University of Heidelberg

Department of Economics

Discussion Paper Series | No. 680

Decision-making with partial information

Jürgen Eichberger and Illia Pasichnichenko

March 2020

Decision-making with partial information*

Jürgen Eichberger† and Illia Pasichnichenko‡

March 24, 2020

Abstract

In this paper, we study choice under uncertainty with belief functions on a set of outcomes as objects of choice. Belief functions describe what is objectively known about the probabilities of 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 which characterize a preference representation for belief functions that captures what is (objectively) known about the likelihood of outcomes and combines it with subjective beliefs about unknown probabilities according to the *"principle of insufficient reason"*. The approach is novel in its treatment of partial information and in its axiomatization of the uniform distribution in the case of ignorance. Moreover, our treatment of partial information yields a natural distinction between ambiguity and ambiguity attitude.

1 Introduction

In economics, decision analysis under uncertainty has almost exclusively focused on the two extreme cases of 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). With few exceptions[1], the more recent literature on decision making under ambiguity takes a purely subjective perspective. Many economic decision problems are, however, characterized by knowledge about the frequencies or probabilities of some events, yet not about all. Giraud and Tallon (2011) raise this issue and point to belief functions as a formal concept allowing one to combine objective, that is inter-subjectively verifiable, information about events, with purely subjective beliefs implicit in an individual's preferences.

*We would like to thank Hannes Rau for helpful comments on the early drafts of this paper. Illia Pasichnichenko acknowledges funding for this research by the Alexander von Humboldt Foundation.

†Alfred-Weber-Institute for Economics, Heidelberg University. E-mail: juergen.eichberger@awi.uni-heidelberg.de.

‡Alfred-Weber-Institute for Economics, Heidelberg University. E-mail: io.pasich@gmail.com.

[1] In a couple of papers, Chateauneuf and Vergnaud (2000), Gajdos, Hayashi, Tallon, and Vergnaud (2008), Gajdos, Tallon, and Vergnaud (2004, 2008) study the impact of objective information on belief functions. We will discuss this literature in more detail in Section 6 below.

In this paper, we will study belief functions on a set of outcomes as the decision maker's object of choice. In the spirit of Shafer (1976), these belief functions describe what is objectively known about the probability distributions over outcomes. We assume that decision makers can order these belief functions and deduce a preference representation from a set of axioms. Hence, two acts leading to the same belief function on outcomes are treated as equal. The preference representation will reflect both the objective information about the likelihood of events embodied in the belief function and, in case of unknown likelihoods, a subjective belief of the decision maker in the spirit of the *"principle of indifference"* advanced in Keynes (1921, Chapter IV). The approach is novel in its treatment of partial information and in its axiomatization of the uniform distribution as a representation of beliefs in case of ignorance.

In the following section, we will introduce and illustrate by examples how objective information can be captured by a belief function. Section 3 will propose a representation of preferences which uses the available partial information but resorts to the principle of indifference for events without any information. Section 4 provides and discusses a set of axioms characterizing the representation studied in Section 3. Section 5 summarizes essential properties of the representation functional, in particular the representation of ambiguity attitudes, and Section 6 compares the approach advanced in this paper to related representations in the literature.

2 Partial information and belief functions

Faced with uncertainty about the outcome of an action, decision makers base their choices on beliefs about the likelihood of the outcomes. Traditionally, these beliefs have been represented by probability distributions. von Neumann and Morgenstern (1944) were the first to consider lotteries over outcomes as objects of choice. In contrast, Savage (1954) took state-contingent outcomes as primitives and deduced subjective probabilities over outcomes from preferences over these state-contingent outcomes. The former approach assumes that information about 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 usually 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 suitable for integrating partial information in a formal approach. Belief functions are a special case of capacities for which the Choquet integral provides a natural way of forming an expected value. Capacities and the Choquet integral have been extensively studied in the literature (for example in Grabisch, 2016). The advantage of the special case of belief functions follows from the fact that the associated Möbius inverse, called *mass* by Shafer (1976), is a probability distribution over subsets of the set of outcomes. Hence, partial information about the probability of an event can be directly associated with the respective event in a mass distribution. The

Möbius transform converts this mass distribution in 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 a mass distribution putting weight only on singleton events. The case of *no information* (Savage, 1954) or *complete ignorance*, on the other hand, holds if the mass distribution puts a weight of one on the set of all possible outcomes and, hence, a weight of zero to all other subsets.

2.1 Belief functions and the Choquet integral

A decision maker's information and actions induce a probability distribution over outcomes and the decision maker chooses an action, that is a probability distribution over outcomes, yielding the highest expected utility. If the available information is insufficient, however, then the probability distribution over outcomes arising from an action is not determined completely. Partial information about the unknown probability distribution over outcomes can be described by a belief function. Before going to examples, we give formal definitions of a belief function and the Choquet integral.

Consider a finite set of outcomes X and the set 2^X of all subsets of X.

Definition 1. A function $m : 2^X \to [0,1]$ is called a *mass distribution on* X, if $m(\varnothing) = 0$ and $\sum_{A \subseteq X} m(A) = 1$.

Hence, a mass distribution on X shows the proportion of evidence that supports each event $A \subseteq X$. For example, if the decision maker knows only that an outcome belongs to the event A then the mass $m(A) = 1$ and $m(B) = 0$ for all other subsets of X, is a distribution on 2^X representing this state of information. Or if the probabilities of the outcomes x and y are known to be p_x and p_y, respectively, with $p_x + p_y = 1$, then $m(\{x\}) = p_x$, $m(\{y\}) = p_y$ and $m(B) = 0$ for all other $B \subseteq X$ is the mass distribution representing the probability distribution (p_x, p_y). More examples follow in Section 2.3.

A belief function $\mu(A)$ aggregates masses of subevents $B \subseteq A$ indicating the overall degree of evidence that A is true.

Definition 2. Let m be a mass distribution on X. We call $\mu : 2^X \to [0,1]$ a *belief function on* X, if $\mu(A) = \sum_{B \subseteq A} m(B)$.

A belief function on X resembles a probability distribution on X, although it is not necessary additive. In fact, it is a monotone of all orders capacity.[2] Given a belief function, the underlying mass distribution can be recovered uniquely. Hence, there is a one-to-one correspondence between belief functions and mass distributions. To highlight this connection, we will sometimes write μ^m.

The Choquet integral of a belief function μ^m can be obtained as the expected value of the minimal outcome in each event, $\min\{u(x) | x \in A\}$, with respect to the mass distribution.[3]

[2] Monotonicity of order 2 is convexity, i.e. $\mu(A \cup B) \geq \mu(A) + \mu(B) - \mu(A \cap B)$. See Chateauneuf and Jaffray (1989) for details. On the contrary, a mass distribution is a set function that is neither normalized nor monotone, hence no capacity.

[3] For a detailed discussion, see Gilboa and Schmeidler (1994). They show that the Choquet integral with respect to μ^m equals this average of minimums. We use the average of minimums as a definition.

Definition 3. Let μ^m be a belief function on X. We call

$$V^C\left(\mu^m\right) = \sum_{A \subseteq X} m(A) \min\left\{u(x) | x \in A\right\}. \tag{1}$$

the *Choquet integral of* μ^m. Abusing notation, we will also write $V^C(m)$ instead of $V^C\left(\mu^m\right)$.

Note that V^C depends on decision maker's risk preferences embodied in the von Neumann-Morgenstern utility function u.

We assume that the preferences of the decision maker are defined on the set of belief functions $\mathcal{M}$ over outcomes in X. By assumption, the decision maker is indifferent between any two acts leading to the same belief function on outcomes. In the next subsection, we discuss how information about states and an act can be converted into a belief function on X.

2.2 States and acts

In economic applications, information may be available for states s in a set of states S. In this case, *actions* (or *acts*) of a decision maker are functions $f : S \to X$ associating outcomes to states. In this case, information about states is modeled by a mass distribution m on S. This information together with an act f induces a distribution of mass $m * f$ on the set of outcomes X,

$$m * f\ (A) \ = \sum_{E \subseteq S : f(E) = A} m(E) \tag{2}$$

for any $A \subseteq X$. One checks easily that $m * f$ is a mass distribution on X. For the corresponding belief functions, we have

$$\mu^{m*f}(A) = \mu^m\left(f^{-1}(A)\right)$$

For the Choquet integral (1), one has

$$V^C(m * f) = \sum_{E \subseteq S} m(E) \min\{u(f(s)) | s \in E\},$$

which is easy to check using formula (2).

2.3 Leading examples

W.l.o.g., we will assume throughout this subsection that the von Neumann-Morgenstern utility index $u : X \to \mathbb{R}$ is a strictly increasing function.

The first example shows how one can easily compute the Choquet integral of an act given a belief function derived from a mass distribution. It illustrates also that the Choquet integral represents an extremely pessimistic or conservative evaluation of ambiguity.

Example 1. Consider a decision maker who can choose one of two actions $\{a, b\}$. The set of possible outcomes is $X = \{0, 90, 95, 100\}$. Action a yields an outcome from the set $A = \{0, 90, 100\}$ and action b from the set $B = \{0, 90, 95, 100\}$. Notice that different actions may yield different sets of outcomes. Assume that from previous experience it is known that, for both actions, there is at least a 40% chance of obtaining an outcome greater than 90. Without further information, this implies a mass distribution $m * a\,(\{90, 100\}) = 0.4$ and $m * a\,(\{0\}) = 0.6$ and $m * a = 0$ for all other subsets of X. For action b one obtains the mass distribution $m * b\,(\{90, 95, 100\}) = 0.4$ and $m * b\,(\{0\}) = 0.6$ and $m * b = 0$ for all other subsets of X. The Choquet integrals of the corresponding belief functions are the weighted averages of the worst possible outcomes in each event:

$$
\begin{aligned}
V^C(m * a) &= m * a\,(\{90, 100\}) \min\{u(x) | x \in \{90, 100\}\} + m * a\,(\{0\})u(0) \\
&= 0.4u(90) + 0.6u(0) \\
&= m * b\,(\{90, 95, 100\}) \min\{u(x) | x \in \{90, 95, 100\}\} + m * b\,(\{0\})u(0) \\
&= V^C(m * b).
\end{aligned}
$$

Hence, according to the Choquet evaluation the decision maker will be indifferent between the two acts a and b, since $V^C(m * a) = V^C(m * b)$.

Partial information about the probability distribution over outcomes often arises from information about marginal distributions. The second example illustrates how the partial information about the outcomes of acts may induce different mass distributions and therefore different belief functions over sets of outcomes.

Example 2. An investor considers two investments a and b. Investment a concerns a firm with markets in Europe (Region 1) and Asia (Region 2). State hh corresponds to good sales in both regions, state ll to bad sales in both regions, and states lh and hl to bad sales in one region and bad sales in the other region, respectively. Investment b has a safe return y. For investment a the probability distribution over outcomes $X = \{0, r, R\}$ with $0 < r < R$ is induced by the probability distribution over the set of states $\{hh, hl, lh, ll\}$.

		Region 2	
		h	l
Region 1	h	R	r
	l	r	0

	states			
S :	hh	hl	lh	ll
a :	R	r	r	0
	outcomes			

Suppose it is known from previous studies that the probability of high sales in Region 2 is $\Pr(h) = p$. Hence, it is known that the probability of the event $E_{2h} = \{hh, lh\}$ equals p and the probability of the event $E_{2l} = \{hl, ll\}$ is $1 - p$. There is no information about the other events.

This information induces the mass distribution m on events of the state space $S = \{hh, hl, lh, ll\}$ with $m(E_2) = p$ for $E_2 = \{hh, lh\}$, $= 1 - p$ for $E_2 = \{hl, ll\}$ and $= 0$ otherwise. This mass distribution m on the events of the state space S translates into an action-dependent mass distribution $m * a$ on the events of the outcome space X:

	mass distribution over outcomes						
a	R	r	0	$\{r, R\}$	$\{0, R\}$	$\{0, r\}$	$\{0, r, R\}$
$m * a$	0	0	0	p	0	$1-p$	0

The mass distribution of investment a yields the Choquet expected utility

$$\begin{aligned} V^C(m*a) &= m(\{hl, ll\}) \min\{u(a(s))|s \in \{hl, ll\}\} \\ &+ m(\{hh, lh\}) \min\{u(a(s))|s \in \{hh, lh\}\} \\ &= m*a(\{0, r\}) \min\{u(0), u(r)\} + m*a(\{r, R\}) \min\{u(r), u(R)\} \\ &= (1-p)u(0) + pu(r). \end{aligned}$$

Similarly, for the investment b with the safe return y, one obtains $V^C(m*b) = u(y)$. Hence, the decision maker will be willing to invest in a if $V^C(m*a) > V^C(m*b)$ or, equivalently, if $(1-p)u(0) + pu(r) > u(y)$.

Notice that the decision maker in Example 2 chooses extremely cautiously, disregarding the best possible outcome R. Note also that risk attitudes, i.e., the curvature of u will matter for the optimal choice. For example, a risk-loving decision maker with a convex von Neumann-Morgenstern function u may choose investment a even if its expected return $p \cdot r$ is less than the safe return y of investment b. Such a decision maker may be extremely uncertainty averse with respect to unknown outcomes but risk-loving with respect to probabilistic gambles. This feature can explain seemingly contradictory behavior of a consumer buying lottery tickets while insuring against uncertain risks (see Friedman and Savage, 1948).

The following example interprets the well-known Ellsberg three-color urn paradox (Ellsberg, 1961) as a consequence of partial information about the composition of the urn.

Example 3 (Three-color urn (Ellsberg, 1961))**.** Consider the version of the Ellsberg paradox where the decision maker can bet on the color of balls drawn form an opaque urn containing 30 red balls and 60 black or yellow balls in unknown proportion. The possible bets and corresponding prizes are given in the table.

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

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, $m * a(\{0, 100\}) = 0$. If all mass value can be allocated to singleton events, the corresponding belief function is a probability measure.
In contrast, for bet b, one knows that there is a chance of one third of receiving nothing (when the drawn ball is red), hence $m * b(\{0\}) = 1/3$, and of two thirds of getting 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 corresponding to the four acts in the table below. We would get the same result by formally applying equation (2).

	{0}	{100}	{0,100}	V^C
$m * a$	2/3	1/3	0	1/3
$m * b$	1/3	0	2/3	0
$m * a'$	0	1/3	2/3	1/3
$m * b'$	1/3	2/3	0	2/3

Assuming, w.l.o.g., a von Neumann-Morgenstern utility function $u : X \to R$ with $u(0) = 0$ and $u(100) = 1$, one can calculate the expected value of the minimal outcomes of each event with respect to these mass distributions, i.e.

$$V^C(m * f) = m * f(\{0\})u(0) + m * f(\{100\})u(100) + m * f(\{0, 100\}) \min\{u(0), u(100)\} \quad (3)$$

for $f \in \{a, b, a', b'\}$. The Choquet expected values of the four bets support the choices observed by Ellsberg (1961): $V^C(m * a) > V^C(m * b)$ and $V^C(m * a') < V^C(m * b')$.

Example 3 shows that careful consideration of the partial information about events together with the strong ambiguity aversion, implicit in the Choquet integral, i.e., in taking the minimal values over events, suffices to justify the choices in the Ellsberg three-colors paradox. Notice that the curvature of u, that is risk attitudes of the decision maker do not matter for this result.

As a final example, we consider the standard textbook insurance case.

Example 4 (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 for n_U cases there is no information recorded. From this data base, the insurer and the potential insuree can derive the following mass distribution over events $\{W\}$, $\{W - L\}$, and $\{W, W - L\}$

$$m(\{W - L\}) = \frac{n_L}{N}, \quad m(\{W\}) = \frac{n_W}{N}, \quad m(\{W - L, W\}) = \frac{n_U}{N},$$

with $N = n_L + n_W + n_U$. Denote by $\pi_L = \frac{n_L}{n_L + n_W}$ the loss proportion and by $\gamma = \frac{n_L + n_W}{N}$ the proportion of recorded outcomes, which may be interpreted as a degree of confidence in the frequency distribution $(\pi_L,\, 1 - \pi_L)$. Rewriting the mass distribution in terms of the parameters π_L and γ, one obtains

$$m(\{W - L\}) = \gamma \pi_L, \quad m(\{W\}) = \gamma(1 - \pi_L), \quad m\left(\{W - L, W\}\right) = 1 - \gamma.$$

Hence, the consumer's Choquet expected utility *without insurance* is

$$\begin{aligned} V^C(m * 0) &= m(\{W - L\})u(W - L) + m(\{W\})u(W) \\ &\quad + m\left(\{W - L, W\}\right) \min\left\{u(W - L), u(W)\right\} \\ &= \gamma \pi_L u(W - L) + \gamma(1 - \pi_L)u(W) + (1 - \gamma) \min\left\{u(W - L), u(W)\right\} \\ &= \left[\gamma \pi_L + (1 - \gamma)\right] u(W - L) + \gamma(1 - \pi_L)u(W). \end{aligned}$$

Suppose insurance is available at a premium $Q = qL$. With insurance, the wealth in case of a loss would be $W - L + L - qL = W - qL$ and in case of no loss $W - qL$. Thus, *complete insurance* yields a Choquet expected utility of

$$\begin{aligned} V^C(m * L) = m(\{W - L\})u(W - qL) + m(\{W\})u(W - qL) \\ + m\left(\{W - L, W\}\right) \min\left\{u(W - qL), u(W - qL)\right\} = u(W - qL). \end{aligned}$$

The consumer will insure completely if

$$V^C(m * L) > V^C(m * 0). \tag{4}$$

For the inequality to be true, it is sufficient that u is concave and $q < \gamma \pi_L + (1 - \gamma)$.

Notice that the consumer in Example 4 may take out full insurance even for an unfair insurance premium of $\pi_L < q < \gamma \pi_L + (1 - \gamma)$ since $\pi_L < \gamma \pi_L + (1 - \gamma)$ holds, if the lack of data induces a low degree of confidence γ less than 1. In principle, inequality (4) can hold for a convex u provided that γ is sufficiently small. Hence, this consumer may also buy lottery tickets at unfair odds.

3 Choice over belief functions

In most applied cases of decision making under uncertainty, complete ignorance, that is situations with no information about the likelihood of events, is an extreme case. Usually, some information about the probability of events is available from empirical studies. In many situations under uncertainty, objective information about the probability of some events can be gleaned from observed frequencies. Such information is, however, mostly incomplete or available only for imprecisely specified states and events. Belief functions, as was shown in the previous section, provide a useful method for modeling partial or imprecise information.

The Choquet integral of a belief function μ^m is the expected value of the minimal outcomes in each event, $\min\{u(x)|x \in A\}$, with respect to the mass distribution m of the belief function μ^m,

$$V^C(\mu^m) = \sum_{A \subseteq X} m(A) \min\{u(x)|x \in A\}. \tag{5}$$

For the special case of a mass distribution assigning positive weight only to singleton sets, the Choquet integral (5) is the well-known expected utility value with respect to the probability distribution m over outcomes. At the other extreme, if the mass distribution assigns all weight to a single subset A then $V^C(\mu^m) = \min\{u(x)|x \in A\}$ equals the worst possible outcome in the event A. Hence, one can interpret the Choquet integral associated with a mass distribution as an expected utility value reflecting the available information about events by the weight given to the events in the mass distribution while evaluating possible outcomes within an event by the worst outcome in the respective event.

This evaluation of an event by its worst outcome can be interpreted as *pure pessimism* or as (extremely) *precautionary* behavior. In a seminal paper, Jaffray (1989) takes belief functions as primitive objects of choice and assumes that decision makers have preferences over these belief functions. Preferences over belief functions can be interpreted as a combined evaluation of outcomes and objective information regarding the probabilities of events.

Applying the von Neumann and Morgenstern (1944) axioms to belief functions yields an expected utility value over sets of outcomes for which probabilities are known. In Jaffray (1989), a further axiom then implies an evaluation for sets of outcomes depending only on the minimum and the maximum utility of the outcomes in the event. A special case of Jaffray (1989)'s representation is the Hurwicz functional (Hurwicz, 1951). Another special case is the evaluation by the Choquet integral in equation (5).

Evaluating uncertain outcomes by the Choquet integral, however, takes a very pessimistic perspective on the evaluation of outcomes whose probabilities are unknown. Wald (1955) and Hurwicz (1951) proposed such rules for situations where no information about the probability of events is available[4]. An alternative view by Savage (1954) takes into account also non-extreme intermediate outcomes by weighting them with a purely subjective probability distribution.

For the case of complete ignorance, Keynes (1921, Chapter IV) advocates a *"principle of indifference"*, or *"principle of insufficient reason"* reaching back to Bernoulli (1713). This principle suggests an uninformative prior distribution, i.e. a uniform distribution over states in case of complete ignorance about probabilities, and to compute expected utility with respect to this uniform distribution.

In contrast to Savage (1954), who provides behavioral axioms for expected utility with an arbitrary subjective probability distribution, but remains silent about what may determine the subjective probabilities, the *"principle of insufficient reason"* suggests a

[4]Compare Luce and Raiffa (1957) for a detailed discussion of these and other decision rules under uncertainty.

uniform distribution, appealing to the symmetry of ignorance.[5]

There have been behavioral arguments put forward in favor of the *"principle of insufficient reason"*, e.g., Eva (2019) and Sinn (1980), but to our knowledge there is no axiomatization of the uniform distribution from behavioral axioms about preferences. In the next section, we will provide such an axiomatization in the framework of Jaffray (1989).

More specifically, let $|A|$ be the number of elements in A. We will assume that a decision maker's evaluation of a belief function μ^m can be described by a representation

$$V(\mu^m) = \sum_{A \subseteq X} m(A) M_\phi \left(u_1, \ldots, u_{|A|}\right), \tag{6}$$

where $M_\phi \left(u_1, \ldots, u_{|A|}\right)$ is a *quasi-arithmetic mean*[6] of utilities $u_1 = u(x_1), \ldots, u_{|A|} = u(x_{|A|})$ over the outcomes in A,

$$M_\phi \left(u_1, \ldots, u_{|A|}\right) = \phi^{-1} \left(\frac{\phi(u_1) + \cdots + \phi(u_{|A|})}{|A|} \right), \tag{7}$$

for some increasing function ϕ. Under risk, i.e. when the mass distribution m is positive only for singletons, the representation in equation (6) reduces to the expected-utility functional, because $M_\phi(u_1) = u_1$ in case $|A| = 1$.

More generally, the function ϕ reflects the decision maker's attitude towards complete ignorance about the likelihood of outcomes in A. The more concave is ϕ, the closer M_ϕ approaches the lowest utility of an outcome in A representing a pessimistic attitude when facing unknown chances. In contrast, a convex ϕ represents an optimistic attitude, pushing M_ϕ towards the highest utility among the outcomes in A.[7]

Before providing an axiomatic characterization for the representation in equations (6) and (7) above, we will briefly reconsider the examples of Section 2 and compare the evaluation by the Choquet expected value (5) with the evaluation based on the principle of insufficient reason in equations (6) and (7).

3.1 Leading examples revisited

Example 5 illustrates the main difference between the principle of insufficient reason and the more pessimistic approach of Choquet expected utility. While the Choquet expected utility value disregards all outcomes other than the worst, the principle of insufficient reason gives all uncertain outcomes the same weight.

[5] Keynes (1921, p. 41) quotes the moral philosopher Bernard Bosanquet with the following interchange of two character, "Absolute" and "Sir Anthony": ABSOLUTE. Sure, Sir, this is not very reasonable, to summon my affection for a lady I know nothing of.' SIR ANTHONY: I am sure, Sir, 'tis more unreasonable in you to object to a lady you know nothing of.'

[6] Quasi-arithmetic mean was first characterized by Kolmogorov (1930). We rely on the characterization in Matkowski and Páles (2015) which we cite in the appendix for convenience.

[7] We will discuss the issue of ambiguity attitude more formally and in more detail below in Section 5.

Example 5 (Example 1 resumed)**.** Reconsider the decision maker of Example 1 who can choose one of two actions a or b yielding outcomes from the sets $A = \{0, 90, 100\}$ in case of a or $B = \{0, 90, 95, 100\}$ in case of b. Given the mass distributions $m * a$ and $m * b$ derived before, a Choquet expected utility maximizer will be indifferent between the two actions, $V^C(m * a) = V^C(m * b)$.

In contrast, assuming the principle of insufficient reason with respect to the uncertain outcomes in the events, one obtains

$$\begin{aligned} V(m * a) &= m * a\,(\{90, 100\}) M_\phi(u(90), u(100)) + m * a\,(\{0\}) u(0) \\ &= 0.4 M_\phi(u(90), u(100)) + 0.6 u(0) \\ &< 0.4 M_\phi(u(90), u(95), u(100)) + 0.6 u(0) \\ &= m * b\,(\{90, 95, 100\}) M_\phi(u(90), u(95), u(100)) + m * b\,(\{0\}) u(0) \\ &= V(m * b). \end{aligned}$$

where the inequality assumes a strict increasing and concave function u and a linear function ϕ. We will argue below in Section 5 that a linear function ϕ corresponds to ambiguity neutrality. Hence, risk aversion, represented by the concave function u, suffices to make action b preferable to action a, in contrast to the result for the Choquet integral.

Partial information about the probability distribution over outcomes often arises from information about marginal distributions.

Example 6 (Example 2 resumed)**.** As shown before, investing in asset a yields the mass distribution $m * a(A) = p$ for $A = \{R, r\}$, $= 1 - p$ for $A = \{r, 0\}$, and 0 otherwise. For the principle of insufficient reason, this mass distribution is evaluated as

$$\begin{aligned} V(m * a) = m * a\,(\{0, r\}) M_\phi(u(0), u(r)) + m * a\,(\{r, R\}) M_\phi(u(r), u(R)) \\ = (1 - p) M_\phi(u(0), u(r)) + p M_\phi(u(r), u(R)). \end{aligned}$$

Since the quasi-arithmetic mean of two distinct numbers is always larger than the smallest of the two, one has $V(m * a) > (1 - p)u(0) + pu(r) = V^C(m * a)$. Hence, a decision maker who values uncertainty by the principle of insufficient reason will be willing to pay more for the investment opportunity than a decision maker with the Choquet evaluation. Notice that no assumption about risk attitudes, the curvature of u, is necessary for this result.

Ambiguity aversion also suffices for obtaining the usual behavior in the Ellsberg three-color urn paradox.

Example 7 (Example 3 resumed: Three-color urn (Ellsberg, 1961))**.** According to the principle of insufficient reason, the value of a bet $f \in \{a, b, a', b'\}$ is given by the formula

$$V(m * f) = m * f\,(\{0\}) u(0) + m * f\,(\{100\}) u(100) + m * f\,(\{0, 100\}) M_\phi(u(0), u(100)).$$

Assuming $u(0) = 0$ and $u(100) = 1$, as before, and also, w.l.o.g., $\phi(0) = 0$ and $\phi(1) = 1$, we get

$$V(m * f) = m * f(\{100\}) + m * f(\{0, 100\}) \phi^{-1}\left(\tfrac{1}{2}\right).$$

The values for the four bets are listed in the table below. For easier comparison, we also reproduce the values obtained for the Choquet expected utility (3).

	V	V^C
$m * a$	$\frac{1}{3}$	$\frac{1}{3}$
$m * b$	$\frac{2}{3}\phi^{-1}\left(\frac{1}{2}\right)$	0
$m * a'$	$\frac{1}{3} + \frac{2}{3}\phi^{-1}\left(\frac{1}{2}\right)$	$\frac{1}{3}$
$m * b'$	$\frac{2}{3}$	$\frac{2}{3}$

If ϕ is concave, then $\phi^{-1}\left(\frac{1}{2}\right) < \frac{1}{2}$. Hence, $V(m * a) > V(m * b)$ and $V(m * a') < V(m*b')$ follow. Preferences with a concave ϕ are usually interpreted as ambiguity averse. Similarly, a convex function ϕ will represent ambiguity prone preferences. Finally, for linear ϕ, the decision maker would be indifferent between a and b, and between a' and b'.

Notice that it is impossible to obtain the preferences $V(m * a) > V(m * b)$ and $V(m*a') > V(m*b')$ by just varying ϕ, holding the mass distribution m fixed. Varying ϕ influences only the evaluation of non-singleton events with some "ignorance" component. E.g., the mass of $\{0, 100\}$ is the same for the bets b and a'. Any increase in the value of the bet b due to a change in the shape of ϕ leads to an equal increase in the value of a'. Thus, preferences for a and a' imply that decision maker's information must be different from that represented by m.

Example 7 shows that even with a uniform distribution in case of complete ignorance about the outcomes in an event ambiguity attitude can enter the decision via the transformation function ϕ.

Finally, we reconsider the insurance case.

Example 8 (Example 4 resumed: Insurance)**.** Recall the consumer with wealth W facing a potential loss L. For the case of ignorance about events due to missing data, a consumer subscribing to the principle of insufficient reason evaluates the initial allocation without insurance as

$$V(m * 0) = \gamma\pi_L u(W - L) + \gamma(1 - \pi_L)u(W) + (1 - \gamma)M_\phi(u(W - L), u(W))$$

and, for the case with insurance at a premium $Q = qL$, one has $V(m * L) = u(W - qL)$. If $V(m * L) - V(m * 0) > 0$ 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

$$\phi^{-1}\left(\tfrac{1}{2}[\phi(u(W - L)) + \phi(u(W))]\right) < \tfrac{1}{2}[u(W - L) + u(W)] < u\left(W - \tfrac{1}{2}L\right),$$
$$\pi_L u(W - L) + (1 - \pi_L)u(W) < u(W - \pi_L L),$$

and, therefore,

$$V(m*0) < \gamma u(W - \pi_L L) + (1-\gamma) u\left(W - \tfrac{1}{2}L\right) < u\left(W - \left(\gamma\pi_L + (1-\gamma)\tfrac{1}{2}\right) L\right).$$

Hence, the consumer will insure completely if

$$q \leq \gamma\pi_L + (1-\gamma)\tfrac{1}{2}$$

holds, since in this case one has

$$V(m*L) - V(m*0) > u(W - qL) - u(W - (\gamma\pi_L + (1-\gamma)\tfrac{1}{2})L) \geq 0.$$

If $\gamma < 1$, the decision maker following the principle of insufficient reason will have to weigh the frequency information of π_L against the equal probability in those cases where no recorded information is available. Notice that if $\pi_L > \frac{1}{2}$, the decision maker may be not willing to buy full insurance at a fair premium $q = \pi_L$. On the other hand, when $\pi_L < \frac{1}{2}$, full insurance is bought even at an unfair premium $\pi_L < q \leq \gamma\pi_L + (1-\gamma)\frac{1}{2}$. Hence, in contrast to the evaluation according to the Choquet integral where ambiguity aversion was sufficient for this result, one needs also a loss probability π_L which is smaller than $\frac{1}{2}$, i.e., the default probability of this event due to the principle of insufficient reason.

4 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, 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[8] 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 4. 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 $\mu^m = \sum_{A \in \mathcal{X}} m(A) e_A$.[9]

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 $\mathcal{M}$:

$$V(\mu^m) = \sum_{A \in \mathcal{X}} m(A) V(e_A), \tag{8}$$

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

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

[9] 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.

order to specify $V(e_A)$ in equation (8), additional assumptions need to be made about decision maker's evaluation of such situations.[10]

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 (7).[11]

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 $A \succcurlyeq A \cup B \succcurlyeq B$.[12]

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.

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,

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

[11]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).

[12]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.

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

$$V(\mu^m) = \sum_{A \in \mathcal{X}} m(A)\, \phi^{-1}\left(\frac{1}{|A|}\sum_{x \in A} \phi(u(x))\right), \tag{9}$$

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 straight forward 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.

5 General properties of the representation

In this section, we will discuss some properties of the representation (9). Our focus will be on how ambiguity and ambiguity attitudes are captured by this representation.

5.1 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.

In general, a mass distribution assigns a weight to both singleton and non-singleton events $A \in \mathcal{X}$. The special case of a probability distribution over outcomes arises if only singleton events have positive mass. Information is ambiguous if some non-singleton events carry positive mass. Hence, a natural measure of ambiguity is the total mass assigned to non-singleton events.

Definition 5 (*Degree of ambiguity*)**.** For a mass distribution m on $\mathcal{X}$, define the degree of ambiguity δ_m as the total mass assigned to non-singletons,[13]

$$\delta_m = \sum_{A \in \mathcal{X}, |A|>1} m(A).$$

According to Definition 5, there is no ambiguity for $\delta_m = 0$, which implies that the belief function μ^m is a probability distribution with probability $p(x)$ of outcome x equal to $\mu^m(\{x\}) = m(\{x\})$ for all $x \in X$. It is easy to check that $\sum_{x \in X} p(x) = \sum_{x \in X} m(\{x\}) = 1$. On the other hand, in case of complete ignorance, no mass is assigned to singletons, and, hence, one has $\delta_m = 1$. In the general case when mass is assigned to both singleton and non-singleton events, one has $0 < \delta_m < 1$.

The following example from a laboratory experiment by Kops and Pasichnichenko (2020) illustrates this idea.

Example 9 (This example describes the setup of an experiment run by Kops and Pasichnichenko (2020))**.** Consider an urn containing 21 balls which are either green or blue. Subjects were given the information that n_g balls were green and n_b balls were blue and the remaining $21 - n_g - n_b$ balls were either green or blue. Clearly, for $n_g + n_b = 21$ and $n_g + n_b = 0$, one obtains the well-known Ellsberg two-urn case. Arguably, there is more ambiguity if $n_g + n_b$ is close to zero than if $n_g + n_b$ is close to 21. The mass distribution for the urn would assign $m(\{g\}) = \frac{n_g}{21}$, $m(\{b\}) = \frac{n_b}{21}$, and $m(\{g,b\}) = \frac{21-n_g-n_b}{21}$. Hence, $\delta = \frac{21-m(\{g\})-m(\{b\})}{21} = m(\{g,b\})$. For $\delta = 0$, one has the case of pure risk, Ellsberg's *"unambiguous urn"*, and for $\delta = 1$ the case of complete ambiguity, Ellsberg's *"ambiguous urn"* (Ellsberg, 1961).

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.

[13] Recall that following the convention of Section 4 m is positive only for a finite number of finite sets in $\mathcal{X}$.

For a formal definition of ambiguity attitude it is necessary to define a reference point. Ambiguity is reflected by the weights of the mass function for non-singleton events $A \in \bar{\mathcal{X}}$. The subjective evaluation of this ambiguity is captured by the value a decision maker assigns to the elementary belief function e_A reflecting the situation where the decision maker is certain that the true outcome x is in A but has no idea about its probability, $V(e_A) = \phi^{-1}\left(\frac{1}{|A|}\sum_{x\in A}\phi(u(x))\right)$, where $\frac{1}{|A|}\sum_{x\in A}\phi(u(x))$ reflects the principle of insufficient reason and ϕ the weight associated with this ambiguous "expected utility". This evaluation has to be assessed with respect to the lottery (without ambiguity) $\bar{\ell}_A$ which gives equal weight to each outcome in A,

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

Following Ellsberg (1961), for any ambiguous event A, the definition of ambiguity attitude compares the decision maker's evaluation of the uniform distribution in case of complete ignorance $V(e_A)$, with the value of the uniform lottery $V(\bar{\ell}_A)$ in this event.

Definition 6 (Ambiguity attitudes)**.** A decision maker is

- *ambiguity averse* if $\bar{\ell}_A \succcurlyeq e_A$ for all $A \in \bar{\mathcal{X}}$,
- *ambiguity neutral* if $\bar{\ell}_A \sim e_A$ for all $A \in \bar{\mathcal{X}}$, and
- *ambiguity loving* if $\bar{\ell}_A \preccurlyeq e_A$ for all $A \in \bar{\mathcal{X}}$.

For the representation (9), ambiguity attitude is measured by the function $\phi : \mathbb{R} \to \mathbb{R}$.

Proposition 1. *A decision maker is ambiguity averse (resp. loving, neutral) if and only if 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 5.2 below, we will try to provide more intuition by studying the case of two outcomes.

Example 10 (two-outcome case)**.** Consider the case of two outcomes, $X = \{x_1, x_2\}$ and a belief function μ^m. In this case, one can write the representation (9) as

$$V(\mu^m) = m(\{x_1\})u(x_1) + m(\{x_2\})u(x_2) + m(\{x_1, x_2\})\phi^{-1}\left(\frac{1}{2}\left[\phi(u(x_1)) + \phi(u(x_2))\right]\right).$$

Or, for notational convenience, we will write

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

with $m_i = m(\{x_i\}), m_{kl} = m(\{x_k, x_l\})$, etc. denoting the parameters of the mass distribution. Note that $0 \le m_1, m_2, m_{12} \le 1$ and $m_1 + m_2 + m_{12} = 1$. 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

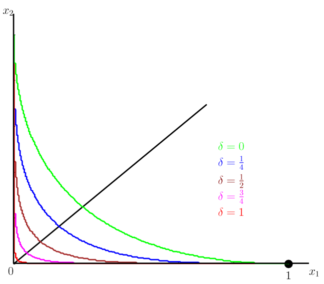
1. certainty implies independence from risk and ambiguity attitudes, hence, the marginal rate of substitution equals the known odds: $-\frac{m_1}{m_2}$,

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 $\frac{m_1 + \frac{1}{2} m_{12}}{m_2 + \frac{1}{2} m_{12}} \neq \frac{m_1}{m_2}$.

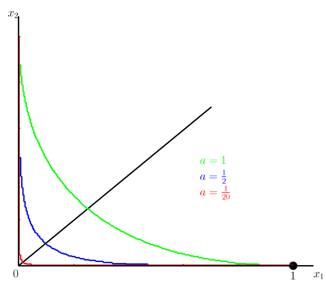
The following diagrams show indifference curves of the preference representation V over outcome combinations (x_1, x_2) for different degrees of ambiguity δ (1a) and an exponential ambiguity attitude function $\phi(x) = x^a$ with different degrees of ambiguity attitude a (1b).

Figure 1: **Ambiguity** δ and **Ambiguity attitudes** a



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

Figure 1a



$\delta = \frac{3}{4}$

Figure 1b

$$V(x_1, x_2) = (1 - \delta)\left[pu(x_1) + (1 - p)u(x_2)\right] + \delta\phi^{-1}\left(\frac{1}{2}[\phi(u(x_1) + \phi(u(x_2))]\right) = V(1, 0),$$
$$u(x) = \sqrt{x} \text{ and } \phi(x) = x^a,$$
$$p = \tfrac{1}{2}.$$

Figure 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 $\bar{\ell}_X$ 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, $\delta \to 1$.

Figure 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, $a = 1$, $V(\mu^m) = V(\bar{\ell}_X)$ for all degrees of ambiguity δ.

The diagrammatic illustrations of the effects of ambiguity and ambiguity attitude in Figure 1 can be generalized to outcome sets X with an arbitrary number of outcomes, to arbitrary representations u of risk attitudes, and to arbitrary ambiguity attitude functions ϕ. In the following section, we will derive these properties in full generality.

5.2 Comparison of ambiguity attitudes

Definition 6 defines ambiguity attitude for ambiguous events. This notion can be applied to any belief function $\mu^m \in \mathcal{M}$ in a straightforward way. Given the partial information of the belief function μ^m and the evaluation of uncertain events by the principle of insufficient reason, an ambiguity averse decision maker will prefer the expected utility from the average distribution

$$p^*(x) = \sum_{A \ni x} \frac{m(A)}{|A|}, \quad x \in X,$$

over the belief function μ^m,

$$V(\mu^m) \leq \sum_{x \in X} p^*(x) u(x)$$

(see Proposition 6 in Section 6). Moreover, a "more concave" ϕ corresponds to a higher degree of ambiguity aversion or pessimism. In this section, we formalize this idea.

Let $\boldsymbol{i}$ and $\boldsymbol{j}$ be two decision makers with preferences on $\mathcal{M}$ satisfying Axioms 1-7.

Definition 7 (More ambiguity averse)**.** We say decision maker $\boldsymbol{i}$ *is more ambiguity averse than decision maker* $\boldsymbol{j}$ if the two share the same risk preferences and 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 a concave transformations of ϕ.[14]

Proposition 2. *Decision maker* $\boldsymbol{i}$ *is more ambiguity averse than decision maker* $\boldsymbol{j}$ *if and only if* $u_{\boldsymbol{i}} = a u_{\boldsymbol{j}} + b$ *for some numbers* $a > 0$ *and* b *and* $\phi_{\boldsymbol{i}} = h \circ \phi_{\boldsymbol{j}}$ *for some strictly increasing and concave function* h*.*

Condition $u_{\boldsymbol{i}} = a u_{\boldsymbol{j}} + b$ means that both decision makers share the same risk preferences since the two Bernoulli utility functions are linear affine transformations of each other.

It follows from Proposition 2 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 3. *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 both share the same von Neumann-Morgenstern utility function* u *and, for any* $r \in U$,

$$-\frac{\phi''_{\boldsymbol{i}}(r)}{\phi'_{\boldsymbol{i}}(r)} \geq -\frac{\phi''_{\boldsymbol{j}}(r)}{\phi'_{\boldsymbol{j}}(r)}. \tag{10}$$

[14] 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, Marinacci, and Mukerji (2005)

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 (9) converges to the the Choquet integral V^C in equation (5).

Proposition 4. *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 $\boldsymbol{j}$ that shares with $\boldsymbol{i}_1, \boldsymbol{i}_2, \ldots$ the same von Neumann-Morgenstern utility function, 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 $+\infty$.

6 Related literature

In this section, we discuss two closely related approaches in the literature. Gajdos, Hayashi, Tallon, and Vergnaud (2008) model *partial information* about probabilities by restricting the set of probabilities to a subset $P \subseteq \Delta(X)$ [15] that is interpreted as the set of beliefs consistent with the information available. In another approach, Klibanoff, Marinacci, and Mukerji (2005) model information about probabilities by a (second-order) probability distribution ν on the set of probabilities $\Delta(X)$.[16] Taking the support of the (second-order) probability distribution ν of Klibanoff, Marinacci, and Mukerji (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.

6.1 Multiple priors and partial information

In order to study *partial information* about probabilities, Gajdos, Hayashi, Tallon, and Vergnaud (2008) 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 $P \subseteq \Delta(S)$. If the set P is a singleton, then information is perfect. In extreme contrast, if there is no objective information, P equals the set of all probability distributions $\Delta(S)$, that is, there is complete ignorance. Cases of partial information are modeled by a set P which is neither a singleton not the universal set.

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

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

[16] Both Gajdos, Hayashi, Tallon, and Vergnaud (2008) and Klibanoff, Marinacci, and Mukerji (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.

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, Hayashi, Tallon, and Vergnaud (2008) 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 Steiner point of P as the subjective set of beliefs $\varphi(P)$ (Gajdos, Hayashi, Tallon, and Vergnaud, 2008, 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,

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

and that the Choquet integral of a belief function satisfies the following relationship

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

Hence, 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, i.e.,

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

Therefore, the objective information encoded in the belief function μ defines uniquely a set of probability distributions compatible with the information in μ. This links the approach in Gajdos, Hayashi, Tallon, and Vergnaud (2008) to the Choquet integral of a belief function. Hence, one may view the Choquet integral of a belief function as a special case of the approach advanced in Gajdos, Hayashi, Tallon, and Vergnaud (2008).

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 5. *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 Proposition 6 we show that in case of ambiguity neutrality, p is the *centroid* of $\text{core}(\mu)$, which we define in the next paragraph.

A set $A \subseteq X$ such that $m(A) > 0$ is called a *focal element* of μ^m. We call p an extreme probability in $\text{core}(\mu^m)$, if for any focal element A of μ^m there exists $x \in A$ such that $p(x) = m(A)$. Intuitively, an extreme probability assigns all mass $m(A)$ to the element $x \in A$. By choosing different $x \in A$, we can produce $|A|$ extreme probabilities. Hence, if $A_1, \ldots, A_k$ are focal elements of μ^m, then there are $n = |A_1| \cdots |A_k|$ extreme probabilities $p_1, \ldots, p_n$ in $\text{core}(\mu^m)$ (which are not necessarily distinct). The average probability distribution p^* on X, given by

$$p^*(x) = \sum_{i=1}^{n} p_i(x)$$

for all $x \in X$, is called the *centroid* of $\text{core}(\mu^m)$.[17] It is not difficult to show (see the proof of Proposition 6) that

$$p^*(x) = \sum_{A \ni x} \frac{m(A)}{|A|} \tag{11}$$

for all $x \in X$. The next proposition shows connection between ambiguity attitude and the notion of centroid.

Proposition 6. *For any $\mu \in \mathcal{M}$ and centroid p^* of* $\text{core}(\mu)$*, if a decision maker is ambiguity averse (resp. loving, neutral), then*

$$V(\mu) \leq V(p^*) \quad (\textit{resp. } V(\mu) \geq V(p^*),\ V(\mu) = V(p^*))$$

.

The following example illustrates these results.

Example 11. Consider the following belief function on $X = \{x, y, z\}$ defined by the mass distribution $m(\{x\}) = m_x, m(\{y, z\}) = m_{yz}, m(\{x, y\}) = m_{xy}$ with $m_x + m_{xy} + m_{yz} = 1$. Hence, one obtains the belief function

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

Figure 2 illustrates the core of μ^m,

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

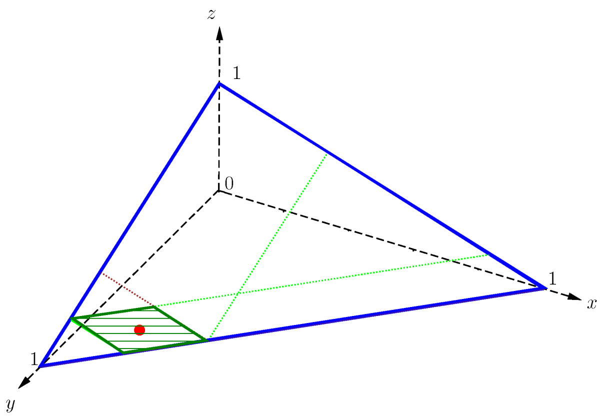
[17] It is also called the Shapley value or the Steiner point of $\text{core}(\mu^m)$.

In this case, the subsets $\{x\}, \{x, y\}, \{y, z\}$ are focal elements of μ^m. According to formula (11), the centroid of core(μ^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 Figure 2.

Figure 2: Core of the belief function μ^m and its centroid



In case of ambiguity neutrality, we have

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

as claimed in Proposition 6.

6.2 The smooth model

In economic applications, the smooth model of decision making under uncertainty, suggested and axiomatized by Klibanoff, Marinacci, and Mukerji (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,

$$V(\mu^m) = \sum_{A \subseteq X} m(A) \phi^{-1} \left(\frac{1}{|A|} \sum_{x \in A} \phi(u(x)) \right), \tag{12}$$

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, Marinacci, and Mukerji (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,

$$\hat{u}(x) = \phi(u(x))$$

for all $x \in X$. Then representation (12) becomes

$$V(\mu^m) = \sum_{A \subseteq X} m(A)\, \phi^{-1} \left(\frac{1}{|A|} \sum_{x \in A} \hat{u}(x) \right) = \mathbb{E}_\sigma \phi^{-1}(\mathbb{E}_\pi \hat{u}), \tag{13}$$

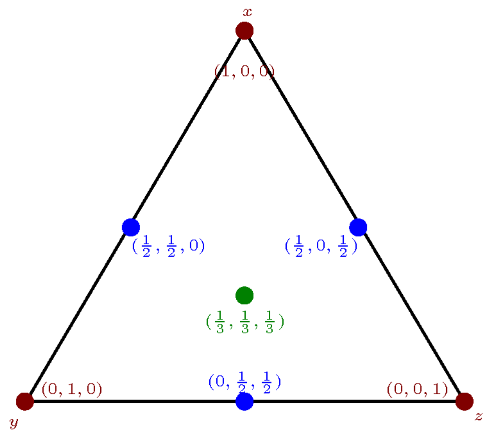
where $\mathbb{E}$ denotes the expectation operator, $\sigma(\bar{\pi}_A) = m(A)$ for all $A \subseteq X$ and $\bar{\pi}_A$ is the uniform distribution over the event (face) A of $\Delta(X)$. The last expression in (13) corresponds to the smooth model of Klibanoff, Marinacci, and Mukerji (2005). The following Figure 3 for $X = \{x, y, z\}$ illustrates this interpretation.

Notice that according to formula (11) the centroid p^* of $\text{core}(\mu^m)$ is the convex combination of the points in $\Delta(X)$ with weights given by the mass m.

7 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

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



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).[18] However, such a representation has not received enough attention in modern axiomatic models.

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

References

Anscombe, F. J., and R. J. Aumann (1963): "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, 34, 199–205. 5.1

Bernoulli, J. (1713): *Ars conjectandi*, vol. 107 of *Ostwalds Klassiker der exakten Wissenschaften, Band 107 (1999).* Wissenschaftlicher Verlag Harri Deutsch, Frankfurt am Main. 3

Chateauneuf, A., and J. Y. Jaffray (1989): "Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion," *Mathematical Social Sciences*, 17(3), 263–283. 2

Chateauneuf, A., and J.-C. Vergnaud (2000): "Ambiguity reduction through new statistical data," *International Journal of Approximate Reasoning*, 24, 283–299. 1

Ellsberg, D. (1961): "Risk, ambiguity and the Savage axioms," *Quarterly Journal of Economics*, 75(4), 643–669. 2.3, 3, 3, 7, 9, 5.1

Eva, B. (2019): "Principles of Indifference," *Journal of Philosophy*, forthcoming. 3

Friedman, M., and L. J. Savage (1948): "Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, 56(4), 279–304. 2.3

Gajdos, T., T. Hayashi, J.-M. Tallon, and J.-C. Vergnaud (2008): "Attitude toward imprecise information," *Journal of Economic Theory*, 14(1), 27–65. 1, 6, 6.1, 16

Gajdos, T., J.-M. Tallon, and J.-C. Vergnaud (2004): "Decision making with imprecise probabilistic information," *Journal of Mathematical Economics*, 40, 647–681. 1

——— (2008): "Representation and aggregation of preferences under uncertainty," *Journal of Economic Theory*, 141(1), 68–99. 1

Gilboa, I., and D. Schmeidler (1994): "Additive representations of non-additive measures and the Choquet integral," *Annals of Operations Research*, 52(1), 43–65. 3, 6.1

Giraud, R., and J.-M. Tallon (2011): "Are beliefs a matter of taste? A case for objective imprecise information," *Theory and Decision*, 71, 23–31. 1

Grabisch, M. (2016): *Set Functions, Games and Capacities in Decision Making*, vol. 46 of *Theory and Decision Library C*. Springer, game theory, social choice, decision theory, and optimization edn. 2

Hurwicz, L. (1951): "Optimiality Criteria for Decision Making under Ignorance," Discussion paper, Cowles Comission. 3

Jaffray, J. Y. (1989): "Linear utility theory for belief functions," *Operation Research Letters*, 8(2), 107–112. 3, 4, 8, 10

Ke, S. (2019): "Boundedly rational backward induction," *Theoretical Economics*, 14(1), 103–134. 11

Keynes, J. M. (1921): *A Treatise on Probability.* Dover Publications, New York, 2004 edn. 1, 3, 5, 7

Klibanoff, P., M. Marinacci, and S. Mukerji (2005): "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, 73(6), 1849–1892. 14, 6, 16, 6.2, 6.2, 6.2

Knight, F. H. (1921): *Risk, Uncertainty, and Profit.* Houghton Mifflin, New York. 5.1

Kolmogorov, A. N. (1930): "On the Notion of Mean," in *Selected Works of A.N. Kolmogorov: Volume I: Mathematics and Mechanics (1991)*, p. 144–146. Springer Science & Business Media. 6

Kops, C., and I. Pasichnichenko (2020): "A Test of Information Aversion," Discussion paper, University of Heidelberg, Alfred Weber Institut. 5.1, 9

Luce, R. D., and H. Raiffa (1957): *Games and Decisions. Introduction and Critical Survey.* Wiley, New York. 4

Machina, M. J., and M. Siniscalchi (2014): "Chapter 13 - Ambiguity and Ambiguity Aversion," in *Handbook of the Economics of Risk and Uncertainty*, ed. by M. Machina, and K. Viscusi, vol. 1 of *Handbook of the Economics of Risk and Uncertainty*, p. 729–807. North-Holland. 5.1, 7, 18

Matkowski, J., and Z. Páles (2015): "Characterization of Generalized Qasi-Arithmetic Means," *arXiv:1501.02857v2 [math.CA] 21 Jan 2015.* 6, 4, 7, 7

Savage, L. J. (1954): *Foundations of Statistics.* Wiley, New York. 1, 2, 3, 5.1

Shafer, G. (1976): *A Mathematical Theory of Evidence.* Princeton Univ. Press. 1, 2

Sinn, H.-W. (1980): "A Rehabilitation of the Principle of Insufficient Reason," *Quarterly Journal of Economics*, 94(3), 493–506. 3

von Neumann, J., and O. Morgenstern (1944): *The Theory of Games and Economic Behavior.* Princeton University Press, New Jersey. 1, 2, 3

Wald, A. (1955): "Basic ideas of a general theory of statistical decision rules," in *Selected Papers in Statistics and Probability*, ed. by A. Wald, p. 656–668, New York. McGraw-Hill. 3

Yaari, M. E. (1987): "The dual theory of choice under risk," *Econometrica: Journal of the Econometric Society*, p. 9–115. 14

Appendix: Proofs

The main Theorem 1 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 1

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 (9), we have to show that there exists a continuous strictly increasing function $\phi : U \to \mathbb{R}$ such that

$$V(e_A) = \phi^{-1}\left(\frac{1}{|A|}\sum_{x\in A}\phi(u(x))\right) \tag{14}$$

holds for any $A \in \bar{\mathcal{X}}$. To do this, we first prove (14) 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 $x^* \succcurlyeq x \succcurlyeq x_*$ for all $x \in A$, then $x^* \succcurlyeq A \succcurlyeq x_*$ 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

$$M\left(M(r,s), M(t,k)\right) = M\left(M(r,t), M(s,k)\right) \tag{15}$$

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 (15) 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 (15) 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

$$M(r,s) = \phi_0^{-1}\left(\frac{\phi_0(r)+\phi_0(s)}{2}\right) \tag{16}$$

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 (16). Thus, we proved (14) for two-element sets.

Now we extend (14) to an arbitrary $A \in \bar{\mathcal{X}}$. Suppose that (14) is true for all sets with no more than n elements, $n \geq 2$, and prove (14) 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 (14), i.e.

$$V\left(e_A\right) = \phi^{-1}\left(\frac{\phi\left(u\left(c_{\{x_1,\dots,x_k\}}\right)\right)+\phi\left(u\left(c_{\{x_{k+1},\dots,x_{2k}\}}\right)\right)}{2}\right). \tag{17}$$

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

$$\phi\left(u\left(c_{\{x_1,\dots,x_k\}}\right)\right) = \phi\left(V\left(e_{\{x_1,\dots,x_k\}}\right)\right) = \frac{1}{k}\sum_{i=1}^{k}\phi\left(u(x_i)\right) \tag{18}$$

and similarly

$$\phi\left(u\left(c_{\{x_{k+1},\dots,x_{2k}\}}\right)\right) = \frac{1}{k}\sum_{i=k+1}^{2k}\phi\left(u(x_i)\right). \tag{19}$$

Substituting the two terms in equation (17), by the right-hand sides of equations (18) and (19) 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 (14) 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 1

We provide our proof only for the case of ambiguity aversion. By Definition 6 and Theorem 1, if a decision maker is ambiguity averse, then $V\left(\bar{\ell}_{\{x_1,x_2\}}\right) \geq V\left(e_{\{x_1,x_2\}}\right)$ for any $x_1, x_2 \in X$, i.e.,

$$\frac{u(x_1)+u(x_2)}{2} \geq \phi^{-1}\left(\frac{\phi(u(x_1))+\phi(u(x_2))}{2}\right). \tag{20}$$

Since ϕ is a strictly increasing function, we can apply it to both sides of inequality (20). We know that ϕ is defined on an interval and continuous, and we just showed that for any r_1 and r_2 from the 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. The converse statement follows from Jensen's inequality.

Proposition 2

For the "if" part, assume w.l.o.g. $u_{\boldsymbol{i}} = u_{\boldsymbol{j}} = u$ and $\phi_{\boldsymbol{i}} = h \circ \phi_{\boldsymbol{j}}$. Then for any $A \in \bar{\mathcal{X}}$, we obtain

$$\frac{1}{|A|}\sum_{x\in A} h \circ \phi_{\boldsymbol{j}}(u(x)) \leq h\left(\frac{1}{|A|}\sum_{x\in A}\phi_{\boldsymbol{j}}(u(x))\right)$$

by Jensen's inequality. The inequality still holds when we apply a strictly increasing function $\phi_{\boldsymbol{i}}^{-1} = \phi_{\boldsymbol{j}}^{-1} \circ h^{-1}$, so that

$$\phi_{\boldsymbol{i}}^{-1}\left(\frac{1}{|A|}\sum_{x\in A}\phi_{\boldsymbol{i}}(u(x))\right) \leq \phi_{\boldsymbol{j}}^{-1}\left(\frac{1}{|A|}\sum_{x\in A}\phi_{\boldsymbol{j}}(u(x))\right),$$

which means $V_{\boldsymbol{i}}(e_A) \leq V_{\boldsymbol{j}}(e_A)$. Together with $V_{\boldsymbol{i}}(l) = V_{\boldsymbol{j}}(l)$, this allows us to conclude that $l \succ_{\boldsymbol{j}} e_A$ implies $l \succ_{\boldsymbol{i}} e_A$.

For the "only if" part, we need to prove that $\phi_{\boldsymbol{i}}$ and $\phi_{\boldsymbol{j}}$ are connected by a concave transformation. For any $A \in \bar{\mathcal{X}}$, there exists a lottery l on A such that $e_A \sim_{\boldsymbol{i}} l$, as was shown in the proof of Theorem 1. Thus, we have $e_A \succsim_{\boldsymbol{j}} l$ by Definition 7, which implies

$$V_{\boldsymbol{j}}(e_A) \geq V_{\boldsymbol{j}}(l) = V_{\boldsymbol{i}}(l) = V_{\boldsymbol{i}}(e_A).$$

Take a set A consisting of two outcomes x_1 and x_2. From $V_{\boldsymbol{j}}(e_A) \geq V_{\boldsymbol{i}}(e_A)$, it follows that

$$\phi_{\boldsymbol{j}}^{-1}\left(\frac{\phi_{\boldsymbol{j}}(u(x_1)) + \phi_{\boldsymbol{j}}(u(x_2))}{2}\right) \geq \phi_{\boldsymbol{i}}^{-1}\left(\frac{\phi_{\boldsymbol{i}}(u(x_1)) + \phi_{\boldsymbol{i}}(u(x_2))}{2}\right).$$

If we define $h = \phi_{\boldsymbol{i}} \circ \phi_{\boldsymbol{j}}^{-1}$, then the previous inequality implies

$$h\left(\frac{\phi_{\boldsymbol{j}}(u(x_1)) + \phi_{\boldsymbol{j}}(u(x_2))}{2}\right) \geq \frac{h(\phi_{\boldsymbol{j}}(u(x_1))) + h(\phi_{\boldsymbol{j}}(u(x_2)])}{2}.$$

Therefore, h is concave, because it is defined on an interval, continuous, and for any r_1 and r_2 from the interval, $h(\frac{1}{2}r_1 + \frac{1}{2}r_2) \geq \frac{1}{2}h(r_1) + \frac{1}{2}h(r_2)$.

Proposition 3

By Proposition 2, we only need to show that expression (10) is equivalent to $\phi_{\boldsymbol{i}} = h \circ \phi_{\boldsymbol{j}}$ for a strictly increasing and concave function h. The proof is standard. By differentiating equation $\phi_{\boldsymbol{i}}(r) = h(\phi_{\boldsymbol{j}}(r))$ twice and rearranging terms, we get

$$h''(\phi_{\boldsymbol{j}}(r)) = \frac{\phi_{\boldsymbol{i}}'(r)}{(\phi_{\boldsymbol{j}}'(r))^2}\left[\frac{\phi_{\boldsymbol{i}}''(r)}{\phi_{\boldsymbol{i}}'(r)} - \frac{\phi_{\boldsymbol{j}}''(r)}{\phi_{\boldsymbol{j}}'(r)}\right],$$

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

Proposition 4

From the assumptions and Proposition 2 it follows that $\boldsymbol{i}_1, \boldsymbol{i}_2, \ldots$ share the same von Neumann-Morgenstern utility function u and 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 (9), $V_n(\mu)$ is a convex combination of 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 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 3 and Assumption (ii) that $-\frac{\phi_n''}{\phi_n'}$ converges uniformly to $+\infty$.

Proposition 5

Since for any $A \subseteq X$,

$$\min_{x\in A} u(x) \leq \phi^{-1}\left(\frac{1}{|A|}\sum_{x\in A}\phi(u(x))\right) \leq \max_{x\in A} u(x), \tag{21}$$

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,

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

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 6

Let $A_1, \ldots, A_k$ be focal elements of μ^m and $n = |A_1| \cdots |A_k|$. For an extreme probability p_i, $i \in \{1, \ldots, n\}$, we have $p_i(x) > 0$ if and only if $p_i(x) = m(A)$ for some focal element $A \ni x$. For a fixed A, there are $\frac{n}{|A|}$ such extreme probabilities p_i. Therefore,

$$p^*(x) = \frac{1}{n} \sum_i p_i(x) = \frac{1}{n} \sum_{A \ni x:\, m(A) > 0} \frac{n \cdot m(A)}{|A|} = \sum_{A \ni x} \frac{m(A)}{|A|}.$$

On the other hand, if a decision maker is ambiguity averse, then ϕ is concave by Proposition 1, 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 other cases is similar.