ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 198 (2021) 105369

JOURNAL OF
Economic
Theory

www.elsevier.com/locate/jet

Decision-making with partial information [☆]

Jürgen Eichberger [a,b], Illia Pasichnichenko [c,*]

[a] *Alfred-Weber-Institute, Heidelberg University, Germany*
[b] *HSE University Perm, Russia*
[c] *School of Economics, University of Bristol, United Kingdom of Great Britain and Northern Ireland*

Received 14 September 2020; final version received 2 September 2021; accepted 10 October 2021
Available online 18 October 2021

Abstract

In this paper, we study choice under uncertainty with belief functions. Belief functions can capture partial information by describing what is objectively known about the probabilities of events. State-contingent acts together with a belief function over states induce belief functions over outcomes. We assume that decision makers have preferences over belief functions that reflect both their valuation of outcomes and the information available about the likelihood of outcomes. We provide axioms characterizing a preference representation for belief functions that captures what is (objectively) known about the likelihood of outcomes and combines it with subjective beliefs according to the *"principle of insufficient reason"* whenever the likelihood of events is unknown. This treatment of partial information yields a natural distinction between ambiguity and ambiguity attitudes. The approach is novel in its treatment of partial information and in its axiomatization of the uniform distribution in case of ignorance.

JEL classification: D81

Keywords: Principle of insufficient reason; Ambiguity; Ambiguity attitude; Information about events; Belief functions

---

[☆] We would like to thank Yehuda Izhakian, David Kelsey, Hannes Rau, Peter Wakker for comments and information about related literature unknown to us when we wrote the first drafts of this paper. We are particularly grateful to Peter Wakker for sharing with us his views on Jaffray's and Dempster's perspective on uncertainty and belief functions. We thank also two anonymous referees and the editor of this journal for their constructive criticism that helped us improve the exposition of our paper. Illia Pasichnichenko acknowledges funding for this research by the Alexander von Humboldt Foundation.

[*] Corresponding author.
E-mail addresses: juergen.eichberger@awi.uni-heidelberg.de (J. Eichberger), illia.pasichnichenko@bristol.ac.uk (I. Pasichnichenko).

https://doi.org/10.1016/j.jet.2021.105369

1. Introduction

In economics, decision analysis under uncertainty has almost exclusively focused on the extreme cases of *no information about probabilities* and, hence, purely subjective probabilities derived from a decision maker's preferences (Savage, 1954), or *perfect information about probabilities* (objective lotteries) analyzed by von Neumann and Morgenstern (1944). Yet, providing subjects with partial information about the probabilities of some events, Ellsberg (1961) observed a systematic preference for betting on events where probabilities were known, for example on red balls and on blue-or-yellow balls in the three-color urn paradox.[1]

As in the Ellsberg paradoxes, many economic decision problems are characterized by partial knowledge about the frequencies or probabilities of events. As Shafer (1990, p. 2) notes

> *"I might have conclusive evidence, for example, that I lost my wallet in one of three places, without any clue as to which one. This calls for a belief function that assigns a degree of belief of 100% to the three places as a set, but a degree of belief of 0% to each of the three individually."*

In his *Mathematical Theory of Evidence*, Shafer (1976) suggests a *basic probability assignment* to the events of an outcome space as primitive concept reflecting information about the probabilities of events. Given a finite space of outcomes X and the set of subsets (events) 2^X, a *basic probability assignment,* or a *mass distribution* as it is more commonly called (Grabisch, 2016, p. 380), is a probability distribution m on the finite set of events 2^X. If a mass distribution assigns positive weights to singleton events only, then it is a probability distribution over X. In general, a mass distribution is the *Moebius transform* of a *belief function*.

Even if probabilities of states are unknown in general, the probabilities of some events may be known from publicly available information, as in the case of the Ellsberg paradoxes[2] or from incomplete data collection as suggested by Dempster (1967).[3]

Mass distributions are a natural generalization of probability distributions, encoding available information about the probabilities of the events in 2^X. In general, there will be events $A \in 2^X$ for which the probability, i.e., the weight of the mass distribution $m(A)$, is positive but the probabilities of all subevents $F \subset A$ are zero. Since mass distributions correspond to belief functions, a special case of capacities, a natural notion of expected utility is given by Choquet expected utility (CEU),

$$V^C(m) = \sum_{A \in 2^X} m(A) \min \{u(x) \mid x \in A\}. \tag{1}$$

In a seminal paper, Jaffray (1989) assumed that decision makers have preferences over mass distributions on a set of outcomes X. A mass distribution represents the information about the probability distribution over outcomes that an individual has at the point of decision making. A preference order over a set of mass distributions assumes, therefore, that decision makers can compare outcome distributions with differing partial information, ranging from the case of

---

[1] We discuss this case in detail in Example 3.

[2] As in Ellsberg (1961), instructions of behavioral experiments often provide information about the probabilities of some events.

[3] We will discuss this approach in more detail in Section 2.2.2.

complete information, when m is a probability distribution, to the other extreme when there is no information except that outcomes are elements of the set X, $m(X) = 1$. Since every mass distribution m defines a belief function μ^m,

$$\mu^m(E) = \sum_{F \subseteq E} m(F)$$

for all $E \subseteq S$, one can associate a unique set of probability distributions over outcomes with it, namely, the core of μ^m,

$$\text{core}(\mu^m) = \{p \in \Delta(X) \mid p(A) \geq \mu^m(A) \text{ for any } A \subseteq X\}.$$

Hence, a mass distribution corresponds to a set of probability distributions over outcomes.[4]

In the spirit of von Neumann and Morgenstern (1944), Jaffray (1989) provides axioms for a preference order over mass distributions that characterize a representation V satisfying

$$V(m) = \sum_{A \in 2^X} m(A) V(e_A), \tag{2}$$

where e_A denotes the *elementary mass distribution* assigning a weight of 1 to the event A and, hence, a weight of 0 to all other events. The elementary mass distribution e_A represents the case of partial information where it is known that the event A occurs with certainty but there is no information available regarding other events. If A is a singleton, e.g., $A = \{x\}$, then it is natural to identify $V(e_A)$ with the utility of the outcome $u(x)$. Thus, the formula in Equation (2) coincides with the expected utility representation of von Neumann and Morgenstern (1944) whenever m is a probability distribution, i.e., assigns positive weight to singleton events only.[5]

The representation in Equation (2) leaves open the question of how to evaluate elementary mass distributions e_A for events that are not singletons. Further axioms are needed in order to determine the value $V(e_A)$ of elementary mass distributions. Jaffray (1989) axiomatizes values for elementary mass distributions depending only on the maximal and the minimal outcome in these events.[6] The CEU of the belief function induced by a mass distribution in Equation (1) is a special case of the representation derived in Jaffray (1989).

In this paper, we will propose an alternative representation for preferences over mass distributions that embodies the *Principle of Insufficient Reason (PIR)*. This representation incorporates the objective partial information of a mass distribution as in Equation (2) but assumes a *generalized arithmetic mean*, i.e., a uniform distribution for ambiguous events, i.e., for elementary mass distributions e_A,

$$V(e_A) = \phi^{-1}\left(\frac{1}{|A|}\sum_{x \in A} \phi(u(x))\right). \tag{3}$$

---

[4] In this respect, modeling partial information by a mass distribution is similar to the approach of Gajdos et al. (2008b) who represent partial information by sets of probability distributions. We will discuss this relationship in more detail in Section 5.1 below.

[5] Such application of the linear utility theory is restricted to belief functions, since the Moebius transform of a capacity which is not a belief function is negative for some events (Chateauneuf and Jaffray, 1989).

[6] The α-MEU and CEU functionals are special cases.

Notice that the von Neumann-Morgenstern utility function u is subject to a further transformation ϕ that vanishes for singleton events A but, as we will argue below, will reflect the decision makers' attitudes towards ambiguity when events contain more than one element.[7]

Combining the representation for elementary mass distributions in Equation (3) with the representation of partial information by a mass distribution m in Equation (2) yields a representation for a general mass distribution m based on PIR

$$V(\mu^m) = \sum_{A \in \mathcal{X}} m(A)\, \phi^{-1}\left(\frac{1}{|A|} \sum_{x \in A} \phi(u(x))\right), \tag{4}$$

that we will study in this paper.

In particular, we will

1. show that the outcome-distribution based theories of Dempster (1967), Shafer (1976), and Jaffray (1989) can be embedded in a Savage framework of acts f mapping states in a set S to outcomes in a set X when partial information is modeled by a mass distribution over events in S (rather than X directly);
2. provide an axiomatic foundation of the PIR-based smooth representation in Equation (4);
3. study ambiguity and ambiguity attitudes using mass distributions and the representation in Equation (4); and
4. show that the representation in Equation (4) provides a unifying perspective relating the notion of partial information modeled by sets of priors in Gajdos et al. (2008a) and the smooth preference representation advanced in Klibanoff et al. (2005).

1.1. Relation to the literature

Modeling partial information by a mass distribution is conceptionally, although not formally, close to the idea of modeling partial information by sets of probability distributions advanced in Gajdos et al. (2008a).[8] Gajdos et al. (2008a) assume that "*objective but imprecise information*" determines a set of probability distributions over states.[9] A decision maker has preferences over pairs (P, f) consisting of a set of probability distributions over states P reflecting objective partial information and a Savage act f relating states to outcomes. In this framework, the authors derive a maxmin expected utility (MEU) representation. In contrast to Gajdos et al. (2008a) we study a decision maker who evaluates ambiguity over outcomes according to a quasi-arithmetic mean in the spirit of the PIR.[10]

From a formal point of view our approach is closely related to Jaffray (1989). In particular, Jaffray (1989) was the first to suggest a preference order on the set of belief functions over an outcome space and to apply the axioms ofvon Neumann and Morgenstern (1944) in order to derive the representation in Equation (2). From this point on, however, Jaffray (1989) takes another route for completing his representation. While he axiomatized an evaluation for ambiguous

---

[7] The representation in Equation (3) resembles the representation of the smooth model proposed in Klibanoff et al. (2005), a relationship we will explore in more detail in Section 5.2 below.

[8] Indeed, Giraud and Tallon (2011) suggest the possibility of modeling partial information about the probabilities of events by belief functions but do not follow up on this idea.

[9] Gajdos et al. (2008a, p. 27).

[10] We will provide a more detailed discussion of this approach in Section 5.1.

events based on the maximum and minimum values of the ambiguous outcomes, we propose to evaluate ambiguous outcomes by a generalized arithmetic mean.

In the literature, the distinction of ambiguity and ambiguity attitude became an important topic. Two approaches emerged from this literature: one by Ghirardato et al. (2004) identifies ambiguity neutrality with Subjective Expected Utility (SEU)[11] and a second one by Epstein and Zhang (2001) with probabilistic sophistication.[12] The former approach is bound to confound risk attitudes and ambiguity attitudes and for the latter one has to distinguish ambiguous from unambiguous events.[13] Evaluating ambiguous outcomes according to PIR, one can compare this subjective average due to ambiguity with the objective average of a uniform probability distribution over these outcomes.

In a recent paper, Grant et al. (2021) provide an axiomatic treatment of preferences over acts in the framework of Savage (1954). In this context, they derive a *subjective* belief function for ambiguous sets of outcomes. Moreover, the *implicit linear utility representation* characterized in their paper allows for a generalized arithmetic average as a special case.

1.2. Economic example: insurance

Example 1 *(Insurance).* Consider a consumer with wealth W facing a potential loss L. From a data set of similar cases, it is supposed to be known that a loss occurs in n_L cases, no loss in n_W cases, and that for n_U cases there is no information recorded. From this data base, the insurer and the potential insuree can derive the mass distribution $m(\{W-L\}) = \frac{n_L}{n_L+n_W+n_U}$, $m(\{W\}) = \frac{n_W}{n_L+n_W+n_U}$ and $m(\{W-L, W\}) = \frac{n_U}{n_L+n_W+n_U}$. Denote by $\pi = \frac{n_L}{n_L+n_W}$ the loss proportion and by $\gamma = \frac{n_L+n_W}{n_L+n_W+n_U}$ the proportion of recorded outcomes, which may be interpreted as a degree of confidence in the frequency distribution $(\pi,\ 1-\pi)$. Rewriting the mass distribution in terms of the parameters π and γ, one obtains $m(\{W-L\}) = \gamma\pi$, $m(\{W\}) = \gamma(1-\pi)$, $m(\{W-L, W\}) = 1-\gamma$. Evaluating mass distributions over outcomes by the representation based upon PIR in Equation (4), the value of the initial allocation without insurance is

$$V(m) = \gamma\pi u(W-L) + \gamma(1-\pi)u(W) + (1-\gamma)\phi^{-1}\left(\tfrac{1}{2}[\phi(u(W-L)) + \phi(u(W))]\right).$$

For the case with insurance at a premium $Q = qL$, one has $u(W-qL)$ with certainty. If $V(m) < u(W-qL)$ holds, the consumer will buy full insurance against the loss L. For a risk-averse consumer with a concave von Neumann-Morgenstern utility index u who is also ambiguity averse, i.e., with a concave function ϕ, one has $\phi^{-1}\left(\frac{1}{2}[\phi(u(W-L)) + \phi(u(W))]\right) < \frac{1}{2}[u(W-L) + u(W)] < u\left(W - \frac{1}{2}L\right)$ and $\pi u(W-L) + (1-\pi)u(W) < u(W-\pi L)$. Hence, $V(m) < \gamma u(W-\pi L) + (1-\gamma)u\left(W-\frac{1}{2}L\right) < u\left(W - \left(\gamma\pi + (1-\gamma)\frac{1}{2}\right)L\right)$. If $q \leq \gamma\pi + (1-\gamma)\frac{1}{2}$ holds, then

$$u(W-qL) \geq u(W-(\gamma\pi + (1-\gamma)\tfrac{1}{2})L) > V(m),$$

---

[11] More recently, Peter and Toquebeuf (2020) discussed the implications of this approach for several non-expected utility representations widely used in economic applications.

[12] Nehring (2006) provides a lucid discussion of this issue.

[13] Machina and Siniscalchi (2014, Section 13.6, pp. 777-782) provide an excellent survey on this controversy. Baillon et al. (2018) suggest an empirical test for measuring distinct degrees of ambiguity and ambiguity attitude.

which implies that the consumer will buy full insurance. For $\gamma < 1$, following the principle of insufficient reason the decision maker will have to weigh the frequency information of π against the equal probability in those cases where no recorded information is available. Notice that for $\pi > \frac{1}{2}$, the decision maker may be not willing to buy full insurance at the fair premium $q = \pi$. On the other hand, for $\pi < \frac{1}{2}$, full insurance is bought even at an unfair premium $\pi < q \leq \gamma\pi + (1-\gamma)\frac{1}{2}$. This reflects the influence of ambiguity and ambiguity attitude.

2. Partial information, mass distributions, and the PIR

In this section, we propose a formal framework for studying partial information over states in a Savage framework.

2.1. The formal framework: states, acts and mass distributions

Consider a finite state space S and an outcome space X. As in Savage (1954), *actions* (or *acts*) are mappings $f : S \to X$ associating states with outcomes. Partial information about the probabilities of the states will be modeled by a *mass distribution* m, a probability distribution over the finite set of events $\mathcal{E} = 2^S$ in the state space S, that is $m(E) \geq 0$ for all $E \in \mathcal{E}$ and $\sum_{E \in \mathcal{E}} m(E) = 1$. We will denote by $\mathcal{F}$ the set of all acts and by M_S the set of all mass distributions on the state space S.

Special cases of a mass distribution m are a *probability distribution* if $m(E) = 0$ holds for all events E with $|E| > 1$ and an *elementary mass distribution* if $m(E) = 1$ for some $E \in \mathcal{E}$ (and, hence by implication), $m(F) = 0$ for all $F \neq E$.

Remark 2. A *mass distribution* m on a set of events $\mathcal{E}$ is the Moebius transform of a *belief function*, or *totally monotone capacity*, μ^m defined by $\mu^m(E) = \sum_{F \subseteq E} m(F)$ for all $E \in \mathcal{E}$. A capacity μ is a *belief function* if and only if its Moebius transform, i.e., the vector of coefficients $m_\mu(E)$, $E \in \mathcal{E}$, is non-negative and sums to 1, $\sum_{E \in \mathcal{E}} m_\mu(E) = 1$. Belief functions are convex capacities. Note that a mass distribution m is a set function but not a capacity, since it is not monotone and not normalized. It is well-known, e.g., Gilboa and Schmeidler (1994, Theorem 4.3), that the Choquet integral of an act $f \in \mathcal{F}$ with respect to a belief function μ^m can be expressed in terms of its mass distribution as $\int f d\mu^m = \sum\limits_{E \in \mathcal{E} \setminus \varnothing} m(E)\big[\min_{s \in E} f(s)\big]$. Since a belief function μ^m is a convex capacity, the set of probability distributions that are eventwise dominated by it, the core of μ^m, is not empty. Grabisch (2016, Chapter 7, pp. 377-437) provides a comprehensive treatment of mass distributions, in particular, results regarding belief functions, their properties, and the relationship with other set functions.

In the Savage framework, the outcome of an action $f : S \to X$ will depend on the state $s \in S$ that will be realized. Therefore, uncertainty over outcomes arises from uncertainty about states. Relevant information, therefore, concerns the probability of states. We will model partial information about states by a mass distribution m on S.

A tuple (f, m) of an act f and a mass distribution m induces a mass distribution $m * f$ on the set of outcomes X. For any set of outcomes $A \subseteq X$, the mass $m * f$ allocated to the event A by the act f is

Table 1
Ellsberg's three-color urn.

2^S	$\varnothing$	$\{R\}$	$\{B\}$	$\{Y\}$	$\{R, B\}$	$\{R, Y\}$	$\{B, Y\}$	$\{R, B, Y\}$
$m(E)$	0	$\frac{1}{3}$	0	0	0	0	$\frac{2}{3}$	0
$\mu^m(E)$	0	$\frac{1}{3}$	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	1

mass distribution on states

	30	60	
	R	B	Y
a	\$100	0	0
b	0	\$100	0
a'	\$100	0	\$100
b'	0	\$100	\$100

actions

2^X	{0}	{100}	{0, 100}
$m * a$	$\frac{2}{3}$	$\frac{1}{3}$	0
$m * b$	$\frac{1}{3}$	0	$\frac{2}{3}$
$m * a'$	0	$\frac{1}{3}$	$\frac{2}{3}$
$m * b'$	$\frac{1}{3}$	$\frac{2}{3}$	0

mass distribution on outcomes

$$m * f\,(A) = \sum_{E \subseteq S : f(E) = A} m(E). \tag{5}$$

One checks easily that $m * f$ is a mass distribution on X. For the special case of a probability distribution, this yields the probability of the event $A \subseteq X$. The belief function over outcomes in X corresponding to mass distribution $m * f$ is given by $\mu^{m*f}(A) = \mu^m\left(f^{-1}(A)\right)$.

Example 3 *(Ellsberg 1961, three-color urn).* Consider the three-color version of the Ellsberg paradox where the decision maker can bet on the color of balls drawn from an opaque urn containing 30 red balls and 60 balls that are either black or yellow in unknown proportion. In this case, the state space is $S = \{R, B, Y\}$ and the outcomes $x \in X$ are monetary payments. There is no precise information regarding the probabilities of the colors. The information about the proportion of red and yellow or blue balls, respectively, suggests a probability of $\frac{1}{3}$ for the event $\{R\}$ that a red ball is drawn from the urn and a probability of $\frac{2}{3}$ for the event $\{B, Y\}$. This partial information can be represented by a mass distribution m. The top panel of Table 1 records the mass distribution m corresponding to the partial information about the urn in its first row and the associated belief function μ^m. Notice that the value of the mass distribution for events with unknown probabilities has been set to 0.

In Ellsberg's experiment, decision makers had to choose first on whether to *bet on red*, action a, or *to bet on black*, action b, and, secondly, between *betting on red or yellow*, action a', and *betting on black or yellow*, action b'. Possible acts and corresponding prizes are given in the left panel of Table 1.

There are two possible outcomes, $X = \{0, 100\}$, and, omitting the empty set, three possible events: outcome 100 occurs $\{100\}$, outcome zero occurs $\{0\}$, and either 100 or zero occurs $\{0, 100\}$. For bet a, the decision maker has complete information about her chances to win \$100. This follows from the fact that the proportion of red balls in the urn is known. Hence, for betting on red, one can assign a mass value of one third to the event that an outcome of 100 occurs, $m * a\,(\{100\}) = 1/3$, and a value of two thirds to the event that the outcome is zero, $m * a\,(\{0\}) = 2/3$. In this case, there is no ambiguity about the event that either 0 or 100 may obtain, $m * a\,(\{0, 100\}) = 0$. For bet b, one knows that there is a chance of one third of receiving nothing (when the ball drawn is red), hence $m * b\,(\{0\}) = 1/3$, and of two thirds of getting either 0 or 100, $m * b\,(\{0, 100\}) = 2/3$. This bet illustrates a case of partial information: the probability of two thirds for the event of obtaining 0 or \$100 cannot be subdivided between 0 and \$100, because the proportions of black and yellow balls are unknown. The corresponding belief function is not additive and reflects the ambiguity due to the lack of information about the proportions of black and yellow balls. Applying the same reasoning to the bets a' and b' yields the mass distributions over outcomes in the right panel of Table 1.[14]

In summary, for any $(f, m) \in \mathcal{F} \times M_S$ where $f : S \to X$ is a Savage act and m is a mass distribution on 2^S, Equation (5) yields a mass distribution $m * f$ on the set of outcomes. Let M_X denote the set of mass distributions on X, then $m * f$ defines a mapping from $\mathcal{F} \times M_S$ to the set of mass distribution on outcomes M_X, $m * f : \mathcal{F} \times M_S \to M_X$. The mass distribution on outcomes $m * f$ combines the information from the mass distribution on states m with the information from the act $f \in \mathcal{F}$. We assume that a decision maker evaluates combinations (f, m) in the same way as the induced mass distributions $m * f$.[15] Hence, preferences of a decision maker over acts, given the partial information of a mass distribution over states, will be represented by a preference order $\succcurlyeq$ over mass distributions on outcomes in M_X or the corresponding belief functions.[16]

2.2. *Partial information and mass distributions*

Faced with uncertainty about the outcome of an action, decision makers base their choices on beliefs about the likelihood of the outcomes. Traditionally, such beliefs have been represented by probability distributions. von Neumann and Morgenstern (1944) were the first to assume a preference order over outcome lotteries and to derive an expected-utility representation from axioms on these preferences. In contrast, Savage (1954) took state-contingent outcomes as primitive objects of choice and deduced from preferences over these state-contingent outcomes both a subjective probability distribution over outcomes and an expected-utility representation. The former approach assumes that the likelihood of outcomes is completely specified by objective probabilities (lotteries), while the latter assumes complete ignorance about the likelihood of events and views probabilities as purely subjective. We will argue in this section that real decision situations are typically characterized by more or less information about the likelihood of events, that is by *partial information* about the actual probability distribution.

In his *Mathematical Theory of Evidence*, Shafer (1976) suggests *belief functions* (or *totally monotone capacities*) as a concept capable of capturing partial information, evidence as he calls it. Partial information that is available for events, i.e., probabilities known for events, can be directly attributed to these events as in Example 3, while a probability of zero is assigned to events for which no information is available. The Moebius transform converts the mass distribution into a belief function with a well-defined Choquet integral. The special case of perfect information (von Neumann and Morgenstern, 1944) corresponds to the limiting case of probability distributions. The case of *no information* (Savage, 1954) or *complete ignorance*, on the other hand, arises if the mass distribution puts all weight on the set of all possible outcomes.

---

[14] Applying Equation (5), one obtains the same result.

[15] Notice the analogy to the approach of Gajdos et al. (2008a) who assume that a decision maker orders pairs of acts and sets of priors (f, P).

[16] This is also the approach chosen in the seminal paper by Jaffray (1989).

We find several approaches in the literature for how information determines a belief function or mass distribution. While Dempster (1967) views belief functions as a formal concept for capturing incomplete statistical observations, Shafer (1976) regards them as a systematic method for aggregating information from sources differing in reliability.[17] Other interpretations have been suggested in the literature.[18] A comprehensive survey of these interpretations is beyond the scope of this paper. We will, however, provide two examples in order to illustrate the two most prominent approaches advanced by Dempster (1967) and Shafer (1976).

2.2.1. *Shafer‘s (1976) theory of evidence*

As the title of his book suggests Shafer (1976) sees his approach as a theory of evidence where statements about the likelihood of events have to be judged by their reliability. The following example from Shafer (1990, Chapter 7, p. 2) illustrates this approach.

Example 4. In 1990, Shafer (1990, Chapter 7, p. 2) writes:

> *“We can derive degrees of belief for statements made by witnesses from subjective probabilities for the reliability of these witnesses.*
>
> *Degrees of belief obtained in this way differ from probabilities in that they may fail to add to 100%. Suppose, for example, that Betty tells me a tree limb fell on my car. My subjective probability that Betty is reliable is 90%; my subjective probability that she is unreliable is 10%. Since they are probabilities, these numbers add to 100%. But Betty’s statement, which must be true if she is reliable, is not necessarily false if she is unreliable. From her testimony alone, I can justify a 90% degree of belief that a limb fell on my car, but only a 0% (not 10%) degree of belief that no limb fell on my car. (This 0% does not mean that I am sure that no limb fell on my car, as a 0% probability would; it merely means that Betty’s testimony gives me no reason to believe that no limb fell on my car.) The 90% and the 0%, which do not add to 100%, together constitute a belief function.”*

Denote by l the state that “a tree limb fell on the car” and by nl the state that “no tree limb fell on the car”. Moreover, let π be the likelihood of l stated by the witness “Betty” and by δ the degree of confidence in her statement, then the assessed likelihood of state l is $m(\{l\}) = \delta\pi(l)$, $m(\{nl\}) = \delta\pi(nl)$, and $m(\{l, nl\}) = 1 - \delta$. The following table summarizes the mass distribution m and its associated belief function μ^m.[19]

2^S	$\varnothing$	$\{l\}$	$\{nl\}$	$\{l, nl\}$
$m(E)$	0	$\frac{9}{10}$	0	$\frac{1}{10}$
$\mu^m(E)$	0	$\frac{9}{10}$	0	1

---

[17] Shafer (1976) typically illustrates his theory with examples where sources of information are statements of witnesses or experts of differing reliability.

[18] Grabisch (2016, Chapter 7, 377-337) provides a survey and comparison of the approaches of *“Upper and Lower Probabilities”* by Dempster (1967), of the *“Evidence Theory”* by Shafer (1976) and of the *“Random Sets”* approach by Matheron (1975) and Kendall (1974). See also Pearl (1988, Chapter 9) for further references.

[19] The “Three Prisoners Puzzle”, discussed in Pearl (1988, 417-421), provides a similar example.

2.2.2. *Dempster's 1967 theory of lower probabilities*

In a seminal paper, Dempster (1967) suggests the following model for aggregating partial information. He considers a pair of spaces Ω and S and a multi-valued mapping Γ assigning a subset $\Gamma(\omega) \subset S$ to every $\omega \in \Omega$.[20] The probability distribution μ over Ω is supposed to be known, but for S no probability information is available.

> *Suppose, however, that it is known that an element* $\omega \in \Omega$ *"corresponds to an uncertain outcome* $s \in \Gamma(\omega)$*, what probability judgments may be made about an uncertain outcome* $s \in S$*?" (Dempster, 1967, p. 325)*

For finite sets Ω and S, $\{\omega \in \Omega \mid \emptyset \neq \Gamma(\omega) \subset S\}$ is the domain of Γ. Assuming $\mu(\{\omega \in \Omega \mid \emptyset \neq \Gamma(\omega) \subset S\}) = 1$, one can assign a *lower probability* $P_*(T) = \mu\left(\{\omega \in \Omega \mid \emptyset \neq \Gamma(\omega) \subset T\}\right)$ to every event $T \subseteq S$. The lower probability distribution is a *belief function* and its dual $P^*(T) = 1 - P_*(T^c) = \mu\left(\{\omega \in \Omega \mid \Gamma(\omega) \cap T \neq \emptyset\}\right)$ is the upper probability called *possibility function*.

Γ is a correspondence from states in Ω for which probabilities are known to subsets of states in S. The tuple (μ, Γ) of a probability distribution over states in Ω and the correspondence Γ capturing the information about the unknown states in S defines a belief function.

The following example from Mukerji (1997) illustrates how partial information from a correspondence, mapping from a space with known probabilities into a space with unknown probabilities, induces a lower probability (belief function) on the space with unknown probabilities.

Example 5. Mukerji (1997, pp. 27-29) considers futures trades in two types of oranges, t_1 and t_2. The payoff of the futures will depend on the spot market conditions in the future. Assuming that there are four payoff-relevant states of the world $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ corresponding to "a glut for both types of oranges" θ_1, "a glut of t_1-oranges together with a scarcity for t_2-oranges" θ_2, "a glut of t_2-oranges together with a scarcity of t_1-oranges" θ_3, and "scarcity of both types of oranges" θ_4. The probability of these payoff-relevant states is assumed to be unknown.

The trader consults with an expert who provides the trader with a probability distribution over five weather states $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$ and a description of their impact on the payoff-relevant states described by the implication mapping $\Gamma : \Omega \to 2^{\Theta}$ given in the following table.

Ω	ω_1	ω_2	ω_3	ω_4	ω_5
$\Gamma(\omega)$	$\{\theta_1\}$	$\{\theta_2, \theta_3\}$	$\{\theta_4\}$	$\{\theta_3\}$	$\{\theta_1, \theta_2, \theta_3, \theta_4\}$

The correspondence Γ induces a mass distribution on the space of payoff-relevant states: for all $E \subseteq \Theta$, $m(E) = \mu(\Gamma^{-1}(E))$ for $E \in \Gamma(\Omega)$ and 0 otherwise.

Example 5 illustrates the idea of Dempster's approach to deduce probability ranges, that is upper and lower probabilities, for events in a space with unknown probabilities from knowledge about the relationship of these events with states for which probabilities are known.

---

[20] Dempster (1967) uses the notation X instead of Ω.

The approach of Dempster (1967) can be easily applied to statistical models with a joint probability distribution over states and signals where the correspondence Γ maps signals into sets of states. Indeed, Wakker (2000) views Dempster (1967)'s notion of a belief function as resulting from a two-stage process where

> "*in a first stage a random message is received, designating a subset of the state space that will contain the true state of nature*" (the correspondence Γ) and where "*uncertainty of the first stage, regarding the random message to be received,* [...]*can be expressed in terms of probabilities*" (the probability distribution on Ω), while "*in the second, final stage, one element of a set of states of nature will obtain, the true state of nature.*" (Wakker, 2000, p. 271)

2.3. *Ambiguity and the* Principle of Insufficient Reason *(PIR)*

In Section 2.1, we showed that a mass distribution m on a state space representing a decision maker's information together with actions f mapping from states in S into outcomes in X induce a mass distribution $m * f$ on X. Following Jaffray (1989), we assume that preferences of the decision maker order the set M_X of all mass distributions on the outcome set X (equivalently, the corresponding set of belief functions). In the following Section 3, we will provide a set of axioms characterizing the representation in Equation (4). In our axiomatisation, we depart from Jaffray (1989) in the treatment of events of complete uncertainty. Recall that an elementary mass distribution e_A represents an information situation of complete ignorance where it is known that the event A has occured, but where is complete ambiguity about the likelihood $m(B) = 0$ of all subevents $B \subset A$. In case of unknown probabilities, Luce and Raiffa (1957, Section 13.2, 278-286) discuss four decision criteria, among them (i) Hurwicz's "*optimism-pessimism index criterion*", and (ii) Laplace's "*principle of insufficient reason*". Adding a monotonicity axiom Jaffray (1989) derives a value for elementary mass distributions depending only on the maximum and the minimum outcome of an ambiguous event. Special cases of this representation are MEU and α-MEU.

In contrast, in this paper, we will axiomatize an evaluation of elementary mass distributions $V(e_A)$ by a *generalized arithmetic mean* as in Equation (3), that we repeat here for convenience, $V(e_A) = \phi^{-1}\left(\frac{1}{|A|}\sum_{x \in A} \phi(u(x))\right)$. Inspired by the *Principle of Insufficient Reason (PIR)*, this *generalized arithmetic mean* evaluates ambiguous events e_A by a transformed arithmetic average of outcomes in A.

The arguments for and against the Principle of Insufficient Reason (PIR) have been discussed extensively in the literature since it has been put forward by James Bernoulli.[21] The main argument in favor of PIR appeals to the balanced weighting of the uncertain outcomes: if there is complete uncertainty regarding the outcomes in a set then there is no reason to give special weight to a particular outcome. Considering only the best and the worst outcomes as in the Hurwicz criterion may distort the value of uncertain outcomes unreasonably according to a subjective optimistic or pessimistic perspective. In particular, a possibly large subset of outcomes in the uncertain event may be disregarded completely. As Keynes wrote in 1921,

---

[21] Keynes (1921, p. 41) attributes the principle to James Bernoulli and provides an extensive discussion of this rule that he calls "Principle of Indifference".

> *"equal probabilities must be assigned to each of several arguments, if there is an absence of positive ground for assigning unequal ones. [...] however, the plausibility of this principle will be most easily shaken by an exhibition of the contradictions which it involves."* (Keynes, 1921, p. 42)

The main argument against PIR is the dependence of the uniform probability on the precise specification of the state space. Luce and Raiffa (1957, p. 284) write:

> *"Suppose we are confronted with a real problem in decision making under uncertainty, then our first task is to give a mutually exclusive and exhaustive listing of the possible states of nature. The rub is that many such listings are possible, and in general these different abstractions of the same problem will, when resolved by the principle of insufficient reason, yield different real solutions."*

Refining the state space implies a larger number of states and, therefore, according to the PIR a reduced likelihood of each state. The following Example 6 from Savage (1954, p. 65) illustrates this fact.[22]

Example 6 *(Savage 1954, p. 65).*

> *"Suppose [...] it is known of an urn only that it contains either two white balls, two black balls, or a white ball and a black ball. The principle of insufficient reason has been invoked to conclude that the three possibilities are equally probable, so that in particular the probability of one white and one black ball is concluded to be* $1/3$*. But the principle has also been applied to conclude that there are four equally probable possibilities, namely, that the first ball is white and the second black, etc. On this basis, the probability of one white and one black ball is, of course,* $1/2$*. Personally, I do not try to arbitrate between the two conclusions but consider that the existence of the pair of them reflects doubt on the notion that a person's knowledge relevant to any matter admits any full and precise description in terms of propositions he knows to be true and others about which he knows nothing." (Savage, 1954, p. 65)*

Whether one assesses the likelihood of a black and a white ball being drawn as $\frac{1}{3}$ or $\frac{1}{2}$ depends on how one defines the states of the world in this example. The information about the states that have to be distinguished does not suffice for defining a unique state space. Savage considers two possible descriptions of this situation: case (i): $S_1 = \{(b,b),(w,w),(b,w)\}$ and case (ii): $S_2 = \{b,w\} \times \{b,w\}$. Since there are three states in case (i) and four in case (ii) the PIR would assign a probability of $\frac{1}{3}$ to each state in case (i) and of $\frac{1}{4}$ to each state in case (ii). Note, however, that in both cases it is assumed that there is complete uncertainty about the probability of events other than the state space itself.

If the decision maker faces a bet (b,w) on a white and a black ball being drawn from the urn, however, the PIR will evaluate the probability distribution over outcomes arising in both cases in the same way. In case (i) the elementary mass distribution is e_{S_1} and the bet

---

[22] See also Sinn (1980, p. 505).

$f(s) = 1$ for $s = (b, w)$ and 0 otherwise yields the mass distribution $e_{S_1} * f = e_{\{0,1\}}$ assigning weight 1 to the event $\{0, 1\}$ and weight 0 to the singleton events $\{0\}$ and $\{1\}$. Hence, according to Equation (3) $V(e_{\{0,1\}}) = \phi^{-1}\left(\frac{1}{2}[\phi(u(0)) + \phi(u(1))]\right)$. Similarly, in case (ii), the elementary mass distribution e_{S_2} and the bet f induce the mass distribution $e_{S_2} * f = e_{\{0,1\}}$ on the set of outcomes $\{0, 1\}$ yielding the same value $V(e_{\{0,1\}}) = \phi^{-1}\left(\frac{1}{2}[\phi(u(0)) + \phi(u(1))]\right)$.

The following Lemma shows that the observation in Example 6 is in fact quite general. If there is complete ignorance about the states in a set $A \subseteq S$ then, for each action $f \in \mathcal{F}$, there will be complete ignorance about the outcomes in its image $f(A)$.

Proposition 7. *Given an event $A \subseteq S$ and an action $f \in \mathcal{F}$, then $e_A * f = e_{f(A)}$.*

Proof. The result follows immediately from Equation (5). □

Given that actions, e.g., bets, map often on the same set of outcomes, Lemma 7 implies that in case of complete uncertainty the number of states in the set of states with unknown probabilities does not matter. A careful specification of the actions $f \in \mathcal{F}$ will identify the set of outcomes for which there is complete ambiguity.[23]

Representing information over the probabilities of a state space by a mass distribution restricts the PIR to elementary mass distributions. Hence, specifying partial information by a mass distribution shows that the exact specification of the unknown state space is less important for the correct *uniform distribution over outcomes*. We take this observation as a supporting argument for considering mass distributions over outcomes in X as the relevant domain for the preferences of a decision maker.

3. Axioms and representation

In this section, we assume an infinite set of consequences X together with an algebra $\mathcal{X}$ of subsets of X containing all finite subsets. Denote $\bar{\mathcal{X}}$ the set of finite subsets of X.

Let $\mathcal{M}$ be the set of belief functions on $\mathcal{X}$ that are concentrated on a finite subset. In other words, for any belief function $\mu^m \in \mathcal{M}$ there exists a finite number of sets $A_1, \ldots, A_n \in \bar{\mathcal{X}}$ such that $\sum_{i=1}^n m(A_i) = 1$. Hence, $\mu^m(D) = 1$ for $D = \cup_{i=1}^n A_i$.

A special case of a belief function in $\mathcal{M}$ is a finitely supported probability distribution or *lottery* l on $\mathcal{X}$. By assumption, the set of consequences X is sufficiently rich to allow for a certainty equivalent in X for any lottery in $\mathcal{M}$. Abusing notation, we make no distinction between consequences in X and the degenerate lotteries in $\mathcal{M}$ concentrated on a singleton subset.

We assume that the decision maker's preferences are given by a binary relation $\succcurlyeq$ on $\mathcal{M}$, that is decision makers can order the set of belief functions $\mathcal{M}$. Preferences over the set of belief functions compare both the values of outcomes and the available information about their likelihood. At one extreme, if a belief function gives only positive weights to singleton events,

---

[23] In this paper, we consider finite sets of states. For infinite state and outcome spaces, Grant et al. (2021, Example 1, p. 10) point out a discontinuity that can arise in context of evaluating acts by PIR. They provide an example with an infinite dimensional outcome space X and consider a sequence of outcome sets $\left\{0, \frac{1}{n}, 1\right\}$ that shrinks in the limit to the set $\{0, 1\}$ as $n \to 0$. In this context acts may no longer be "uniformly continuous" if the probability of the states depends also on n.

one has full information, that is belief functions are probability distributions over outcomes, and at the other extreme, one may know only the set of all possible outcomes if the belief function gives full weight of 1 to a finite set of outcomes. A decision maker is supposed to be able to compare such outcome-information scenarios as, for example, in the Ellsberg two-urn case.

A convex combination of any two belief functions is a belief function again. We interpret a mixture of the two belief functions $\mu, \nu \in \mathcal{M}$ with weight $\lambda \in [0, 1]$, $\lambda\mu + (1-\lambda)\nu \in \mathcal{M}$, as a two-stage lottery over two belief functions. Belief functions are convex capacities and, hence, have a non-empty core, i.e., a set of probability distributions consistent with the information given by the mass distribution. Convex combinations of belief functions represent the sets of probability distributions consistent with the information contained in the convex combination of the mass distributions.

Since convex combinations of belief functions are again belief functions, $\mathcal{M}$ is a mixture space. Hence, one can apply the von Neumann-Morgenstern axiom system[24] to the preference relation $\succcurlyeq$ in order to deduce an "expected utility" representation for a belief function. In this context, Axioms 1 to 3 have the same interpretation as in the von Neumann-Morgenstern setup. In particular, the independence axiom (Axiom 3) allows one to split off common aspects when comparing outcome-information scenarios represented by belief functions.

Axiom 1 *(Weak Order).* $\succcurlyeq$ is a transitive and complete relation on $\mathcal{M}$.

Axiom 2 *(Continuity).* For any $\mu, \nu, \xi \in \mathcal{M}$ such that $\mu \succ \nu \succ \xi$, there exist $0 < \lambda_1, \lambda_2 < 1$ such that $\lambda_1\mu + (1-\lambda_1)\xi \succ \nu \succ \lambda_2\mu + (1-\lambda_2)\xi$.

Axiom 3 *(Independence).* For any $\mu, \nu, \xi \in \mathcal{M}$ and $0 < \lambda < 1$, if $\mu \succ \nu$, then $\lambda\mu + (1-\lambda)\xi \succ \lambda\nu + (1-\lambda)\xi$.

Any belief function in $\mathcal{M}$ can be represented as a convex combination of elementary belief functions.

Definition 8. We call $e_A \in \mathcal{M}$ an *elementary belief function*, if $e_A(B) = 1$ for any $B \supseteq A$ and $e_A(B) = 0$ otherwise.

The mass distribution of an elementary belief function e_A assigns weight 1 to A and 0 to all other events. Any belief function $\mu^m \in \mathcal{M}$ can be written as

$$\mu^m(B) = \sum_{A \in \mathcal{X}} m(A) e_A(B),$$

or simply as $\mu^m = \sum_{A \in \mathcal{X}} m(A) e_A$.[25]

As we will show in the proof of Theorem 1, axioms 1-3 guarantee (see Jaffray, 1989) that there is a linear utility function on $\mathcal{M}$:

$$V(\mu^m) = \sum_{A \in \mathcal{X}} m(A) V(e_A), \tag{6}$$

---

[24] Jaffray (1989) was the first to make this point.

[25] The sum is taken over $A \in \mathcal{X}$ such that $m(A) > 0$. By the definition of $\mu^m \in \mathcal{M}$, there is only a finite number of such sets.

that is an expected value of the elementary belief functions e_A with respect to the mass distribution m of the belief function μ^m. An elementary belief function e_A represents a situation of complete ignorance with respect to the set of outcomes in A. In other words, the decision maker is certain that the true outcome belongs to A, but nothing more. In order to specify $V(e_A)$ in Equation (6), additional assumptions need to be made about decision maker's evaluation of such situations.[26]

In this paper, we would like to derive a representation in which also non-extreme outcomes in A influence the evaluation $V(e_A)$. Moreover, outcomes lacking information about their likelihood should be equally weighted because of their informational symmetry, *the principle of insufficient reason*. Maintaining axioms 1–3, we will propose four additional axioms which will characterize the evaluation of elementary belief functions $V(e_A)$ by a function ϕ and a uniform distribution as in Equation (3).[27]

For notational simplicity, we write $A \succcurlyeq B$ instead of $e_A \succcurlyeq e_B$ for $A, B \in \bar{\mathcal{X}}$, and $A \succcurlyeq x$ instead of $A \succcurlyeq \{x\}$ for $x \in X$.

Axiom 4 *(Monotonicity).* For any $x \in X \setminus A$ and $y \in A$, $x \succcurlyeq y$ if and only if $(A \setminus \{y\}) \cup \{x\} \succcurlyeq A$.

According to Axiom 4, replacing one of the possible outcomes in an event A by a weakly preferred one provides valuable information about the composition of the set A and cannot make the situation of complete ignorance about the outcomes in A worse.

The value of a set of outcomes depends both on the unknown likelihood of the outcomes in the set and on the composition of outcomes in the set. If ignorance about the outcomes in a set B is less important than ignorance over the outcomes in another set A then the ignorance about the combined set $A \cup B$ should not matter more than the ignorance about A nor less than the ignorance about B.

Axiom 5 *(Set Betweenness).* If $A \succcurlyeq B$ and $A \cap B = \varnothing$, then $A \succcurlyeq A \cup B \succcurlyeq B$.[28]

A failure of Set Betweenness, say $A \cup B \prec B$, would imply that adding better outcomes to B makes it less attractive.

Axiom 6 *(Set Continuity).* For any $x_0, y, z \in X$

(a) if $x_0 \succ y$ and $\{x_0, y\} \succ z$, then $\{x_1, y\} \succ z$ for some $x_1 \in X$ such that $x_0 \succ x_1 \succ y$;

(b) if $x_0 \prec y$ and $\{x_0, y\} \prec z$, then $\{x_1, y\} \prec z$ for some $x_1 \in X$ such that $x_0 \prec x_1 \prec y$.

Set Continuity means that strict preference between a situation of complete ignorance and a certain alternative is robust with respect to a minor change in one of the possible outcomes.

---

[26] For example, Jaffray (1989) provides an additional axiom such that $V(e_A)$ depends only on the worst and the best outcomes in A.

[27] The axiomatization of the principle of insufficient reason in this paper was inspired by the axiomatization of the quasi-arithmetic mean as an aggregator of continuation values in Ke (2019). Gravel et al. (2012) use axioms similar to our Axioms 5 and 7 in their expected utility representation with a uniform distribution over a set of outcomes. However, they consider only situations of complete ignorance when no partial information is available.

[28] A weaker version of this axiom is sufficient to derive the representation. Namely, if $B \sim x$ for $B \in \mathcal{X}$ and $x \in X$, then $B \sim B \cup \{x\}$. Nevertheless, we stick to the stronger version because of its intuitive appeal.

Axiom 7 *(Certainty Equivalence Consistency).* For any $x, y \in X$, if $A \cap B = \varnothing$ and $|A| = |B|$, then $A \sim x$ and $B \sim y$ imply $A \cup B \sim \{x, y\}$.

According to Axiom 7, two disjoint events of the same cardinality should have a union which is equivalent to the union of their certainty equivalents. That is, combining two situations of complete ignorance is equivalent to combining their certainty equivalents, provided that the two situations are mutually exclusive and the number of possible outcomes in both cases is the same.

A necessary feature of the principle of insufficient reason is the fact that it does distinguish events with the same number of elements only by the outcomes involved, i.e., for $x, y \notin A$, the set of outcomes $A \cup \{x\}$ must be indifferent to the set of outcomes $A \cup \{y\}$ if $y \sim x$. In fact, it is easy to see that Axioms 4 implies this principle of "equal weights to equivalent outcomes", i.e., if $y \sim x$, then $A \cup \{x\} \sim A \cup \{y\}$.

Denote by U the set of values of u, i.e. $U = u(X)$. In the appendix, we prove the following theorem.

Theorem 9. *Axioms 1-7 hold if and only if there exists a representation V of preferences $\succcurlyeq$ on $\mathcal{M}$ such that*

$$V(\mu^m) = \sum_{A \in \mathcal{X}} m(A)\, \phi^{-1}\left(\frac{1}{|A|} \sum_{x \in A} \phi(u(x))\right), \tag{7}$$

where u is a von Neumann-Morgenstern utility function on X and ϕ is a continuous strictly increasing function on U. Such V is unique up to a positive linear transformation. Given that V is fixed, u is unique and ϕ is unique up to a positive linear transformation.

It is straightforward to check that the function $V(e_A) = \phi^{-1}\left(\frac{1}{|A|} \sum_{x \in A} \phi(u(x))\right)$ satisfies the Axioms A4, A5, A6, and A7. It is not trivial, however, to show that these axioms are also sufficient for a representation by a quasi-arithmetic mean. The proof of this theorem uses a little known theorem characterizing a function as a quasi-arithmetic mean (Matkowski and Páles, 2015, Theorem C). Axioms A4 to A7 imply the important bi-symmetry property of this function.

4. General properties of the representation

In this section, we will discuss some properties of the representation (7). Our focus will be on how ambiguity and ambiguity attitudes are captured by this representation. Since lotteries are a special kind of belief functions, risk attitudes are defined by the restriction of our representation to these lotteries. In contrast, ambiguity attitudes can be defined by comparing belief functions that are lotteries to more general belief functions.[29]

---

[29] Klibanoff et al. (2005), Neilson (2010), and Izhakian (2017) consider two layers of uncertainty in order to obtain distinct measures for risk attitudes and ambiguity attitudes. From an axiomatic perspective, all three are very different to the approach suggested in this paper.

4.1. Ambiguity and ambiguity attitudes

Most of the literature on ambiguity (e.g., Machina and Siniscalchi, 2014, pp. 730-732) distinguishes situations of risk, where the decision maker knows the probabilities of all outcomes, from situations under ambiguity, where the decision maker knows only the outcomes which may occur but not their probabilities. This distinction can be traced back to (Knight, 1921, pp. 224-225). In models of ambiguity where probabilities are subjective and unknown (Savage, 1954; Anscombe and Aumann, 1963), there is no obvious criterion for classifying a situation as "more or less ambiguous". Hence, for purely subjective beliefs derived from preferences, most attempts to distinguish ambiguity from ambiguity attitude have failed (see Machina and Siniscalchi, 2014, p. 750).

In the approach advanced in this paper, the uncertainty which a decision maker faces is closely related to the information available. Belief functions provide a natural framework for distinguishing ambiguity and ambiguity attitudes: ambiguity is a property of the objective information embodied in belief functions and ambiguity attitude is a property of the subjective preferences over these belief functions.

Assuming that ambiguity is a property of the objective information embodied in a belief function and in the corresponding mass distribution, we formalize increasing ambiguity by analogy with the increasing risk literature originated by Rothschild and Stiglitz (1970). Our Definition 10 will be based on the following notion of centroid.

For a mass distribution m and the corresponding belief function μ^m, define the *centroid* p^* of the core of μ^m as

$$p^*(x) = \sum_{A \ni x} \frac{m(A)}{|A|} \tag{8}$$

for all $x \in X$. It is not difficult to show that p^* is the average of the extreme probabilities in the core of μ^m. Denneberg (2000, p. 64) refers to p^* as the Shapley measure of μ^m, while Gajdos et al. (2008a) call it the Steiner point of the core.

Definition 10. For two mass distributions, we say that m is obtained from n by an *elementary increase in ambiguity* if m and n both share the same centroid and $m(E) > n(E)$ for exactly one set $E \subseteq X$. We say that m is a *spread* of n if it is obtained from n by a sequence of elementary increases in ambiguity.

To understand this definition, let us look at the sets $F_i \subseteq X$ such that $m(F_i) < n(F_i)$. In other words, these are the sets that lose from the elementary increase in ambiguity. Note that $\cup_i F = E$. Indeed, if x is in E but not in $\cup_i F$, the centroid loses a positive mass at x, and if x is in $\cup_i F$ but not in E, the centroid gains a positive mass at x, so the two sets must coincide to preserve the centroid.

A simple example of the elementary increase in ambiguity for $X = \{1, 2, 3\}$ is given by a situation of complete ignorance with $m(123) = 1$ and a uniform lottery $n(1) = n(2) = n(3) = \frac{1}{3}$. Both have $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ as their centroid and $m(E) > n(E)$ for exactly one set $E = 123$. Another example is given by $n'(123) = \frac{1}{3}$, $n'(12) = n'(23) = \frac{2}{9}$, $n'(1) = n'(3) = \frac{1}{9}$, where m is also obtained from n' by an elementary increase in ambiguity. Note how the centroid plays here a role similar to the mean in decision analysis under risk.

The following example provides further illustration.

Example 11 *(This example describes the setup of an experiment run by Kops and Pasichnichenko (2021)).* Consider an urn containing 21 balls which are either green or blue. Subjects were given the information that k balls were green and k balls were blue and the remaining $21 - 2k$ balls were either green or blue in an unknown proportion. Arguably, there is more ambiguity if $2k$ is close to zero than if $2k$ is close to 21. The mass distribution for the urn would assign $m(\{g\}) = m(\{b\}) = \frac{k}{21}$ and $m(\{g, b\}) = \frac{21-2k}{21}$. Let m_1 and m_2 be two mass distributions corresponding to different values of k, i.e., $k_1 < k_2$. Then m_1 represents an elementary increase in ambiguity with respect to m_2.

The mass distribution m reflects the available information of a decision maker and, hence, defines ambiguity objectively. *Ambiguity attitude,* on the other hand, is a property of the subjective preferences of the decision maker. Hence, while ambiguity is embedded in the belief function, the evaluation of the belief function should reflect the (subjective) ambiguity attitude of the decision maker, since it is derived axiomatically from preferences over belief functions.

Since spreads represent objective increases in ambiguity, ambiguity attitudes may be defined as attitudes to spreads.

Definition 12 *(Ambiguity attitudes).* A decision maker is *ambiguity averse (resp. loving, neutral)* if $\mu^m \preccurlyeq \mu^n$ for all mass distributions m and n such that m is a spread of n. (resp. $\mu^m \succcurlyeq \mu^n$, $\mu^m \sim \mu^n$).

On the other hand, following Ellsberg (1961), it is possible to define ambiguity attitude by comparing a situation of complete ignorance, e_A, with the corresponding uniform lottery (without ambiguity), $\bar{\ell}_A$, which gives equal weight to each outcome in A,

$$\bar{\ell}_A(x) = \begin{cases} \frac{1}{|A|} & \text{for } x \in A \\ 0 & \text{otherwise.} \end{cases}$$

The following proposition shows that both definitions of ambiguity attitudes are, in fact, equivalent for representation (7). Moreover, we see that ambiguity attitude is measured by the function $\phi : \mathbb{R} \to \mathbb{R}$.

Proposition 13. *The following statements are equivalent.*

(i) *A decision maker is ambiguity averse (resp. loving, neutral).*
(ii) *A decision maker always prefers the uniform lottery to the situation of complete ignorance with respect to the same set of outcomes (resp. prefers ignorance, indifferent).*
(iii) *ϕ is a concave (resp. convex, linear) function.*

All proofs are given in the appendix.

Before providing some more general results on ambiguity and ambiguity attitude in Subsection 4.2 below, we will try to provide more intuition by studying the case of two outcomes.

Example 14 *(Two-outcome case).* Consider the case of two outcomes, $X = \{x_1, x_2\}$ and a belief function μ^m. In this case, writing $V(x_1, x_2)$ instead of $V(\mu^m)$ one obtains from the representation (7)

$$V(x_1,x_2) = m_1 u(x_1) + m_2 u(x_2) + m_{12}\phi^{-1}\left(\frac{1}{2}\left[\phi(u(x_1)) + \phi(u(x_2))\right]\right)$$

with $m_i = m(\{x_i\}), m_{kl} = m(\{x_k, x_l\})$, etc. denoting the parameters of the mass distribution. Assuming that the functions u and ϕ are differentiable for all x_1 and x_2, the equation $V(x_1, x_2(x_1)) = c$ defines implicitly a function $x_2(x_1)$. By the implicit function theorem, we get the slope of the function $x_2(x_1)$, $s(x_1,x_2)$, as the tangent line to the indifference curve at point (x_1,x_2),

$$s(x_1,x_2) = -\frac{\frac{\partial V(x_1,x_2)}{\partial x_1}}{\frac{\partial V(x_1,x_2)}{\partial x_2}} = -\frac{m_1 + \frac{1}{2}m_{12}\rho(x_1,x_2)\phi'(u(x_1))}{m_2 + \frac{1}{2}m_{12}\rho(x_1,x_2)\phi'(u(x_2))} \cdot \frac{u'(x_1)}{u'(x_2)},$$

where

$$\rho(x_1,x_2) = (\phi^{-1})'\left(\tfrac{1}{2}[\phi(u(x_1)) + \phi(u(x_2))]\right) = \frac{1}{\phi'\left(\phi^{-1}\left(\frac{1}{2}[\phi(u(x_1))+\phi(u(x_2))]\right)\right)}.$$

The partial derivative of the representation V,

$$\frac{\partial V(x_1,x_2)}{\partial x_1} = m_1 u'(x_1) + \tfrac{1}{2}m_{12}\rho(x_1,x_2)\phi'(u(x_1))u'(x_1),$$

has a first term corresponding to the risk part of the representation and a second term corresponding to the ambiguity part. Hence, in general, ambiguity attitudes and risk attitudes jointly determine the evaluation of an outcome. The function $\rho(x_1,x_2)$ measures the ambiguity attitude at the average expected utility with respect to the uniform distribution which is the default distribution in case of ambiguity by the principle of insufficient reason.

As special cases, we obtain:

- *no ambiguity (pure risk):* $m_{12} = 0$,

  $$s(x_1,x_2) = -\tfrac{m_1}{m_2} \cdot \tfrac{u'(x_1)}{u'(x_2)},$$

- *complete ambiguity (complete ignorance):* $m_{12} = 1$ ($\Rightarrow m_1 = m_2 = 0$),

  $$s(x_1,x_2) = -\tfrac{\phi'(u(x_1))}{\phi'(u(x_2))} \cdot \tfrac{u'(x_1)}{u'(x_2)},$$

- *ambiguity neutrality:* ϕ is a linear function,

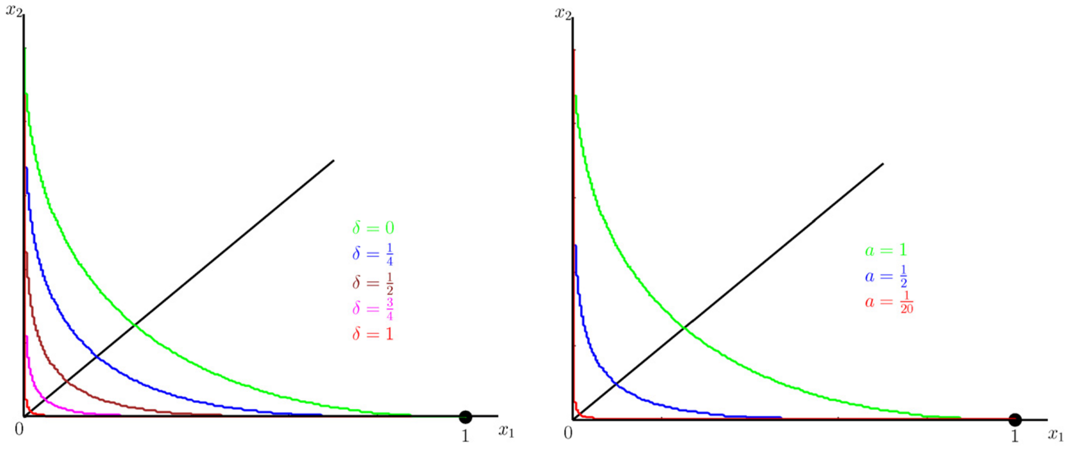
  $$s(x_1,x_2) = -\tfrac{m_1+\frac{1}{2}m_{12}}{m_2+\frac{1}{2}m_{12}},$$

- *certainty:* $x_1 = x_2$,

  $$s(x_1,x_2) = -\tfrac{m_1}{m_2}.$$

Note that

1. certainty implies independence from risk and ambiguity attitudes, hence, the marginal rate of substitution equals the known odds: $-\frac{m_1}{m_2}$,
2. for no ambiguity, $\delta = 0$, only risk attitudes, as measured by the Bernoulli utility function u, matter: $-\frac{m_1}{m_2} \cdot \frac{u'(x_1)}{u'(x_2)}$, and
3. ambiguity neutrality is not equivalent to pure risk, that is $\frac{m_1+\frac{1}{2}m_{12}}{m_2+\frac{1}{2}m_{12}} \neq \frac{m_1}{m_2}$.



$a = \frac{1}{3}$
Figure 1a

$\delta = \frac{3}{4}$
Figure 1b

Fig. 1. Ambiguity δ and Ambiguity attitudes a.

The following diagrams show indifference curves of the preference representation V over outcome combinations (x_1, x_2) for different degrees of ambiguity δ (1a) and an exponential ambiguity attitude function $\phi(x) = x^a$ with different degrees of ambiguity attitude a (1b).

Fig. 1a illustrates the effect of ambiguity on the evaluation of belief functions by an ambiguity averse decision maker. Without ambiguity (green indifference curve), $\delta = 0$, the lottery $\overline{\ell}_X$ yields the highest utility and with complete ambiguity (red indifference curve), $\delta = 1$, utility is at the lowest level. The indifference curves for intermediate degrees of ambiguity ($\delta = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$) lie between these two extremes. Notice that both bets $(x_1, x_2) = (1, 0)$ and $(x_1, x_2) = (0, 1)$ are valued the same for any δ, but are valued increasingly lower as ambiguity increases, $\delta \to 1$.

Fig. 1b shows indifference curves for different degrees of ambiguity aversion. As the degree of ambiguity aversion decreases, $a \to 1$, the utility of the bets $(x_1, x_2) = (1, 0)$ and $(x_1, x_2) = (0, 1)$ increases. For ambiguity neutrality, $a = 1$, $V(\mu^m) = V(\overline{\ell}_X)$ for all degrees of ambiguity δ.

4.2. *Comparison of ambiguity attitudes*

In this section, we formalize the idea that a "more concave" ϕ corresponds to a higher degree of ambiguity aversion or pessimism.

Let $\boldsymbol{i}$ and $\boldsymbol{j}$ be two decision makers with preferences on $\mathcal{M}$ satisfying Axioms 1-7. We assume that both decision makers share the same preferences on X. For the next definition, let $CE(l)$ be a certainty equivalent of a lottery l.

Definition 15. Decision maker $\boldsymbol{i}$ is *weakly more risk averse* than decision maker $\boldsymbol{j}$ if $CE_{\boldsymbol{i}}(l) \preccurlyeq CE_{\boldsymbol{j}}(l)$ for any lottery l.

If a less ambiguity averse decision maker prefers risk to ambiguity, then the same must be true for a more ambiguity averse decision maker. This intuition forms the basis for the comparison of ambiguity attitudes in the next definition.

Definition 16 *(More ambiguity averse).* Decision maker $\boldsymbol{i}$ is more ambiguity averse than decision maker $\boldsymbol{j}$ if for any $A \in \bar{\mathcal{X}}$ and any lottery l on A, $l \succ_{\boldsymbol{j}} e_A$ implies $l \succ_{\boldsymbol{i}} e_A$.

Now we can characterize this relation in terms of concave transformations of ϕ.[30] Although our characterization includes restrictions on risk preferences, we do not assume that the two decision makers must share the same utility function u for risk for their ambiguity preferences to be comparable.

Proposition 17. *The following holds.*

(i) *If $\boldsymbol{i}$ is more ambiguity averse than $\boldsymbol{j}$ and $\boldsymbol{i}$ is weakly more risk averse than $\boldsymbol{j}$, then $\phi_{\boldsymbol{i}} \circ u_{\boldsymbol{i}} = h \circ \phi_{\boldsymbol{j}} \circ u_{\boldsymbol{j}}$ for some strictly increasing and concave function h.*
(ii) *If $\phi_{\boldsymbol{i}} \circ u_{\boldsymbol{i}} = h \circ \phi_{\boldsymbol{j}} \circ u_{\boldsymbol{j}}$ for some strictly increasing and concave function h and $\boldsymbol{j}$ is weakly more risk averse than $\boldsymbol{i}$, then $\boldsymbol{i}$ is more ambiguity averse than $\boldsymbol{j}$.*

There is a corollary to Proposition 17 where “more ambiguity averse”, “more risk averse” and “concave” is replaced by “more ambiguity loving”, “more risk loving” and “convex”, respectively. In general, however, such comparative-static analysis is difficult if risk attitudes and ambiguity attitudes do not match. An exhaustive and careful study of this relationship in the spirit of the study by Chateauneuf et al. (2005) for the rank-dependent model is beyond the scope of this paper.

In order to gain further insights, however, we will assume for the rest of this section that all decision makers share the same von Neumann-Morgenstern utility function. The next statement follows immediately from Proposition 17 and, therefore, given without a proof.

Proposition 18. *Decision maker $\boldsymbol{i}$ is more ambiguity averse than decision maker $\boldsymbol{j}$ if and only if $\phi_{\boldsymbol{i}} = h \circ \phi_{\boldsymbol{j}}$ for some strictly increasing and concave function h.*

It follows from Proposition 18 that one can define a *coefficient of ambiguity aversion* $-\frac{\phi''}{\phi'}$ in analogy to the degree of risk aversion in standard expected utility theory under pure risk.

Proposition 19. *Suppose $\phi_{\boldsymbol{i}}$ and $\phi_{\boldsymbol{j}}$ are twice continuously differentiable functions. Decision maker $\boldsymbol{i}$ is more ambiguity averse than decision maker $\boldsymbol{j}$ if and only if for any $r \in U$,*

$$-\frac{\phi_{\boldsymbol{i}}''(r)}{\phi_{\boldsymbol{i}}'(r)} \geq -\frac{\phi_{\boldsymbol{j}}''(r)}{\phi_{\boldsymbol{j}}'(r)}. \tag{9}$$

Once again in analogy to the theory of decision making under pure risk, the next proposition will show that with increasing ambiguity aversion our representation V based on the principle of insufficient reason in Equation (7) converges to the Choquet integral V^C in Equation (1).

Proposition 20. *Let $\boldsymbol{i}_1, \boldsymbol{i}_2, \ldots$ be a sequence of decision makers such that*
(i) for any n, $\boldsymbol{i}_{n+1}$ is more ambiguity averse than $\boldsymbol{i}_n$,
(ii) for any decision maker $\boldsymbol{j}$, there exists $\tilde{n}$ such that $\boldsymbol{i}_{\tilde{n}}$ is more ambiguity averse than $\boldsymbol{j}$, then for any $\mu \in \mathcal{M}$, $V_n(\mu)$ converges to the Choquet integral of μ.

If in addition each ϕ_n is twice continuously differentiable, then $-\frac{\phi_n''}{\phi_n'}$ converges uniformly to $+\infty$.

---

[30] This procedure is similar to what one does in the theory of decision-making under risk (Yaari, 1987) and in the smooth model of Klibanoff et al. (2005).

5. Multiple priors and smooth preferences

In this section, we discuss two closely related approaches in the literature. Gajdos et al. (2008a) model *partial information* about probabilities by restricting the set of probabilities to a subset $P \subseteq \Delta(X)$[31] that is interpreted as the set of beliefs consistent with the information available. In another approach, Klibanoff et al. (2005) model information about probabilities by a (second-order) probability distribution ν on the set of probabilities $\Delta(X)$.[32] Taking the support of the (second-order) probability distribution ν of Klibanoff et al. (2005) as a constraint on the set of probability distributions in $\Delta(X)$, that is, taking $P = \text{supp}(\nu)$ highlights the similarity of these approaches.

For simplicity, we consider a finite set of outcomes X in this section.

5.1. Multiple priors and partial information

In order to study *partial information* about probabilities,Gajdos et al. (2008a) consider choice situations under uncertainty where decision makers have to rank Savage acts $f : S \to X$, mapping from a set of states S into a set of outcomes X, in the light of objective information about the possible probabilities of states. Objective information constrains the set of priors $P \subseteq \Delta(S)$.

Decision makers are assumed to be able to compare decision situations (P, f), consisting of information about probabilities P and acts f. In this framework, the authors derive a MEU representation

$$U(P, f) = \min_{p \in \varphi(P)} \sum_{s \in S} u(f(s))P(s)$$

where the set of relevant priors $\varphi(P)$ is a selection from the set of information-consistent probabilities P. This set of endogenously derived probability distributions $\varphi(P)$ is interpreted as the set of *subjective* beliefs of the decision maker in the light of the objective partial information P.

Gajdos et al. (2008a) also derive a notion of comparative precision and a concept of *imprecision aversion* in terms of properties of the selection mapping φ. For example, one may take all probabilities in a neighborhood of the centroid of the core of P as the subjective set of beliefs $\varphi(P)$ (Gajdos et al., 2008a, Section 4, p. 42). The size of the neighborhood can be interpreted as a measure of imprecision.

It is well-known (see, e.g., Gilboa and Schmeidler, 1994, Theorem E, p. 202) that a belief function μ has a non-empty core and that the Choquet integral of a belief function satisfies the following relationship

$$V^C(\mu) = \min_{p \in \text{core}(\mu)} \sum_{x \in X} u(x)p(x).$$

We can interpret the Choquet integral of a belief function as a MEU functional over the probability distributions in the core of the capacity μ. Therefore, the objective information encoded in the belief function μ defines uniquely a set of probability distributions compatible with the information in μ. Hence, one may view the Choquet integral of a belief function as a special case of the approach advanced in Gajdos et al. (2008a).

---

[31] $\Delta(X)$ is the set of all probability distributions over X.

[32] Both Gajdos et al. (2008a) and Klibanoff et al. (2005) consider probability distributions over a set of states $\Delta(S)$, rather than over outcomes $\Delta(X)$. Applying their reasoning to the framework in this paper is, however, straightforward.

In contrast to the Choquet integral of the belief function μ, the representation based on the principle of insufficient reason advanced in this paper can be viewed as a weighted average of the extreme probability distributions of the core of μ. The following proposition shows that our representation $V(\mu)$ can be interpreted as an expected utility with respect to a probability distribution $p \in \text{core}(\mu)$.

Proposition 21. *For any $\mu \in \mathcal{M}$, there exists a probability $p \in \text{core}(\mu)$ such that*

$$V(\mu) = \sum_{x \in X} p(x)u(x).$$

In other words, $V(\mu) = V(p)$ for some probability $p \in \text{core}(\mu)$. In the next proposition, we show that for the case of ambiguity neutrality, p is the centroid of $\text{core}(\mu)$. Moreover, the centroid serves as a benchmark for ambiguity attitude.

Proposition 22. *For any $\mu \in \mathcal{M}$ and centroid p^* of $\text{core}(\mu)$, if a decision maker is ambiguity averse (resp. loving, neutral), then*

$$V(\mu) \le V(p^*) \quad \big(\textit{resp. } V(\mu) \ge V(p^*),\ V(\mu) = V(p^*)\big).$$

The following example illustrates these results.

Example 23. Consider the following belief function on $X = \{x, y, z\}$ defined by the mass distribution $m_x + m_{xy} + m_{yz} = 1$. Hence, one obtains the belief function

$$\mu^m(A) = \begin{cases} 1 & \text{for } A = X \\ m_{yz} & \text{for } A = \{y, z\} \\ m_x + m_{xy} & \text{for } A = \{x, y\} \\ m_x & \text{for } A = \{x, z\} \\ m_x & \text{for } A = \{x\} \\ 0 & \text{otherwise} \end{cases}.$$

Fig. 2 illustrates the core of μ^m,

$$\text{core}(\mu^m) = \left\{ p \in \Delta^3 \mid p_x \ge m_x, p_x + p_y \ge m_x + m_{xy}, p_y + p_z \ge m_{yz} \right\}.$$

According to formula (8), the centroid of $\text{core}(\mu^m)$ is $(p^*(x), p^*(y), p^*(z)) = \left(m_x + \frac{m_{xy}}{2}, \frac{m_{xy}+m_{yz}}{2}, \frac{m_{yz}}{2}\right)$. It is easy to check that $p^*(x) + p^*(y) + p^*(z) = 1$. The core and its centroid are also indicated in Fig. 2.

In case of ambiguity neutrality, we have

$$\begin{aligned} V(\mu^m) &= m_x u(x) + m_{xy} \left[\frac{1}{2} (u(x) + u(y)) \right] + m_{yz} \left[\frac{1}{2} (u(y) + u(z)) \right] \\ &= p^*(x)u(x) + p^*(y)u(y) + p^*(z)u(z), \end{aligned}$$

as claimed in Proposition 22.

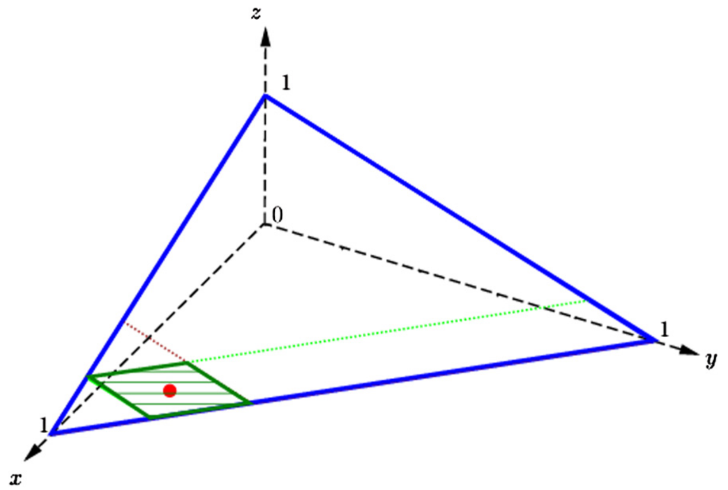


Fig. 2. Core of the belief function μ^m and its centroid.

5.2. *The smooth model*

In economic applications, the smooth model of decision making under uncertainty, suggested and axiomatized by Klibanoff et al. (2005), has proved very useful since it allows to analyze optimization problems with standard differential calculus. The smooth model represents uncertainty about probability distributions by a secondary probability distribution ν on $\Delta(S)$ and ambiguity attitude by a real-valued function Φ on the expected values of the unknown probabilities,

$$V(f) = \int_{\Delta(S)} \Phi\left(\sum_{s \in S} u(f(s))\pi(s)\right) d\nu.$$

In economic applications, one can interpret the utility function u applied to outcomes as measuring the decision-maker's attitude towards risk and the function Φ applied to the expected values of the probability distributions $\pi \in \Delta(S)$ as the decision-maker's attitude towards ambiguity.

The representation proposed in this paper shares some similarities with but also some differences to the smooth representation which we would like to discuss in this section. Regarding economic applications, however, the representation is equally smooth and can, therefore, be analyzed by differential calculus methods.

In contrast, to the smooth model of Klibanoff et al. (2005), we do not consider a probability distribution over all elements of $\Delta(X)$ but only uniform distributions over all probability distribution in the faces B of the simplex $\Delta(X)$. The uniform distribution over probabilities in a face $1/|A|$ reflects the *principle of insufficient reason* with respect to the uncertainty about the probabilities in the event (face) A, while the probability distribution over the faces of the simplex, i.e. the mass m, represents the (partial) information of the decision maker. Clearly, if there is complete uncertainty, $m(X) = 1$, the decision maker applies the principle of insufficient reason to all outcomes, $1/|X|$. If there is full information, that is, if $\sum_{x \in X} m(\{x\}) = 1$ then the decision maker knows the probability of each outcome x and the principle of insufficient reason becomes irrelevant.

More formally, define a utility function $\hat{u}$ over outcomes as $\hat{u}(x) = \phi(u(x))$ for all $x \in X$. Then representation (7) becomes

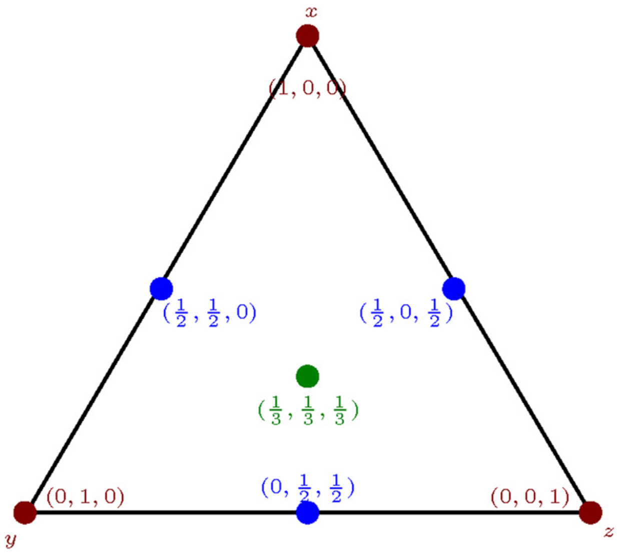


Fig. 3. Uniform distributions on the subsets of $X = \{x, y, z\}$.

$$V(\mu^m) = \sum_{A \subseteq X} m(A)\, \phi^{-1}\left(\frac{1}{|A|} \sum_{x \in A} \hat{u}(x) \right) = \mathbb{E}_\sigma \phi^{-1}(\mathbb{E}_\pi \hat{u}), \tag{10}$$

where $\mathbb{E}$ denotes the expectation operator, $\sigma(\bar{\pi}_A) = m(A)$ for all $A \subseteq X$ and $\bar{\pi}_A$ is the uniform distribution over the event (face) A of $\Delta(X)$. The last expression in (10) corresponds to the smooth model of Klibanoff et al. (2005). The following Fig. 3 for $X = \{x, y, z\}$ illustrates this interpretation.

Notice that according to formula (8) the centroid p^* of core(μ^m) is the convex combination of the points in $\Delta(X)$ with weights given by the mass m.

6. Concluding remarks

In this paper, we suggest and axiomatize a representation for a preference order over belief functions. Belief functions capture the (partial) information about probabilities over outcomes that a decision maker may have. In this framework, ambiguity is an objective feature of the information available in a decision situation. The preference order captures subjective features of the decision maker such as the evaluation of outcomes, risk attitudes and attitudes towards ambiguity. This approach allows for a clear separation of ambiguity as a feature of the information embodied in a belief function and ambiguity attitude as a feature of the preference relation. Hence, this approach resolves an important problem of decision making under uncertainty as discussed in Machina and Siniscalchi (2014).

Moreover, in our axiomatization, we provide a characterization of a decision maker evaluating the outcomes of an ambiguous event according to a generalized average, that is a uniform distribution over the outcomes in the ambiguous event. Uniform distributions in case of ignorance have a long tradition in economic and statistical decision theory as “Principle of Non-Sufficient Reason”, attributed to James Bernoulli by Keynes (1921, p. 41) or “Principle of Indifference” (Keynes, 1921).[33] However, such a representation has not received enough attention in modern axiomatic models.

---

[33] See also Machina and Siniscalchi (2014) Section 13.2.5.

Appendix. Proofs

The main Theorem 9 uses the following result of Matkowski and Páles (2015) which we quote here for the reader's convenience.

Matkowski and Páles (2015) define a quasi-arithmetic mean as follows:

"The notion of *quasi-arithmetic mean* was introduced in the book of Hardy, Littlewood and Pólya in [12][a] as a function $A_f : \cup_{n=1}^{\infty} I^n \to I$ defined by

$$A_f(x_1, ..., x_n) := f^{-1}\left(\frac{f(x_1) + \cdots + f(x_n)}{n}\right) \; (n \in \mathbb{N}, x_1, ..., x_n \in I)$$

where $I \subseteq \mathbb{R}$ denotes a non-degenerated interval (also in the rest of this paper) and $f : I \to \mathbb{R}$ is a continuous strictly monotone function. The mean A_f is said to be the quasi-arithmetic mean generated by f. The restriction of A_f to I^n will be called the n-variable quasi-arithmetic mean generated by f."

Matkowski and Páles (2015) prove the following theorem.

Theorem C. Let $n \geq 2$ and let $M : I^n \to I$. Then M is an *n-variable quasi-arithmetic mean*, that is, $M = A_f|_{I^n}$ for some continuous strictly monotone function $f : I \to \mathbb{R}$ if and only if
(i) M is a continuous and symmetric function on I^n which is strictly increasing in each of its variables;
(ii) M is reflexive;
(iii) M bisymmetric, that is, for all $x_{i,j} \in I$ $(i, j \in 1, ..., n)$, we have

$$M(M(x_{1,1}, ..., x_{1,n}), ..., M(x_{n,1}, ..., x_{n,n})) = M(M(x_{1,1}, ..., x_{n,1}), ..., M(x_{1,n}, ..., x_{n,n})).$$

---

[a] G. H. Hardy, J. E. Littlewood, and G. Pólya. Inequalities. Cambridge University Press, Cambridge, 1934. (first edition), 1952 (second edition).

Theorem 9

Axioms 1–3 imply that there is a linear utility function V on $\mathcal{M}$, i.e.

$$V(\mu^m) = \sum_{A \in \mathcal{X}} m(A) V(e_A).$$

Let u be defined by $u(x) = V(e_{\{x\}})$ for each $x \in X$. Since for any lottery $l^m \in \mathcal{M}$, we have $m(A) > 0$ only if $|A| = 1$, function $V(l^m)$ has an expected utility form,

$$V(l^m) = \sum_{x \in X} m(\{x\}) u(x).$$

Recall that X contains a certainty equivalent of any lottery in $\mathcal{M}$, for example, for any $x, y \in X$ and $0 < \lambda < 1$ there is $z \in X$ such that $u(z) = \lambda u(x) + (1 - \lambda) u(y)$. Therefore, U is an (open, closed, half-closed, finite or infinite) real interval of positive length (omitting the trivial case of complete indifference). To construct representation (7), we have to show that there exists a continuous strictly increasing function $\phi : U \to \mathbb{R}$ such that

$$V(e_A) = \phi^{-1}\left(\frac{1}{|A|}\sum_{x\in A}\phi(u(x))\right) \tag{11}$$

holds for any $A \in \bar{\mathcal{X}}$. To do this, we first prove (11) for two-element sets and then generalize the result to arbitrary finite sets.

Let function $M : U^2 \to \mathbb{R}$ be defined by

$$M\left(u(x), u(y)\right) = V\left(e_{\{x,y\}}\right).$$

Note that M is well-defined, because if $u(x_1) = u(x_2)$ and $u(y_1) = u(y_2)$, then $\{x_1, y_1\} \sim \{x_2, y_2\}$ by Monotonicity, therefore $V\left(e_{\{x_1,y_1\}}\right) = V\left(e_{\{x_2,y_2\}}\right)$. In what follows we study properties of this function.

For each $r \in U$ by richness of X there exist $x \neq y$ such that $u(x) = u(y) = r$. Since $x \sim y$, Set Betweenness implies $\{x\} \sim \{x, y\}$. Since

$$M\left(u(x), u(y)\right) = V\left(e_{\{x,y\}}\right) = V\left(e_{\{x\}}\right) = u(x),$$

we get $M(r, r) = r$, i.e., M is reflexive.

Let $r_1, r_2, s \in U$ and $r_1 < r_2$. Take different $x_1, x_2, y \in X$ such that $u(x_1) = r_1$, $u(x_2) = r_2$ and $u(y) = s$. By Monotonicity $\{x_1, y\} \prec \{x_2, y\}$, thus $M(r_1, s) < M(r_2, s)$. Therefore, M is strictly increasing in both variables.

For the next step, we have to show first that for any $A \in \bar{\mathcal{X}}$ there exists a certainty equivalent c_A, i.e. $c_A \in X$ and $c_A \sim A$. Indeed, if $x^*, x_* \in A$ and $x^* \succsim x \succsim x_*$ for all $x \in A$, then $x^* \succsim A \succsim x_*$ by Set Betweenness. We can find $0 \leq \lambda \leq 1$ such that $V(e_A) = \lambda u(x^*) + (1-\lambda)u(x_*)$. Therefore, the certainty equivalent of the lottery $\lambda x^* + (1-\lambda)x_*$ is also a certainty equivalent of A.

Now we would like to prove that

$$M\left(M(r, s), M(t, k)\right) = M\left(M(r, t), M(s, k)\right) \tag{12}$$

for arbitrary $r, s, t, k \in U$ (bisymmetry). To do this, take different $x, y, z, w \in X$ such that $u(x) = r$, $u(y) = s$ etc. Since $M(r, s) = u\left(c_{\{x,y\}}\right)$, we have

$$M\left(M(r, s), M(t, k)\right) = M\left(u\left(c_{\{x,y\}}\right), u\left(c_{\{z,w\}}\right)\right).$$

Axiom 7 implies that $\{c_{\{x,y\}}, c_{\{z,w\}}\} \sim \{x, y, z, w\} \sim \{c_{\{x,z\}}, c_{\{y,w\}}\}$, which leads to

$$M\left(u\left(c_{\{x,y\}}\right), u\left(c_{\{z,w\}}\right)\right) = M\left(u\left(c_{\{x,z\}}\right), u\left(c_{\{y,w\}}\right)\right)$$

from which (12) follows.

The fact that M is continuous follows from Set Continuity. To show this, we take two interior points r_0 and s_0 in U and prove that if $|r - r_0| < \delta$, then $|M(r, s_0) - M(r_0, s_0)| < \varepsilon$. Let $r_0 > s_0$, $x_0, y_0 \in X$, $u(x_0) = r_0$, and $u(y_0) = s_0$. Since $u(y_0) < M(r_0, s_0) < u(x_0)$, we can take $\varepsilon > 0$ such that the ε-neighborhood of $M(r_0, s_0)$ entirely lies between $u(y_0)$ and $u(x_0)$. Take $z \in X$ such that $u(z) = M(r_0, s_0) - \varepsilon$. Since $\{x_0, y_0\} \succ z$, by part (a) of Set Continuity $\{x_1, y_0\} \succ z$ for some $x_0 \succ x_1 \succ y_0$. Let $\delta_1 = r_0 - u(x_1) > 0$. If $r > r_0 - \delta_1$, then for $x \in X$ such that $u(x) = r$ we have $x \succ x_1$ implying $\{x, y_0\} \succ \{x_1, y_0\} \succ z$, which means $M(r, s_0) > M(r_0, s_0) - \varepsilon$. Using part (b) of Set Continuity and following a similar argument, we can find $\delta_2 > 0$ such that if $r < r_0 + \delta_2$, then $M(r, s_0) < M(r_0, s_0) + \varepsilon$. By taking $\delta = \min\{\delta_1, \delta_2\}$, we finish the proof that M is continuous in the first variable. For the second variable, the proof is similar.

Thus, we proved that M is continuous and strictly increasing in both variables, satisfies (12) and $M(r, r) = r$ for all $r \in U$. According to the theorem characterizing the quasi-arithmetic mean (see Matkowski and Páles (2015), Theorem C), M satisfies these conditions if and only if there exists a continuous and strictly monotonic function $\phi_0 : U \to \mathbb{R}$ with which

$$M(r,s) = \phi_0^{-1}\left(\frac{\phi_0(r)+\phi_0(s)}{2}\right) \tag{13}$$

holds for each $r, s \in U$. Define $\phi = \phi_0$ if ϕ_0 is an increasing function, and $\phi = -\phi_0$ otherwise. Therefore, ϕ is a continuous strictly increasing function satisfying (13). Thus, we proved (11) for two-element sets.

Now we extend (11) to an arbitrary $A \in \bar{\mathcal{X}}$. Suppose that (11) is true for all sets with no more than n elements, $n \geq 2$, and prove (11) for a set $A = \{x_1, \dots, x_{n+1}\}$.

If $n+1$ is an even number, then by Axiom 7 we have

$$A \sim \{c_{\{x_1,\dots,x_k\}}, c_{\{x_{k+1},\dots,x_{2k}\}}\},$$

where $2k = n+1$. The later set has only two elements, so we can apply representation (11), i.e.

$$V\left(e_A\right) = \phi^{-1}\left(\frac{\phi\left(u\left(c_{\{x_1,\dots,x_k\}}\right)\right)+\phi\left(u\left(c_{\{x_{k+1},\dots,x_{2k}\}}\right)\right)}{2}\right). \tag{14}$$

Since $k \leq n$, using the induction hypothesis we can rewrite

$$\phi\left(u\left(c_{\{x_1,\dots,x_k\}}\right)\right) = \phi\left(V\left(e_{\{x_1,\dots,x_k\}}\right)\right) = \frac{1}{k}\sum_{i=1}^{k}\phi\left(u(x_i)\right) \tag{15}$$

and similarly

$$\phi\left(u\left(c_{\{x_{k+1},\dots,x_{2k}\}}\right)\right) = \frac{1}{k}\sum_{i=k+1}^{2k}\phi\left(u(x_i)\right). \tag{16}$$

Substituting the two terms in Equation (14), by the right-hand sides of equations (15) and (16) we obtain

$$V\left(e_A\right) = \phi^{-1}\left(\frac{1}{2k}\sum_{i=1}^{2k}\phi\left(u(x_i)\right)\right),$$

as claimed.

Now suppose that $n+1$ is an odd number. For a certainty equivalent c_A by Set Betweenness we have

$$A \sim \{x_1, \dots, x_{n+1}, c_A\}.$$

The later set has $n+2$ elements, which is an even number. By repeating the previous argument we obtain

$$\phi\left(V\left(e_A\right)\right) = \frac{1}{2k}\sum_{i=1}^{2k-1}\phi\left(u(x_i)\right) + \frac{1}{2k}\phi\left(u(c_A)\right).$$

Since $u(c_A) = V(e_A)$,

$$\frac{2k-1}{2k}\phi\left(V\left(e_A\right)\right) = \frac{1}{2k}\sum_{i=1}^{2k-1}\phi\left(u(x_i)\right),$$

therefore,

$$\phi\left(V\left(e_A\right)\right)=\frac{1}{n+1}\sum_{i=1}^{n+1}\phi\left(u(x_i)\right).$$

Thus, we proved (11) for a set with $n+1$ elements.

The proof that the representation implies the axioms is straightforward. The same is true for the uniqueness results.

This proves the theorem.

Proposition 13

We prove only the case of ambiguity aversion since the other cases are similar.

(i) ⇒ *(ii)* is trivial because the situation of complete ignorance is obtained from the corresponding uniform lottery by an elementary increase in ambiguity.

(ii) ⇒ *(iii)* Since $V\left(\overline{\ell}_{\{x_1,x_2\}}\right)\geq V\left(e_{\{x_1,x_2\}}\right)$ for any $x_1,x_2\in X$, we have

$$\frac{u(x_1)+u(x_2)}{2}\geq\phi^{-1}\left(\frac{\phi(u(x_1))+\phi(u(x_2))}{2}\right). \tag{17}$$

Because ϕ is a strictly increasing function, we can apply it to both sides of inequality (17). We know that ϕ is defined on an interval and continuous, and we have just showed that for any r_1 and r_2 from that interval, $\phi(\frac{1}{2}r_1+\frac{1}{2}r_2)\geq\frac{1}{2}\phi(r_1)+\frac{1}{2}\phi(r_2)$. Hence, it is concave.

(iii) ⇒ *(i)* Assume first that m is obtained from n by an elementary increase in ambiguity, i.e. by shifting some positive mass to set E from a finite number of sets F_i in such a way that the centroid is preserved. Let p_i be the fraction of the mass that is shifted away from F_i, $\sum_i p_i=1$. Note that $\mu^m\preccurlyeq\mu^n$ if and only if $V(E)\leq\sum_i p_i V(F_i)$. Because the centroid is preserved, for any $x\in E$ it must be that

$$\sum_{i:x\in F_i}\frac{p_i}{|F_i|}=\frac{1}{|E|},$$

otherwise the centroid would have been shifted. Denoting $\phi(u(x))=\phi_x$, we obtain

$$\begin{aligned}V(E)&=\phi^{-1}\left(\sum_{x\in E}\frac{1}{|E|}\phi_x\right)=\phi^{-1}\left(\sum_{x\in E}\left(\sum_{i:x\in F_i}\frac{p_i}{|F_i|}\right)\phi_x\right)\\&=\phi^{-1}\left(\sum_i\sum_{x\in F_i}\frac{p_i}{|F_i|}\phi_x\right)=\phi^{-1}\left(\sum_i p_i\left(\sum_{x\in F_i}\frac{1}{|F_i|}\phi_x\right)\right)\\&\leq\sum_i p_i\,\phi^{-1}\left(\sum_{x\in F_i}\frac{1}{|F_i|}\phi_x\right)\\&=\sum_i p_i V(F_i),\end{aligned}$$

where the inequality follows from the convexity of ϕ^{-1}. Thus, $\mu^m\preccurlyeq\mu^n$ when m is obtained from n by an elementary increase in ambiguity.

If m is a spread of n, then m is obtained from n by a sequence of elementary increases in ambiguity. Thus, we have $\mu^m\preccurlyeq\mu^n$ because the preference relation is transitive, and, therefore, the decision maker is ambiguity averse.

Proposition 17

(i) Let h be defined by the equation $\phi_i \circ u_i = h \circ \phi_j \circ u_j$. Note that h is well-defined, because $\phi_j \circ u_j(x) = \phi_j \circ u_j(y)$ implies $\phi_i \circ u_i(x) = \phi_i \circ u_i(y)$. We must show that h is concave. Pick arbitrary r_1 and r_2 among the values of ϕ_j. There exist $x, y \in X$ such that $\phi_j \circ u_j(x) = r_1$ and $\phi_j \circ u_j(y) = r_2$. Moreover, there exist lotteries l^i and l^j on $\{x, y\}$ such that $l^i \sim_i e_{\{x,y\}}$ and $l^j \sim_j e_{\{x,y\}}$ (see the proof of Theorem 9). Then

$$h(\tfrac{1}{2}r_1 + \tfrac{1}{2}r_2) = h(\tfrac{1}{2}\phi_j \circ u_j(x) + \tfrac{1}{2}\phi_j \circ u_j(y)) = h \circ \phi_j \circ u_j \circ CE_j(l^j)$$

Note that $l^j \succsim_j l^i$, since otherwise we would have $l^i \succ_j l^j \sim_j e_{\{x,y\}}$ and, therefore, $l^i \succ_i e_{\{x,y\}}$, a contradiction. This implies that $CE_j(l^j) \succsim CE_j(l^i)$. Because $\boldsymbol{i}$ is weakly more risk averse than $\boldsymbol{j}$, we have $CE_j(l^i) \succsim CE_i(l^i)$. Thus,

$$\begin{aligned} h(\tfrac{1}{2}r_1 + \tfrac{1}{2}r_2) \geq h \circ \phi_j \circ u_j \circ CE_i(l^i) = \phi_i \circ u_i \circ CE_i(l^i) \\ = \tfrac{1}{2}\phi_i \circ u_i(x) + \tfrac{1}{2}\phi_i \circ u_i(y) = \tfrac{1}{2}h(r_1) + \tfrac{1}{2}h(r_2) \end{aligned}$$

Since h is a continuous function defined on an interval, this implies that it is concave.

(ii) Assuming $l \succ_j e_A$, we must show that $l \succ_i e_A$. Note that

$$\begin{aligned} \phi_i^{-1}(V_i(l)) &= \phi_i(\textstyle\sum_{x\in A} l(x)u_i(x)) = \phi_i \circ u_i \circ CE_i(l) \\ &\geq \phi_i \circ u_i \circ CE_j(l) = h \circ \phi_j \circ u_j \circ CE_j(l) \\ &= h \circ \phi_j(\textstyle\sum_{x\in A} l(x)u_j(x)) \\ &> h(\tfrac{1}{|A|}\textstyle\sum_{x\in A} \phi_j \circ u_j(x)) \\ &\geq \tfrac{1}{|A|}\textstyle\sum_{x\in A} h \circ \phi_j \circ u_j(x) \\ &= \tfrac{1}{|A|}\textstyle\sum_{x\in A} \phi_i \circ u_i(x) \\ &= \phi_i^{-1}(V_i(e_A)), \end{aligned}$$

where the first inequality follows from $CE_i(l) \succsim CE_j(l)$, the second follows from $l \succ_j e_A$, and the third follows from Jensen's inequality. Since ϕ_i^{-1} is monotone, this implies $l \succ_i e_A$. Thus, $\boldsymbol{i}$ is more ambiguity averse than $\boldsymbol{j}$.

Proposition 19

By Proposition 18, we only need to show that expression (9) is equivalent to $\phi_i = h \circ \phi_j$ for a strictly increasing and concave function h. The proof is standard. By differentiating equation $\phi_i(r) = h(\phi_j(r))$ twice and rearranging terms, we get

$$h''(\phi_j(r)) = \frac{\phi_i'(r)}{(\phi_j'(r))^2}\left[\frac{\phi_i''(r)}{\phi_i'(r)} - \frac{\phi_j''(r)}{\phi_j'(r)}\right],$$

which is non-positive if and only if the term in square brackets is non-positive.

Proposition 20

From the assumptions and Proposition 18 it follows that for any n there exists a strictly increasing and concave function h_n such that $\phi_{n+1} = h_n \circ \phi_n$. Moreover, for any continuous strictly increasing function ψ on U there exist $\tilde{n}$ and a strictly increasing and concave function $\tilde{h}$ such that $\psi = \tilde{h} \circ \phi_{\tilde{n}}$.

Take an arbitrary $\mu \in \mathcal{M}$. According to representation (7), $V_n(\mu)$ is a convex combination of the terms

$$\Phi(A;\phi_n)=\phi_n^{-1}\left(\frac{1}{|A|}\sum_{x\in A}\phi_n(u(x))\right),$$

whereas the Choquet integral of μ is a convex combination of the terms

$$\underline{\Phi}(A)=\min\{u(x)|\,x\in A\}$$

with the same weights $m(A)$. It is clear that $\underline{\Phi}(A)\leq\Phi(A;\phi_n)$ for all n. By Jensen's inequality, it is also true that $\Phi(A;\phi_{n+1})\leq\Phi(A;\phi_n)$ for all n. Therefore, we only need to show that $\Phi(A;\phi_n)$ is close to $\underline{\Phi}(A)$ for a sufficiently large n. We can do this by choosing ψ such that $\Phi(A;\psi)-\underline{\Phi}(A)\leq\varepsilon$ for all A. Then

$$\underline{\Phi}(A)\leq\Phi(A;\phi_n)\leq\Phi(A;\psi)\leq\underline{\Phi}(A)+\varepsilon$$

for any $n\geq\tilde{n}$, which means that $V_n(\mu)$ converges to the Choquet integral of μ.

Since we can always pick ψ with a sufficiently large coefficient of ambiguity aversion, it follows from Proposition 19 and Assumption (ii) that $-\frac{\phi_n''}{\phi_n'}$ converges uniformly to $+\infty$.

Proposition 21

Since for any $A\subseteq X$,

$$\min_{x\in A}u(x)\leq\phi^{-1}\left(\frac{1}{|A|}\sum_{x\in A}\phi(u(x))\right)\leq\max_{x\in A}u(x),\tag{18}$$

we have

$$\phi^{-1}\left(\frac{1}{|A|}\sum_{x\in A}\phi(u(x))\right)=\sum_{x\in A}\alpha^A(x)u(x)$$

for some probability distribution α^A over A. Therefore,

$$\begin{aligned}V(\mu^m)&=\sum_{A\subseteq X}m(A)\phi^{-1}\left(\frac{1}{|A|}\sum_{x\in A}\phi(u(x))\right)=\sum_{A\subseteq X}m(A)\sum_{x\in A}\alpha^A(x)u(x)\\&=\sum_{x\in X}u(x)\sum_{A\ni x}m(A)\alpha^A(x).\end{aligned}$$

For each $x\in X$, define

$$p(x)=\sum_{A\ni x}m(A)\alpha^A(x),$$

and show that p is a probability in the core of μ^m. Since $m\geq 0$ and $\alpha^A\geq 0$, we have $p\geq 0$. Also,

$$\begin{aligned}\sum_{x\in X}p(x)=\sum_{x\in X}\sum_{A\ni x}m(A)\alpha^A(x)&=\sum_{A\subseteq X}\sum_{x\in A}m(A)\alpha^A(x)\\&=\sum_{A\subseteq X}m(A)\sum_{x\in A}\alpha^A(x)=\sum_{A\subseteq X}m(A)=1.\end{aligned}$$

Now show that p is in the core of μ^m. For any $B\subseteq X$, we have

$$p(B)=\sum_{x\in B}p(x)=\sum_{x\in B}\sum_{A\ni x}m(A)\alpha^A(x)=\sum_{A\subseteq X}\sum_{x\in B\cap A}m(A)\alpha^A(x)$$
$$=\sum_{A\subseteq B}\sum_{x\in A}m(A)\alpha^A(x)+\sum_{A\setminus B\neq\emptyset}\sum_{x\in B\cap A}m(A)\alpha^A(x).$$

Take the first term,

$$\sum_{A\subseteq B}\sum_{x\in A}m(A)\alpha^A(x)=\sum_{A\subseteq B}m(A)=\mu^m(B).$$

Since the second term is non-negative, $p(B)\geq\mu^m(B)$.

Proposition 22

If a decision maker is ambiguity averse, then ϕ is concave by Proposition 13, which implies

$$V(\mu^m)=\sum_{A\subseteq X}m(A)\phi^{-1}\left(\frac{1}{|A|}\sum_{x\in A}\phi(u(x))\right)\leq\sum_{A\subseteq X}m(A)\frac{1}{|A|}\sum_{x\in A}u(x)$$
$$=\sum_{x\in X}u(x)\sum_{A\ni x}\frac{m(A)}{|A|}=\sum_{x\in X}u(x)p^*(x).$$

The argument for the other cases is similar.

References

Anscombe, F.J., Aumann, R.J., 1963. A definition of subjective probability. Ann. Math. Stat. 34, 199–205.

Baillon, A., Huang, Z., Selim, A., Wakker, P., 2018. Measuring ambiguity attitudes for all (natural) events. Econometrica 86 (5), 1839–1858.

Chateauneuf, A., Cohen, M., Meilijson, I., 2005. More pessimism than greediness: a characterization of monotone risk aversion in the rank-dependent expected utility model. Econ. Theory 25, 649–667.

Chateauneuf, A., Jaffray, J.Y., 1989. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. Math. Soc. Sci. 17 (3), 263–283.

Dempster, A.P., 1967. Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. 38 (2), 325–339.

Denneberg, D., 2000. Non-additive measure and integral, basic concepts and their role for applications. In: Grabisch, M., Murofushi, T., Sugeno, M. (Eds.), Fuzzy Measures and Integrals. Physica-Verlag, pp. 289–313.

Ellsberg, D., 1961. Risk, ambiguity and the Savage axioms. Q. J. Econ. 75 (4), 643–669.

Epstein, L.G., Zhang, J.-K., 2001. Subjective probabilities on subjectively unambiguous events. Econometrica 69, 265–306.

Gajdos, T., Hayashi, T., Tallon, J.-M., Vergnaud, J.-C., 2008a. Attitude toward imprecise information. J. Econ. Theory 14 (1), 27–65.

Gajdos, T., Tallon, Jean-Marc, Vergnaud, J.-C., 2008b. Representation and aggregation of preferences under uncertainty. J. Econ. Theory 141 (1), 68–99.

Ghirardato, P., Maccheroni, F., Marinacci, M., 2004. Differentiating ambiguity and ambiguity attitude. J. Econ. Theory 118, 133–173.

Gilboa, I., Schmeidler, D., 1994. Additive representations of non-additive measures and the Choquet integral. Ann. Oper. Res. 52 (1), 43–65.

Giraud, R., Tallon, J.-M., 2011. Are beliefs a matter of taste? A case for objective imprecise information. Theory Decis. 71, 23–31.

Grabisch, M., 2016. Set Functions, Games and Capacities in Decision Making. Springer, New York.

Grant, S., Roorda, B., Yang, J., 2021. Coherent Rich Beliefs, Decomposable Splits, and Dynamically Consistent Choice. Discussion paper. Australian National University, Canberra.

Gravel, N., Marchant, T., Sen, A., 2012. Uniform expected utility criteria for decision making under ignorance or objective ambiguity. J. Math. Psychol. 56 (5), 297–315.

Izhakian, Y., 2017. Expected utility with uncertain probabilities theory. J. Math. Econ. 69, 91–103.

Jaffray, J.Y., 1989. Linear utility theory for belief functions. Oper. Res. Lett. 8 (2), 107–112.

Ke, S., 2019. Boundedly rational backward induction. Theor. Econ. 14 (1), 103–134.

Kendall, D., 1974. Foundations of a theory of random sets. In: Harding, E., Kendall, D. (Eds.), Stochastic Geometry. J. Wiley, New York, pp. 322–376.

Keynes, J.M., 1921. A Treatise on Probability. Dover Publications, New York. 2004 edn.

Klibanoff, P., Marinacci, M., Mukerji, S., 2005. A smooth model of decision making under ambiguity. Econometrica 73 (6), 1849–1892.

Knight, F.H., 1921. Risk, Uncertainty, and Profit. Houghton Mifflin, New York.

Kops, C., Pasichnichenko, I., 2021. Testing Negative Value of Information and Ambiguity Aversion. Discussion paper. School of Economics, University of Bristol.

Luce, R.D., Raiffa, H., 1957. Games and Decisions. Introduction and Critical Survey. Wiley, New York.

Machina, M.J., Siniscalchi, M., 2014. Chapter 13 - Ambiguity and ambiguity aversion. In: Machina, M., Viscusi, K. (Eds.), Handbook of the Economics of Risk and Uncertainty, vol. 1. North-Holland, pp. 729–807.

Matheron, G., 1975. Random Sets and Integral Geometry. J. Wiley, New York.

Matkowski, J., Páles, Z., 2015. Characterization of generalized qasi-arithmetic means. arXiv:1501.02857v2 [math.CA].

Mukerji, S., 1997. Understanding the non-additive probability decision model. Econ. Theory 9, 23–46.

Nehring, K., 2006. Is it Possible to Define Subjective Probabilities in Purely Behavioral Terms? A Comment on Epstein-Zhang (2001). Working Paper.

Neilson, W., 2010. A simplified axiomatic approach to ambiguity aversion. J. Risk Uncertain. 41, 113–124.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference. Morgan Kaufmann Publishers, revised second printing.

Peter, R., Toquebeuf, P., 2020. Separating ambiguity and ambiguity attitude with mean-preserving capacities: Theory and applications. Discussion paper. University of Grenoble-Alpes.

Rothschild, M., Stiglitz, J.E., 1970. Increasing risk: I. A definition. J. Econ. Theory 2 (3), 225–243.

Savage, L.J., 1954. Foundations of Statistics. Wiley, New York.

Shafer, G., 1976. A Mathematical Theory of Evidence. Princeton Univ. Press.

Shafer, G., 1990. Belief functions. Introduction. In: Shafer, G., Pearl, J. (Eds.), Readings in Uncertain Reasoning. Morgan Kaufman Publishers, pp. 1–10, chap. 7.

Sinn, H.-W., 1980. A rehabilitation of the principle of insufficient reason. Q. J. Econ. 94 (3), 493–506.

von Neumann, J., Morgenstern, O., 1944. The Theory of Games and Economic Behavior. Princeton University Press, New Jersey.

Wakker, P., 2000. Dempster belief functions are based on the principle of complete ignorance. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 8 (3), 271–284.

Yaari, M.E., 1987. The dual theory of choice under risk. Econometrica, 9–115.