International Journal of Approximate Reasoning 124 (2020) 194–216

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

An interval-valued utility theory for decision making with Dempster-Shafer belief functions

Thierry Denœux [a,b,c], Prakash P. Shenoy [d,*]

[a] Université de Technologie de Compiègne, CNRS, UMR 7253 Heudiasyc, Compiègne, France
[b] Shanghai University, UTSEUS, Shanghai, China
[c] Institut Universitaire de France, Paris, France
[d] University of Kansas, School of Business, Lawrence, KS, USA

ARTICLE INFO

Article history:
Received 31 August 2019
Received in revised form 2 June 2020
Accepted 17 June 2020
Available online 1 July 2020

Keywords:
Dempster-Shafer theory of evidence
Von Neumann-Morgenstern's utility theory
Interval-valued utility function
Jaffray's linear utility theory
Smets' two-level decision theory
Shafer's constructive decision theory

ABSTRACT

The main goal of this paper is to describe an axiomatic utility theory for Dempster-Shafer belief function lotteries. The axiomatic framework used is analogous to von Neumann-Morgenstern's utility theory for probabilistic lotteries as described by Luce and Raiffa. Unlike the probabilistic case, our axiomatic framework leads to interval-valued utilities, and therefore, to a partial (incomplete) preference order on the set of all belief function lotteries. If the belief function reference lotteries we use are Bayesian belief functions, then our representation theorem coincides with Jaffray's representation theorem for his linear utility theory for belief functions. We illustrate our representation theorem using some examples discussed in the literature, and we propose a simple model for assessing utilities based on an interval-valued pessimism index representing a decision-maker's attitude to ambiguity and indeterminacy. Finally, we compare our decision theory with those proposed by Jaffray, Smets, Dubois et al., Giang and Shenoy, and Shafer.

Published by Elsevier Inc.

1. Introduction

The main goal of this paper is to propose an axiomatic utility theory for lotteries described by belief functions in the Dempster-Shafer (D-S) theory of evidence [5,39]. The axiomatic theory is constructed similar to von Neumann-Morgenstern's (vN-M's) utility theory for probabilistic lotteries [51,29,28,37,34,17]. Unlike the probabilistic case, our axiomatic theory leads to interval-valued utilities, and therefore to a partial (incomplete) preference order on the set of all belief function lotteries. Also, we compare our decision theory to those proposed by Jaffray [30], Smets [48], Dubois et al. [11], Giang and Shenoy [21,22], and Shafer [44].

In the foreword to Glenn Shafer's 1976 monograph [39], Dempster writes: "… I believe that Bayesian inference will always be a basic tool for practical everyday statistics, if only because questions must be answered and decisions must be taken, so that a statistician must always stand ready to upgrade his vaguer forms of belief into precisely additive probabilities." More than 40 years after these lines were written, a lot of approaches to decision-making have been proposed (see the recent review in [7]). However, most of these methods lack a strong theoretical basis. The most important steps toward a decision theory in the D-S framework have been made by Jaffray [30], Smets [48] and Shafer [44]. However, we argue that these proposals are either not sufficiently justified from the point of view of D-S theory, or not sufficiently developed for

* Corresponding author.
E-mail addresses: thierry.denoeux@utc.fr (T. Denœux), pshenoy@ku.edu (P.P. Shenoy).

https://doi.org/10.1016/j.ijar.2020.06.008
0888-613X/Published by Elsevier Inc.

practical use. Our goal is to propose and justify a utility theory that is in line with vN-M's utility theory, but adapted to be used with lotteries whose uncertainty is described by D-S belief functions.

In essence, the D-S theory consists of representations—basic probability assignments (also called mass functions), belief functions, plausibility functions, etc.—together with Dempster's combination rule, and a rule for marginalizing joint belief functions. The representation part of the D-S theory is also used in various other theories of belief functions. For example, in the imprecise probability community, a belief function is viewed as the lower envelope of a convex set of probability mass functions called a credal set. Credal set semantics are also referred to in the literature as *lower probability interpretation* [31–33], and as *generalized probability* [16,27]. Using these semantics, it makes more sense to use the Fagin-Halpern combination rule [16] (also proposed by de Campos et al. [4]), rather than Dempster's combination rule [27,42,43]. The utility theory this article proposes is designed specifically for the D-S belief function theory, and not for the other theories of belief functions. This suggests that Dempster's combination rule should be an integral part of our theory, a property that is not satisfied in the proposals by Jaffray and Smets.

There is a large literature on decision making with a (credal) set of probability mass functions [23] motivated by Ellsberg's paradox [15]. An influential work in this area is the axiomatic framework by Gilboa-Schmeidler [24], where they use Choquet integration [3,25] to compute expected utility. A belief function is a special case of a Choquet capacity. Jaffray's [30] work can also be regarded as belonging to the same line of research, although Jaffray works directly with belief functions without specifying a combination rule. A review of this literature can be found in, e.g., [18], where the authors propose a modification of the Gilboa-Schmeidler [24] axioms. As we said earlier, our focus here is on decision-making with D-S theory of belief functions, and not on decision-making based in belief functions with a credal set interpretation. As we will see, our interval-valued utility functions lead to intervals that are contained in the Choquet lower and upper expected utility intervals.

The remainder of this article is as follows. In Section 2, we sketch vN-M's axiomatic utility theory for probabilistic lotteries as described by Luce and Raiffa [37]. In Section 3, we summarize the basic definitions in the D-S belief function theory. In Section 4, we describe our adaptation of vN-M's utility theory for lotteries in which uncertainty is described by D-S belief functions. Our assumptions lead to an interval-valued utility function, and consequently, to a partial (incomplete) preference order on the set of all belief function lotteries. We also describe a model for assessments of utilities. In Section 5, we compare our utility theory with those described by Jaffray [30], Smets [48], Dubois et al. [11], Giang and Shenoy [21,22], and Shafer [44]. Finally, in Section 6, we summarize and conclude.

2. Von Neumann-Morgenstern's utility theory

In this section, we describe vN-M's utility theory for decision under risk. Most of the material in this section is adapted from [37]. A decision problem can be seen as a situation in which a decision-maker (DM) has to choose a course of action (or *act*) in some set $\mathbf{F}$. An act may have different *outcomes*, depending on the *state of nature* X. Exactly one state of nature will obtain, but this state is unknown. Denoting by $\Omega_X = \{x_1, \ldots, x_n\}$ the set of states of nature and by $\mathbf{O} = \{O_1, \ldots, O_r\}$ the set of outcomes,[1] an act can thus be formalized as a mapping f from Ω to $\mathbf{O}$. In this section, we assume that uncertainty about the state of nature is described by a probability mass function (PMF) p_X on Ω_X. In vN-M's original exposition [51], probabilities on Ω_X are assumed to be objective and to correspond to long-run frequencies. However, the line of reasoning summarized below is also valid with other interpretations of probabilities, such as additive degrees of belief, provided that probabilities are assumed to have been determined beforehand, independently of the decision problem. For instance, in the constructive approach proposed by Shafer [44], probabilities are constructed by comparing a given problem with a scale of examples in which the truth is generated according to known chances.[2]

If the DM selects act f, they will get outcome O_i with probability

$$p_i = \sum_{\{x \in \Omega_X \mid f(x) = O_i\}} p_X(x). \tag{1}$$

To each act f thus corresponds a PMF $\mathbf{p} = (p_1, \ldots, p_r)$ on $\mathbf{O}$. We call $L = [\mathbf{O}, \mathbf{p}]$ a *probabilistic lottery*. As only one state in Ω_X will obtain, a probabilistic lottery will result in exactly one outcome O_i (with probability p_i), and we suppose that the lottery will not be repeated. Another natural assumption is that two acts that induce the same lottery are equivalent: the problem of expressing preference between acts then boils down to expressing preference between lotteries.

We are thus concerned with a DM who has preferences on $\mathcal{L}$, the set of all probabilistic lotteries on $\mathbf{O}$, and our task is to find a real-valued *utility function* $u : \mathcal{L} \to \mathbb{R}$ such that the DM strictly prefers L to L' if and only if $u(L) > u(L')$, and the DM is indifferent between L and L' if and only if $u(L) = u(L')$. We write $O_i \succ O_j$ if the DM strictly prefers O_i to O_j, write $O_i \sim O_j$ if the DM is indifferent between (or equally prefers) O_i and O_j, and write $O_i \succsim O_j$ if the DM either strictly prefers O_i to O_j or is indifferent between the two.

[1] The assumption of finiteness of the sets Ω_X and $\mathbf{O}$ is only for ease of exposition. It is unnecessary for the proof of the representation theorem in this section.

[2] Savage [38] derives both probabilities and utilities from a set of axioms. This approach will not be considered in this paper.

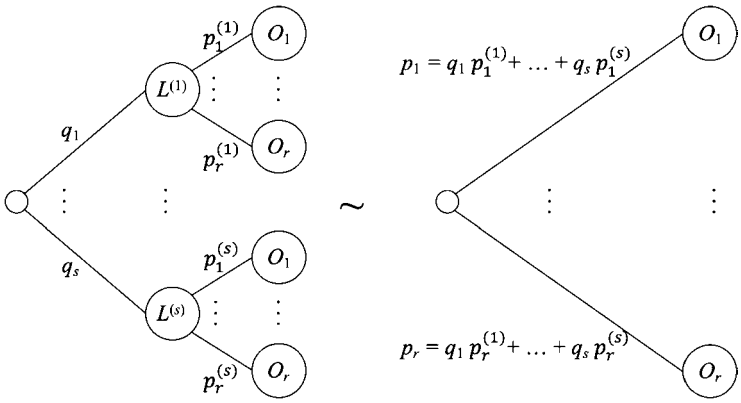

Fig. 1. A two-stage compound lottery reduced to an indifferent simple lottery.

Of course, finding such a utility function is not always possible, unless the DM's preferences satisfy some assumptions. We can then construct a utility function that is *linear* in the sense that the utility of a lottery $L = [\mathbf{O}, \mathbf{p}]$ is equal to its expected utility $\sum_{i=1}^{r} p_i\, u(O_i)$, where O_i is regarded as a degenerate lottery where the only possible outcome is O_i with probability 1. In the remainder of this section, we describe a set of assumptions that lead to the existence of such a linear utility function.

Assumption 2.1 *(Weak ordering of outcomes).* For any two outcomes O_i and O_j, either $O_i \succsim O_j$ or $O_j \succsim O_i$. Also, if $O_i \succsim O_j$ and $O_j \succsim O_k$, then $O_i \succsim O_k$. Thus, the preference relation $\succsim$ over $\mathbf{O}$ is a weak order, i.e., it is complete and transitive.

Given Assumption 2.1, without loss of generality, let us assume that the outcomes are labeled such that $O_1 \succsim O_2 \succsim \cdots \succsim O_r$, and to avoid trivialities, assume that $O_1 \succ O_r$.

Suppose that $\mathbf{L} = \{L^{(1)}, \ldots, L^{(s)}\}$ is a set of s lotteries, where each of the s lotteries $L^j = [\mathbf{O}, \mathbf{p}^{(j)}]$ are over outcomes in $\mathbf{O}$, with PMFs $\mathbf{p}^{(j)}$ for $j = 1, \ldots, s$. Suppose $\mathbf{q} = (q_1, \ldots, q_s)$ is a PMF on $\mathbf{L}$ such that $q_j > 0$ for $j = 1, \ldots, s$, and $\sum_{j=1}^{s} q_j = 1$. Then $[\mathbf{L}, \mathbf{q}]$ is called a *compound* lottery whose outcome is exactly one lottery $L^{(i)}$ (with probability q_i), and lottery $L^{(i)}$ will result in one outcome O_j (with probability $p_j^{(i)}$). Notice that the PMF $\mathbf{p}^{(i)}$ is a conditional PMF for $\mathbf{O}$ in the second stage given that lottery $L^{(i)}$ is realized (with probability $q_i > 0$) in the first stage (see Fig. 1). We can compute the joint PMF for $(\mathbf{L}, \mathbf{O})$, and then compute the marginal $\mathbf{p}$ of the joint for $\mathbf{O}$. The following assumption states that the resulting lottery $[\mathbf{O}, \mathbf{p}]$ is indifferent to the compound lottery $[\mathbf{L}, \mathbf{q}]$.

Assumption 2.2 *(Reduction of compound lotteries).* Any compound lottery $[\mathbf{L}, \mathbf{q}]$, where $L^{(i)} = [\mathbf{O}, \mathbf{p}^{(i)}]$, is indifferent to a simple (non-compound) lottery $[\mathbf{O}, \mathbf{p}]$, where

$$p_i = q_1\, p_i^{(1)} + \ldots + q_s\, p_i^{(s)} \tag{2}$$

for $i = 1, \ldots, r$. PMF $(p_1, \ldots, p_r)$ is the marginal for $\mathbf{O}$ of the joint PMF of $(\mathbf{L}, \mathbf{O})$.

A simple lottery involving only outcomes O_1 and O_r with PMF $(u, 1-u)$, where $0 \le u \le 1$, is called a *reference* lottery, and is denoted by $[\{O_1, O_r\}, (u, 1-u)]$. Let $\mathbf{O}_2$ denote the set $\{O_1, O_r\}$.

Assumption 2.3 *(Continuity).* Each outcome O_i is indifferent to a reference lottery $\widetilde{O}_i = [\mathbf{O}_2, (u_i, 1-u_i)]$ for some u_i, where $0 \le u_i \le 1$, i.e., $O_i \sim \widetilde{O}_i$.

Assumption 2.4 *(Weak order).* The preference relation $\succsim$ for lotteries in $\mathcal{L}$ is a weak order, i.e., it is complete and transitive.

Assumption 2.4 generalizes Assumption 2.1 for outcomes, which can be regarded as degenerate lotteries.

Assumption 2.5 *(Substitutability).* In any lottery $L = [\mathbf{O}, \mathbf{p}]$, if we substitute an outcome O_i by the reference lottery $\widetilde{O}_i = [\mathbf{O}_2, (u_i, 1-u_i)]$ that is indifferent to O_i, then the result is a compound lottery that is indifferent to L (see Fig. 2), i.e.,

$$[(O_1, \ldots, O_{i-1}, O_i, O_{i+1}, \ldots, O_r), \mathbf{p}] \sim [(O_1, \ldots, O_{i-1}, \widetilde{O}_i, O_{i+1}, \ldots, O_r), \mathbf{p}].$$

From Assumptions 2.1–2.5, given any lottery $L = [\mathbf{O}, \mathbf{p}]$, it is possible to find a reference lottery $\widetilde{L} = [\mathbf{O}_2, (u, 1-u)]$ that is indifferent to L (see Fig. 3). This is expressed by Theorem 2.1 below.

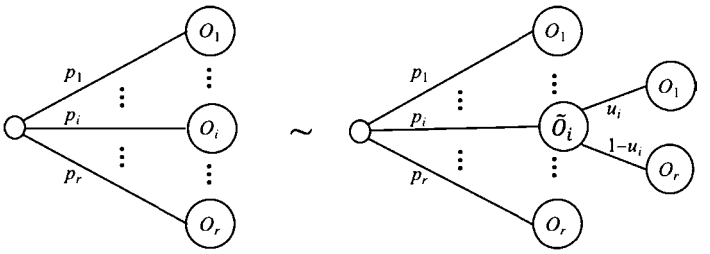

Fig. 2. The substitutability Assumption 2.5.

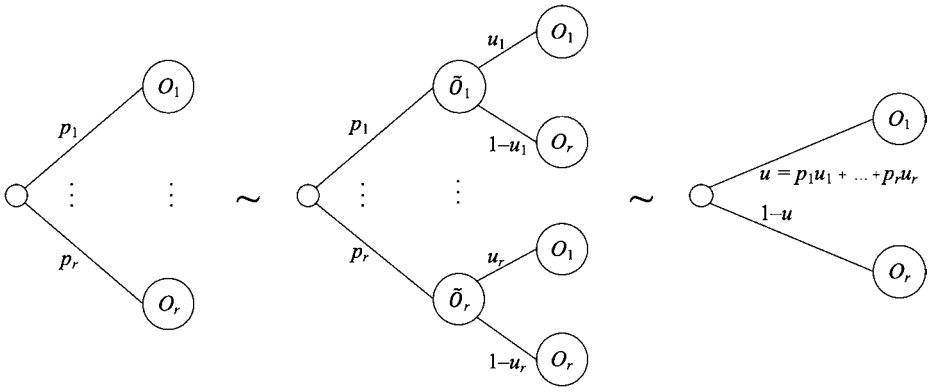

Fig. 3. Reducing a lottery to an indifferent compound lottery and then to an indifferent reference lottery.

Theorem 2.1 *(Reducing a lottery to an indifferent reference lottery). Under Assumptions 2.1-2.5, any lottery $L = [\boldsymbol{O}, \boldsymbol{p}]$ is indifferent to a reference lottery $\widetilde{L} = [\boldsymbol{O}_2, (u, 1-u)]$ with*

$$u = \sum_{i=1}^{r} p_i\, u_i. \tag{3}$$

Proof. ([37]) First, we replace each O_i by $\widetilde{O}_i$ for $i = 1, \ldots, r$. Assumption 2.3 (continuity) states that these indifferent lotteries exist, and Assumption 2.5 (substitutability) says that they are substitutable without changing the preference relation. So by using Assumption 2.4 serially, $[\mathbf{O}, \mathbf{p}] \sim [\widetilde{\mathbf{O}}, \mathbf{p}]$. Now if we apply Assumption 2.2 (reduction of compound lotteries), then $[\mathbf{O}, \mathbf{p}] \sim [\mathbf{O}_2, (u, 1-u)]$, where u is given by Eq. (3). □

Assumption 2.6 *(Monotonicity).* A reference lottery $L = [\mathbf{O}_2, (u, 1-u)]$ is preferred or indifferent to reference lottery $L' = [\mathbf{O}_2, (u', 1-u')]$ if and only if $u \geq u'$.

As $O_1 \sim \widetilde{O}_1 = [\mathbf{O}, (u_1, 1-u_1)]$ and $O_r \sim \widetilde{O}_r = [\mathbf{O}, (u_r, 1-u_r)]$, Assumptions 2.4 and 2.6 imply that $u_1 = 1$ and $u_r = 0$. Also, from $O_1 \succsim O_2 \succsim \cdots \succsim O_r$, we can deduce that $1 = u_1 \geq u_2 \geq \cdots \geq u_r = 0$.

Assumptions 2.1–2.6 allow us to define the utility of a lottery as the probability of the best outcome O_1 in an indifferent reference lottery, and this utility function for lotteries on $\mathbf{O}$ is linear. This is stated by the following theorem.

Theorem 2.2 *([37]). If the preference relation $\succsim$ on $\mathcal{L}$ satisfies Assumptions 2.1–2.6, then there are numbers u_i associated with outcomes O_i for $i = 1, \ldots, r$, such that for any two lotteries $L = [\mathbf{O}, \mathbf{p}]$, and $L' = [\mathbf{O}, \mathbf{p}']$, $L \succsim L'$ if and only if*

$$\sum_{i=1}^{r} p_i\, u_i \geq \sum_{i=1}^{r} p'_i\, u_i. \tag{4}$$

Thus, we can define the utility of lottery $L = [\boldsymbol{O}, \boldsymbol{p}]$ as $u(L) = \sum_{i=1}^{r} p_i\, u_i$, where $u_i = u(O_i)$. Also, such a linear utility function is unique up to a strictly increasing affine transformation, i.e., if $u'_i = a\, u_i + b$, where $a > 0$ and b are real constants, then $u(L) = \sum_{i=1}^{r} p_i\, u'_i$ also qualifies as a utility function.

3. Basic definitions in the D-S belief function theory

In this section, we review the basic definitions in the D-S theory of belief functions. Like various uncertainty theories, D-S belief function theory includes functional representations of uncertain knowledge, and basic operations for making inferences from such knowledge. These will be recalled, respectively, in Section 3.1 and 3.2. Conditional belief functions and the notion of conditional embedding are then introduced in Section 3.3, and the semantics of belief functions in D-S theory is discussed in Section 3.4. Most of the material in this section (except Section 3.4) is taken from [36]. For further details, the reader is referred to [39] and to [8] for a recent review.

3.1. Representations of belief functions

Belief functions can be represented in several different ways, including as basic probability assignments, plausibility functions and belief functions.[3] These are briefly discussed below.

Definition 1 *(Basic probability assignment).* Suppose X is an unknown quantity (variable) with possible values (states) in a finite set Ω_X called the state space of X. We assume that X takes one and only one value in Ω_X, but this value is unknown. Let 2^{Ω_X} denote the set of all subsets of Ω_X. A basic probability assignment (BPA) m_X for X is a function $m_X : 2^{\Omega_X} \rightarrow [0,1]$ such that

$$\sum_{\mathsf{a}\subseteq\Omega_X} m_X(\mathsf{a}) = 1, \quad \text{and} \quad m_X(\emptyset) = 0. \tag{5}$$

The subsets $\mathsf{a} \subseteq \Omega_X$ such that $m_X(\mathsf{a}) > 0$ are called *focal sets* of m_X. An example of a BPA for X is the *vacuous* BPA for X, denoted by ι_X, such that $\iota_X(\Omega_X) = 1$. We say that m_X is *deterministic* if m_X has a single focal set (with mass 1). Thus, the vacuous BPA for X is deterministic with focal set Ω_X. If all focal sets of m_X are singleton subsets (of Ω_X), then we say that m_X is *Bayesian*. In this case, m_X is equivalent to the PMF P_X for X such that $P_X(x) = m_X(\{x\})$ for each $x \in \Omega_X$.

Definition 2 *(Plausibility function).* The information in a BPA m_X can be represented by a corresponding plausibility function Pl_{m_X} defined as follows:

$$Pl_{m_X}(\mathsf{a}) = \sum_{\{\mathsf{b}\subseteq\Omega_X | \mathsf{b}\cap\mathsf{a}\neq\emptyset\}} m_X(\mathsf{b}) \quad \text{for all } \mathsf{a} \subseteq \Omega_X. \tag{6}$$

For an example, suppose $\Omega_X = \{x, \bar{x}\}$. Then, the plausibility function Pl_{ι_X} corresponding to BPA ι_X is given by $Pl_{\iota_X}(\emptyset) = 0$, $Pl_{\iota_X}(\{x\}) = 1$, $Pl_{\iota_X}(\{\bar{x}\}) = 1$, and $Pl_{\iota_X}(\Omega_X) = 1$.

Definition 3 *(Belief function).* The information in a BPA m_X can also be represented by a corresponding belief function Bel_{m_X} that is defined as follows:

$$Bel_{m_X}(\mathsf{a}) = \sum_{\{\mathsf{b}\subseteq\Omega_X | \mathsf{b}\subseteq\mathsf{a}\}} m_X(\mathsf{b}) \quad \text{for all } \mathsf{a} \subseteq \Omega_X. \tag{7}$$

For the example above with $\Omega_X = \{x, \bar{x}\}$, the belief function Bel_{ι_X} corresponding to BPA ι_X is given by $Bel_{\iota_X}(\emptyset) = 0$, $Bel_{\iota_X}(\{x\}) = 0$, $Bel_{\iota_X}(\{\bar{x}\}) = 0$, and $Bel_{\iota_X}(\Omega_X) = 1$. For any proposition $\mathsf{a} \in 2^{\Omega_X}$, it is easy to see that $Bel_{m_X}(\mathsf{a}) \leq Pl_{m_X}(\mathsf{a})$. Thus, if a DM's belief in proposition a is an interval, say $[p, p+q]$, where $p, q \geq 0$ and $p + q \leq 1$, then such beliefs can be represented by a BPA m_X such that $m_X(\mathsf{a}) = p$, $m_X(\Omega_X \setminus \mathsf{a}) = 1 - p - q$, and $m_X(\Omega_X) = q$. For such a BPA, $Bel_{m_X}(\mathsf{a}) = p \leq p + q = Pl_{m_X}(\mathsf{a})$.

All three representations—BPA, belief and plausibility functions—have exactly the same information, as any one of them allows us to recover the other two [39].

Next, we describe the two main operations for making inferences.

3.2. Basic operations in the D-S theory

There are two main operations in the D-S theory—Dempster's combination rule and marginalization.

Dempster's combination rule. In the D-S theory, we can combine two BPAs m_1 and m_2 representing distinct pieces of evidence by Dempster's rule [5] and obtain the BPA $m_1 \oplus m_2$, which represents the combined evidence. Dempster refers to this rule

[3] Belief functions can also be mathematically represented by a convex set of PMFs called a *credal set*, but the semantics of such a representation are incompatible with Dempster's combination rule [40,42,43,27]. For these reasons, we skip a credal set representation of a belief function.

as the product-intersection rule, as the product of the BPA values are assigned to the intersection of the focal sets, followed by normalization. Normalization consists of discarding the mass assigned to $\emptyset$, and normalizing the remaining values so that they add to 1. In general, Dempster's rule of combination can be used to combine two BPAs for arbitrary sets of variables.

Let $\mathcal{X}$ denote a finite set of variables. The state space of $\mathcal{X}$ is $\bigtimes_{X\in\mathcal{X}}\Omega_X$. Thus, if $\mathcal{X}=\{X,Y\}$ then the state space of $\{X,Y\}$ is $\Omega_X\times\Omega_Y$. Projection of states simply means dropping extra coordinates; for example, if (x,y) is a state of $\{X,Y\}$, then the projection of (x,y) to X, denoted by $(x,y)^{\downarrow X}$, is simply x, which is a state of X. Projection of subsets of states is achieved by projecting every state in the subset. Suppose $\mathsf{b}\in 2^{\Omega_{\{X,Y\}}}$. Then $\mathsf{b}^{\downarrow X}=\{x\in\Omega_X:(x,y)\in\mathsf{b}\}$. Notice that $\mathsf{b}^{\downarrow X}\in 2^{\Omega_X}$.

Vacuous extension of a subset of states of $\mathcal{X}_1$ to a subset of states of $\mathcal{X}_2$, where $\mathcal{X}_2\supseteq\mathcal{X}_1$, is a cylinder set extension, i.e., if $\mathsf{a}\in 2^{\mathcal{X}_1}$, then $\mathsf{a}^{\uparrow\mathcal{X}_2}=\mathsf{a}\times\Omega_{\mathcal{X}_2\setminus\mathcal{X}_1}$. Thus, if $\mathsf{a}\in 2^{\Omega_X}$, then $\mathsf{a}^{\uparrow\{X,Y\}}=\mathsf{a}\times\Omega_Y$.

Definition 4 *(Dempster's rule using BPAs).* Suppose m_1 and m_2 are BPAs for $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. Then $m_1\oplus m_2$ is a BPA for $\mathcal{X}_1\cup\mathcal{X}_2=\mathcal{X}$, say, given by $(m_1\oplus m_2)(\emptyset)=0$ and

$$(m_1\oplus m_2)(\mathsf{a})=K^{-1}\sum_{\{\mathsf{b}_1,\mathsf{b}_2\subseteq\Omega_{\mathcal{X}}|\mathsf{b}_1\cap\mathsf{b}_2=\mathsf{a}\}} m_1(\mathsf{b}_1^{\downarrow\mathcal{X}_1})\,m_2(\mathsf{b}_2^{\downarrow\mathcal{X}_2}), \tag{8}$$

for all $\mathsf{a}\subseteq\Omega_{\mathcal{X}}$, where K is a normalization constant given by

$$K=1-\sum_{\{\mathsf{b}_1,\mathsf{b}_2\subseteq\Omega_{\mathcal{X}}|\mathsf{b}_1\cap\mathsf{b}_2=\emptyset\}} m_1(\mathsf{b}_1^{\downarrow\mathcal{X}_1})\,m_2(\mathsf{b}_2^{\downarrow\mathcal{X}_2}). \tag{9}$$

The definition of Dempster's rule assumes that the normalization constant K is non-zero. If $K=0$, then the two BPAs m_1 and m_2 are said to be in *total conflict* and cannot be combined. If $K=1$, we say m_1 and m_2 are *non-conflicting*.

Marginalization. Marginalization in D-S theory is addition of values of BPAs.

Definition 5 *(Marginalization).* Suppose m is a BPA for $\mathcal{X}$. Then, the marginal of m for $\mathcal{X}_1$, where $\mathcal{X}_1\subset\mathcal{X}$, denoted by $m^{\downarrow\mathcal{X}_1}$, is a BPA for $\mathcal{X}_1$ such that for each $\mathsf{a}\subseteq\Omega_{\mathcal{X}_1}$,

$$m^{\downarrow\mathcal{X}_1}(\mathsf{a})=\sum_{\{\mathsf{b}\subseteq\Omega_{\mathcal{X}}|\mathsf{b}^{\downarrow\mathcal{X}_1}=\mathsf{a}\}} m(\mathsf{b}). \tag{10}$$

3.3. Conditional belief functions

In probability theory, it is common to construct joint PMFs for a set of discrete variables by using conditional probability distributions. For example, we can construct joint PMF for (X,Y) by first assessing PMF P_X of X, and conditional PMFs $P_{Y|x}$ for each $x\in\Omega_X$ such that $P_X(x)>0$. The pointwise multiplication of $P_{Y|x}$ for all $x\in\Omega_X$ is called a CPT, and denoted by $P_{Y|X}$. Then, $P_{X,Y}=P_X\otimes P_{Y|X}$. We can construct joint BPA for $\{X,Y\}$ in a similar manner.

Suppose that there is a BPA for Y expressing our belief about Y if we know that $X=x$, and denote it by $m_{Y|x}$. Notice that $m_{Y|x}:2^{\Omega_Y}\to[0,1]$ is such that $\sum_{\mathsf{b}\in 2^{\Omega_Y}} m_{Y|x}(\mathsf{b})=1$. We can embed this conditional BPA for Y into a BPA for $\{X,Y\}$, which is denoted by $m_{x,Y}$, such that the following three conditions hold. First, $m_{x,Y}$ tells us nothing about X, i.e., $m_{x,Y}^{\downarrow X}(\Omega_X)=1$. Second, $m_{x,Y}$ tells us nothing about Y, i.e., $m_{x,Y}^{\downarrow Y}(\Omega_Y)=1$. Third, if we combine $m_{x,Y}$ with the deterministic BPA $m_{X=x}$ for X such $m_{X=x}(\{x\})=1$ using Dempster's rule, and marginalize the result to Y we obtain $m_{Y|x}$, i.e., $(m_{x,Y}\oplus m_{X=x})^{\downarrow Y}=m_{Y|x}$. The least committed way to obtain such an embedding, called *conditional embedding*, was derived by Smets [46,47] (see also [41]). It consists of taking each focal set $\mathsf{b}\in 2^{\Omega_Y}$ of $m_{Y|x}$, and converting it to a corresponding focal set of $m_{x,Y}$ (with the same mass) as follows: $(\{x\}\times\mathsf{b})\cup(\overline{\{x\}}\times\Omega_Y)$, where $\overline{\{x\}}$ denotes the complement of $\{x\}$ in Ω_X. It is easy to confirm that this method of embedding satisfies the three conditions mentioned above, and $m_{x,Y}$ is the least committed (minimally informative) BPA verifying this property.

Example 1 *(Conditional embedding).* Consider discrete variables X and Y, with $\Omega_X=\{x,\bar{x}\}$ and $\Omega_Y=\{y,\bar{y}\}$. Suppose that m_X is a BPA for X such that $m_X(x)>0$. If we have a conditional BPA $m_{Y|x}$ for Y given $X=x$ as follows:

$$\begin{aligned} &m_{Y|x}(y)=0.8, \text{ and}\\ &m_{Y|x}(\Omega_Y)=0.2, \end{aligned} \tag{11}$$

then its conditional embedding into BPA $m_{x,Y}$ for $\{X,Y\}$ is

$$\begin{aligned} &m_{x,Y}(\{(x,y),(\bar{x},y),(\bar{x},\bar{y})\})=0.8, \text{ and}\\ &\qquad m_{x,Y}(\Omega_{\{X,Y\}})=0.2. \end{aligned} \tag{12}$$

There are some differences with conditional probability distributions. First, in probability theory, $P_{Y|X}$ consists of *all* conditional distributions $P_{Y|x}$ that are well-defined, i.e., for all $x \in \Omega_X$ such that $P_X(x) > 0$. In D-S belief function theory, we do not have similar constraints. We can include only those non-vacuous conditionals $m_{Y|x}$ such that $m_X(\{x\}) > 0$. Also, if we have more than one conditional BPA for Y, given, say for $X = x_1$, and $X = x_2$ (assuming $m_X(\{x_1\}) > 0$, and $m_X(\{x_2\}) > 0$), we embed these two conditionals for Y to get BPAs $m_{x_1,Y}$ and $m_{x_2,Y}$ for $\{X, Y\}$, and then combine them using Dempster's rule of combination to obtain one conditional BPA $m_{Y|X} = m_{x_1,Y} \oplus m_{x_2,Y}$, which corresponds to $P_{Y|X}$ in probability theory.

Second, given any joint PMF $P_{X,Y}$ for $\{X, Y\}$, we can always factor this into $P^{\downarrow X}_{X,Y} = P_X$ for X, and $P_{Y|X}$ for $\{X, Y\}$, such that $P_{X,Y} = P_X \otimes P_{Y|X}$. This is not true in D-S belief function theory. Given a joint BPA $m_{X,Y}$ for $\{X, Y\}$, we cannot always find a BPA $m_{Y|X}$ for $\{X, Y\}$ such that $m_{X,Y} = m^{\downarrow X}_{X,Y} \oplus m_{Y|X}$. However, we can always *construct* joint BPA $m_{X,Y}$ for $\{X, Y\}$ by first assessing m_X for X, and assessing conditionals $m_{Y|x_i}$ for Y for those x_i that we have knowledge about and such that $m_X(\{x_i\}) > 0$, embed these conditionals into BPAs for $\{X, Y\}$, and combine all such BPAs to obtain the BPA $m_{Y|X}$ for $\{X, Y\}$. An implicit assumption here is that BBAs $m_{x_i,Y}$ are distinct, and it is acceptable to combine them using Dempster's rule. We can then construct $m_{X,Y} = m_X \oplus m_{Y|X}$.

3.4. Semantics of D-S belief function

In D-S theory, belief functions are representations of an agent's state of knowledge based on some evidence. As explained by Shafer [40], such representations can be constructed by comparing the available evidence with a hypothetical situation in which we receive a coded message, the meaning of which is random. More precisely, assume that a source sends us an encrypted message using a code selected at random from a set of codes $C = \{c_1, \ldots, c_n\}$ with known probabilities $p_1, \ldots, p_n$. If we decode the message with code c_i, we get a decoded message of the form "$X \in \Gamma(c_i)$", where Γ is a multi-valued mapping from C to 2^{Ω_X}. For any nonempty subset a of Ω_X, the probability that the meaning of the original message is "$X \in \mathsf{a}$" is

$$m(\mathsf{a}) = \sum_{i=1}^{n} p_i I(\Gamma(c_i) = \mathsf{a}),$$

where $I(\cdot)$ is the indicator function. The random message metaphor thus provides a way to construct BPAs m. The fundamental assumption of D-S theory is that such metaphors provide a scale of canonical examples to which any piece of evidence (or, at least, most pieces of evidence encountered in practice) can be meaningfully compared.

The random set metaphor accounts for the use of Dempster's rule, which can be easily derived from the assumption that the two BPAs m_1 and m_2 are induced by stochastically independent randomly coded messages. Two bodies of evidence are considered as independent if "(they) are sufficiently unrelated that pooling them is like pooling stochastically independent randomly coded messages" [40, Section 5.1].

In D-S theory, any belief function can thus be thought of as being induced by a multi-valued mapping from a probability space to the power set of the frame of discernment. Such multi-valued mappings were already explicitly constructed from a statistical model in Dempster's original application of belief functions to statistical inference [5]. A statistical model is no more "real" than a random code canonical examples: both are idealizations that allow us to formalize our knowledge and make inferences based on reasonable assumptions.

4. A utility theory for D-S belief function theory

In this section, we describe a new utility theory for lotteries where the uncertainty is described by D-S belief functions. These lotteries, called *belief function lotteries*,[4] will be introduced in Section 4.1. We present and discuss assumptions in Section 4.2 and state a representation theorem in Section 4.3. In Section 4.4, we show that an additional assumption leads to a simpler model, and we state the corresponding representation theorem. Finally, in Section 4.5, we describe an even simpler practical model allowing us to assess the utility of a belief function lottery based on a limited number of parameters.

4.1. Belief function lotteries

We generalize the decision framework outlined in Section 2 by assuming that uncertainty about the state of nature X is described by a BPA m_X for X. The probabilistic framework is recovered as a special case when m_X is Bayesian. The BPA m_X is assumed to be given, and is assumed to be a meaningful representation of the DM's state of knowledge about X at a given time, with the semantics described in Section 3.4. As before, we define an act as a mapping f from Ω_X to the set $\mathbf{O}$ of outcomes. Mapping f pushes m_X forward from Ω_X to $\mathbf{O}$, transferring each mass $m_X(\mathsf{a})$ for $\mathsf{a} \in 2^{\Omega_X}$ to the image of subset a by f, denoted as $f[\mathsf{a}] = \{f(x) : x \in \mathsf{a}\}$. The resulting BPA m for $\mathbf{O}$ is then defined as

[4] This notion was previously introduced in [7] under the name "evidential lottery."

$$m(\mathsf{b}) = \sum_{\{\mathsf{a}\in 2^{\Omega_X} \mid f[\mathsf{a}]=\mathsf{b}\}} m_X(\mathsf{a}), \tag{13}$$

for all $\mathsf{b} \subseteq \mathbf{O}$ [8]. Eq. (13) clearly generalizes Eq. (1). The pair $[\mathbf{O}, m]$ will be called a belief function (bf) lottery. It is a representation of the DM's subjective beliefs about the outcome that will obtain as a result of selecting act f. As noted in [7], a bf lottery can also arise from a BPA m_X on Ω_X and a *nondeterministic* act f, defined as mapping from Ω_X to $2^{\mathbf{O}}$. This formalism may be useful to account for under-specified decision problems in which, for instance, the set of acts or the state space Ω_X are too coarsely defined to allow for a precise description of the consequences of an act [19].

As before, we assume that two acts can be compared from what we believe their outcomes will be, irrespective of the evidence on which we base our beliefs. This assumption is a form of what Wakker [52] calls the *principle of complete ignorance* (PCI). It implies that two acts resulting in the same bf lottery are equivalent. The problem of expressing preferences between acts becomes that of expressing preferences between bf lotteries.

Remark 1. As a consequence of the PCI, preferences between acts do not depend on the cardinality of the state space Ω_X in case of complete ignorance. For instance, assume that we define $\Omega_X = \{x_1, x_2\}$, and we are completely ignorant of the state of nature, so that our belief state is described by the vacuous BPA $m_X(\Omega_X) = 1$. Consider two acts f_1 and f_2 that yield \$100 if, respectively, x_1 or x_2 occurs, and \$0 otherwise. These two acts induce the same vacuous bf lottery $m(\mathbf{O}) = 1$ with $\mathbf{O} = \{\$100, \$0\}$: consequently, they are equivalent according to the PCI. Now, assume that we decide to express the states of nature with finer granularity and we refine state x_1 into two states x_{11} and x_{12}. Let $\Omega_{X'} = \{x_{11}, x_{12}, x_2\}$ denote the refined frame. We still have $m_{X'}(\Omega_{X'}) = 1$ and $m(\mathbf{O}) = 1$, so that our preferences between acts f_1 and f_2 are unchanged. We note that a Bayesian DM applying Laplace's principle of indifference (PI) would reach a different conclusion: before the refinement, the PI implies $p_X(x_1) = p_X(x_2) = 1/2$, which results in the same probabilistic lottery $\mathbf{p} = (1/2, 1/2)$ on $\mathbf{O} = \{\$100, \$0\}$ for the two acts, but after the refinement the same principle gives us $p_X(x_{11}) = p_X(x_{12}) = p_X(x_2) = 1/3$; this results in two different lotteries $\mathbf{p}_1 = (2/3, 1/3)$ for act f_1 and $\mathbf{p}_2 = (1/3, 2/3)$ for act f_2, which makes f_1 strictly preferable to f_2. Considering that the granularity of the state space is often partly arbitrary (as discussed by Shafer in [39]), we regard this property of invariance to refinement under complete ignorance as a valuable feature of a decision theory based on D-S belief functions.

We are thus concerned with a DM who has preferences on $\mathcal{L}_{bf}$, the set of all bf lotteries. Our task is to find a utility function $u : \mathcal{L}_{bf} \to [\mathbb{R}]$, where $[\mathbb{R}]$ denotes the set of closed real intervals, such that the $u(L) = [u, 1 - v]$ is viewed as an interval-valued utility of L. The interval-valued utility can be interpreted as follows: u and v are, respectively, the degrees of belief of receiving the best and the worst outcome in a bf reference lottery equivalent to L (and $1 - v$ is, consequently, the degree of plausibility of receiving the best outcome). Given two lotteries L and L', L is preferred to L' if and only if $u \geq u'$ and $v \leq v'$. This leads to incomplete preferences on the set of all bf lotteries. If we assume $u = 1 - v$ for all bf lotteries, then we have a real-valued utility function on $\mathcal{L}_{bf}$, and consequently, complete preferences.

Example 2 *(Ellsberg's Urn).* Ellsberg [15] describes a decision problem that questions the adequacy of the vN-M axiomatic framework. Suppose we have an urn with 90 balls, of which 30 are red, and the remaining 60 are either black or yellow. We draw a ball at random from the urn. Let X denote the color of the ball drawn, with $\Omega_X = \{r, b, y\}$. Notice that the uncertainty of X can be described by a BPA m_X for X such that $m_X(\{r\}) = 1/3$, and $m_X(\{b, y\}) = 2/3$.

First, we are offered a choice between Lottery L_1: \$100 on red, and Lottery L_2: \$100 on black, i.e., in L_1, you get \$100 if the ball drawn is red, and \$0 if the ball drawn is black or yellow, and in L_2, you get \$100 if the ball drawn is black and \$0 if the ball drawn is red or yellow. Choice of L_1 can be denoted by alternative $f_1 : \Omega_X \to \{\$100, \$0\}$ such that $f_1(r) = \$100$, $f_1(b) = f_1(y) = \$0$. Similarly, choice of L_2 can be denoted by alternative $f_2 : \Omega_X \to \{\$100, \$0\}$ such that $f_2(b) = \$100$, $f_2(r) = f_2(y) = \$0$. L_1 can be represented by the BPA m_1 for $\mathbf{O} = \{\$0, \$100\}$ as follows: $m_1(\{\$100\}) = 1/3$, $m_1(\{\$0\}) = 2/3$. L_2 can be represented by BPA m_2 for $\mathbf{O}$ as follows: $m_2(\{\$0\}) = 1/3$, $m_2(\{\$0, \$100\}) = 2/3$. Notice that L_1 and L_2 are bf lotteries. Ellsberg notes that a frequent pattern of response is L_1 preferred to L_2.

Second, we are offered a choice between L_3: \$100 on red or yellow, and L_4: \$100 on black or yellow, i.e., in L_3 you get \$100 if the ball drawn is red or yellow, and \$0 if the ball drawn is black, and in L_4, you get \$100 if the ball drawn is black or yellow, and \$0 if the ball drawn is red. L_3 can be represented by BPA m_3 as follows: $m_3(\{\$100\}) = 1/3$, and $m_3(\{\$0, \$100\}) = 2/3$, and L_4 can be represented by the BPA m_4 as follows: $m_4(\{\$0\}) = 1/3$, $m_4(\{\$100\}) = 2/3$. L_3 and L_4 are also belief function lotteries. Ellsberg notes that L_4 is often strictly preferred to L_3. Also, the same subjects who prefer L_1 to L_2, prefer L_4 to L_3. Table 1 is a summary of the four bf lotteries.

Thus, if the outcomes of a lottery are based on the states of a random variable X, which is described by a BPA m_X for X, then we have a belief function lottery. In this example, we have only two outcomes, \$100, and \$0. L_1 and L_4 can also be regarded as probabilistic lotteries as the corresponding BPAs are Bayesian. L_2 and L_3 have BPAs with non-singleton focal sets. Thus, these two lotteries can be considered as involving "ambiguity" as the exact distribution of the probability (of $2/3$) between outcomes \$100 and \$0 is unknown. Regardless of how the probability of $2/3$ is distributed between b and y, the preferences of subjects violate the tenets of vN-M utility theory.

Table 1
Four belief function lotteries in Example 2.

Lottery	$m_i(\{\$100\})$	$m_i(\{\$0\})$	$m_i(\{\$100, \$0\})$
L_1 (\$100 on r)	1/3	2/3	
L_2 (\$100 on b)		1/3	2/3
L_3 (\$100 on r or y)	1/3		2/3
L_4 (\$100 on b or y)	2/3	1/3	

4.2. Assumptions of our framework

As in the probabilistic case, we will assume that a DM's preferences for bf lotteries are reflexive and transitive. However, unlike the probabilistic case (Assumption 2.4), we do not assume that these preferences are complete. In the probabilistic case, incomplete preferences are studied in [1], and in the case of sets of utility functions, in [13].

Our first assumption is identical to Assumption 2.1.

Assumption 4.1 *(Weak ordering of outcomes).* The DM's preferences $\succsim$ for outcomes in $\mathbf{O} = \{O_1, \ldots, O_r\}$ are complete and transitive.

This allows us to label the outcomes such that

$$O_1 \succsim O_2 \succsim \cdots \succsim O_r, \quad \text{and} \quad O_1 \succ O_r. \tag{14}$$

Let $\mathcal{L}_{bf}$ denote the set of all bf lotteries on $\mathbf{O} = \{O_1, \ldots O_r\}$, where the outcomes satisfy Eq. (14). As every BPA m on $\mathbf{O}$ is a bf lottery, $\mathcal{L}_{bf}$ is essentially the set of all BPAs on $\mathbf{O}$. As the set of all BPAs include Bayesian BPAs, the set $\mathcal{L}_{bf}$ is a superset of $\mathcal{L}$, i.e., every probabilistic lottery on $\mathbf{O}$ can be considered a bf lottery.

Consider a compound lottery $[\mathbf{L}, m]$, where $\mathbf{L} = \{L_1, \ldots, L_s\}$, m is a BPA for $\mathbf{L}$, and $L_j = [\mathbf{O}, m_{j|L_j}]$ is a bf lottery on $\mathbf{O}$, where $m_{j|L_j}$ is a conditional BPA for $\mathbf{O}$ in the second stage given that lottery L_j is realized in the first stage. We thus have $s+1$ pieces of evidence represented by BPAs m and $m_{1|L_1}, \ldots, m_{s|L_s}$. Assuming these pieces of evidence to be independent, they can be combined by Dempster's rule (8), after conditionally embedding the conditional BPAs $m_{j|L_j}$ (see Section 3.3). Marginalizing the orthogonal sum on $\mathbf{O}$, we get a BPA m'. Assumption 4.2 below posits that the resulting simple lottery $[\mathbf{O}, m']$ is equally preferred to the original compound lottery $[\mathbf{L}, m]$, i.e., we can reduce the compound lottery to a simple bf lottery on $\mathbf{O}$ using the D-S calculus.

Assumption 4.2 *(Reduction of compound lotteries).* Suppose $[\mathbf{L}, m]$ is a compound lottery as described in the previous paragraph. Then, $[\mathbf{L}, m] \sim [\mathbf{O}, m']$, where

$$m' = \left(m \oplus \left(\bigoplus_{j=1}^{s} m_{L_j, j} \right) \right)^{\downarrow \mathbf{O}}, \tag{15}$$

and $m_{L_j,j}$ is a BPA for $(\mathbf{L}, \mathbf{O})$ obtained from $m_{j|L_j}$ by conditional embedding, for $j = 1, \ldots, s$.

Example 3. Consider two urns: Urn 1 contains 90 balls, 30 of which are black, and 60 are red or yellow. Urn 2 is identical to Ellsberg's urn in Example 2: it contains 90 balls, 30 of which are red, and 60 are black or yellow. Assume that, in the first stage, you are allowed to draw one ball B_1 from Urn 1:

- If B_1 is black or red, then you will be allowed to draw one ball from Urn 2 in the second stage, and you will get \$100 if it is red, and \$0 otherwise (lottery L_1 of Example 2);
- If B_1 is yellow, you will be allowed to draw one ball from Urn 2, and you will get \$100 if it is black, and \$0 otherwise (lottery L_2 of Example 2).

Here, we have a compound bf lottery with outcome space $\mathbf{L} = \{L_1, L_2\}$. We get L_1 or L_2 depending on the color X_1 of ball B_1 drawn from Urn 1. We have $m_{X_1}(\{b\}) = 1/3$ and $m_{X_1}(\{r, y\}) = 2/3$, and the act f defined by $f(b) = f(r) = L_1$, $f(y) = L_2$. The BPA on $\mathbf{L}$ is, thus, $m(\{L_1\}) = 1/3$ and $m(\{L_1, L_2\}) = 2/3$. Now, L_1 and L_2 are bf lotteries on $\mathbf{O} = \{\$100, \$0\}$, with BPAs

$$m_{1|L_1}(\{\$100\}) = 1/3, \quad m_{1|L_1}(\{\$0\}) = 2/3$$

and

$$m_{2|L_2}(\{\$0\}) = 1/3, \quad m_{2|L_2}(\{\$0, \$100\}) = 2/3.$$

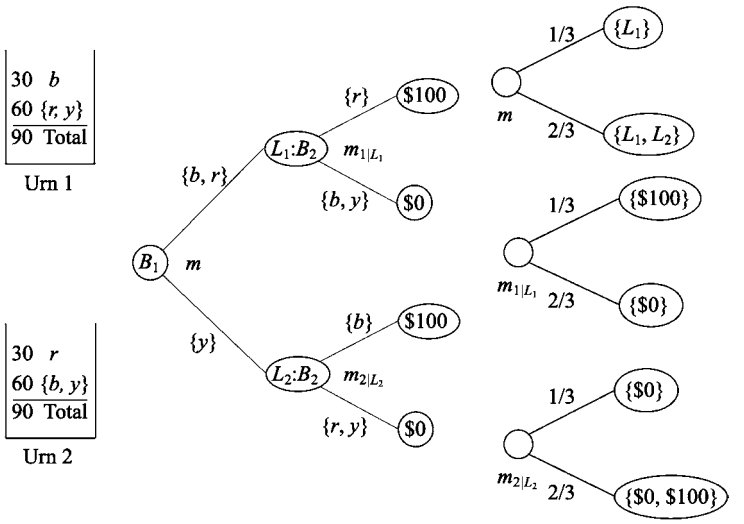

Fig. 4. A 2-stage compound belief function lottery. B_1 is drawn from Urn 1, and B_2 is drawn from Urn 2. The corresponding BPAs are shown on the right. In the first stage, m is a BPA for $\{L_1, L_2\}$. In the second stage, $m_{1|L_1}$ and $m_{2|L_2}$ are BPAs for $\{\$100, \$0\}$.

Here, $m_{j|L_j}$ is a conditional BPA on $\mathbf{O}$, given that L_j is obtained in the first stage (see Fig. 4). The conditional embeddings of $m_{1|L_1}$ and $m_{2|L_2}$ are BPAs on $\mathbf{L} \times \mathbf{O}$ equal, respectively, to

$$m_{L_1,1}(\{(L_1, \$100), (L_2, \$100), (L_2, \$0)\}) = 1/3,$$
$$m_{L_1,1}(\{(L_1, \$0), (L_2, \$100), (L_2, \$0)\}) = 2/3,$$

and

$$m_{L_2,2}(\{(L_2, \$0), (L_1, \$100), (L_1, \$0)\}) = 1/3,$$
$$m_{L_2,2}(\mathbf{L} \times \mathbf{O}) = 2/3,$$

Their orthogonal sum is

$$(m_{L_1,1} \oplus m_{L_2,2})(\{(L_1, \$100), (L_2, \$0)\}) = 1/9,$$
$$(m_{L_1,1} \oplus m_{L_2,2})(\{(L_1, \$0), (L_2, \$0)\}) = 2/9,$$
$$(m_{L_1,1} \oplus m_{L_2,2})(\{(L_1, \$100), (L_2, \$100), (L_2, \$0)\}) = 2/9,$$
$$(m_{L_1,1} \oplus m_{L_2,2})(\{(L_1, \$0), (L_2, \$100), (L_2, \$0)\}) = 4/9.$$

Combining it with m, we get

$$(m \oplus m_{L_1,1} \oplus m_{L_2,2})(\{(L_1, \$100)\}) = 3/27,$$
$$(m \oplus m_{L_1,1} \oplus m_{L_2,2})(\{(L_1, \$0)\}) = 6/27,$$
$$(m \oplus m_{L_1,1} \oplus m_{L_2,2})(\{(L_1, \$100), (L_2, \$0)\}) = 2/27,$$
$$(m \oplus m_{L_1,1} \oplus m_{L_2,2})(\{(L_1, \$0), (L_2, \$0)\}) = 4/27,$$
$$(m \oplus m_{L_1,1} \oplus m_{L_2,2})(\{(L_1, \$100), (L_2, \$100), (L_2, \$0)\}) = 4/27,$$
$$(m \oplus m_{L_1,1} \oplus m_{L_2,2})(\{(L_1, \$0), (L_2, \$100), (L_2, \$0)\}) = 8/27.$$

Marginalizing on $\mathbf{O}$, we get $m' = (m \oplus m_{L_1,1} \oplus m_{L_2,2})^{\downarrow \mathbf{O}}$ equal to

$$m'(\{\$100\}) = 3/27 = 1/9,$$
$$m'(\{\$0\}) = 10/27,$$
$$m'(\{\$100, \$0\}) = 14/27.$$

According to Assumption 4.2, a rational DM should be indifferent between receiving the compound bf lottery $[\mathbf{L}, m]$, or receiving the bf lottery $[\mathbf{O}, m']$, i.e., a prize about which the only information he has is given by a randomly coded message

whose meaning can be "The value of the prize is \$100" with probability 1/9, "The value of the prize is \$0" with probability 10/27, and "The value of the prize is unknown" with probability 14/27.

The following proposition states that Assumption 4.2 generalizes Assumption 2.2.

Proposition 1. *Let $\boldsymbol{L} = \{L_1, \ldots, L_s\}$ be a set of bf lotteries, with $L_j = [\boldsymbol{O}, m_{j|L_j}]$, in which $m_{j|L_j}$ is a Bayesian conditional BPA for $\boldsymbol{O}$ such that $m_{j|L_j}(\{O_i\}) = p_i^{(j)}$ and $\sum_{i=1}^{r} p_i^{(j)} = 1$ for $j = 1, \ldots, s$. Let $[\boldsymbol{L}, m]$ be a compound lottery in which m is a Bayesian BPA for $\boldsymbol{L}$ such that $m(\{L_j\} = q_j$ for $j = 1, \ldots, s$ with $\sum_{j=1}^{s} q_j = 1$. Then BPA m' defined by (15) is Bayesian and it verifies*

$$m'(\{O_i\}) = \sum_{j=1}^{s} q_j p_i^{(j)} \tag{16}$$

for $i = 1, \ldots, r$.

Proof. The conditional embedding of $m_{j|L_j}$ is given by

$$m_{L_j,j}\left(\{(L_j, O_i)\} \cup \overline{\{L_j\}} \times \mathbf{O}\right) = p_i^{(j)}, \quad i = 1, \ldots, r.$$

Let $m_0 = \bigoplus_{j=1}^{s} m_{L_j,j}$. It is a BPA for $\mathbf{L} \times \mathbf{O}$ defined by

$$m_0\left(\{(L_1, O_{i_1}), \ldots, (L_s, O_{i_s})\}\right) = p_{i_1}^{(1)} \ldots p_{i_s}^{(s)}$$

for all $(i_1, \ldots, i_s) \in \{1, \ldots, r\}^s$. Combining m_0 with m, we get a Bayesian BPA m_0' on $\mathbf{L} \times \mathbf{O}$ such that

$$m_0'(\{(L_j, O_i)\}) = q_j p_i^{(j)}.$$

After marginalizing on $\mathbf{O}$, we finally get Eq. (16). □

Next, we define a *bf reference* lottery $[\mathbf{O}_2, m]$ as a bf lottery on $\mathbf{O}_2 = \{O_1, O_r\}$. A bf reference lottery has three parameters $u = m(\{O_1\})$, $v = m(\{O_r\})$, and $w = m(\mathbf{O}_2)$, which are all non-negative and sum to 1. It can be equivalently denoted as $[\mathbf{O}_2, (u, v, w)]$. The first and second elements of the triple are, respectively, the degrees of belief of receiving the best and worst outcomes, while the third element can be seen as a degree of ignorance. Obviously, the degrees of plausibility of receiving the best and the worst outcomes are, respectively, $1 - v$ and $1 - u$. The following assumption states that any deterministic bf lottery is equally preferred to some bf reference lottery.

Assumption 4.3 *(Continuity).* Any subset of outcomes $\mathsf{a} \subseteq \mathbf{O}$ (considered as a deterministic bf lottery) is indifferent to a bf reference lottery $\widetilde{\mathsf{a}} = [\mathbf{O}_2, (u_\mathsf{a}, v_\mathsf{a}, w_\mathsf{a})]$ for some $u_\mathsf{a}, v_\mathsf{a}, w_\mathsf{a} \geq 0$ such that $u_\mathsf{a} + v_\mathsf{a} + w_\mathsf{a} = 1$. Furthermore, $w_\mathsf{a} = 0$ if $\mathsf{a} = \{O_i\}$ is a singleton subset.

For singleton subsets, the equivalent bf reference lottery is Bayesian: this ensures that Assumption 4.3 is a generalization of Assumption 2.3. For non-singleton subsets a of outcomes, we may have $w_\mathsf{a} > 0$, i.e., the bf reference lottery may not be Bayesian. In other words, we do not assume that ambiguity can be resolved by selecting an equivalent probabilistic reference lottery.

Example 4. Consider lottery $L_2 = [\{\\$100, \\$0\}, m_2]$ in Example 2, where $m_2(\{\\$0\}) = 1/3$, and $m_2(\{\\$100, \\$0\}) = 2/3$. Suppose we wish to assess the utility of focal set $\{\\$100, \\$0\}$ using a probabilistic reference lottery $[\{\\$100, \\$0\}, (p, 1-p)]$. A DM may have the following preferences. For any $p \leq 0.2$ she prefers $\{\\$100, \\$0\}$ to the probabilistic reference lottery, and for any $p \geq 0.3$, she prefers the probabilistic reference lottery to $\{\\$100, \\$0\}$. However, she is unable to give us a precise p such that $\{\\$100, \\$0\} \sim [\{\\$100, \\$0\}, (p, 1-p)]$. For such a DM, we can assess a bf reference lottery $[\{\\$100, \\$0\}, (0.2, 0.7, 0.1)]$ such that $Bel_{m_\mathsf{a}}(\{\\$100\}) = 0.2$ and $Pl_{m_\mathsf{a}}(\{\\$100\}) = 0.3$.

Assumption 4.4 *(Quasi-order).* The preference relation $\succsim$ for bf lotteries on $\mathcal{L}_{bf}$ is a quasi-order, i.e., it is reflexive and transitive.

In contrast with the probabilistic case (Assumption 2.4), we do not assume that $\succsim$ is complete. There are many reasons we may not wish to assume completeness. It is not descriptive of human behavior. Even from a normative point of view, it is questionable that a DM has complete preferences on all possible lotteries. The assumption of incomplete preferences is consistent with the D-S theory of belief functions where we have non-singleton focal sets. Several authors, such as Aumann [1], and Dubra et al. [13] argue why the assumption of complete preferences may not be realistic in many circumstances.

The substitutability assumption is similar to the probabilistic case (Assumption 2.5)—we replace an outcome in the probabilistic case by a focal set of m in the bf case.

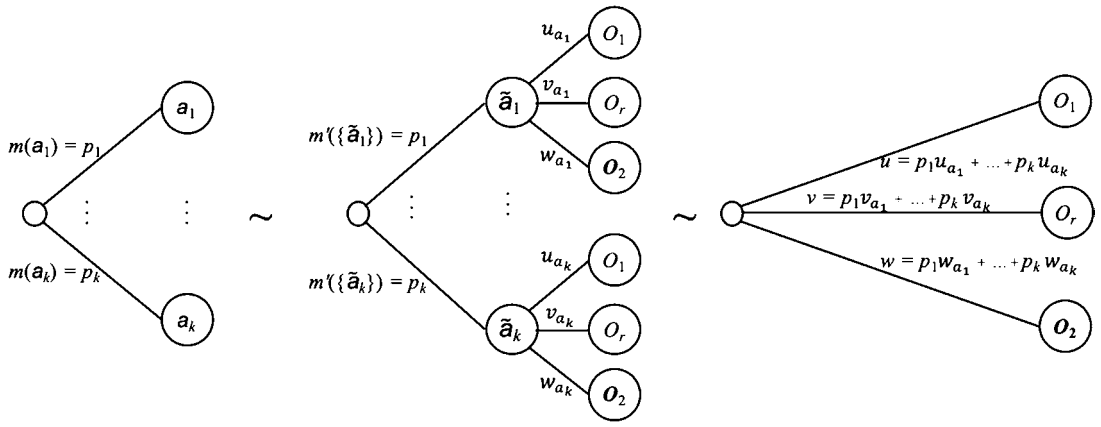

Fig. 5. Reducing a bf lottery to a bf reference lottery.

Assumption 4.5 *(Substitutability).* In any bf lottery $L = [\mathbf{O}, m]$, if we substitute a focal set a of m by an equally preferred bf reference lottery $\widetilde{\mathsf{a}} = [\mathbf{O}_2, m_{\mathsf{a}}]$, then the result is a compound lottery that is equally preferred to L.

It follows from Assumptions 4.1–4.5 that given any bf lottery, we can reduce it to an equally preferred bf reference lottery. This is stated as Theorem 4.1 below.

Theorem 4.1 *(Reducing a bf lottery to an indifferent bf reference lottery). Under Assumptions 4.1–4.5, any bf lottery $L = [\mathbf{O}, m]$ with focal sets $a_1, \ldots, a_k$ is indifferent to a bf reference lottery $\widetilde{L} = [\mathbf{O}_2, \widetilde{m}]$, such that*

$$\widetilde{m}(\{O_1\}) = \sum_{i=1}^{k} m(a_i)\, u_{a_i}, \tag{17a}$$

$$\widetilde{m}(\{O_r\}) = \sum_{i=1}^{k} m(a_i)\, v_{a_i}, \quad \textit{and} \tag{17b}$$

$$\widetilde{m}(\mathbf{O}_2) = \sum_{i=1}^{k} m(a_i)\, w_{a_i}, \tag{17c}$$

where u_{a_i}, v_{a_i} and w_{a_i} are the masses assigned, respectively, to $\{O_1\}$, $\{O_r\}$ and $\mathbf{O}_2$ by the bf reference lottery $\widetilde{\mathsf{a}}_i$ equivalent to a_i.

Proof. From Assumption 4.3 (continuity), we can replace each focal set a_i of m one at a time by an indifferent bf reference lottery $\widetilde{\mathsf{a}}_i = [\mathbf{O}_2, m_{i|\widetilde{\mathsf{a}}_i}]$, yielding a sequence of compound lotteries. From Assumptions 4.5 (substitutability) and 4.4 (quasi-order), these compound lotteries are all indifferent to L (see Fig. 5). Let $L' = [\mathbf{L}, m']$ be the compound lottery obtained after all focal sets a_i have been substituted, with $m'(\{\widetilde{\mathsf{a}}_i\}) = m(\mathsf{a}_i)$. From Assumption 4.2 (reduction of compound lotteries), L' can be reduced to a reference bf lottery $\widetilde{L} = [\mathbf{O}_2, \widetilde{m}]$, by considering each BBA $m_{i|\widetilde{\mathsf{a}}_i}$ as a conditional BPA and applying the rules of D-S calculus. The reduced bf lottery $\widetilde{L} = [\mathbf{O}_2, \widetilde{m}]$ is then given by

$$\widetilde{m} = \left(m' \oplus \left(\bigoplus_{i=1}^{k} m_{\widetilde{\mathsf{a}},i} \right) \right)^{\downarrow \mathbf{O}_2},$$

where $m_{\widetilde{\mathsf{a}},i}$ is the BPA for $\mathbf{O}_2$ obtained from $m_{i|\widetilde{\mathsf{a}}_i}$ by conditional embedding. If we have $m_{i|\widetilde{\mathsf{a}}_i}(\{O_1\}) = u_{\mathsf{a}_i}$, $m_{i|\widetilde{\mathsf{a}}_i}(\{O_r\}) = v_{\mathsf{a}_i}$ and $m_{i|\widetilde{\mathsf{a}}_i}(\mathbf{O}_2) = w_{\mathsf{a}_i}$, then after conditional embedding BPA $m_{\widetilde{\mathsf{a}}_i,i}$ is as follows:

$$m_{\widetilde{\mathsf{a}}_i,i}(\{(\widetilde{\mathsf{a}}_i, O_1)\} \cup (\overline{\{\widetilde{\mathsf{a}}_i\}} \times \mathbf{O}_2)) = u_{\mathsf{a}_i}$$
$$m_{\widetilde{\mathsf{a}}_i,i}(\{(\widetilde{\mathsf{a}}_i, O_r)\} \cup (\overline{\{\widetilde{\mathsf{a}}_i\}} \times \mathbf{O}_2)) = v_{\mathsf{a}_i}$$
$$m_{\widetilde{\mathsf{a}}_i,i}(\mathbf{L} \times \mathbf{O}_2) = w_{\mathsf{a}_i}.$$

Let m_0 denote $\bigoplus_{i=1}^{k} m_{\widetilde{\mathsf{a}}_i,i}$. The focal sets of m_0 are of the form

$$\mathsf{a}_{(I_1,I_2,I_3)} = \left(\bigcup_{i \in I_1} \widetilde{\mathsf{a}}_i \times \{O_1\} \right) \cup \left(\bigcup_{i \in I_2} \widetilde{\mathsf{a}}_i \times \{O_r\} \right) \cup \left(\bigcup_{i \in I_3} L_i \times \mathbf{O}_2 \right),$$

for all partitions (I_1, I_2, I_3) of $\{1, \ldots, k\}$, and the corresponding value of m_0 is:

$$m_0(\mathsf{a}_{(I_1,I_2,I_3)}) = \left(\prod_{i\in I_1} u_{\mathsf{a}_i}\right)\left(\prod_{i\in I_2} v_{\mathsf{a}_i}\right)\left(\prod_{i\in I_3} w_{\mathsf{a}_i}\right).$$

Next, we combine Bayesian BPA m' for $\mathbf{L}$ with m_0. The focal sets of $m' \oplus m_0$ are of the form $\{\widetilde{\mathsf{a}}_i\} \times \{O_1\}$, $\{\widetilde{\mathsf{a}}_i\} \times \{O_r\}$, and $\{\widetilde{\mathsf{a}}_i\} \times \mathbf{O}_2$ depending on whether $i \in I_1$ or $i \in I_2$, or $i \in I_3$, respectively, with mass $m(\mathsf{a}_i)m_0(\mathsf{a}_{(I_1,I_2,I_3)})$. Finally we marginalize $m' \oplus m_0$ to $\mathbf{O}_2$. The mass assigned to each focal set $\{\widetilde{\mathsf{a}}_i\} \times \{O_1\}$ is $m_{\mathsf{a}_i} u_{\mathsf{a}_i}$. Thus, $\widetilde{m}(\{O_1\}) = \sum_{i=i}^{k} m(\mathsf{a}_i) u_{\mathsf{a}_i}$. Similarly, $\widetilde{m}(\{O_r\}) = \sum_{i=i}^{k} m(\mathsf{a}_i) v_{\mathsf{a}_i}$, and $\widetilde{m}(\mathbf{O}_2) = \sum_{i=i}^{k} m(\mathsf{a}_i) w_{\mathsf{a}_i}$. □

Next, we formulate a monotonicity assumption to generalize Assumption 2.6. Given two bf reference lotteries, if any of them assigns a higher degree of belief to the best outcome O_1 and a lower degree of belief to the worst outcome O_r (or, equivalently, higher degrees of belief and plausibility to O_1), then it should arguably be preferred. If this is not the case, i.e., if the best and the worst outcomes both have a higher degree of belief for one lottery (or, equivalently, if the belief-plausibility intervals for O_1 in the two lotteries are strictly nested), then there does not seem to be any solid ground for preference, and the two lotteries can be considered as incomparable. This line of reasoning is formalized in Assumption 4.6 below.

Assumption 4.6 *(Monotonicity).* Suppose $L = [\mathbf{O}_2, (u, v, w)]$ and $L' = [\mathbf{O}_2, (u', v', w')]$ are bf reference lotteries. Then, $L \succsim L'$ if and only if $u \geq u'$ and $1 - v \geq 1 - v'$.

It is clear that $\succsim$ as defined in Assumption 4.6 is reflexive and transitive. The corresponding indifference relation is $L \sim L'$ if and only if $u = u'$ and $v = v'$, i.e., if and only if $L = L'$. Also, the preference relation $\succsim$ on the set of all bf reference lotteries is obviously incomplete: two lotteries L and L' are *incomparable* if not $L \succsim L'$ and not $L' \succsim L$, i.e., if the intervals $[u, 1 - v]$ and $[u', 1 - v']$ are strictly nested.

The preference relation defined in Assumption 4.6 can equivalently be expressed as $L \succsim L'$ if and only if $Bel_m(\{O_1\}) \geq Bel_{m'}(\{O_1\})$ and $Pl_m(\{O_1\}) \geq Pl_{m'}(\{O_1\})$ (meaning that the best outcome O_1 is deemed both more credible and more plausible under L than it is if under L').

We note that a stronger notion of preference would be to prefer L over L' if and only if $Bel_m(\{O_1\}) \geq Pl_{m'}(\{O_1\})$, i.e., if and only if $u \geq 1 - v'$. This alternative preference relation is arguably too strict, which would lead to a more incomplete preference order on lotteries.

To conclude this section, we note that Assumptions 4.1, 4.3 and 4.6 imply the following consistency constraints between the reference bf lotteries equivalent to single outcomes:

$$1 = u_{\{O_1\}} \geq u_{\{O_2\}} \geq \ldots \geq u_{\{O_r\}} = 0. \tag{18}$$

4.3. Representation theorem

Theorem 4.2 *(Interval-valued utility function). Suppose $L = [\boldsymbol{O}, m]$ and $L' = [\boldsymbol{O}, m']$ are bf lotteries on $\boldsymbol{O}$. If the preference relation $\succsim$ on $\mathcal{L}_{bf}$ satisfies Assumptions 4.1–4.6, then there are intervals $[u_a, 1 - v_a]$ associated with nonempty subsets $a \subseteq \boldsymbol{O}$ such that $L \succsim L'$ if and only if*

$$\sum_{\emptyset \neq a \subseteq \boldsymbol{O}} m(a)\, u_a \geq \sum_{\emptyset \neq a \subseteq \boldsymbol{O}} m'(a)\, u_a \tag{19a}$$

and

$$\sum_{\emptyset \neq a \subseteq \boldsymbol{O}} m(a)\, v_a \leq \sum_{\emptyset \neq a \subseteq \boldsymbol{O}} m'(a)\, v_a. \tag{19b}$$

Thus, for a bf lottery $L = [\boldsymbol{O}, m]$, we can define

$$[u](L) = [u, 1 - v] \tag{20}$$

as an interval-valued utility of L, with

$$u = \sum_{\emptyset \neq a \subseteq \boldsymbol{O}} m(a)\, u_{a_i} \quad \text{and} \quad v = \sum_{\emptyset \neq a \subseteq \boldsymbol{O}} m(a)\, v_a. \tag{21}$$

Also, such a utility function is unique up to a strictly increasing affine transformation, i.e., if $u' = a\,u + b$, and $v' = a\,v + b$, where $a > 0$, and b are real constants, then

$$[u'](L) = [u', 1 - v']$$

also qualifies as an interval-valued utility function.

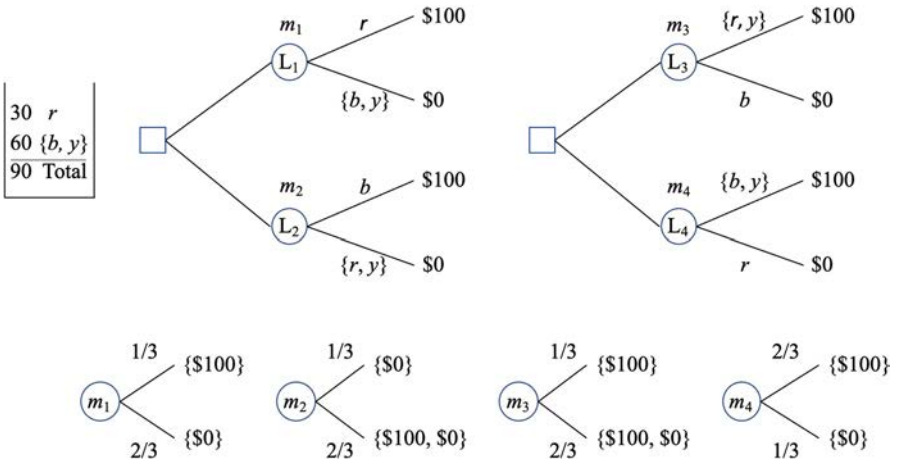

Fig. 6. Ellsberg's urn, choice of lotteries, and the corresponding belief function lotteries.

Proof. The proof is immediate from Theorem 4.1 and Assumption 4.6 (monotonicity). □

A special case of Theorem 4.2 is if we use Bayesian bf reference lotteries for the continuity assumption, i.e., $w_{\mathsf{a}} = 0$ for all focal sets a of m. In this case, Theorem 4.2 implies Corollary 4.1 below where we have a real-valued utility function and a complete ordering on $\mathcal{L}_{bf}$.

Corollary 4.1 *(Real-valued utility function). Suppose $L = [\boldsymbol{O}, m]$ and $L' = [\boldsymbol{O}, m']$ are bf lotteries on $\boldsymbol{O}$. If the preference relation $\succsim$ on $\mathcal{L}_{bf}$ satisfies Assumptions 4.1–4.6 and if $w_{\mathsf{a}} = 0$ for all focal sets a of m and m', then there are numbers u_{a} associated with nonempty subsets $\mathsf{a} \subseteq \boldsymbol{O}$ such that $L_1 \succsim L_2$ if and only if*

$$\sum_{\emptyset \neq \mathsf{a} \subseteq \boldsymbol{O}} m(\mathsf{a})\, u_{\mathsf{a}} \geq \sum_{\emptyset \neq \mathsf{a} \subseteq \boldsymbol{O}} m'(\mathsf{a})\, u_{\mathsf{a}}.$$

Thus, for a bf lottery $L = [\boldsymbol{O}, m]$, we can define

$$u(L) = \sum_{\emptyset \neq \mathsf{a} \subseteq \boldsymbol{O}} m(\mathsf{a})\, u_{\mathsf{a}} \tag{22}$$

as the utility of L. Also, such a utility function is unique up to a strictly increasing affine transformation, i.e., if $u'_{\mathsf{a}} = a\, u_{\mathsf{a}} + b$, where $a > 0$, and b are real constants, then

$$u'(L) = \sum_{\emptyset \neq \mathsf{a} \subseteq \boldsymbol{O}} m(\mathsf{a})\, u'_{\mathsf{a}}$$

also qualifies as a utility function.

Proof. The result in Corollary 4.1 follows trivially from Theorem 4.2. □

The utility function in Eq. (22) has exactly the same form as Jaffray's linear utility [30]. This is discussed further in Section 5.1.

Next, we illustrate the application of Theorem 4.2 to some examples: Ellsberg's urn problem described in Example 2, the one red ball problem described in [35], and the 1,000 balls urns described in [2].

Example 5 *(Ellsberg's urn).* Consider the four bf lotteries described in Example 2 (also in Table 1, and in Fig. 6). Given a vacuous bf lottery $[\{\$100, \$0\}, \iota]$, where ι is the vacuous BPA on $\boldsymbol{O} = \{\$100, \$0\}$, what is an indifferent bf reference lottery? For an ambiguity-averse DM,

$$[\{\$100, \$0\}, (1/2, 1/2, 0)] \succ [\{\$100, \$0\}, \iota].$$

For such a DM, we thus have $1 - v_{\{\$100, \$0\}} < 1/2$.

For the first choice problem between L_1 (\$100 on r) and L_2 (\$100 on b), using Eq. (22), $[u](L_1) = [1/3, 1/3]$, and

$$[u](L_2) = \frac{2}{3}\left[u_{\{\$100, \$0\}}, 1 - v_{\{\$100, \$0\}}\right].$$

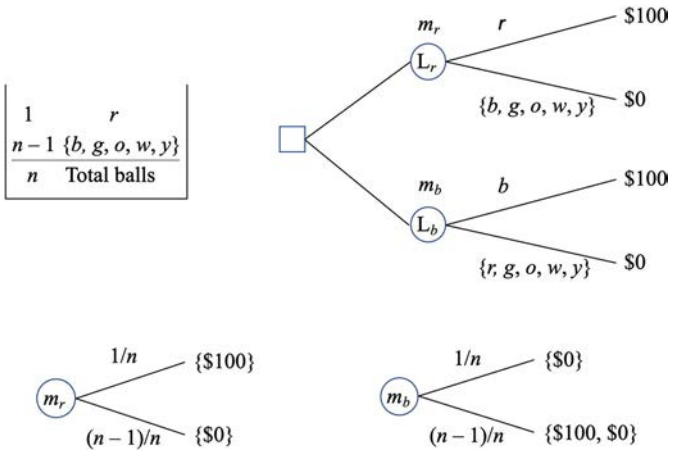

Fig. 7. One red ball urn, choices, and the corresponding belief function lotteries.

Thus, an ambiguity-averse DM would choose L_1. This result is valid as long as $1 - v_{\{\$100,\,\$0\}} < 1/2$ and is consistent with Ellsberg's findings. For the second choice problem between L_3 (\$100 on r or y), and L_4 (\$100 on b or y),

$$[u](L_3) = \frac{1}{3}(1) + \frac{2}{3}\left[u_{\{\$100,\,\$0\}},\, 1 - v_{\{\$100,\,\$0\}}\right],$$

and $[u](L_4) = [2/3, 2/3]$. An ambiguity-averse DM would choose L_4, as

$$\frac{1}{3} + \frac{2}{3}(1 - v_{\{\$100,\,\$0\}}) < \frac{2}{3}$$

as long as $1 - v_{\{\$100,\,\$0\}} < 1/2$, a result that is also consistent with Ellsberg's empirical findings.

Example 6 *(One red ball).* Consider the following example called 'one red ball' in [35] (see Fig. 7). An urn possibly contains balls of six colors: red (r), blue (b), green (g), orange (o), white (w), and yellow (y). One ball is drawn at random from the urn. We are informed that the urn has a total of n balls, where n is a positive integer, and that there is exactly one red ball in the urn. Suppose random variable X denotes the color of the ball drawn from the urn. Then $\Omega_X = \{r, b, g, o, w, y\}$, and m_X is a BPA for X such that $m_X(\{r\}) = 1/n$, and $m_X(\{b, g, o, w, y\}) = (n-1)/n$. First, you choose a color, and then you draw a ball at random from the urn. You win \$100 if the color of the ball drawn from the urn matches the color you chose, and you win \$0 if it doesn't. What color do you choose? In [35], the authors describe some informal experiments where all respondents chose red for $n \leq 7$, and for $n \geq 8$, several respondents preferred a color different from red.

Suppose you choose r. The bf lottery L_r based on m_X is as follows: $[\{\$100, \$0\}, m_r]$, where $m_r(\{\$100\}) = 1/n$, and $m_r(\{\$0\}) = (n-1)/n$. If the color you pick is b, then the bf lottery L_b is $[\{\$100, \$0\}, m_b]$, where $m_b(\{\$0\}) = 1/n$, and $m_b(\{\$100, \$0\}) = (n-1)/n$. Thus, we have $[u](L_r) = [1/n, 1/n]$, and

$$[u](L_b) = \frac{n-1}{n}\left[u_{\{\$100,\,\$0\}},\, 1 - v_{\{\$100,\,\$0\}}\right].$$

So, L_b is strictly preferred to L_r whenever

$$\frac{n-1}{n} u_{\{\$100,\,\$0\}} > \frac{1}{n},$$

i.e., whenever $u_{\{\$100,\,\$0\}} > 1/(n-1)$, and L_r is strictly preferred to L_b whenever

$$\frac{n-1}{n}\left(1 - v_{\{\$100,\,\$0\}}\right) < \frac{1}{n},$$

i.e., whenever $1 - v_{\{\$100,\,\$0\}} < 1/(n-1)$. Hence, L_b is increasingly preferred to L_r when n increases, which is consistent with the findings reported in [35]. In our model, when

$$u_{\{\$100,\,\$0\}} < \frac{1}{n-1} < 1 - v_{\{\$100,\,\$0\}},$$

the two lotteries L_r and L_b are incomparable. If forced to choose, the DM might just choose arbitrarily. As the experiment reported in [35] did not allow the respondents to express inability to choose between the two lotteries, it does not provide any evidence for or against our model.

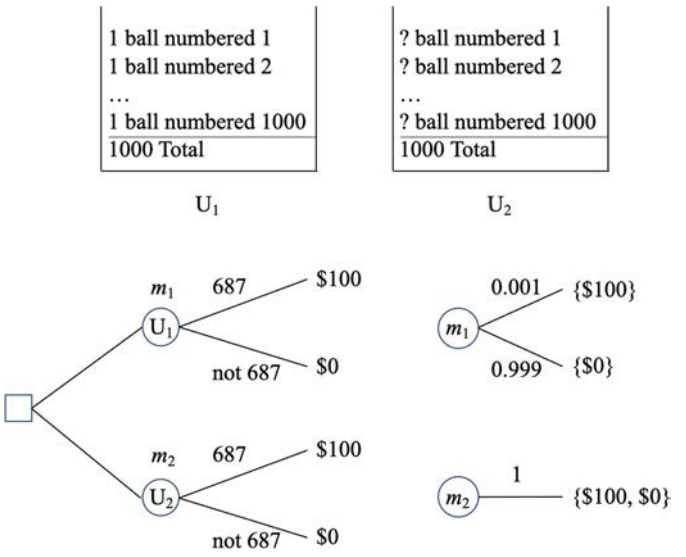

Fig. 8. Two urns with 1,000 balls, choices, and the corresponding belief function lotteries.

Example 7 *(Urns with 1,000 balls).* The following example is discussed in [2], where it is credited to Ellsberg in an oral conversation (with the authors of [2]). It is also discussed in [14]. There are two urns, each with 1,000 balls, numbered from $1-1,000$. Urn 1 has exactly one ball for each number, and there is no ambiguity. Urn 2 has unknown number of balls of each number, and there is much ambiguity. One ball is to be chosen at random from an urn of your choosing. If the number on the ball matches a specific number, e.g., 687, you win \$100, and if not, you win nothing (\$0). Which one of the two urns will you choose? This choice problem is shown in Fig. 8.

It is reported in [2] that many respondents chose Urn 2. Why? Urn 1 has only one ball numbered 687, and therefore, the probability of winning \$100 if the choice is Urn 1 is very small, 0.001. Urn 2 could possibly have anywhere from 0 to 1,000 balls numbered 687. Thus, the choice of Urn 2, although ambiguous, is appealing. Let's analyze this problem using Theorem 4.2.

Let X_1 denote the number on the ball chosen Urn 1, and let X_2 denote the number on the ball chosen Urn 2. $\Omega_{X_1} = \Omega_{X_2} = \{1, \ldots, 1000\}$. Function m_{X_1} is a BPA for X_1 as follows: $m_{X_1}(\{1\}) = \ldots = m_{X_1}(\{1000\}) = 0.001$. BPA m_{X_2} is vacuous, i.e., $m_{X_2}(\Omega_{X_2}) = 1$.

Lottery L_1 corresponding to choice of Urn 1 (say, alternative f_1) is $[\{\$100, \$0\}, m_1]$, where m_1 is a BPA for $\{\$100, \$0\}$ such that $m_1(\{\$100\}) = 0.001$, and $m_1(\{\$0\}) = 0.999$. L_1 is a bf reference lottery, and thus, $[u](L_1) = [0.001, 0.001]$. Lottery L_2 corresponding to choice of Urn 2 (say, alternative f_2) is $[\{\$100, \$0\}, m_2]$, where m_2 is a vacuous BPA for $\{\$100, \$0\}$. The utility of L_2 is

$$[u](L_2) = \left[u_{\{\$100,\,\$0\}}, 1 - v_{\{\$100,\,\$0\}}\right].$$

Consequently, L_2 is preferred to L_1 as long as

$$u_{\{\$100,\,\$0\}} \geq 0.001,$$

a condition that is easily satisfied. This may explain why many DMs appear to be ambiguity-seeking in this context, i.e., prefer L_2 to L_1.

4.4. An additional assumption and the corresponding representation theorem

Whereas Theorem 4.2 guarantees the existence of an interval valued utility function for bf lotteries, there remains the problem of practical elicitation of utilities. The maximum number of utilities to be elicited increases exponentially with the number r of outcomes. As the utilities of the worst and the best outcomes are, by construction, 0 and 1, and $w_{\{o_i\}} = 0$ for each other single outcome o_i, the actual number of utilities to be elicited is, at most, $2(2^r - 1 - r) + r - 2 = 2^{r+1} - r - 4$. By making one more reasonable assumption, we can drastically limit the number of parameters to be elicited.

Assumption 4.7 below has no counterpart in the vN-M theory, but it is rooted in decision-making under ignorance [30,52]. For any nonempty subset of consequences $\mathsf{a} \subseteq \mathbf{O}$, let $\underline{o}_{\mathsf{a}}$ and $\overline{o}_{\mathsf{a}}$ denote, respectively, the worst and the best outcome in a. To simplify the notations, we assimilate a deterministic BPA with its focal set, and we write $\mathsf{a} \gtrsim \mathsf{b}$ to mean that the deterministic lottery with focal set a is preferred or indifferent to the deterministic lottery with focal set b. Then our last assumption can be stated as follows.

Table 2
Acts in Example 8.

	ω_1	ω_2	ω_3	ω_4
f_1	0	1	1000	1000
f_2	0	0	999	1000
f_1'	0	1	1	1000
f_2'	0	999	1000	1000

Assumption 4.7 *(Dominance).* For all nonempty subsets a and b of $\mathbf{O}$, if $\underline{O}_{\mathsf{a}} \succsim \underline{O}_{\mathsf{b}}$ and $\overline{O}_{\mathsf{a}} \succsim \overline{O}_{\mathsf{b}}$, then $\mathsf{a} \succsim \mathsf{b}$.

Assumption 4.7 implies that the preference between two deterministic lotteries with focal sets a and b is determined only by the best and worst outcomes in a and b. In particular, when $\underline{O}_{\mathsf{a}} = \underline{O}_{\mathsf{b}}$ and $\overline{O}_{\mathsf{a}} = \overline{O}_{\mathsf{b}}$, then $\mathsf{a} \sim \mathsf{b}$. Although counter-intuitive at first glance, this assumption cannot be avoided if we accept the PCI, i.e., if we accept that any two acts yielding the same bf lottery are equivalent, as shown by the following example.

Example 8. Assume that the set of outcomes is $\mathbf{O} = \{\$0, \$1, \$999, \$1,000\}$. According to Assumption 4.7, a DM would be indifferent between receiving one of the prizes in $\mathsf{a} = \{\$0, \$1, \$1,000\}$, without any further information, and receiving one of the prizes in $\mathsf{b} = \{\$0, \$999, \$1,000\}$. It may be argued that most DMs would strictly prefer b to a. Yet, assume that the state space is $\Omega_X = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, we are in a state of complete ignorance, i.e., $m_X(\Omega_X) = 1$, and the deterministic lotteries a and b are generated by the acts f_1 and f_2 shown in Table 2. It is clear that f_1 dominates f_2 (it yields at least as desirable consequences for all states of nature, and strictly preferred consequences for some states of nature), so it would be paradoxical to strictly prefer f_2 over f_1, i.e., to strictly prefer b over a. But a and b might also have been generated by acts f_1' and f_2' in Table 2 and, as f_2' dominates f_1', it would also be paradoxical to strictly prefer f_1' over f_2', i.e., to strictly prefer a over b. Consequently, indifference between a and b seems to be the only rational option in this case.

Generalizing Example 8, Jaffray [30] shows that, whenever $\underline{O}_{\mathsf{a}} \succsim \underline{O}_{\mathsf{b}}$ and $\overline{O}_{\mathsf{a}} \succsim \overline{O}_{\mathsf{b}}$, we can always construct a state space Ω_X and two acts f_1 and f_2 such that $f_1[\Omega_X] = \mathsf{a}$, $f_2[\Omega_X] = \mathsf{b}$ and, for any $\omega \in \Omega_X$, $f_1(\omega) \succsim f_2(\omega)$. As f_1 yields at least as desirable outcomes as f_2 under any state of nature, it should be preferred whatever our beliefs on Ω_X, and in particular when $m_X(\Omega_X) = 1$. Hence, we should have $\mathsf{a} \succsim \mathsf{b}$.

Assumption 4.7 implies that $\mathsf{a} \succsim \underline{O}_{\mathsf{a}}$ and $\overline{O}_{\mathsf{a}} \succsim \mathsf{a}$. From Assumption 4.6, we thus have $u_{\mathsf{a}} \geq u_{\underline{O}_{\mathsf{a}}}$ and $1 - v_{\mathsf{a}} \leq u_{\overline{O}_{\mathsf{a}}}$. Consequently, the utility bounds u_{a} and $1 - v_{\mathsf{a}}$ of subset a can be written as convex combinations of the utilities of its worst and best outcomes:

$$u_{\mathsf{a}} = \alpha(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}})\, u_{\underline{O}_{\mathsf{a}}} + \left(1 - \alpha(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}})\right) u_{\overline{O}_{\mathsf{a}}} \tag{23a}$$

$$1 - v_{\mathsf{a}} = \beta(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}})\, u_{\underline{O}_{\mathsf{a}}} + \left(1 - \beta(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}})\right) u_{\overline{O}_{\mathsf{a}}}, \tag{23b}$$

where $\alpha(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}})$ and $\beta(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}})$ are two coefficients depending only on the best and worst outcomes in a, such that $0 \leq \alpha(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}}) \leq \beta(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}}) \leq 1$. In Jaffray's framework [30], $w_{\mathsf{a}} = 0$ (see Section 5.1) and $\alpha(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}})$ is called a *local pessimism index*. In our framework, we can see the interval $\left[\alpha(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}}), \beta(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}})\right]$ as an interval-valued local pessimism index reflecting both the DM's attitude to ambiguity and indeterminacy. Assumption 4.7 thus brings the maximum number of parameters to be elicited from $2^{r+1} - r - 4$ down to $r(r-1) + r - 2 = r^2 - 2$. The above discussion can be summarized in the form of the following representation theorem (generalizing Theorem 2 in [30]).

Theorem 4.3 *(Interval-valued local pessimism index). Suppose $L = [\mathbf{O}, m]$ and $L' = [\mathbf{O}, m']$ are bf lotteries on $\mathbf{O}$. If the preference relation $\succsim$ on $\mathcal{L}_{bf}$ satisfies Assumptions 4.1–4.7, then there are numbers u_O associated with outcomes $O \in \mathbf{O}$ and two mappings α and β from*

$$\mathcal{O} = \{(O, O') \in \mathbf{O}^2 : O' \succsim O\}$$

to $[0, 1]$, with $\alpha \leq \beta$, such that $L \succsim L'$ if and only if

$$\sum_{\emptyset \neq a \subseteq \mathbf{O}} m(a) \left[\alpha(\underline{O}_a, \overline{O}_a) u_{\underline{O}_a} + (1 - \alpha(\underline{O}_a, \overline{O}_a)) u_{\overline{O}_a}\right] \geq \sum_{\emptyset \neq a \subseteq \mathbf{O}} m'(a) \left[\alpha(\underline{O}_a, \overline{O}_a) u_{\underline{O}_a} + (1 - \alpha(\underline{O}_a, \overline{O}_a)) u_{\overline{O}_a}\right]$$

and

$$\sum_{\emptyset \neq a \subseteq \mathbf{O}} m(a) \left[\beta(\underline{O}_a, \overline{O}_a) u_{\underline{O}_a} + (1 - \beta(\underline{O}_a, \overline{O}_a)) u_{\overline{O}_a}\right] \geq \sum_{\emptyset \neq a \subseteq \mathbf{O}} m'(a) \left[\beta(\underline{O}_a, \overline{O}_a) u_{\underline{O}_a} + (1 - \beta(\underline{O}_a, \overline{O}_a)) u_{\overline{O}_a}\right],$$

where $\underline{O}_a$ *and* $\overline{O}_a$ *are, respectively, the worst and the best outcomes in* $a \subseteq \boldsymbol{O}$*. Thus, for a bf lottery* $L = [\boldsymbol{O}, m]$*, we can define*

$$[u](L) = [u, 1 - v]$$

as an interval-valued utility of L, with

$$u = \sum_{\emptyset \neq a \subseteq \boldsymbol{O}} m(a) \left[\alpha(\underline{O}_a, \overline{O}_a) u_{\underline{O}_a} + (1 - \alpha(\underline{O}_a, \overline{O}_a)) u_{\overline{O}_a} \right]$$

and

$$1 - v = \sum_{\emptyset \neq a \subseteq \boldsymbol{O}} m(a) \left[\beta(\underline{O}_a, \overline{O}_a) u_{\underline{O}_a} + (1 - \beta(\underline{O}_a, \overline{O}_a)) u_{\overline{O}_a} \right].$$

Also, this utility function is unique up to a strictly increasing affine transformation.

4.5. A simpler model

Strat [50] proposes independently, but without any axiomatic justification, a criterion similar to that of Theorem 4.3 but with a constant parameter $\alpha(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}}) = \beta(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}}) = \alpha$ that does not depend on the subset a. In a similar way, we can assume that the lower and upper pessimism indices take on constant values: $\alpha(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}}) = \alpha$ and $\beta(\underline{O}_{\mathsf{a}}, \overline{O}_{\mathsf{a}}) = \beta$, with $0 \leq \alpha \leq \beta \leq 1$. This simple model depends on only r parameters: the utilities of the single outcomes $u_{\{O_i\}}$ for $i = 2, \ldots, r-1$, and the two coefficients α and β. It allows us to recover some existing decision criteria as special cases:

- When $\alpha = \beta$, the utility interval $[u](L)$ is reduced to a point $u(L)$ and we get Strat's criterion, also called the generalized Hurwicz criterion in [7], which is a special case of the real-valued utility (22) in Corollary 4.1;
- In particular, when $\alpha = \beta = 0$, then

$$u(L) = \sum_{\mathsf{a} \subseteq \boldsymbol{O}} m(\mathsf{a}) \min_{O \in \mathsf{a}} u_O, \tag{24}$$

  which is the lower expected utility $\underline{u}_m$ with respect to m [5,40,6]. As shown by Gilboa and Schmeidler [25], $\underline{u}_m$ is also the Choquet expected utility [3] with respect to the belief function Bel_m corresponding to m. The preference relation between bf lotteries then corresponds to the maximin criterion, which reflects a pessimistic attitude of the DM.
- Similarly, when $\alpha = \beta = 1$, we get

$$u(L) = \sum_{\mathsf{a} \subseteq \boldsymbol{O}} m(\mathsf{a}) \max_{O \in \mathsf{a}} u_O, \tag{25}$$

  which is the upper expected utility $\overline{u}_m$, or the Choquet expected utility with respect to the plausibility function Pl_m corresponding to m [25]. The corresponding decision strategy corresponds to the maximax criterion, which models an optimistic attitude of the DM.
- When $\alpha = 0$ and $\beta = 1$, then the interval-valued utility is equal to the lower-upper expected utility interval

$$[u](L) = \left[\underline{u}_m, \overline{u}_m\right].$$

  The corresponding preference relation is then the *interval bound dominance* relation [10,7], defined by

$$L \gtrsim L' \Leftrightarrow \left(\underline{u}_m \geq \underline{u}_{m'} \quad \text{and} \quad \overline{u}_m \geq \overline{u}_{m'}\right). \tag{26}$$

In the general case, we have

$$\underline{u}_m \leq u \leq 1 - v \leq \overline{u}_m, \tag{27}$$

where u and v are as in Eq. (21). Thus, the interval-valued utility $[u](L)$ of lottery $[\mathbf{O}, m]$ as defined in Theorem 4.2 is always included in the lower-upper expected utility interval, and the preference relation induced by our interval-valued utilities compares more bf lotteries than the interval dominance relation (26). The lower and upper expectations defined by Eqs. (24)-(25) can thus be seen as lower and upper bounds of the interval utility of a lottery $L = [\mathbf{O}, m]$ and could be used as conservative estimates if parameters α and β cannot be elicited.

Example 9. Assume that the set of outcome is $\mathbf{O} = \{\$0, \$10, \$50, \$100\}$. The full model (without Assumption 4.7) requires the assessment of 24 parameters: $u_{\{\$10\}}$, $u_{\{\$50\}}$, and the u and v values for the 11 subsets of $\mathbf{O}$ with cardinality strictly greater than 1. With Assumption 4.7, the number of parameters to be elicited is down to 14: $u_{\{\$10\}}$, $u_{\{\$50\}}$, and the α and β values for the following pairs of worst and best outcomes: ($\$0, \$10$), ($\$0, \$50$), ($\$0, \$100$), ($\$10, \$50$), ($\$10, \$100$) and ($\$50, \$100$). Assuming α and β to be constant brings the number of parameters to only 4.

A practical elicitation procedure. Whatever the simplifying assumptions made, the trickiest part for eliciting the interval-valued utility of a bf lottery resides in the determination of the equivalent bf reference lottery for any non-singleton focal set a (if α and β are assumed to be constant, this determination needs to be done for only one non-singleton focal set). Let $\widetilde{\mathsf{a}} = [\mathbf{O}_2, (u_\mathsf{a}, v_\mathsf{a}, w_\mathsf{a})]$ be the bf reference lottery equivalent to a (assumed to exist from Assumption 4.3). For any a probabilistic reference lottery $L = [\mathbf{O}_2, (u, 1-u)]$, there are three cases:

1. If $u \geq 1 - v_\mathsf{a}$, then $L \succ \widetilde{\mathsf{a}}$;
2. If $u \leq u_\mathsf{a}$, then $\widetilde{\mathsf{a}} \succ L$;
3. If $u_\mathsf{a} < u \leq 1 - v_\mathsf{a}$ or $u_\mathsf{a} \leq u < 1 - v_\mathsf{a}$, then $\widetilde{\mathsf{a}}$ and L are incomparable.

To determine u_a and v_a, we can thus start with $u = 0$ and gradually increase u until $\widetilde{\mathsf{a}}$ and L become incomparable, which gives us u_a, and then gradually decrease u from $u = 1$ until $\widetilde{\mathsf{a}}$ and L become incomparable, which gives us v_a. Parameters α and β are then obtained by solving Eqs. (23). This procedure was used implicitly in Example 4.

5. Comparison with some existing decision theories

In this section, we compare our utility theory to Jaffray's linear utility theory [30], Smets' two-level decision theory [48], decision theories for possibility theory [11,21] and partially consonant belief functions [22], and Shafer's constructive decision theory [44].

5.1. Comparison with Jaffray's axiomatic theory

Jaffray's axiomatic theory is based on considering the set of all belief functions on $\mathbf{O}$ as a mixture set as follows. Suppose m_1 and m_2 are BPAs for $\mathbf{O}$, and suppose $\lambda \in [0, 1]$. Then m defined as:

$$m(\mathsf{a}) = \lambda\, m_1(\mathsf{a}) + (1 - \lambda) m_2(\mathsf{a}) \tag{28}$$

for all $\mathsf{a} \in 2^{\mathbf{O}}$, is a BPA for $\mathbf{O}$. BPA m can be written as $m = \lambda\, m_1 + (1 - \lambda) m_2$, and called a *mixture* of m_1 and m_2. Using the Jensen-version [34] of vN-M axiom system, Jaffray uses the following axioms, all of which are expressed using mixture BPA functions:

Assumption 5.1 *(Completeness and transitivity).* The relation $\succsim$ is complete and transitive over $\mathcal{L}_{bf}$.

Assumption 5.2 *(Independence).* For all $L_1 = [\mathbf{O}, m_1]$ and $L_2 = [\mathbf{O}, m_2]$ in $\mathcal{L}_{bf}$, and $\lambda \in (0, 1)$, $L_1 \succ L_2$ implies $[\mathbf{O}, \lambda\, m_1 + (1 - \lambda)\, m] \succ [\mathbf{O}, \lambda\, m_2 + (1 - \lambda)\, m]$.

Assumption 5.3 *(Continuity).* For all $L_1 = [\mathbf{O}, m_1]$, $L_2 = [\mathbf{O}, m_2]$, and $L_3 = [\mathbf{O}, m_3]$ in $\mathcal{L}_{bf}$ such that $L_1 \succ L_2 \succ L_3$, there exists λ and μ in $(0, 1)$ such that

$$[\mathbf{O}, \lambda\, m_1 + (1 - \lambda)\, m_3] \succ [\mathbf{O}, m_2] \succ [\mathbf{O}, \mu\, m_1 + (1 - \mu)\, m_3].$$

Theorem 5.1 *(Jaffray's representation theorem [30]). The preference relation $\succsim$ on $\mathcal{L}_{bf}$ satisfies Assumptions 5.1–5.3 if and only if there exists a utility function $u : \mathcal{L}_{bf} \to \mathbb{R}$ such that for any lottery $L = [\boldsymbol{O}, m]$ in $\mathcal{L}_{bf}$,*

$$u(L) = \sum_{\emptyset \neq a \subseteq \boldsymbol{O}} m(a)\, u_a \tag{29}$$

where $u_a = u([\boldsymbol{O}, m^d_a])$, and m^d_a is a deterministic BPA for $\boldsymbol{O}$ such that $m^d_a(a) = 1$.

Thus, Jaffray's axioms result in the same solution as that of Corollary 4.1, which is a special case of Theorem 4.2. As Jaffray's axioms do not use Dempster's rule explicitly, it is not clear whether Eq. (29) applies to the D-S framework or not. The mixture BPA m derived from BPAs m_1 and m_2 using Eq. (28) is not Dempster's combination rule, although Eq. (28) can be derived from a belief function model using Dempster's rule. In [32], Jaffray writes:

> "It has been shown by [30,31] that, in the *lower probability interpretation* of belief functions, the axioms of von Neumann-Morgenstern linear utility theory could be justified with the same arguments as in the case of risk (probabilized uncertainty)" (emphasis added).

Also, in [33], Jaffray and Wakker write:

"Given the widespread use of belief functions, it is remarkable that only recently were decision criterion for the above-mentioned type of situations proposed and axiomatized in [30]. He uses as a primitive axiom the independence condition with respect to mixtures of belief functions over the outcomes to generalize expected utility: [31] justifies this condition by means of a *lower-probability interpretation* of belief functions." (emphasis added).

Thus, it is clear that Jaffray has in mind the credal set semantics of belief functions, which are inconsistent with Dempster's combination rule.[5]

In comparison, our representation theorem is based on a set of axioms making use of the basic constructs of DS theory (namely, Dempster's combination rule, marginalization, and conditional embedding), we provide more compelling arguments supporting Eq. (29) as a natural definition of the real-valued utility of a bf lottery in the D-S theory.

Also, there is no explicit notion of a bf reference lottery in Jaffray's framework. Thanks to our continuity axiom (Assumption 4.3), the interval-valued utility $[u_{\mathsf{a}}, 1 - v_{\mathsf{a}}]$ in our framework receives a simple interpretation as an interval-valued probability of a best outcome O_1, in a bf reference lottery $[\mathbf{O}_2, m_{\mathsf{a}}]$ that is indifferent to a and such that $m_{\mathsf{a}}(\{O_1\}) = u_{\mathsf{a}}$, $m_{\mathsf{a}}(\{O_r\}) = v_{\mathsf{a}}$, and $m_{\mathsf{a}}(\mathbf{O}_2) = 1 - u_{\mathsf{a}} - v_{\mathsf{a}}$. We believe that this simple interpretation can be very helpful when eliciting utilities from DMs, as discussed in Section 4.5.

5.2. Comparison with Smets' decision theory

Smets' decision theory [48] is a two-level framework where beliefs, represented by belief functions, are held at a credal level. When a DM has to make a decision, the marginal belief function for a variable of interest is transformed into a PMF, and the Bayesian expected utility framework is then used to make a decision.

Smets uses a transformation called the *pignistic* transform to transform belief functions into PMFs. This transform is justified in [49] using a mixture property as follows. Let T denote the belief-PMF transformation. Smets [49] argues that this transformation should be linear, i.e., we should have, for any $\lambda \in [0, 1]$,

$$T(\lambda m_1 + (1 - \lambda) m_2) = \lambda T(m_1) + (1 - \lambda) T(m_2). \tag{30}$$

The unique transformation T verifying (30) is the pignistic transformation defined as $T(m) = BetP_m$ with

$$BetP_m(O) = \sum_{\mathsf{a} \subseteq \mathbf{O}} \frac{m(\mathsf{a})}{|\mathsf{a}|} I(O \in \mathsf{a}) \tag{31}$$

for all $O \in \mathbf{O}$. The pignistic PMF $BetP_m$ is mathematically identical to the Shapley value in cooperative game theory [45]. In [49], Smets attempts to derive Eq. (30) from the maximum expected utility principle. The argument, however, is quite technical and not very compelling.

Given the definition in Eq. (31), the expected utility of a bf lottery $L = [\mathbf{O}, m]$ according to the pignistic PMF is

$$u_{BetP}(L) = \sum_{O \in \mathbf{O}} BetP_m(O)\, u_{\{O\}} \tag{32a}$$

$$= \sum_{O \in \mathbf{O}} \left(\sum_{\mathsf{a} \subseteq \mathbf{O}} \frac{m(\mathsf{a})}{|\mathsf{a}|} I(O \in \mathsf{a}) \right) u_{\{O\}} \tag{32b}$$

$$= \sum_{\mathsf{a} \subseteq \mathbf{O}} m(\mathsf{a}) \left(\frac{1}{|\mathsf{a}|} \sum_{O \in \mathsf{a}} u_{\{O\}} \right). \tag{32c}$$

It is a special case of Eq. (22), with

$$u_{\mathsf{a}} = \frac{1}{|\mathsf{a}|} \sum_{O \in \mathsf{a}} u_{\{O\}}.$$

Smets' decision theory thus amounts to assuming that a DM is indifferent between a bf lottery that gives them an outcome in a for sure, and a bf reference lottery in which the probability of the best outcome is equal to the average utilities of the outcomes in a. This is consistent with our Assumptions 4.1–4.6, but it is inconsistent with Assumption 4.7. For instance, in Example 8, a DM using the pignistic criterion would strictly prefer b to a, even though the act f_1 generating a dominates the act f_2 generating b. Moreover, this restricted model does not have any parameter to represent a DM's attitude toward ambiguity. As a result, it is unable to explain Ellsberg's paradox and the ambiguity aversion of human DMs as described, e.g., in the examples presented in Section 4.3.

[5] It is possible that in 1989, it was not well understood that credal set semantics of belief functions were incompatible with Dempster's combination rule. This was apparently clarified in the early 1990s by Shafer in [42,43] and also by Fagin and Halpern in [16,27].

5.3. Comparison with other axiomatic theories

In this subsection, we compare our axiomatic decision theory with other axiomatic decision theories for D-S belief functions.

Dubois et al. [11] describe an axiomatic decision theory for the case where uncertainty is described by a possibility distribution, which is a special case of a belief function with nested focal elements (such a belief function is said to be consonant). Dubois et al.'s decision theory consists of two sets of axioms, one for the pessimistic case, and one for the optimistic case. In contrast, Giang and Shenoy [21] propose an axiomatic theory for possibility theory with one set of axioms, and the utility function is binary-valued (binary-valued utilities are possibility distribution values of O_1 and O_r for possibilistic reference lotteries). The two axiomatic theories for possibility theory are compared in detail in [21]. The latter theory is generalized in terms of partially consonant belief functions in [22]. A partially consonant belief function is a belief function where the set of focal elements can be divided into groups such that (a) the focal elements in different groups are disjoint, and (b) the focal elements in the same group are nested. The family of partially consonant belief functions include Bayesian belief functions and consonant belief functions.

Giang [20] compares the Giang-Shenoy decision theory for partially consonant belief functions with Jaffray's axiomatic decision theory for general belief functions. Similar to Jaffray's theory, our decision theory is for the case of general belief functions. While our utility is interval-based, leading to incomplete preferences, Jaffray's theory for general belief functions, and Giang-Shenoy's theory for partially consonant belief functions based on binary utility, result in complete preferences, which is a special case of our theory. Walley [53] argues that partially consonant belief functions is the only class of D-S belief functions that is consistent with the likelihood principle of statistics, but this argument applies only to statistical inference, and not to uncertain reasoning in general.

5.4. Comparison with Shafer's constructive decision theory

Shafer [44] argues for a decision theory that allows us to construct both goals and beliefs in response to a decision. In the vN-M utility theory, we start with a probabilistic lottery, and construct a utility function that reflects a DM's risk attitude. Thus, probabilities and utilities are separate constructs that are then combined for the computation of expected utility. In many situations, we have neither objective nor subjective probabilities. For such situations, Shafer argues for constructing belief functions from available evidence, and constructing a set of consistent and monotonic goals. Given a set of actions, we examine which goals each of the actions will achieve. We use belief functions to make judgments based on evidence about what will happen if an action is taken. We then use these belief functions to compute the expected number of goals that an action will satisfy, and pick an action that satisfies the most goals. This can be generalized to the case where not all goals are equally weighted, some are weighted more than others.

Our utility theory is more in line with vN-M utility theory than Shafer's constructive decision theory. There is considerable literature in many domains about the use of utility theory for decision making. While Shafer's constructive decision theory is intriguing and may indicate an interesting direction to explore, there is much to be done before we can apply it in many domains for which we have a decision theory in the vN-M style.

6. Summary and conclusions

In this section, we summarize our proposal and sketch some future work. We start with Luce and Raiffa's version of the vN-M utility theory for probabilistic lotteries. We then consider bf lotteries, lotteries when our beliefs about the state of the world is described by DS belief functions. We use a similar set of axioms as vN-M, but first we replace each singleton outcome in a probabilistic lottery by a focal set of a BPA. Second, we replace the reduction of compound lotteries with a corresponding axiom that uses Dempster's combination rule and belief function marginalization in place of probabilistic combination (pointwise multiplication followed by normalization) and probabilistic marginalization (addition). Third, we use a bf reference lottery with two independent parameters. The axioms lead to a decision theory that involves assessing the utility of each focal element of a BPA as an interval-valued utility. Interval-valued utilities lead to a partial preference relation on the set $\mathcal{L}_{bf}$ of all bf lotteries. If we use Bayesian bf reference lotteries with a single parameter, then our axiomatic framework leads to a real-valued utility function that is exactly the same as in Jaffray's linear utility theory [30].

The decision theory that results from our axioms is more general than that proposed by Jaffray [30], which can be construed as a decision theory for belief functions interpreted as generalized probabilities. Jaffray's axiomatic theory is based on a set of mixture BPAs. A mixture of two BPAs is not the same as a Dempster's combination of two BPAs, although we could construct a belief function model where the mixture BPA is obtained by Dempster's rule. Thus, it is not clear if Jaffray's linear utility theory is applicable to D-S belief function lotteries or not. Our utility theory confirms that this is indeed the case. Our bf reference lotteries lead to interval-valued utilities, and consequently, a partial preference relation on the set of all bf lotteries.

We also compare our axiomatic theory to Smets' two-level framework [48,49], and note that his framework is too constrained to explain ambiguity-aversion or ambiguity-seeking behavior of human DMs. Other axiomatic decision theories proposed by Dubois et al. [11] and Giang and Shenoy [21,22] are restricted to consonant or quasi-consonant belief functions. Shafer [44] has recently published his constructive decision theory where he rejects the separation of beliefs and utilities.

He proposes, instead, constructing a set of consistent and monotonic goals, and measuring the utility of each choice by the number of goals (or weighted goals) achieved by the choice. Shafer's constructive decision theory needs to be fleshed out before it can be applied to practical decision-making situations.

In practice, implementing the most general form of our axiomatic theory may need assessment of $2k$ parameters, where k is the number of focal sets of a bf lottery. In the worst case, k can be as large as $2^{|\mathbf{O}|} - 1$. Based on additional assumptions, we propose a model based on only two parameters, which can be interpreted as reflecting both the DM's attitude to ambiguity and their indeterminacy. This model, as well as others, will have to be further studied and developed. More generally, a rigorous methodology to elicit interval-valued utilities remains to be designed and validated experimentally.

Finally, in this paper, we start from the assumption that the D-S formalism is an adequate model of an agent's state of knowledge, and we derive a corresponding decision theory from a set of rationality requirements. Thus, a belief function on the state space is assumed to be given, and we generate interval-valued expected utilities for bf lotteries. A further step would be to justify not only utilities, but also the D-S calculus itself (including belief functions and Dempster's rule) from properties of the DM's preference relation over acts, similar to what Savage [38] did to provide a foundation for decision-making with probability theory, similar to what Dubois et al. [12] did to justify decision-making with qualitative possibility theory, and similar to what Gul and Pesendorfer [26], and Zhou et al. [54] did for decision-making with a theory of belief functions where the belief functions are interpreted as credal sets. This task remains to be done.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was done during a visit by the second author to Université de Technologie de Compiègne (UTC) in Spring 2019. This research was funded partially by a sabbatical from the University of Kansas, and supported by the Labex MS2T, which is funded by the French Government through the program "Investments for the future" by the National Agency for Research (reference ANR-11-IDEX-0004-02). Thanks to Sébastien Destercke for many discussions on the topic of this manuscript, and to Suzanna Emelio for proofreading this manuscript. A short version of this paper appeared as [9], and we are grateful to the reviewers of our manuscript submitted to ISIPTA-2019 for their comments and suggestions. Thanks also to the three reviewers of our original manuscript submitted to IJAR for their comments and questions, our revision has benefitted from their comments.

References

[1] R. Aumann, Utility theory without the completeness axiom, Econometrica 30 (3) (1962) 445–462.
[2] S.W. Becker, F.O. Brownson, What price ambiguity? or the role of ambiguity in decision-making, J. Polit. Econ. 72 (1) (February 1964) 62–73.
[3] G. Choquet, Theory of capacities, Ann. Inst. Fourier 5 (1953) 131–295.
[4] L.M. de Campos, M.T. Lamata, S. Moral, The concept of conditional fuzzy measure, Int. J. Intell. Syst. 5 (3) (1990) 237–246.
[5] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (2) (1967) 325–339.
[6] A.P. Dempster, A. Kong, Comment to "Probabilistic expert systems in medicine: practical issues in handling uncertainty", Stat. Sci. 2 (1) (1987) 32–36.
[7] T. Denoeux, Decision-making with belief functions: a review, Int. J. Approx. Reason. 109 (6) (2019) 87–110.
[8] T. Denoeux, D. Dubois, H. Prade, Representations of uncertainty in artificial intelligence: beyond probability and possibility, in: P. Marquis, O. Papini, H. Prade (Eds.), A Guided Tour of Artificial Intelligence Research, vol. 1, Springer Verlag, 2020, pp. 119–150, chapter 4.
[9] T. Denoeux, P.P. Shenoy, An axiomatic utility theory for Dempster-Shafer belief functions, in: J. de Bock, C.P. de Campos, G. de Cooman, E. Quaeghebeur, G. Wheeler (Eds.), Proceedings of Machine Learning Research: Eleventh International Symposium on Imprecise Probabilities: Theories and Applications, vol. 103, JMLR, 2019, pp. 145–155.
[10] S. Destercke, A decision rule for imprecise probabilities based on pair-wise comparison of expectation bounds, in: C. Borgelt, G. González-Rodríguez, W. Trutschnig, M.A. Lubiano, M.Á. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), Combining Soft Computing and Statistical Methods in Data Analysis, Springer, Berlin, Heidelberg, 2010, pp. 189–197.
[11] D. Dubois, L. Godo, H. Prade, A. Zapico, On the possibilistic decision model: from decision under uncertainty to case-based decision, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 7 (6) (1999) 631–670.
[12] D. Dubois, H. Prade, R. Sabbadin, Decision-theoretic foundations of qualitative possibility theory, Eur. J. Oper. Res. 128 (3) (2001) 459–478.
[13] J. Dubra, F. Maccheroni, E. Ok, Expected utility theory without the completeness axiom, J. Econ. Theory 115 (1) (2004) 118–133.
[14] H.J. Einhorn, R.M. Hogarth, Decision making under ambiguity, J. Bus. 59 (4) (1986) S225–S250.
[15] D. Ellsberg, Risk, ambiguity and the Savage axioms, Q. J. Econ. 75 (2) (1961) 643–669.
[16] R. Fagin, J.Y. Halpern, A new approach to updating beliefs, in: P. Bonissone, M. Henrion, L. Kanal, J. Lemmer (Eds.), Uncertainty in Artificial Intelligence 6, North-Holland, Amsterdam, 1991, pp. 347–374.
[17] P.C. Fishburn, The Foundations of Expected Utility, Reidel, Dordrecht, 1982.
[18] T. Gajdos, T. Hayashi, J.-M. Tallon, J.-C. Vergnaud, Attitude toward imprecise information, J. Econ. Theory 140 (1) (2008) 23–56.
[19] P. Ghirardato, Coping with ignorance: unforeseen contingencies and non-additive uncertainty, Econ. Theory 17 (2001) 247–276.
[20] P.H. Giang, Decision with Dempster-Shafer belief functions: decision under ignorance and sequential consistency, Int. J. Approx. Reason. 53 (7) (2012) 38–53.
[21] P.H. Giang, P.P. Shenoy, Two axiomatic approaches to decision making using possibility theory, Eur. J. Oper. Res. 162 (2) (2005) 450–467.
[22] P.H. Giang, P.P. Shenoy, A decision theory for partially consonant belief functions, Int. J. Approx. Reason. 52 (3) (2011) 375–394.
[23] I. Gilboa, M. Marinacci, Ambiguity and the Bayesian paradigm, in: H. Arló-Costa, V. Hendricks, J. van Nenthem (Eds.), Readings in Formal Epistemology, in: Springer Graduate Texts in Philosophy, vol. 1, Springer, Cham, 2016, pp. 385–439.

[24] I. Gilboa, D. Schmeidler, Maxmin expected utility with a non-unique prior, J. Math. Econ. 18 (2) (1989) 141–153.
[25] I. Gilboa, D. Schmeidler, Additive representations of non-additive measures and the Choquet integral, Ann. Oper. Res. 52 (1994) 43–65.
[26] F. Gul, W. Pesendorfer, Expected uncertain utility theory, Econometrica 82 (1) (2014) 1–39.
[27] J.Y. Halpern, R. Fagin, Two views of belief: belief as generalized probability and belief as evidence, Artif. Intell. 54 (3) (1992) 275–317.
[28] M. Hausner, Multidimensional utilities, in: R.M. Thrall, C.H. Coombs, R.L. Davis (Eds.), Decision Processes, Wiley, New York, NY, 1954, pp. 167–180.
[29] I.N. Herstein, J. Milnor, An axiomatic approach to measurable utility, Econometrica 21 (2) (1953) 291–297.
[30] J.-Y. Jaffray, Linear utility theory for belief functions, Oper. Res. Lett. 8 (2) (1989) 107–112.
[31] J.-Y. Jaffray, Linear utility theory and belief functions: a discussion, in: A. Chikan (Ed.), Theory and Decision Library, in: Progress in Decision, Utility and Risk Theory, vol. 13, Springer, Dordrecht, 1991.
[32] J.-Y. Jaffray, Dynamic decision making with belief functions, in: R.R. Yager, M. Fedrizzi, J. Kacprzyk (Eds.), Advances in the Dempster-Shafer Theory of Evidence, Wiley, 1994, pp. 331–352.
[33] J.-Y. Jaffray, P. Wakker, Decision making with belief functions: compatibility and incompatibility with the sure-thing principle, J. Risk Uncertain. 7 (3) (1993) 255–271.
[34] N.E. Jensen, An introduction to Bernoullian utility theory: a discussion, Swed. J. Econ. 69 (3) (1967) 163–183.
[35] R. Jirousek, P.P. Shenoy, Ambiguity aversion and a decision-theoretic framework using belief functions, in: Proceedings of 2017 IEEE Symposium Series on Computational Intelligence, (SSCI), IEEE, Piscataway, NJ, 2017, pp. 326–332.
[36] R. Jiroušek, P.P. Shenoy, On properties of a new decomposable entropy of Dempster-Shafer belief functions, Int. J. Approx. Reason. 119 (4) (2020) 260–279.
[37] R.D. Luce, H. Raiffa, Games and Decisions: Introduction and Critical Survey, Wiley, New York, NY, 1957.
[38] L.J. Savage, The theory of statistical decision, J. Am. Stat. Assoc. 46 (1951) 55–67.
[39] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[40] G. Shafer, Constructive probability, Synthese 48 (1) (1981) 1–60.
[41] G. Shafer, Belief functions and parametric models, J. R. Stat. Soc., Ser. B 44 (3) (1982) 322–352.
[42] G. Shafer, Perspectives on the theory and practice of belief functions, Int. J. Approx. Reason. 4 (5–6) (1990) 323–362.
[43] G. Shafer, Rejoinders to comments on “Perspectives on the theory and practice of belief functions”, Int. J. Approx. Reason. 6 (3) (1992) 445–480.
[44] G. Shafer, Constructive decision theory, Int. J. Approx. Reason. 79 (12) (2016) 45–62.
[45] L.S. Shapley, A value for *n*-person games, in: H. Kuhn, A.W. Tucker (Eds.), Contributions to the Theory of Games, vol. 2, Princeton University Press, Princeton, NJ, 1953, pp. 307–317.
[46] P. Smets, Un modèle mathématico-statistique simulant le processus du diagnostic médical, PhD thesis, Université Libre de Bruxelles, 1978.
[47] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, Int. J. Approx. Reason. 9 (1) (1993) 1–35.
[48] P. Smets, Decision making in a context where uncertainty is represented by belief functions, in: R.P. Srivastava, T.J. Mock (Eds.), Belief Functions in Business Decisions, Physica-Verlag, Heidelberg, 2002, pp. 316–332.
[49] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, Int. J. Approx. Reason. 38 (2) (2005) 133–147.
[50] T.M. Strat, Decision analysis using belief functions, Int. J. Approx. Reason. 4 (5–6) (1990) 391–417.
[51] J. von Neumann, O. Morgenstern, Theory of Games and Economic Behavior, 2nd edition, Princeton University Press, Princeton, NJ, 1947.
[52] P.P. Wakker, Dempster belief functions are based on the principle of complete ignorance, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 8 (3) (2000) 271–284.
[53] P. Walley, Belief function representation of statistical evidence, Ann. Stat. 15 (4) (1987) 1439–1465.
[54] C. Zhou, B. Qin, X. Du, A Savage-style utility theory for belief functions, in: Proceedings of the twenty-seventh International Joint Conference on Artificial Intelligence, (IJCAI-18), AAAI Press, 2018, pp. 5135–5141.