0	0.4

4. Utilities

In this section we study decision making when information about the uncertain states on which decision consequence depends is represented by a partially consonant belief function. The basic object of study is the preference relation over lotteries on the prize set that are induced by acts. Since the pcb includes both probability and possibility functions as special cases, it is natural to expect that the decision theory to be developed for pcb will

is the same and (2) among strata, possibilities are ordered in descending manner. Given that stratification, the innermost focus is A_{i1} and its mass is the difference between its possibility and one of the next stratum i.e., $\Pi_i(A_{i1}) - \Pi_i(A_{i2})$. The next focus is $A_{i1} \cup A_{i2}$ with mass $\Pi_i(A_{i2}) - \Pi_i(A_{i3})$ and so on.

subsume the theories for probabilistic and possibilistic uncertainty. We start by reviewing these special cases and then move to the general case.

4.1. Decision problem, acts and lotteries

A decision problem under uncertainty is a tuple $(S, \Delta, \mathbf{A}, \mathcal{U})$ where S is assumed to be a finite set. Δ is a uncertainty measure over algebra of subsets of S. We assume that Δ is equipped with a conditionalization operator that for each event $E \subseteq S$ maps Δ into a new uncertainty measure Δ_E on the algebra of subsets of E. $\mathcal{U}$ is the set of prizes. For the sake of clarity, we make the following assumption about prizes.

Assumption 1. (1) *$\mathcal{U}$ is identified to the real unit interval* $[0, 1]$ *and* (2) *the value of prize is measured in the risk-adjusted utility unit.*

This assumption allows us to ignore the scaling and risk attitude issues and focus on the ambiguity. In discussion, we show how risk attitude can be accounted in this framework. Thus, 1 is the best (most desirable) and 0 - the worst (least desirable) prizes in $\mathcal{U}$.

$\mathbf{A}$ is the set of acts which is defined recursively. Each prize $w \in \mathcal{U}$ is an act (called *constant* act or act of depth 0) i.e. $\mathcal{U} = \mathbf{A}^0$. A *simple* act is a mapping of the form $d : S \to \mathbf{A}^0$. The semantics of acts is obtained through the gambling interpretation. Act d is a gamble that if s is the true state of nature then $d(s)$ is the prize. The set of simple acts is denoted by $\mathbf{A}^1$. Clearly, $\mathbf{A}^0 \subset \mathbf{A}^1$ because a prize w can be identified mapping $\forall s \in S, s \mapsto w$.

Another notation for an act is by a set of rules. A rule is denoted by a hook arrow $\hookrightarrow$ to differentiate from a mapping denoted by a straight arrow $\to$. $d = \{E_i \hookrightarrow w_i | i \in I\}$ where I is an index set, $\{E_i | i \in I\}$ is a partition of S and $E_i = d^{-1}(w_i)$ - the event whose occurrence triggers prize w_i.

An act of depth k is defined as a set of rules $\{E_i \hookrightarrow d_i | i \in I\}$ where $d_i \in \mathbf{A}^{k-1}$ - an act of depth no more than $k - 1$. It is a gamble that instead of a prize in $\mathcal{U}$ will deliver a ticket to play another gamble. The act set $\mathbf{A}$ is defined as $\mathbf{A}^\infty$.

Visually, an act is a *rooted tree* [4]. A rule $E_i \hookrightarrow d_i$ corresponds to an edge labeled with E_i from the root to a node which is the root of the tree representing d_i. The leaves of the tree are prizes in $\mathcal{U}$. An edge in a rooted tree has a natural orientation which we use the convention to denote the direction *away* from the root. For example, the tree in figure 2 represents act $\{E_1 \hookrightarrow \{F_1 \hookrightarrow w_1, F_2 \hookrightarrow w_2\}, E_2 \hookrightarrow w_3\}$. The reading of this example is that if E_1 occurs then the decision maker receives a gamble (if F_1 occurs then w_1 and if F_2 then w_2); if E_2 occurs then w_3. Logically, it is equivalent to 3 rules: if E_1 and F_1 then w_1; if E_1 and F_2 then w_2; and if E_2 then w_3 i.e. $\{E_1 \cap F_1 \hookrightarrow w_1, E_1 \cap F_2 \hookrightarrow w_2, E_2 \hookrightarrow w_3\}$.

In general, for a given leaf w_i, suppose the labels on the path from the root to the leaf are $E_1^{(i)}, E_2^{(i)}, \ldots E_k^{(i)}$. The operational semantics of the rules consisting the path reads: "if $E_1^{(i)}$ and $E_2^{(i)}$ and ... $E_k^{(i)}$ occur then prize w_i is delivered" which means a rule $\cap_{j=1}^{k} E_j^{(i)} \hookrightarrow w_i$. Repeating the argument for each leaf, we demonstrate the fact that every act is logically equivalent to a simple act (of depth 1). This is referred to as the *principle of equivalence.* The difference between an act of depth k and its equivalent version of depth 1 is in the order the information about the true state of nature is revealed.

Definition 6. *Two acts f and g are equivalent if the final prizes they deliver are the same no matter which state obtains i.e. $f(s) = g(s)$ for any $s \in S$.*

Example 2 In the fig. 2, S describes possible outcomes of a dice rolling i.e. natural numbers from 1 to 6. E_1 is "an odd number", E_2 is "an even number", F_1 is "greater than 4" and F_2 is "less than or equal to 4". For the gamble on the left, the outcome of rolling is revealed in two steps: the first is whether the number is odd or even, the second is about the magnitude. In the case of even number, the prize is $w_3 = 0$, in the case of odd number, the player gets a ticket to a gamble which resolved on the same dice roll, if the roll is higher than 4 then player gets $w_1 = \$1$, if the number is 4 or less then the player gets $w_2 = \$0.5$.

The gamble on the right has the rules: if the number is even then player gets $w_3 = 0$, if the number is 5 then $w_1 = \$1$, if the number is 1 or 3 then player gets $w_2 = \$0.5$. It is easy to verify that no matter what the outcome the dice rolling turns out, both gambles deliver the same prize and therefore they are equivalent. Note that this equivalence is established before the issue of uncertainty is considered. We insist that the equivalence will hold under any uncertainty measure. ■

In a rooted tree, each node is associated with (a) an event which is the conjunction of events on the path from the tree's root to the node and (b) a subtree that consists of the node itself and all the nodes down the stream.

Suppose d_i is a subtree of d (notation $d[d_i]$) and d_i' is another tree. $d[d_i/d_i']$ denotes a tree obtained by replacing subtree d_i by d_i'.

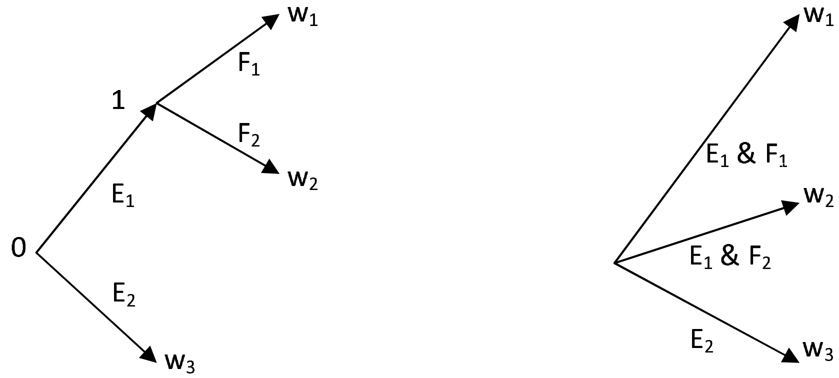

Figure 2: Equivalent acts.

With introduction of uncertainty Δ, a rooted tree (act) gets uncertainty annotation. Each node N is associated with a conditional measure $\Delta_{E_N}(.)$ or $\Delta(.|E_N)$ where E_N is the conjunction of the events labeling the edges on the path from the root to the node. The rationale for this requirement is that a decision maker is moving along a tree as she learns events, at each node, she would update her belief based on information learned so far. In the example of the left tree in fig. 2, after being told that the dice roll is an odd number (E_1) she would move from the root (0) to node 1. At this point, she would condition her belief on that fact.

Each edge labeled with F emanating from N is associated with $\Delta_{E_N}(F)$. For example in the left tree in fig. 2, the root is associated with set S and uncertainty Δ. Node 1 is associated with event E_1 and conditional Δ_{E_1}. The edge leading to 1 is associated with $\Delta(E_1)$, the edge from 1 to w_1 with $\Delta_{E_1}(F_1)$ and so on.

Notice that the path leading to w_1 on the left tree of fig. 2 is associated with $\Delta(E_1)$ and $\Delta_{E_1}(F_1)$ while on the right tree, associated with w_1 is $\Delta(E_1 \& F_1)$. All the measures of uncertainty considered subsequently have "chain rule" property $\Delta(E_1 \& F_1) = \Delta(E_1).\Delta_{E_1}(F_1)$. This is another argument for the principle of equivalence.

When uncertainty Δ is a pcb, a node is called *ambiguous* (*unambiguous*) if the event associated with it is ambiguous (unambiguous).

Lottery is a concept closely related to acts. If in a tree representing an act, all the events labeling edges are removed leaving only their uncertainty labels then a lottery is obtained. For example, instead of E_1 to label the edge from the root to node 1 is $\Delta(E_1)$ and in place of F_1 is $\Delta_{E_1}(F_1)$ and so on. In other words, in a lottery, the nature of events are abstracted away. It does not matter to the decision maker if E_i is "dice rolling outcome" or "precipitation quantity at some location on earth" as long as the uncertainty measured by Δ is the same. This assumption is justified by an assumption that uncertainty measure Δ captures all information relevant to the decision problem. For example, act $\{E_1 \hookrightarrow \{F_1 \hookrightarrow w_1, F_2 \hookrightarrow w_2\}, E_2 \hookrightarrow w_3\}$ corresponds to lottery $[\Delta(E_1)/[\Delta_{E_1}(F_1)/w_1, \Delta_{E_1}(F_2)/w_2], \Delta(E_2)/w_3]$. Note the notational difference: in a lottery square brackets '[', ']' are used instead of curly brackets for boundary, and slash '/' is used instead of $\hookrightarrow$.

4.2. Linear utility for probabilistic lottery

When Δ is a probability function on S, the lotteries are called probabilistic. For example $[p_1/w_1, p_2/w_2, \ldots p_k/w_k]$ where $\sum_{i=1}^{k} p_i = 1$. The set of probabilistic lotteries is denoted by $\mathcal{L}_P$.

Von Neumann and Morgenstern's utility theory (the exposition by Luce and Raiffa [17]) considers a preference relation[4] $\succeq_P$ on $\mathcal{L}_P$ that is assumed to satisfy the following axioms.

- P1 (*Strong order on prizes*) Preference $\succeq_P$ on $\mathcal{U}$ is identical with $\geq$ ("greater than or equal to") relation on reals i.e., $w_1 \succeq_P w_2$ iff $w_1 \geq w_2$. In particular, $1 \succ_P 0$ and for all $w \in \mathcal{U}$, $1 \succeq_P w$ and $w \succeq_P 0$.
- P2 (*Reduction of compound lotteries*) Any compound lottery is indifferent to a simple lottery with prizes in $\mathcal{U}$. Moreover, the probability attached to each prize in the simple lottery is the sum of the probabilities of the paths leading to the prize in the compound lottery.
- P3 (*Continuity*) Each prize $w \in \mathcal{U}$ is indifferent to a canonical lottery involving just 1 and 0.
- P4 (*Substitutability*) In any lottery, each prize can be replaced by the canonical lottery that is indifferent to it.
- P5 (*Transitivity*) $\succeq_P$ on $\mathcal{L}_P$ is transitive.
- P6 (*Monotonicity*) $[p/1, (1-p)/0] \succeq_P [p'/1, (1-p')/0]$ iff $p \geq p'$.

Note that the formulation of $P1$ is stronger than the corresponding version found in [17] in the sense that it makes the $\succeq_P$ restricted on $\mathcal{U}$ a *strong* rather than *weak* order. This formulation is the consequence of our assumption that $\mathcal{U}$ measures the risk-adjusted utility.

[4]Strict preference ($\succ_P$) and indifference ($\sim_P$) are defined from $\succeq_P$ as usual.

The first part of $P2$ is justified by the principle of equivalence. The second part of $P2$ says that if a lottery L (a rooted tree) has k paths from the root leading to prize w with path probabilities p_i $1 \leq i \leq k$ then L is indifferent to a simple lottery L' that has probability associated with w: $p_w = \sum_{i=1}^{k} p_i$. This is a result of additivity of probability.

The concept of canonical lottery is used in $P3$, $P4$ and $P6$. A lottery that has the form $[a_1/1, a_2/0]$ where 1 is the best and 0 is the worst prize in $\mathcal{U}$ is called canonical.

Theorem 3 (von Neumann & Morgenstern [24]). *$\succeq_P$ satisfies axioms $P1 - P6$ iff there exists a utility function $u : \mathcal{L}_P \rightarrow [0,1]$ such that $L_1 \succeq L_2$ iff $u(L_1) \geq u(L_2)$. In particular, it satisfies*

$$u([p_1/w_1, p_2/w_2, \ldots, p_n/w_n]) = \sum_{i=1}^{n} p_i.w_i. \tag{24}$$

From (24), $u(w) = w$ for all $w \in \mathcal{U}$. This is the case because the risk attitude has been assumed away and the utility scale is fixed to the unit interval.

4.3. Binary utility for possibilistic lottery

Following vNM approach, in [8, 11] a utility theory is developed for the case when Δ is a possibility function. For example, $[\pi_1/w_1, \pi_2/w_2, \ldots \pi_k/w_k]$ is a possibilistic lottery. π_i is the possibility of getting prize w_i with $\max_i \pi_i = 1$. The set of possibilistic lotteries is denoted by $\mathcal{L}_\Pi$. A preference relation $\succeq_\Pi$ on $\mathcal{L}_\Pi$ satisfies the following axioms.

PP1 (*Strong order on prizes*) Preference $\succeq_\Pi$ on $\mathcal{U}$ is identical with $\geq$ ("greater than or equal to") relation on reals i.e., $w_1 \succeq_\Pi w_2$ iff $w_1 \geq w_2$. In particular, $1 \succ_\Pi 0$ and for all $w \in \mathcal{U}$, $1 \succeq_\Pi w$ and $w \succeq_\Pi 0$.

PP2 (*Reduction of compound lotteries*) Any compound lottery is indifferent to a simple lottery with prizes in $\mathcal{U}$. Moreover, the possibility of a prize in the simple lottery is the maximum of the possibilities of the paths leading to the prize in the compound lottery.

PP3 (*Continuity*) Each prize $w \in \mathcal{U}$ is indifferent to a canonical possibilistic lottery involving just 1 and 0.

PP4 (*Substitutability*) In any lottery, each prize can be replaced by the canonical lottery that is indifferent to it.

PP5 (*Transitivity*) $\succeq_\Pi$ on $\mathcal{L}_\Pi$ is transitive.

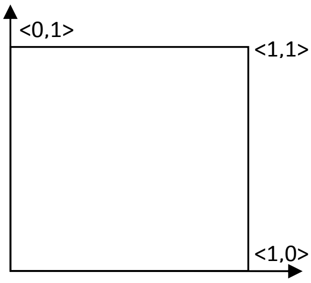

Figure 3: Utility scale Ψ: left component is on horizontal axis, right - vertical.

PP6 (*Monotonicity*) $[\lambda/1, \rho/0] \succeq_\Pi [\lambda'/1, \rho'/0]$ iff $\lambda \geq \lambda'$ and $\rho \leq \rho'$.

This axiom system *PP* has the same structure as the vNM system *P* except for the replacement of probability by possibility which is not additive. In particular, *PP*2 says that if lottery L has k paths from the root leading to prize w with possibilities π_i, $1 \leq i \leq k$ then L is indifferent

Definition 7 (Ordered set of two-component elements).
Let Ψ be the set of two-component (binary) elements such that each component of an element is a real number in the unit interval and the maximum of two components of an element is 1

$$\Psi \stackrel{\text{def}}{=} \{\langle \lambda, \rho \rangle \,|\lambda, \rho \in [0,1], \max(\lambda, \rho) = 1\}$$

An order $\geqq$ together with two operations: component-wise maximization cmax *and product are defined on Ψ as follows:*

$$\langle \lambda, \rho \rangle \geqq \langle \lambda', \rho' \rangle \quad \textit{iff} \;\; \lambda \geq \lambda' \textit{ and } \rho \leq \rho' \tag{25}$$

$$\mathsf{cmax}(\langle \lambda, \rho \rangle, \langle \lambda', \rho' \rangle) \stackrel{\text{def}}{=} \langle \max(\lambda, \lambda'), \max(\rho, \rho') \rangle \tag{26}$$

$$\pi.\, \langle \lambda, \rho \rangle \stackrel{\text{def}}{=} \langle \pi.\lambda, \pi.\rho \rangle \quad \textit{for} \quad 0 \leq \pi \leq 1 \tag{27}$$

Two comments are in order. First, two components of an element can be conveniently thought of as the indices of goodness (left) and badness (right). The more the left index the better and the less the right index the better. The only thing unusual in this case is that the goodness and badness are not additively complementary (if they were a single number would be enough). Thus according to $\geqq$, $\langle 1, 0 \rangle$ is the top element while $\langle 0, 1 \rangle$ is the bottom element. Second, it is *not* necessary $\mathsf{cmax}(\langle \lambda, \rho \rangle, \langle \lambda', \rho' \rangle) \geqq \langle \lambda, \rho \rangle$ nor $\mathsf{cmax}(\langle \lambda, \rho \rangle, \langle \lambda', \rho' \rangle) \geqq \langle \lambda', \rho' \rangle$. Given this notation, there is a representation theorem for possibilistic lottery.

Theorem 4 (Giang & Shenoy 2002). *$\succeq_\Pi$ satisfies axioms $PP1-PP6$ iff there exists a utility function $t : \mathcal{L}_\Pi \rightarrow \Psi$ that satisfies:*

$$(a) \quad \Pi_1, \Pi_2 \in \mathcal{L}_\Pi,\ \Pi_1 \succeq_\Pi \Pi_2 \ \textit{iff}\ t(\Pi_1) \geqq t(\Pi_2) \tag{28}$$

$$(b) \quad t(1) = \langle 1, 0 \rangle\ ; \quad t(0) = \langle 0, 1 \rangle \tag{29}$$

$$(c) \quad t([\pi_1/w_1, \pi_2/w_2, \ldots, \pi_m/w_m]) = \mathsf{c}\max_{1 \leq i \leq m}\{\pi_i.t(w_i)\} \tag{30}$$

The difference between this theorem and one by vNM is that for possibilistic lotteries the binary utility scale Ψ and $\mathsf{c}\max$ are used instead of the real unit interval and summation. The relation between the binary utility and the scalar utility has a parallel in the relation between possibility and probability. In possibility theory, uncertainty of an event is characterized by a pair of necessity and possibility while the probability of an event is a single number. Interested readers are referred to [7] for a detailed discussion.

An issue that needs clarification is the operational semantics of binary utility. In [10, 11] we propose a framework called the *likelihood gamble.* This is a betting framework but instead of using probability, a person uses statistical likelihood information to make her bets. The rules of the game are the following.

R1 A parameter ϵ has two possible values $\{\epsilon_1, \epsilon_2\}$. Each parameter value corresponds to a probability distribution $P_{\epsilon_i}(\cdot)$ on a sample space $\mathcal{X}$. The gambler is told about this information.

R2 A computer selects a value for the parameter and generates an observation according to the corresponding probability distribution. Suppose that the obtained observation is x.

R3 The observation x, but not the value of the parameter, is disclosed to the gambler.

R4 The gambler is offered the following contract:

$$\text{Payoff} \ = \begin{cases} \$1 & \text{if } \epsilon_1 \text{ is the computer selected value} \\ \$0 & \text{if } \epsilon_2 \text{ is the computer selected value} \end{cases}$$

If w is the price that the gambler is willing to pay for the contract then we say she is *indifferent between w and possibilistic lottery* $[Lik_x(\epsilon_1)/1, Lik_x(\epsilon_2)/0]$ where $Lik_x(\epsilon_i)$ is normalized likelihood of observing x if ϵ_i is selected i.e.,

$$Lik_x(\epsilon_i) = \frac{P_{\epsilon_i}(x)}{\max_i(P_{\epsilon_i}(x))}$$

The reason that allows us to make the connection between gambler's choice and her indifference is the likelihood principle. Because likelihood function is sufficient statistic, without loss of information, in rule $R3$, instead of being told about the observation, she can be told about the likelihood function associated with it. And since proportional likelihood functions are equivalent, instead of a likelihood function, one can use its normalized version.

A discernible reader will notice both the similarity and the distinction between this likelihood gamble and the classic coin tossing gamble used to assess linear utility. The main difference is that in coin tossing gamble, the rewards depend on the states (head/tail) for which gambler knows probability while in the likelihood gamble she knows only the likelihood of undisclosed parameter value.

In this likelihood gamble approach, the possibility degree of an event - the plausibility of a consonant belief function - is not treated as an upper betting rate (probability) as suggested by Walley [27] but as the (statistical) likelihood. Obviously, Walley's sure-loss argument against the use of pcb has no force with respect to likelihood gambles.

Another remark is that as it is the case with linear utility for probabilistic lotteries, the possibilistic utility function (eq. (30)) is determined by its values on $\mathcal{U}$. The behavior of t on $\mathcal{U}$ reveals what we call *attitude toward ambiguity.* Let us denote by $t_{\mathcal{U}}$ the restriction of t on $\mathcal{U}$ i.e. $t_{\mathcal{U}} : \mathcal{U} \to \Psi$ defined as $t_{\mathcal{U}}(w) \mapsto t(w)$. $t_{\mathcal{U}}$ can be viewed as a pair of functions $\langle \ell, r \rangle$ where $\ell : \mathcal{U} \to [0, 1]$ and $r : \mathcal{U} \to [0, 1]$. However, not all function pairs can be $t_{\mathcal{U}}$ because of constraint $\max(\ell(x), r(x)) = 1$. In general there is no requirement for ℓ and r to be continuous. For example, the following is perfectly acceptable.

$$\ell(x) = \begin{cases} 1 & \text{if } x \geq 0.5 \\ 1.4x & \text{if } x < 0.5 \end{cases} \quad \text{and} \quad r(x) = \begin{cases} 1.6(1-x) & \text{if } x \geq 0.5 \\ 1 & \text{if } x < 0.5 \end{cases}$$

In this paper, for the sake of regularity, we make the following assumption.

Assumption 2. *t, as a vector function, is continuous in both components i.e., for component functions ℓ and r of t $\lim_{x\to a} \ell(x) = \ell(a)$ and $\lim_{x\to a} r(x) = r(a)$ for $a \in [0, 1]$.*

Lemma 2. *Under continuity assumption of t, $t(\mathcal{U}) = \Psi$.*

Proof: Suppose, on the contrary, there is $\langle \lambda, \rho \rangle$ such that there does not exist any $w \in \mathcal{U} = [0, 1]$ for which $t(w) = \langle \lambda, \rho \rangle$. Consider two sets $\mathcal{U}_1 =$

$\{w \in [0,1] | \ \langle \lambda, \rho \rangle \geqslant t(w)\}$ and $\mathcal{U}_2 = \{w \in [0,1] | \ t(w) \geqslant \langle \lambda, \rho \rangle\}$. Because the order $\geqslant$ on Ψ is complete which is directly derived from the definition, for any $w \in \mathcal{U}$ either $w \in \mathcal{U}_1$ or $w \in \mathcal{U}_2$ i.e. $\mathcal{U}_1 \cup \mathcal{U}_2 = \mathcal{U}$. Also $\mathcal{U}_1 \cap \mathcal{U}_2 = \emptyset$ because if the intersection is not empty then it is equal to $\langle \lambda, \rho \rangle$. It must be the case that $\sup \mathcal{U}_1 = \inf \mathcal{U}_2$ because otherwise a w_3 in between would belong to neither $\mathcal{U}_1$ nor $\mathcal{U}_2$ which contradicts $\mathcal{U}_1 \cup \mathcal{U}_2 = \mathcal{U}$. Denote $w^* = \sup \ \mathcal{U}_1 = \inf \ \mathcal{U}_2$, let $t(w^*) = \langle \lambda', \rho' \rangle$. For sequences w_{1i} in $\mathcal{U}_1$ with $\lim_{i \to \infty} w_{1i} = w^*$ and w_{2i} in $\mathcal{U}_2$ with $\lim_{i \to \infty} w_{2i} = w^*$, because of continuity assumption $\lim_{i \to \infty} \ell(w_{1i}) = \lim_{i \to \infty} \ell(w_{2i}) = \ell(w^*) = \lambda'$. On the one hand, because $w_{1i} \in \mathcal{U}_1$, $\ell(w_{1i}) \leq \lambda$ so at the limit $\lambda' \leq \lambda$. On the other hand, because $w_{2i} \in \mathcal{U}_2$, $\ell(w_{2i}) \geq \lambda$ and at the limit $\lambda' \geq \lambda$. Thus, $\lambda' = \lambda$. Similarly it can be shown that $\rho' = \rho$. So w^* is the point in $\mathcal{U}$ with $t(w^*) = \langle \lambda, \rho \rangle$. This contradicts the supposition that there does not exist any $w \in \mathcal{U} = [0,1]$ for which $t(w) = \langle \lambda, \rho \rangle$ and completes the proof. ■

Recall that $\mathcal{U} \subset \mathcal{L}_\Pi$ because $w \in \mathcal{U}$ can be identified with lottery $[1/w]$. By the representation result, $t(L) = t(w)$ for any $L \in \mathcal{L}_\Pi$ such that $L \sim_\Pi w$. Technically, the inverse of t is undetermined. However, abusing the notation slightly we *define* a function.

Definition 8. *Function $t^{-1} : \Psi \to \mathcal{U}$ is defined by the condition*

$$t^{-1}(t(w)) \mapsto w \quad \textit{for} \quad w \in \mathcal{U}. \tag{31}$$

This implicit function is well defined. For any $\langle \lambda, \rho \rangle \in \Psi$ by lemma 2, there exists $w \in \mathcal{U}$ such that $t(w) = \langle \lambda, \rho \rangle$ and therefore by eq. (31) $t^{-1}(\langle \lambda, \rho \rangle) = w$. Moreover w that satisfies $t(w) = \langle \lambda, \rho \rangle$ is unique because if, on the contrary, there were another $w' \neq w$ such that $t(w') = \langle \lambda, \rho \rangle$. By representation theorem, $w \sim_\Pi w'$. This would violate the strong order on prizes $PP1$.

While t^{-1} is defined to satisfy the cancellation $t^{-1}(t(w)) = w$. The cancellation tt^{-1} also holds. Suppose that $\langle \lambda, \rho \rangle = t(w)$ for some $w \in \mathcal{U}$. Then $t^{-1}(\langle \lambda, \rho \rangle) = t^{-1}(t(w)) = w$. Applying t on both sides we have:

$$t(t^{-1}(\langle \lambda, \rho \rangle)) = t(w) = \langle \lambda, \rho \rangle \tag{32}$$

We use the cancellations to derive the certainty equivalence of a possibilistic lottery. An application of t^{-1} on both sides of (30) yields

$$\begin{aligned} t^{-1}(t([\pi_1/w_1, \pi_2/w_2, \ldots, \pi_m/w_m])) = &\ t^{-1}(\mathsf{c}\max_{1 \leq i \leq m}\{\pi_i . t(w_i)\}) & (33) \\ = &\ u \ \text{ for some } \ u \in \mathcal{U} & (34) \end{aligned}$$

Applying t on both sides of (34) and using tt^{-1} cancellation, we have

$$t([\pi_1/w_1, \pi_2/w_2, \ldots, \pi_m/w_m]) = t(u).$$

By Theorem 4, $[\pi_1/w_1, \pi_2/w_2, \ldots, \pi_m/w_m] \sim_\Pi u$. Thus, we arrive at a useful view: $u = t^{-1}(\mathsf{c}\max_{1\le i\le m}\{\pi_i.t(w_i)\})$ is the *certainty equivalence* or *scalar* utility of $[\pi_1/w_1, \pi_2/w_2, \ldots, \pi_m/w_m]$.

Lemma 3. *t^{-1} is strictly increasing*

$$t^{-1}(\langle\lambda, \rho\rangle) \ge t^{-1}(\langle\lambda', \rho'\rangle) \text{ iff } \langle\lambda, \rho\rangle) \gtrsim \langle\lambda', \rho'\rangle)$$

Proof: For $w, w' \in \mathcal{U}$, $t(w) \gtrsim t(w')$ iff $w \succeq_\Pi w'$ iff $w \ge w'$. The first link is due to Theorem 4, the second is due to axiom $PP1$. Setting $w = t^{-1}(\langle\lambda, \rho\rangle)$ and $w' = t^{-1}(\langle\lambda', \rho'\rangle)$ and using tt^{-1} cancellation we have the lemma. ∎

4.4. Utility representation of pcb lottery

Let us now turn to the situation described by a partially consonant belief function (Δ is a pcb and denoted by Pl). Specifically, Pl is a plausibility function over Θ. For example, $[a_1/w_1, a_2/w_2, \ldots, a_k/w_k]$ denotes a pcb lottery where $a_i = Pl(d^{-1}(w_i))$ is the plausibility of an event whose occurrence triggers prize w_i. Notice that while probabilistic and possibilistic lotteries have normalization conditions ($\sum_i a_i = 1$ and $\max_i a_i = 1$), for a pcb lottery a necessary condition is $\sum_i a_i \ge 1$.

Denote by $\mathcal{L}$ the set of (pcb) lotteries. Since pcb class includes both probability and possibility functions, we have inclusions: $\mathcal{L}_P \subset \mathcal{L}$ and $\mathcal{L}_\Pi \subset \mathcal{L}$. As before we are interested in a preference relation $\succeq$ on $\mathcal{L}$ that is a weak order (complete and transitive). Because $\mathcal{L}_P$ and $\mathcal{L}_\Pi$ are subsets of $\mathcal{L}$ we can consider the restrictions of $\succeq$ on $\mathcal{L}_P$ and $\mathcal{L}_\Pi$ denoted by $\succeq_P$ and $\succeq_\Pi$ respectively. We want $\succeq_P$ to satisfy vNM axioms $P1$ through $P6$ and $\succeq_\Pi$ satisfies axioms $PP1$ through $PP6$. Note that $\mathcal{L}_P \cap \mathcal{L}_\Pi$ is the set of constant lotteries. To see that suppose $[a_1/w_1, a_2/w_2, \ldots, a_k/w_k] \in \mathcal{L}_P$ then the normalization condition is $\sum_{i=1}^k a_i = 1$. At the same time $[a_1/w_1, a_2/w_2, \ldots, a_k/w_k] \in \mathcal{L}_\Pi$ implies that $\max_i a_i = 1$. Since $a_i \ge 0$ it follows that $a_i = 1$ for some i and $a_j = 0$ for $i \ne j$. Thus the $[a_1/w_1, a_2/w_2, \ldots, a_k/w_k]$ is the same a w_i. Naturally, because $\succeq_P$ and $\succeq_\Pi$ are restrictions of the same $\succeq$ on $\mathcal{L}_P$ and $\mathcal{L}_\Pi$, they must be identical on the intersection of $\mathcal{L}_P$ and $\mathcal{L}_\Pi$ which is $\mathcal{U}$ i.e. $w \succeq_P w'$ iff $w \succeq_\Pi w'$ iff $w \succeq w'$ (by definition). Now we describe a list of axioms which amount to pulling together $P1 - P6$ and $PP1 - PP6$ for relation $\succeq$ on $\mathcal{L}$.

B1 (*Strong order on prizes*) Preference $\succeq$ on $\mathcal{U}$ is identical with $\geq$ ("greater than or equal to") relation on reals i.e., $w_1 \succeq w_2$ iff $w_1 \geq w_2$.

B2 (*Reduction of compound lotteries*) Any compound probabilistic (possibilistic) lottery is indifferent to a simple probabilistic (possibilistic) lottery with prizes in $\mathcal{U}$. The plausibilities on the simple probabilistic (possibilistic) are calculated according to belief function calculus.

B3 (*Continuity of prize*) Each prize $w \in \mathcal{U}$ is indifferent to a canonical possibilistic lottery involving just 1 and 0.

B4 (*Substitutability*) In any lottery, a probabilistic lottery on an unambiguous node can be replaced by another probabilistic lottery indifferent to it; a possibilistic lottery can be replaced by an indifferent possibilistic lottery.

B5 (*Transitivity*) $\succeq$ on $\mathcal{L}$ is transitive.

B6 (*Monotonicity*) For possibilistic (probabilistic) canonical lotteries $[\lambda/1, \rho/0] \succeq [\lambda'/1, \rho'/0]$ iff $\lambda \geq \lambda'$ and $\rho \leq \rho'$.

B7 (*Equivalence between two forms*) Each canonical possibilistic lottery is indifferent to a canonical probabilistic lottery.

$B1$, the same as $P1$ or $PP1$, is about the preference order on the set of prizes is the same as the numerical order. The statement of $B2$ is an aggregation of $P2$ and $PP2$. The second part of $B2$ takes advantage of the fact that the belief function calculus subsumes probabilistic and possibilistic calculi. Note that $B2$ does not say how a generic pcb compound lottery is converted to a simple one. The reason for this silence is that in general case, belief function theory does not have an operator that calculates the plausibility of the union of two events directly from the plausibilities of each events.

Axiom $B3$ is identical to $PP3$ which requires that each prize $w \in \mathcal{U}$ is indifferent to a canonical possibilistic lottery.

Axiom $B4$ implies both $P4$ and $PP4$. In a probabilistic lottery, every node is unambiguous, so the unambiguity qualification is not necessary. In the case of pcb lottery, this qualification becomes necessary because besides unambiguous nodes, there are ambiguous nodes for which the substitutability may not hold. Restriction of probabilistic lottery substitutability for unambiguous nodes only is similar to weakening of sure-thing principle considered in [16].

Specifically, if (1) L contains probabilistic L_1 and the node associated with L_1 is unambiguous and (2) L_2 is a probabilistic lottery such that $L_1 \sim L_2$ then $B4$ requires that $L \sim L[L_1/L_2]$ where $L[L_1/L_2]$ is obtained from L by replacing subtree L_1 with tree L_2. Also if L contains possibilistic L_1 and L_2 is a possibilistic lottery such that $L_1 \sim L_2$ then $B4$ requires that $L \sim L[L_1/L_2]$.

$B5$ - transitivity of $\succeq$ is identical to $P5$ and $PP5$. $B6$ is pulling together $P6$ and $PP6$ and ensures the completeness of $\succeq$ which together with $B5$ makes $\succeq$ a weak order. The rationale of $B6$ is summarized as "the more plausible the best the better" and "the less plausible the worst the better". This monotonicity requirement holds for both probabilistic and possibilistic comparisons in the sense that the numbers can be interpreted either as probability or possibility as long as both canonical lotteries are of the same type i.e. they are both possibilistic or probabilistic. $B6$ can not be used to compare a possibilistic lottery with a probabilistic lottery.

$B7$ requires that each canonical possibilistic lottery is indifferent to a canonical probabilistic lottery (a probability distribution on $\{0, 1\}$). The above argument for the equivalence between 1 and possibilistic lottery can be repeated to arrive at the equivalence[5] 1 and probabilistic lottery $[1/1, 0/0]$. Since both possibilistic and probabilistic canonical lotteries span the whole range from 0 to 1, given any canonical possibilistic lottery, there exists a probabilistic canonical lottery equivalent to it.

$B7$ may seems harder to accept if one think of it as the requirement for a decision maker (DM) to be able to switch between the probability and possibility forms of uncertainty attached to the prizes in a canonical lottery. The difficulty, in our opinion, is cognitive rather than intrinsic. A person may be more comfortable thinking in terms of probability than possibility. This is reasonable because of the long history of probability concept and its prevalent exposure in everyday discourse. However this epistemic challenge in no way would void the validity of the assumption. Instead, it points to the need of finding a probability-possibility conversion that is easy to understand.

Our strategy to find utility representation for $\succeq$ is to find representations for $\succeq$ restricted on $\mathcal{L}_\Pi$ and $\mathcal{L}_P$ and then use constant lotteries which are intersection $\mathcal{L}_\Pi \cap \mathcal{L}_P$ to connect the preference. We have a lemma.

Lemma 4. *Suppose preference relation $\succeq$ on $\mathcal{L}$ satisfies $B1 - B7$ then its restriction on $\mathcal{L}_P$ denoted by $\succeq_P$ satisfies $P1 - P6$ and the restriction on $\mathcal{L}_\Pi$ denoted by $\succeq_\Pi$ satisfies $PP1 - PP6$.*

[5] In the canonical lotteries $[0/1, 1/0]$ and $[1/1, 0/0]$ the numbers can be interpreted either as possibility or probability: 0 means impossible/improbable while 1 means sure/certain.

Proof: Let us consider the case of $\succeq_P$. Since $\mathcal{U} \subset \mathcal{L}$, for $w, w' \in \mathcal{U}$, $w \succeq_P w'$ iff $w \succeq w'$. By $B1$, $w \succeq w'$ iff $w \geq w'$. So $\succeq_P$ satisfies $P1$.

$B2$ implies $P2$. Suppose L is a compound lottery and L' is the equivalent lottery obtained from L by collapsing paths from the root to leaves. If (a) L is probabilistic i.e. the uncertainty on each node is described by a (conditional) probability function; (b) the plausibility on each collapsed edge of L' is obtained by multiplying the plausibilities on the path (belief function calculus) then L' is also a probabilistic i.e. the sum of plausibilities on its edges is 1. Since both L and L' are probabilistic, $L \sim_P L'$ iff $L \sim L'$. The latter is guaranteed by $B2$ so $B2$ implies $P2$.

$B3$, $B5$ and $B7$ imply $P3$. For each $w \in \mathcal{U}$, by $B3$ there is a possibilistic canonical C_Π such that $w \sim C_\Pi$. By $B7$ there is a probabilistic canonical C_P such that $C_P \sim C_\Pi$. By transitivity $B5$, $w \sim C_P$.

$B4$ implies $P4$. Suppose $w \sim C_P$ where C_P is a probabilistic canonical lottery, $L[w]$ is a probabilistic lottery containing leaf w and $L[x/C_P]$ is the lottery obtained by substituting w by C_P. Because constant lotteries are probabilistic and any node in a probabilistic lottery is unambiguous, by $B4$, $L[w] \sim L[w/C_P]$. If L is probabilistic then $L[w/C_P]$ is also probabilistic. Therefore, $L[w] \sim_P L[w/C_P]$.

$B5$ implies $P5$. Because $\succeq$ on $\mathcal{L}$ is transitive and $\succeq_P$ is a restriction of $\succeq$ on $\mathcal{L}_P$ then $\succeq_P$ is also transitive.

$B6$ implies $P6$. If $[\lambda/1, \rho/0]$ and $[\lambda'/1, \rho'/0]$ are probabilistic canonical lotteries i.e. $\rho = 1 - \lambda$ and $\rho' = 1 - \lambda'$. By $B6$ $[\lambda/1, \rho/0] \succeq [\lambda'/1, \rho'/0]$ iff $\lambda \geq \lambda'$ and $\rho \leq \rho'$ iff $\lambda \geq \lambda'$ because of additivity. Thus, $[\lambda/1, (1-\lambda)/0] \succeq_P [\lambda'/1, (1-\lambda')/0]$ iff $\lambda \geq \lambda'$.

The case of $\succeq_\Pi$ is similar. ■

Theorem 5. *$\succeq$ on $\mathcal{L}$ satisfies axioms $B1 - B7$ iff (a) and (b) and (c).*

(a) There exists a function $u : \mathcal{L} \to [0,1]$ such that $L \succeq L'$ iff $u(L) \geq u(L')$;

(b) Restricted on $\mathcal{L}_\Pi$, u has the form

$$u([\pi_1/w_1, \pi_2/w_2, \ldots, \pi_m/w_m]) = t^{-1}(\mathsf{cmax}_{1\leq i\leq m}\{\pi_i t(w_i)\})$$

for a function t defined in Theorem 4 and t^{-1} defined in (31);

(c) Suppose $L_i = [\pi_{i1}/w_1, \pi_{i2}/w_2, \ldots \pi_{ik}/w_k]$ for $1 \leq i \leq s$ are possibilistic lotteries, the utility of pcb lottery $[p_1/L_1, p_2/L_2, \ldots p_s/L_s]$ has the form

$$u([p_1/L_1, p_2/L_2, \ldots p_s/L_s]) = \sum_{i=1}^{s} p_i t^{-1}(\mathrm{cmax}_{1\leq j\leq k}\{\pi_{ij} t(w_j)\}) \quad (35)$$

Before proving the theorem, we describe the intuition of the utility expression. First, note that the lottery in part (c) $[p_1/L_1, p_2/L_2, \dots p_s/L_s]$ with p_i are probabilities and L_i are possibilistic lotteries must come from act of the form $\{A_i \hookrightarrow f_{A_i} | 1 \leq i \leq s\}$ where f_{A_i} is an act defined on A_i - the maximal foci of pcb. This is an implication of lemma 1. On the one hand, in order for p_i to be additive each conditioning event must be a union of some foci A_i of the pcb. On the other hand, in order for L_i to be possibilistic, the conditioning event must be a subset of some A_i. To satisfy both conditions the conditioning event must be exact A_i.

Next, we note that unlike its special cases for probabilistic and possibilistic lotteries, for a general pcb lottery $[a_i/w_i]_{i=1}^s$ where a_i are plausibilities the closed form expression of $u([a_i/w_i]_{i=1}^s)$ does not directly include a_i. To compute utility of an arbitrary act (tree), its equivalent simple version is constructed by collapsing the paths leading to prizes. Then formula in eq. (35) is applied for the equivalent version in the form of (c) which is constructed by conditioning the simple act on maximal foci of the pcb.

Third, the key fact that allows a combination two utility representations is that a prize $w \in \mathcal{U}$ has a dual views as probabilistic and possibilistic lottery i.e., $\mathcal{L}_P \cap \mathcal{L}_\Pi = \mathcal{U}$.

Proof: $(\Rightarrow)$ We prove (b), (c) and (a) in that order.

(b) Suppose $\succeq$ on $\mathcal{L}$ satisfies $B1 - B7$. By lemma 4, the restriction on $\mathcal{L}_\Pi$, $\succeq_\Pi$, satisfies $PP1 - PP6$. By Theorem 4, $\succeq_\Pi$ is represented by utility function t given by eq. (30). In particular, $t([\pi_1/w_1, \pi_2/w_2, \dots, \pi_m/w_m]) = \mathsf{c}\max_{1 \leq i \leq m}\{\pi_i t(w_i)\}$. It follows that for $L, L' \in \mathcal{L}_\Pi$, $L \succeq_\Pi L'$ iff $t(L) \geqq t(L')$. By lemma 3, $t(L) \geqq t(L')$ iff $t^{-1}(t(L)) \geq t^{-1}(t(L'))$. So, $L \succeq_\Pi L'$ iff $t^{-1}(t(L)) \geq t^{-1}(t(L'))$. Thus, $t^{-1}(\mathsf{c}\max_{1 \leq i \leq m}\{\pi_i t(w_i)\})$ is the certainty equivalence of $[\pi_1/w_1, \pi_2/w_2, \dots, \pi_m/w_m]$, hence, is its utility.

(c) Suppose $L = [p_1/L_1, p_2/L_2, \dots p_s/L_s]$ where p_i are probabilities i.e., $\sum_{i=1}^s p_i = 1$ and $L_i = [\pi_{i1}/w_1, \pi_{i2}/w_2, \dots \pi_{ik}/w_k]$ are possibilistic lotteries. By part (b), $L_i \sim_\Pi u_i$ for $1 \leq i \leq s$ where $u_i = t^{-1}(\mathsf{c}\max_{1 \leq j \leq k}\{\pi_{ij} t(w_j)\})$. Since $\succeq_\Pi$ is the restriction of $\succeq$, $L_i \sim_\Pi u_i$ also means $L_i \sim u_i$. Because u_i is a possibilistic lottery, axiom $B4$ allows L_i to be substituted by u_i i.e., $[p_1/L_1, p_2/L_2, \dots p_s/L_s] \sim [p_1/u_1, p_2/u_2, \dots p_s/u_s]$. The right hand side is a probabilistic lottery. By lemma 4, $\succeq_P$ the restriction of $\succeq$ on probabilistic lotteries satisfies axioms $P1 - P6$. By Theorem 3, $\succeq_P$ is represented by linear utility function $u([p_1/u_1, p_2/u_2, \dots p_s/u_s]) = \sum_{i=1}^s p_i u_i$. Substituting the expression for u_i we have the certainty equivalence of L i.e.,

$$L \sim \sum_{i=1}^{s} t^{-1}(\mathrm{cmax}_{1\leq i\leq m}\{\pi_i t(w_i)\}) \tag{36}$$

(a) Let us consider a pcb lottery $L_f = [a_1/w_1,\ a_2/w_2,\ \ldots a_k/w_k]$ that is induced by act $f = \{C_1 \hookrightarrow w_1,\ C_2 \hookrightarrow w_2,\ \ldots C_k \hookrightarrow w_k\}$ with $\{C_i\}$ is a partition of Θ and pcb Pl has the focus structure: $A_{i0} \supset A_{i1} \ldots \supset A_{im_i}$ for $1 \leq i \leq s$ such that $\{A_{i0}\}$ is a partition of Θ. The link between the lottery and the act is via $a_i = Pl(C_i)$ for $1 \leq i \leq k$.

Consider the following act: $g = \{A_{10} \hookrightarrow f, A_{20} \hookrightarrow f, \ldots A_{s0} \hookrightarrow f\}$. It is not difficult to show that f and g are equivalent acts. As functions from Θ to $\mathcal{U}$ no matter which $\theta \in \Theta$ is the true parameter, the prize delivered by f and g are the same. Viewing g as a rooted tree, we note (Theorem 2) that restricted on the algebra formed from $\{A_{i0}|1 \leq i \leq s\}$ Pl is additive and conditional plausibilities $Pl(\cdot|A_{i0})$ are possibility functions. So, the lottery induced by g has the form of part (b). Namely, denote $p_i = Pl(A_{i0})$ and $\pi_{ij} = Pl(C_j|A_{i0})$ for $1 \leq i \leq s$ and $1 \leq j \leq k$. The lottery induced from g is $L_g = [p_1/L_1,\ p_2/L_2,\ \ldots p_s/L_s]$ where $L_i = [\pi_{i1}/w_1,\ \pi_{i2}/w_2,\ \ldots \pi_{ik}/w_k]$. By eq. (36) $\sum_{i=1}^{s} t^{-1}(\mathrm{cmax}_{1\leq i\leq m}\{\pi_i t(w_i)\}) \sim L_g$. Due to the principle of equivalence $L_f \sim L_g$ and transitivity of $\succeq$ it follows $\sum_{i=1}^{s} t^{-1}(\mathrm{cmax}_{1\leq i\leq m}\{\pi_i t(w_i)\}) \sim L_f$.

For an arbitrary lottery in $\mathcal{L}$, due to the principle of equivalence, one can find one-stage lottery that is indifferent to. Suppose u and u' are certainty equivalences of single-stage L and L' calculated by eq. (36) i.e. $L \sim u$ and $L' \sim u'$. So $L \succeq L'$ iff $u \succeq u'$. Because of axiom $B1$, $u \succeq u'$ iff $u \geq u'$. Thus, $L \succeq L'$ iff $u \geq u'$.

($\Leftarrow$) Suppose a preference relation $\succeq_u$ is defined by function u given by eq. (35) with continuous t and strictly increasing t^{-1} defined by eq. (31) i.e., $L_1 \succeq_u L_2$ if $u(L_1) \geq u(L_2)$. We show that $\succeq_u$ satisfies B_1 to $B7$. First, we verify the fact that pcb utility (35) subsumes both linear utility for probabilistic lottery and the binary utility for possibilistic utility. If pcb is a probability function, each element $\theta \in \Theta$ is a focus and each L_i is a constant lottery w_i. Because $t^{-1}(t(w_i)) = w_i$, eq. (35) becomes

$$u([p_1/w_1, p_2/w_2, \ldots p_n/w_n]) = \sum_{i=1}^{n} p_i w_i$$

If pcb is a possibility function then $s = 1$ and $p_1 = 1$, eq. (35) reduces to

$$u(L_1) = t^{-1}(\mathrm{cmax}_{1\leq j\leq k}\{\pi_{1j} t(w_j)\})$$

which is the binary utility for a possibilistic lottery wrapped by the strictly increasing t^{-1}. This observation allows us to use Theorems 3 and 4.

$B1$ is satisfied by definition because $u(w) = w$ for $w \in \mathcal{U}$.

$B2$. For a compound probabilistic (possibilistic) lottery L, because $u(L) \in [0, 1]$ as an implication of the observation that u reduces to expected (binary) utility. By definition $L \sim_u u(L)$.

$B3$. By the reverse clause of theorem 4, $\succeq_u$ satisfies $PP3$ which is the same as $B3$.

$B4$. Suppose L contains probabilistic L_1 and the node associated with L_1 is an unambiguous node. L_2 is also a probabilistic lottery such that $L_1 \sim_u L_2$. Because the node in L associated with L_1 is ambiguous, by lemma 1, the event E of that node is a union of some A_i. By definition of $\succeq_u$, $L_1 \sim_u L_2$ implies $u(L_1) = u(L_2)$. $u(L_1)$ and $u(L_2)$ are expected utility because L_1, L_2 are probabilistic. In the expressions of $u(L)$ the part related to L_1 is $p(E) * u(L_1)$. In $u(L[L_1/L_2])$ the part related to L_2 is $p(E) * u(L_2)$. Therefore, $u(L) = u(L[L_1/L_2])$.

Suppose L contains possibilistic L_1^π and L_2^π is another possibilistic such that $L_1^\pi \sim_u L_2^\pi$. Since L_1^π is possibilistic, by lemma 1, the node associated with it is a subset of some A_i. Consider the possibility lottery L' that is obtained from L by conditioning on A_i. Since the node associated with L_1^π is a subset of A_i, L_1^π is a subtree of L'. By Theorem 4, since $L_1^\pi \sim_u L_2^\pi$ and L' is a possibilistic lottery, replacement of L_1^π by L_2^π leads to an indifferent lottery i.e., $L'[L_1^\pi] \sim_u L[L_1^\pi/L_2^\pi]$. So by definition $u(L'[L_1^\pi]) = u(L[L_1^\pi/L_2^\pi])$. In evaluation of $u(L)$ and $u(L[L_1^\pi/L_2^\pi])$, L_1^π and L_2^π enter via $L'[L_1^\pi]$ and $L'[L_1^\pi/L_2^\pi]$ respectively. Thus, $u(L) = u(L[L_1^\pi/L_2^\pi])$ and hence $L \sim_u L[L_1^\pi/L_2^\pi]$.

$B5$. Because u has range in $[0, 1]$ the $\succeq_u$ is transitive.

$B6$. This is an implication of Theorems 3 and 4.

$B7$. Any possibilistic lottery L_π is equivalent to the following canonical probabilistic lottery $[u(L_\pi)/1, (1 - u(L_\pi))/0]$. This completes the proof. ∎

4.5. Binary utility extraction and ambiguity attitude

In practical applications, an operational problem a decision maker (DM) must solve before she can use formula in (35) is how to determine function t. The solution is that it can be determined through the exercising likelihood gambles as described in section 4.3. This activity is analogous to the practice of extracting linear utility function from a DM with an exception that for likelihood gambles the DM should handle the likelihoods obtained from observation, not probability with relative frequency semantics.

As a practical choice, the likelihood gamble game can be repeated for a large enough number of points in Ψ and then the neighbor points are connected to get a whole utility curve.

There is an alternative in which less DM introspection is compensated by an assumption that utility function has a parametric form. Instead of estimating whole empirical utility curve, decision maker's behavioral data is used to estimate the behavioral parameter.

Let us re-examine the likelihood gamble. Suppose that DM is indifferent between value w and having (possibilistic) lottery $[\lambda/1, \rho/0]$ where λ, ρ are normalized likelihoods for (gamble) parameter values θ_1 and θ_2 respectively. This indifference can also be analyzed from a Bayesian point of view which assumes DM has a prior probability, says γ_w, on θ_1. DM's posterior of θ_i would be

$$Pr(\theta_1|x) = \frac{\gamma_w \lambda}{\gamma_w \lambda + (1-\gamma_w)\rho} \qquad Pr(\theta_2|x) = \frac{(1-\gamma_w)\rho}{\gamma_w \lambda + (1-\gamma_w)\rho} \tag{37}$$

Then expected utility principle tells us the indifference implies the following

$$u(w) = Pr(\theta_1|x).u(1) + Pr(\theta_2|x).u(0) = Pr(\theta_1|x). \tag{38}$$

The last equality is due to $u(1) = 1$ and $u(0) = 0$. Substituting (37) into (38), one can solve for γ_w in terms of u, λ and ρ.

$$\gamma_w = \frac{\rho u(w)}{\lambda - (\lambda - \rho)u(w)} = \left(1 + \frac{(1-u(w))\lambda}{u(w)\rho}\right)^{-1} \tag{39}$$

The computed quantity γ_w is called the *implicit prior at* w. In principle the implicit prior can vary for different values of w, but we make an assumption that it is constant i.e., $\gamma_w = \gamma$ for all w. From a practical point of view, the constant-implicit-prior assumption reduces the estimation of (binary) utility curve to the estimation of a single number. Suppose that the likelihood gamble exercise is conducted for k observations x_i for $1 \leq i \leq k$. The corresponding computed implicit priors are γ_i. One of the most convenient estimate suggested by the estimation theory is the average i.e.,

$$\hat{\gamma} = \frac{1}{k}\sum_{i=1}^{k} \gamma_i$$

Given the parameter value γ the relationship between scalar value w and equivalent binary utility $\langle\lambda(w), \rho(w)\rangle$ is as follows (see (37))

$$w = t^{-1}(\langle\lambda(w), \rho(w)\rangle) = \frac{\gamma\lambda(w)}{\gamma\lambda(w) + (1-\gamma)\rho(w)} \tag{40}$$

It follows that

$$\frac{\lambda(w)}{\rho(w)} = \frac{w}{(1-w)}\frac{(1-\gamma)}{\gamma} \tag{41}$$

Taking into account $\max(\lambda(w), \rho(w)) = 1$, one can solve for $0 < w < 1$

$$\lambda(w) = \begin{cases} \frac{w}{1-w}\frac{1-\gamma}{\gamma} & \text{if } \frac{w}{1-w}\frac{1-\gamma}{\gamma} \leq 1 \\ 1 & \text{otherwise} \end{cases} \tag{42}$$

$$\rho(w) = \begin{cases} 1 & \text{if } \frac{w}{1-w}\frac{1-\gamma}{\gamma} \leq 1 \\ \frac{1-w}{w}\frac{\gamma}{1-\gamma} & \text{otherwise} \end{cases} \tag{43}$$

Let us give an interpretation to γ. In particular, from (40), $t^{-1}(\langle 1, 1\rangle) = \gamma$. Recall that $\langle 1, 1\rangle = t([1/1, 1/0])$. In other words, $\langle 1, 1\rangle$ is the binary utility associated with the *fair* likelihood gamble in which the likelihoods of getting 1 and 0 are equal. Using a symmetry argument, one can argue that the "fair" price or the *certain equivalence* for this gamble should be the middle point 0.5 between $u(0) = 0$ and $u(1) = 1$. In this case ($\gamma = 0.5$), we say DM is *ambiguity neutral.* If $\gamma > 0.5$, DM is paying a "premium" price to enjoy the ambiguity in fair gamble. In this case we say that DM is *ambiguity seeking.*[6] In the last case of $\gamma < 0.5$ which is reasonably expected to hold for plurality of decision makers, DM pays a "discount" price for the fair gamble because of ambiguity in it. We say DM is *ambiguity averse.* Thus, γ characterizes the ambiguity attitude of the decision maker.

This classification of ambiguity attitude mirrors the classification of attitudes toward risk (probabilistic uncertainty). Risk seeking behavior is characterized by the convexity of the utility curve; risk averse by concavity and risk neutral by linearity. The differentiating term "ambiguity" signals that it has to do with non-probabilistic uncertainty expressible by DS plausibility,

[6] Ambiguity seeking behavior can exist and be justified by analogous arguments used to justify risk seeking behavior. Moreover, as it will be discussed in connection with Choquet expected utility (CEU) model, ranking according to CEU calculated wrt plausibility function exhibits ambiguity seeking character.

<table>
<tr><td>Dice face</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr>
<tr><td>$m(\{1\})$</td><td>.2</td><td colspan="5"></td></tr>
<tr><td>$m(\{1,2\})$</td><td colspan="2">.1</td><td colspan="4"></td></tr>
<tr><td>$m(\{3,4\})$</td><td colspan="2"></td><td colspan="2">.2</td><td colspan="2"></td></tr>
<tr><td>$m(\{3,4,5,6\})$</td><td colspan="2"></td><td colspan="4">.5</td></tr>
<tr><td>$Pl(.\vert\{1,2\})$</td><td>1</td><td>$\frac{1}{3}$</td><td colspan="4"></td></tr>
<tr><td>$Pl(.\vert\{3,4,5,6\})$</td><td colspan="2"></td><td>1</td><td>1</td><td>$\frac{5}{7}$</td><td>$\frac{5}{7}$</td></tr>
</table>

Table 1: Uncertainty about the dice

statistical likelihood, or fuzzy possibility. For more discussion on ambiguity attitude and its relation to risk attitude readers are referred to [10].

Before we move on to the next section, it is necessary to address an assumption made at the start of section 4 to ignore the DM's risk attitude via a convention to measure the prize $w \in \mathcal{U}$ in risk-adjusted utility. If we want to account for risk explicitly we can do so by introducing a function $r : \$ \to \mathcal{U}$ where \$ is monetary domain. The expression of pcb utility (35) then becomes:

$$u([p_1/L_1, p_2/L_2, \dots p_s/L_s]) = \sum_{i=1}^{s} p_i.t^{-1}(\text{cmax}_{1\le j\le k}\{\pi_{ij}.t(r(w_j))\}) \tag{44}$$

In this utility expression, risk attitude is handled by r, ambiguity attitude by t and uncertainty information by p and π. They are all separate. Notice also that in case pcb is a probability function, because of cancellation $t^{-1}t$, decision maker's ambiguity attitude effect disappears as naturally expected.

5. Examples

Example 3 Consider acts described in fig. 2. The situation is a dice roll with faces numbered from 1 to 6. The act on the right is $f = \{\{2,4,6\} \hookrightarrow 0, \{1,3\} \hookrightarrow 0.5, \{5\} \hookrightarrow 1\}$.

It is known that uncertainty is described by a pcb given in table 1. For example: $m(\{1\}) = .2$, $m(\{1,2\}) = .1$, $m(\{3,4\}) = .2$ and $m(\{3,4,5,6\}) = .5$; the partition size $s = 2$ with $p_1 = Pl(\{1,2\}) = .3$, $p_2 = Pl(\{3,4,5,6\}) = .7$; the conditional plausibility $Pl(1|\{1,2\}) = 1$, $Pl(2|\{1,2\}) = \frac{1}{3}$.

The act f is equivalent to $\{\{1,2\} \hookrightarrow \{2 \hookrightarrow 0, 1 \hookrightarrow 0.5\}, \{3,4,5,6\} \hookrightarrow \{\{4,6\} \hookrightarrow 0, 3 \hookrightarrow .5, 5 \hookrightarrow 1\}\}$. Assume ambiguity aversion with $\gamma = 0.4$,

$t(1) = \langle 1, 0\rangle$, $t(0) = \langle 0, 1\rangle$ and from eqs. (42), (43) it follows $t(0.5) = \langle 1, .67\rangle$.

The possibilistic lottery that corresponds to act $\{2 \hookrightarrow 0, 1 \hookrightarrow 0.5\}$ is $[.33/\langle 0, 1\rangle, 1/\langle 1, 0.67\rangle]$ has binary utility $\langle 1, 0.67\rangle$ and the certainty equivalence is 0.5.

The possibilistic lottery that corresponds to act $\{\{4, 6\} \hookrightarrow 0, 3 \hookrightarrow .5, 5 \hookrightarrow 1\}$ is $[1/\langle 0, 1\rangle, 1/\langle 1, 0.67\rangle, \frac{5}{7}/\langle 1, 0\rangle]$ has binary utility

$$\langle \max(0, 1, .71), \max(1, 0.67, 0)\rangle = \langle 1, 1\rangle$$

and the certainty equivalence by eq. (40) is 0.4.

Finally, the utility of the act or its certainty equivalence is

$$u(f) = 0.3 * 0.5 + 0.7 * 0.4 = 0.43 \blacksquare$$

Example 4 We offer our treatment of Ellsberg's paradox [6]. This kind of experiment is used to demonstrate that rational behavior under ambiguity violates Savage's sure-thing principle. In an urn, there are 90 balls of the same size. The balls are painted one of three colors: `red`, `yellow` and `white`. It is known that 30 balls are `red`. The proportions of `yellow` and `white` are not known.

Ellsberg considers four gambles. Gamble *IA* offers \$1 if a randomly drawn ball is `red`, nothing otherwise. Gamble *IB* offers \$1 if the ball is `yellow`, nothing otherwise. Gamble *IIA* offers \$1 if a randomly drawn ball is `red` or `white`, nothing if the ball is `yellow`. Gamble *IIB* offers \$1 if the ball is `yellow` or `white` and nothing if it is `red`.

Ellsberg discussed findings that a sizable proportion of respondents preferred *IA* to *IB* and, at the same time, preferred *IIB* to *IIA*. This observed preference is not consistent with the sure-thing principle because the pair (IIA, IIB) is different from the pair (IA, IB) only by the level of prize for `white` balls.

The uncertainty in the problem is nicely described by a pcb with 2 foci. $m(\{\texttt{red}\}) = \frac{1}{3}$ and $m(\{\texttt{yellow}, \texttt{white}\}) = \frac{2}{3}$. This pcb decomposes into $P(\{\texttt{red}\}) = \frac{1}{3}$ and $P(\{\texttt{yellow}, \texttt{white}\}) = \frac{2}{3}$ and $\Pi(\texttt{yellow}|\{\texttt{yellow}, \texttt{white}\}) = \Pi(\texttt{white}|\{\texttt{yellow}, \texttt{white}\}) = 1$.

We consider three cases of ambiguity attitude corresponding to ambiguity averse, neutral and seeking.

(1) Ambiguity averse attitude. We assume binary utility function $t_a(\$1) = \langle 1, 0\rangle$, $t_a(\$0) = \langle 0, 1\rangle$ and $t_a(\$.4) = \langle 1, 1\rangle$. The first two equalities are natural since \$1 is the best outcome and \$0 is the worst outcome. The last

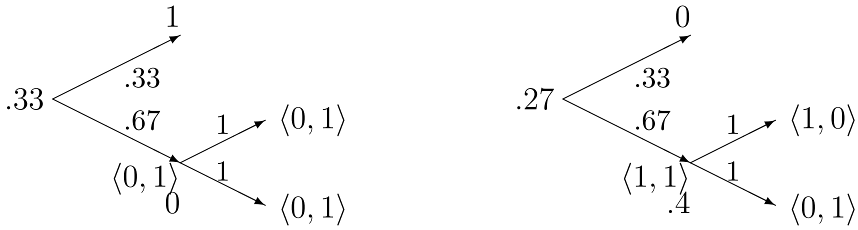

Figure 4: Ellsberg's lotteries IA and IB

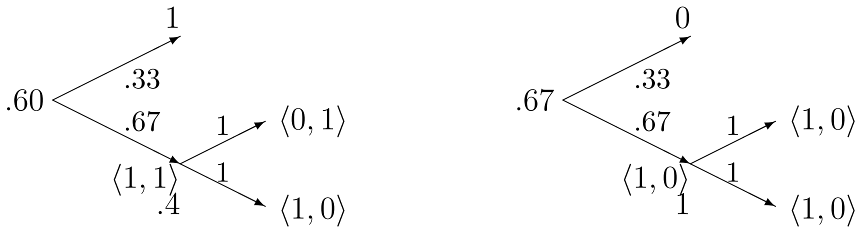

Figure 5: Ellsberg's lotteries IIA and IIB

equality implies that the implicit prior probability is .4 indicating somewhat ambiguity aversion.

In Figures 4 and 5 we show the calculation of mixed utility for the gambles. We have $u_a(IA) = .33$, $u_a(IB) = .27$, $u_a(IIA) = .60$ and $u_a(IIB) = .67$. This means $IIB \succ IIA \succ IA \succ IB$. These preferences are consistent with the observed behavior.

(2) Ambiguity neutrality. We have $t_n(\$1) = \langle 1, 0\rangle$, $t_n(\$0) = \langle 0, 1\rangle$ and $t_n(\$.5) = \langle 1, 1\rangle$ and $u_n(IA) = \frac{1}{3}$, $u_n(IB) = \frac{1}{3}$, $u_n(IIA) = \frac{2}{3}$, $u_n(IIB) = \frac{2}{3}$. This means $IIA \sim IIB \succ IA \sim IB$.

(3) Ambiguity seeking. We have $t_s(\$1) = \langle 1, 0\rangle$, $t_s(\$0) = \langle 0, 1\rangle$ and $t_s(\$.6) = \langle 1, 1\rangle$ and $u_s(IA) = .333$, $u_s(IB) = 0.4$, $u_s(IIA) = 0.733$, $u_s(IIB) = 0.667$. This means $IIA \succ IIB \succ IB \succ IA$. ■

6. Related Works

This work touches two streams of research developed somewhat separately in economics and Artificial Intelligence communities. On the one hand, since the pioneering work of Ellsberg [6] decision making under ambiguity has been an active topic of discussion in economics. Two widely used models are Choquet expected utility (CEU) model by Schmeidler [20] using *capacities* and maximin expected utility with multiple priors by Gilboa and Schmeidler

[12]. In recent years, the focus shifts to dynamic consistency issues [18]. On the other hand, DS belief function theory originated from statistics is actively studied by CS and AI community. DS theory offers expanded uncertainty expressiveness and at the same time provides well defined computational mechanism. It is not our intention to give a detailed review of what is going on in the fields. Rather, the aim is to provide contrast and similarity with our approach.

Smets [23] argues for a two-level process. At the *credal* level, an agent uses belief functions to represent and to reason with uncertainty. When there is a need to make a decision, the agent moves to another *pignistic* level in which the belief function is transformed to a probability function. The vNM expected utility is calculated with respect to this probability function. Specifically, in the pignistic transformation, the mass that assigned to a subset is divided equally to each element in the set. For example, the pignistic transformation of the pcb in Ellsberg's paradox is a probability function

Θ	red	yellow	white
P_{Bet}	.33	.33	.33

With respect to P_{Bet} the utilities of the four lotteries are $u(IA) = u(IB) = .33$ and $u(IIA) = u(IIB) = .66$. This does not explain Ellsberg's paradox. Broadly speaking, DM attitude toward ambiguity (aversion/seeking) is not accounted for in this approach.

Although Smets argued otherwise, the pignistic probability function is obtained by invoking the principle of insufficient reasoning for non-singleton foci. However, this goes against the main motivation for belief functions.

It should be noted however, that this preference would be observed in our framework for ambiguity neutral DM.

Cobb and Shenoy [2] argue that the pignistic transformation is inconsistent with Dempster's rule of combination. They argue, instead, for the use of a plausibility transformation in which a probability function is obtained by normalization of plausibility values of singletons. For instance, the plausibility transformation for the same pcb Ellsberg's example is

Θ	red	yellow	white
P_{Pla}	.20	.40	.40

With respect to P_{Pla}, the utilities of the four lotteries are $u(IA) = .20$, $u(IB) = .40$ and $u(IIA) = .60$, $u(IIB) = .80$ These utilities are not able to explain Ellsberg's paradox. Also, it suggests that $IB \succ IA$, which is contrary to the observed empirical behavior. The plausibility transformation has a number of drawbacks. Notice that the probabilities assigned to singletons in the original pcb are modified downward after translation. The magnitude of distortion depends on (1) the total of masses assigned to non-singletons and (2) the sizes of non-singleton foci. However, it is possible to argue that the plausibility transformation is for the Dempster-Shafer theory of belief functions, in which it is inappropriate to interpret belief and plausibility functions as lower and upper bounds on some true but unknown probabilities since these semantics are inconsistent with Dempster's rule of combination.

Walley [27] studies a class of imprecise probabilities: lower prevision and, its dual, upper prevision. This class includes belief functions as a subclass. He argues that imprecise probability allows only a partial preference ordering among alternatives. The most one can make from such a partial order is to exclude all dominated alternatives. The set of remaining alternatives, which may be large, is left to decision maker to choose by calling in an additional choice mechanism e.g., randomization. This indeterminacy is a significant inconvenience to the decision maker. This approach, developed for imprecise probability in general, also fails to take into account the specific structure offered by pcb. In a disagreement with Walley's approach, our view is that statistical likelihood rather than betting rate semantics is appropriate for uncertainty expressed by belief functions.

Jaffray and Wakker [14], [16] propose a decision model with belief function that combines linear utility with Hurwicz's $\alpha-$criteria for decision under ignorance. For example, a belief function has foci $\{A_1, A_2, \dots A_k\}$. An act f maps a focus into a subset of $\mathcal{U}$ i.e., $f(A_i) \subseteq \mathcal{U}$. Jaffray argued that if the only information is "A_i is true" then nothing is known about how likely a prize in $f(A_i)$ is. Under this ignorance, an application of Hurwicz's criterion finds the certainty equivalence to be the linear combination of the worst in $f(A_i)$ and the best in $f(A_i)$:

$$\alpha * \inf(f(A_i)) + (1 - \alpha) * \sup(f(A_i))$$

Interpreting $m(A_i)$ as the probability of focus A_i, the utility of f is

$$u_J(f) = \sum_{i=1}^{k} m(A_i)(\alpha * \inf(f(A_i)) + (1 - \alpha) * \sup(f(A_i)))$$

In this approach, the ambiguity attitude is expressed by coefficient α with intuition that the more ambiguity averse the higher α, the weight attached to the worst outcome. For example at $\alpha = 0.6$, the Ellsberg's gambles $u_J(IA) = 0.33$, $u_J(IB) = 0.67 * (0.6 * 0 + 0.4 * 1) = 0.268$, $u_J(IIA) = 0.33 + 0.67 * 0.4 = 0.598$ and $u_J(IIB) = 0.67$. These values explain Ellsberg's paradox nicely. Wakker [25] notes that "[Jaffray's] models, developed 20 years ago, achieve a tractability and a separation between risk attitudes, ambiguity attitudes, and ambiguity beliefs that have not yet been obtained in other models popular today". However, a problem with Hurwicz's rule is that it can lead to dynamically inconsistent behavior [15]. Note that our approach possesses the desirable properties mentioned in Wakker's comments. A detailed comparison between Jaffray's approach and one presented in this paper is provided in a upcoming paper.

Schmeidler [20], Sarin & Wakker [19] argue for the use of Choquet expected utility (CEU) for *non-additive probability* or capacity. A real value set function ν on Θ is a *capacity* if it satisfies normalization conditions ($\nu(\emptyset) = 0$, $\nu(\Theta) = 1$) and monotonicity ($\nu(A) \leq \nu(B)$ if $A \subset B$).

For simplicity, assume the prizes are ordered $w_1 > w_2 > \ldots > w_k$ with conventions $w_{k+1} = 0$ and $E_0 = \emptyset$. For a decision d with $d^{-1}(w_i) = E_i$, $1 \leq i \leq k$, CEU wrt capacity ν is defined as

$$\mathit{CEU}(d) = \sum_{i=1}^{k} (w_i - w_{i+1}).\nu(\cup_{j=1}^{i} E_j) \tag{45}$$

$$= \sum_{i=1}^{k} w_i(\nu(\cup_{j=0}^{j=i} E_j) - \nu(\cup_{j=0}^{j=i-1} E_j)) \tag{46}$$

The CEU representation is obtained by relaxing a number of axioms that originally leads to vNM expected utility representation for probabilistic lotteries. The appeal of CEU is that it leads to an (complete) order of alternatives and is supported by intuitive (co-monotonic) axioms. However, in this model the ambiguity attitude is inseparably blent with the uncertainty (ambiguity) information. A capacity is said to be *convex* (*concave*) if $\nu(A \cup B) + \nu(A \cap B)$ is greater than or equal to (less than or equal to) $\nu(A) + \nu(B)$. Schmeidler [20] shows that a decision maker is ambiguity averse (seeking) iff the capacity is convex (concave). Why this inseparability is undesirable? Most useful uncertainty information come from objective sources: experiments, data, observations. For example, in medicine [13] the information obtained via randomized clinical trials is considered most useful, less useful is information obtained via systematic observational studies, the least useful is the kind of anecdotal observation. The subjective character of capacity measures does not facilitate information collection from the objective sources.

For belief functions, there are many equivalent forms: bpa (m), belief (Bel), plausibility (Pl) and commonality (Q). These are equivalent in the sense of information i.e. knowing one form, one can calculate all other forms. If a belief function is obtained from statistical evidence then each of the forms embodies the same information. Among those equivalent forms, Bel and Pl are capacities. As capacities, however, Bel and Pl are opposite. Bel is convex while Pl is concave because $Bel(A \cup B) + Bel(A \cap B) \geq Bel(A) + Bel(B)$ and $Pl(A \cup B) + Pl(A \cap B) \leq Pl(A) + Pl(B)$. That makes CEU with Bel exhibits ambiguity aversion while CEU with Pl exhibits ambiguity seeking behavior.

For an illustration, let us calculate CEU wrt Bel and Pl for Ellsberg's gambles. For IA, $E_1 = \{\texttt{red}\}$, $E_2 = \{\texttt{yellow}, \texttt{white}\}$ because `red` is associated with \$1 prize and `yellow`, `white` with zero. Using (46), we find $CEU_{Bel}(IA) = CEU_{Pl}(IA) = .33$. For IB, $E_1 = \{\texttt{yellow}\}$, $E_2 = \{\texttt{red}, \texttt{white}\}$. We find $CEU_{Bel}(IB) = 0$ and $CEU_{Pl}(IB) = .67$. Therefore, CEU_{Bel} ranks $IA \succ IB$. However, CEU_{Pl} ranks $IB \succ IA$.

In our view, the fact that CEU ranking of lotteries depends on the choice of forms of uncertainty is problematic. Moreover, the lack of an explicit mechanism to express/measure ambiguity attitude is a deficiency. It is unclear how to express/extract ambiguity aversion degree for different decision makers e.g., a DM is more ambiguity averse than another.

7. Summary and Conclusions

In this paper we study decision making for a special class of belief functions called partially consonant belief functions (pcb) introduced by Walley [26]. Pcb is important because (1) it offers a meaningful generalization of both probability and possibility (those are special cases); (2) it is the only subclass of belief functions, which is consistent with the likelihood principle of statistics. Pcb has a nice interpretation - it can be decomposed into a probability function and a number of conditional possibility functions.

We use an axiomatic approach for the problem. A distinct feature of our approach is interpretation of plausibility as statistical likelihood, not as betting rates found in most other works on decision making with DS belief functions. Our axiomatics is similar to and inspired by von Neumann-Morgenstern's linear utility theory [24]. We prove a representation theorem for a preference relation on pcb lotteries. Pcb utility is a mixed construct that subsumes both linear utility for probabilistic lottery and binary utility for possibilistic lottery as special cases.

Our approach is tractable and offers separate modelings of risk attitude, ambiguity attitude and uncertainty information. This separation is important in practice as it allows different factors that influence DM decision to be investigated independently from each other.

An obvious open question is if and how to extend this approach for the general case of DS belief functions. Our speculation is that it can but a clean closed-form expression as in (35) would be unlikely. The reasoning behind this speculation is that a neat expression for utility of an arbitrary belief function lottery could be used to easily compute plausibility/belief of an arbitrary event. Pcb class turns out to be useful enough while still has nice focus structure that enables a closed-form utility expression.

References

[1] Berger, J. O., and Wolpert, R. L. *The Likelihood Principle*, 2 ed. Lecture notes-Monograph series. Institute of Mathematical Statistics, Hayward, California, 1988.

[2] Cobb, B. R., and Shenoy, P. P. On the plausibility transformation method for translating belief function models to probability models. *International Journal of Approximate Reasoning 41*, 3 (2006), 314–340.

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

[4] Diestel, R. *Graph theory*, 2 ed. Springer Verlag, Heidelberg, 2005.

[5] Dubois, D., Nguyen, T. H., and Prade, H. Possibility theory, probability and fuzzy sets. In *Handbook of Fuzzy Sets Series*, D. Dubois and H. Prade, Eds. Kluwer Academic, Boston, 2000, pp. 344–438.

[6] Ellsberg, D. Risk, ambiguity and the Savage's axioms. *Quarterly Journal of Economics 75*, 4 (1961), 643–669.

[7] Giang, P. H. *A Decision Theory for Non-Probabilistic Uncertainty and Its Applications.* PhD thesis, University of Kansas, Lawrence, Kansas, 2003.

[8] Giang, P. H., and Shenoy, P. P. Statistical decisions using likelihood information without prior probabilities. In *Uncertainty in Artificial Intelligence: Proceedings of the Eighteenth Conference (UAI–2002)* (San Francisco, CA, 2002), A. Darwiche and N. Friedman, Eds., Morgan Kaufmann, pp. 170–178.

[9] Giang, P. H., and Shenoy, P. P. Decision making with partially consonant belief functions. In *Uncertainty in Artificial Intelligence: Proceedings of the Nineteenth Conference (UAI–2003)* (San Francisco, CA, 2003), C. Meek and U. Kjaerulff, Eds., Morgan Kaufmann.

[10] Giang, P. H., and Shenoy, P. P. Decision making on the sole basis of statistical likelihood. *Artificial Intelligence 165* (2005), 137–163.

[11] Giang, P. H., and Shenoy, P. P. Two axiomatic approaches to decision making using possibility theory. *European Journal of Operational Research 162*, 2 (2005), 450–467.

[12] Gilboa, I., and Schmeidler, D. Maximin expected utility with a non-unique prior. *Journal of Mathematical Economics 18* (1989), 141–153.

[13] GRADE working group. Grading quality of evidence and strength of recommendations. *BMJ*, 328 (2004), 1490 – 1498.

[14] Jaffray, J.-Y. Linear utility theory for belief functions. *Operations Research Letters 8* (1989), 107–112.

[15] Jaffray, J.-Y., and Jeleva, M. Information processing under imprecise risk with the hurwicz criterion. In *Fifth international symposium on imprecise probability: theories and applications ISIPTA'07* (2007).

[16] Jaffray, J.-Y., and Wakker, P. Decision making with belief functions: Compatibility and incompatibility with the sure-thing principle. *Journal of Risk and Uncertainty 8*, 3 (1994), 255–271.

[17] Luce, R. D., and Raiffa, H. *Games and Decision.* John Wiley & Sons, 1957.

[18] Machina, M. Risk, ambiguity, and the rank-dependence axioms. *American Economic Review 99*, 1 (2009), 385392.

[19] Sarin, R., and Wakker, P. A simple axiomatization of nonadditive expected utility. *Econometrica 60*, 6 (1992), 1255–1272.

[20] Schmeidler, D. Subjective probability and expected utility without additivity. *Econometrica 57*, 3 (1989), 571–587.

[21] Shafer, G. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.

[22] Shafer, G. Belief functions and parametric models. *Journal of the Royal Statistical Society, Series B 44*, 3 (1982), 322–352.

[23] Smets, P. The transferable belief model for quantified belief representation. In *Handbook of Defeasible Reasoning and Uncetainty Management system*, D. M. Gabbay and P. Smets, Eds., vol. 1. Kluwer, Doordrecht, 1998, pp. 267–301.

[24] von Neumann, J., and Morgenstern, O. *Theory of Games and Economic Behavior*, 3 ed. Princeton University Press, Princeton, NJ, 1953.

[25] Wakker, P. Jaffray's ideas on ambiguity. Working paper. Econometric Institute, Erasmus University, Rotterdam, the Netherlands, Available online at http://people.few.eur.nl/wakker/pdf/jaffray.pdf.

[26] Walley, P. Belief function representation of statistical evidence. *The Annals of Statistics 15*, 4 (1987), 1439–1465.

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

[28] Yager, R., and Liu, L. *Classic works on the Dempster-Shafer theory of belief functions*, vol. 219 of *Studies in Fuzziness and Soft Computing*. Springer, Berlin, Heidelberg, 2008.

[29] Zadeh, L. Fuzzy set as a basis for a theory of possibility. *Fuzzy Sets and Systems 1* (1978), 3–28.