

# Ambiguity Aversion, Robustness, and the Variational Representation of Preferences\*

Fabio Maccheroni<sup>a</sup> Massimo Marinacci<sup>b</sup> Aldo Rustichini<sup>c</sup>

<sup>a</sup>Istituto di Metodi Quantitativi and IGIER, Università Bocconi

<sup>b</sup>Dipartimento di Statistica e Matematica Applicata  
and Fondazione Collegio Carlo Alberto, Università di Torino

<sup>c</sup>Department of Economics, University of Minnesota

Revision of March 2006

## Abstract

We characterize in the Anscombe Aumann framework the preferences for which there are a *utility function*  $u$  on outcomes and an *ambiguity index*  $c$  on the set of probabilities on the states of the world such that, for all acts  $f$  and  $g$ ,

$$f \succsim g \Leftrightarrow \min_p \left( \int u(f) dp + c(p) \right) \geq \min_p \left( \int u(g) dp + c(p) \right).$$

The function  $u$  represents the decision maker's risk attitudes, while the index  $c$  captures his ambiguity attitudes.

These preferences include the *multiple priors preferences* of Gilboa and Schmeidler and the *multiplier preferences* of Hansen and Sargent. This provides a rigorous decision-theoretic foundation for the latter model, which has been widely used in macroeconomics and finance.

---

\*An extended version of this paper was previously circulated with the title "Variational Representation of Preferences under Ambiguity" ICER Working Paper 5/2004, March 2004. We thank Erio Castagnoli, Rose-Anne Dana, Larry Epstein, Peter Klibanoff, Bart Lipman, Mark Machina, Jianjun Miao, Sujoy Mukerji, Emre Ozdenoren, and, especially, Andy Postlewaite and four anonymous referees for helpful discussions and suggestions. We also thank several seminar audiences. Part of this research was done while the first two authors were visiting the Department of Economics of Boston University, which they thank for its hospitality. They also gratefully acknowledge the financial support of the Ministero dell'Istruzione, dell'Università e della Ricerca. Rustichini gratefully acknowledges the financial support of the National Science Foundation (grant # SES-0452477).

# 1 Introduction

Ambiguity is a classic issue in Decision Theory since the seminal work of Ellsberg [19]. The fundamental feature of ambiguity pointed out by Ellsberg is that there may be no belief on the states of the world that the decision maker holds and that rationalizes his choices.

A widely used class of preferences that model ambiguity are the *multiple priors preferences* axiomatized by Gilboa and Schmeidler [24] (also known as Maxmin Expected Utility preferences). Agents with these preferences rank payoff profiles  $f$  according to the following criterion

$$V(f) = \min_{p \in C} \int u(f) dp, \quad (1)$$

where  $C$  is a given convex subset of the set  $\Delta$  of all probabilities on states. The set  $C$  is interpreted as a set of priors held by agents, and ambiguity is reflected by the multiplicity of the priors.

In the past few years the possibility that agents may not hold a single belief on the states of the world has widely informed the research in macroeconomics and finance. In particular there has been a growing dissatisfaction in macroeconomics toward the strong uniformity on the agents' beliefs that is imposed by the rational expectations hypothesis. Under this hypothesis all agents share the same probability distribution on some relevant economic phenomenon and each agent has to be firmly convinced that the model he has adopted is the correct one. This is a strong requirement as agents can have different models, each of them being only an approximation of the underlying true model, and they may be aware of the possibility that their model is misspecified. A weakening of this requirement allows agents to entertain different priors on the economy.

A large part of the research has modeled ambiguity with multiple priors models (see, for instance, Epstein and Wang [21] and Chen and Epstein [9]). More recently, a different alternative has been explored, starting with the work of Hansen and Sargent (see, e.g., [26] and [27]), which in turn builds on earlier work in the field of robust control in the engineering and optimal control literature. In this robust preferences approach the objective functions of the agents take into account the possibility that their model  $q$  may not be the correct one, but only an approximation. Specifically, agents rank payoff profiles  $f$  according to the following choice criterion

$$V(f) = \min_{p \in \Delta} \left\{ \int u(f) dp + \theta R(p||q) \right\}, \quad (2)$$

where  $R(\cdot||q) : \Delta \rightarrow [0, \infty]$  is the relative entropy with respect to  $q$  (see Subsection 4.2 for the definition). Preferences represented by criterion (2) are called *multiplier*

preferences.<sup>1</sup>

Agents behaving according to this choice criterion are considering the possibility that  $q$  may not be the appropriate law governing the phenomenon which they are interested in, and for this reason they take into account other possible models  $p$ . The relative likelihood of these alternative models is measured by their relative entropy, while the positive parameter  $\theta$  reflects the weight that agents are giving to the possibility that  $q$  might not be the correct model. As the parameter  $\theta$  becomes larger agents focus more on  $q$  as the correct model, giving less importance to possible alternative models  $p$  (Proposition 22).

Hansen and Sargent [26] have pointed out that this model uncertainty can be viewed as the outcome of ambiguity, possibly resulting from the poor quality of the information on which agents base the choice of their model. For this reason the motivation behind multiplier preferences is closely connected to the one underlying multiple priors preferences. So far, however, this connection has been simply stated as intuitively appealing rather than established on formal grounds. In particular, no behavioral foundation of the preferences in (2) has been provided.

Here we establish this connection by presenting a general class of preferences with common behavioral features and that includes both multiplier and multiple prior preferences as special cases. The nature of the connection between the two main models that we have been discussing so far can be first established formally. The multiple priors criterion (1) can be written as follows

$$V(f) = \min_{p \in \Delta} \left\{ \int u(f) dp + \delta_C(p) \right\}, \quad (3)$$

where  $\delta_C : \Delta \rightarrow [0, \infty]$  is the indicator function of  $C$  (in the sense of Convex Analysis; see [44]) given by

$$\delta_C(p) = \begin{cases} 0 & \text{if } p \in C, \\ \infty & \text{otherwise.} \end{cases}$$

Like the relative entropy, also the indicator function is a convex function defined on the simplex  $\Delta$ .

This reformulation makes the formal connection with the multipliers preferences in (2) clear, and suggests the following general representation

$$V(f) = \min_{p \in \Delta} \left\{ \int u(f) dp + c(p) \right\}, \quad (4)$$

where  $c : \Delta \rightarrow [0, \infty]$  is a convex function on the simplex. In this paper we show that the connection is substantive and not formal, by establishing that the two models have a common behavioral foundation.

---

<sup>1</sup>Hansen and Sargent [27] also consider a class of multiple priors preferences with  $C = \{p \in \Delta : R(p||q) \leq \eta\}$ , which they call *constraint preferences*.

We first axiomatize (Theorem 3) the representation (4) by showing how it rests on a simple set of axioms that generalizes the multiple priors axiomatization of Gilboa and Schmeidler [24]. We then show (Proposition 8) how to interpret in a rigorous way the function  $c$  as an index of ambiguity aversion: the lower is  $c$ , the higher is the ambiguity aversion exhibited by the agent. The relative entropy  $\theta R(p||q)$  and the indicator function  $\delta_C(p)$  can thus be viewed as special instances of ambiguity indices. The assumptions on behavior for this general representation are surprisingly close to those given by Gilboa and Schmeidler in [24], and result from a simple weakening of their Certainty Independence axiom: so the similarity between the different representations has a sound behavioral foundation.

Once we have established this common structure, we can analyze the relationship between ambiguity aversion and probabilistic sophistication. For example, the multiplier preferences used by Hansen and Sargent are probabilistically sophisticated (and the same is true for their constraint preferences). Is this a fundamental property of the robust preferences, that make them profoundly different from multiple priors? In the common structure provided by the representation (4), the answer is simple and clear. The preferences in (4) are probabilistically sophisticated if and only if the ambiguity index  $c$  is symmetric. For multiple priors preferences, this property translates in a symmetry property of the set  $C$ . For multiplier preferences (and for constraint preferences), the condition is reflected by the symmetry of the relative entropy.

The property of being probabilistically sophisticated is therefore a property of the specific ambiguity index, not of the model. Being a symmetry property, this property is fragile: small perturbations destroy it and produce Ellsberg type behavior. Even if one adopts the view that such behavior is necessary for ambiguity aversion, the preferences in (4) are typically ambiguity averse (in the precise sense of Theorem 14 below).

Though our original motivation came from multiple priors and multiplier preferences, the class of preferences we axiomatize goes well beyond these two classes of preferences. In particular, multiplier preferences are a very special example of a new class of preferences, called divergence preferences, that we introduce and study in the paper. These preferences are able to accommodate Ellsberg-type behavior and, unlike multiple priors preferences, they are in general smooth (Proposition 23), an important feature for applications. This new class of preferences can provide a tractable alternative to multiple priors preferences in economic applications dealing with ambiguity, and the works of Hansen and Sargent can be viewed as an instance of this (see Theorem 16).

This claim is further substantiated by the observation that divergence preferences include as a special case a third classic class of preferences, the *mean-variance preferences* of Markowitz [40] and Tobin [49]. Recall that mean-variance preferences are

represented by the preference functional

$$V(f) = \int f dq - \frac{1}{2\theta} \text{Var}(f). \quad (5)$$

In Theorem 24 we show that, on the domain of monotonicity of  $V$ , the following equality holds:

$$\int f dq - \frac{1}{2\theta} \text{Var}(f) = \min_{p \in \Delta} \left\{ \int f dp + \theta G(p||q) \right\},$$

where  $G(\cdot||q) : \Delta \rightarrow [0, \infty]$  is the relative Gini concentration index (see Subsection 4.3 for the definition). As a result, the mean-variance preference functional (5) is a special case of our representation (4). Interestingly, the associated index of ambiguity aversion is the relative version of the classic Gini index. After Shannon's entropy, a second classic concentration index thus comes up in our analysis.

Summing up, in this paper we generalize a popular class of preferences dealing with ambiguity, the multiple priors preferences, and in this way we are able to introduce divergence preferences, a large new class of preferences under ambiguity that are in general smooth and that include as special cases two widely used classes of preferences, the multiplier preferences of Hansen and Sargent and the mean-variance preferences of Markowitz and Tobin. We are thus able to provide both a rigorous decision-theoretic foundation on two widely used classes of preferences, and a setting in which the two most classic concentration indices, Shannon's entropy and Gini's index, have a natural decision-theoretic interpretation. We finally characterize the preferences in this class that are probabilistically sophisticated, and show that this property coincides with a symmetry property of the cost function. This property is fragile, and so the preferences we characterize are typically not probabilistically sophisticated.

## 1.1 Ambiguity Aversion

Besides Schmeidler [48]'s original notion based on preference for randomization (see axiom A.5 in Section 3), there are two main notions of ambiguity aversion in the literature, proposed by Epstein [20] and Ghirardato and Marinacci [23]. The key difference in the two approaches lies in the different notion of ambiguity neutrality they use: while [23] identifies ambiguity neutrality with subjective expected utility, [20] more generally identifies ambiguity neutrality with probabilistic sophistication. In a nutshell, [23] claims that, unless the setting is rich enough, probabilistically sophisticated preferences may be compatible with behavior that intuitively can be viewed as generated by ambiguity. For this reason, they only consider subjective expected utility preferences as ambiguity neutral preferences.<sup>2</sup>

---

<sup>2</sup>We refer to [20] and [23] for a detailed presentation and motivation of their approaches. Notice that [23]'s notion of ambiguity aversion is what provides a foundation for the standard comparative

For the general class of preferences we axiomatize, the relative merits of these two notions of ambiguity aversion are the same as for the special case represented by multiple priors preferences. Though in the paper we adopt the view and terminology of [23], the appeal of our analysis does not depend on this choice; in particular, we expect that the ambiguity features of our preferences can be studied along the lines of [20], in the same way as it has been done in [20] for multiple priors preferences.

To clarify further this issue, in Subsection 3.5 we study the form that probabilistic sophistication takes in our setting. As we already mentioned, we show that our preferences are probabilistically sophisticated once their ambiguity indices satisfy a symmetry property. As a result, probabilistic sophistication is not peculiar to some particular specification of our preferences, but, to the contrary, all examples of our preferences will include special cases of probabilistically sophisticated ones. For instance, both multiplier and mean-variance preferences are easily seen to be examples of probabilistically sophisticated divergence preferences, though Example 17 shows that this is not the case for general divergence preferences, even for ones that are very close to multiplier and mean-variance preferences.

We close by briefly mentioning a possible alternative interpretation of our preferences. Consider an agent that has to make choices having only limited information, without a full understanding of what is going on. Some recent psychological literature (e.g., Keren and Gerritsen [30], Kühberger and Perner [31]) suggests that in this case the agent may behave as if he were playing against an informed opponent who might take advantage of this uncertainty and turn it against him.

This is a psychological attitude that can be relevant in many choice situations – including Ellsberg-type choice situations – and our preference functional (4) can be viewed as modelling such psychological treat. In fact, agents ranking payoff profiles according to (4) can be viewed as believing they playing a zero-sum game against (a malevolent) Nature. This view, however, has not been firmly established in the psychological and neuroscience literatures and it is the object of current research (see Hsu, Bhatt, Adolphs, Tranel and Camerer [29] and Rustichini [46]); for this reason, we do not expatiate further on this interpretation.

The paper is organized as follows. After introducing the set up in Section 2, we present the main representation result in Section 3. In the same section we discuss the ambiguity attitudes featured by the preferences we axiomatize, and we give conditions that make them probabilistically sophisticated. In Section 4 we study two important examples of our preferences, that is, the multiple priors preferences of Gilboa and

---

statics exercises in ambiguity for multiple priors preferences that are based on the size of the set of priors.

Schmeidler [24] and divergence preference, a new class of preferences that includes as special cases multiplier preferences and mean-variance preferences. Proofs and related material are collected in the Appendices.

## 2 Set Up

Consider a set  $S$  of *states of the world*, an algebra  $\Sigma$  of subsets of  $S$  called *events*, and a set  $X$  of *consequences*. We denote by  $\mathcal{F}$  the set of all the (*simple*) *acts*: finite-valued functions  $f : S \rightarrow X$  which are  $\Sigma$ -measurable. Moreover, we denote by  $B_0(\Sigma)$  the set of all real-valued  $\Sigma$ -measurable simple functions, so that  $u(f) \in B_0(\Sigma)$  whenever  $u : X \rightarrow \mathbb{R}$ , and by  $B(\Sigma)$  the supnorm closure of  $B_0(\Sigma)$ .

Given any  $x \in X$ , define  $x \in \mathcal{F}$  to be the constant act such that  $x(s) = x$  for all  $s \in S$ . With the usual slight abuse of notation, we thus identify  $X$  with the subset of the constant acts in  $\mathcal{F}$ . If  $f \in \mathcal{F}$ ,  $x \in X$ , and  $A \in \Sigma$ , we denote by  $xAf \in \mathcal{F}$  the act yielding  $x$  if  $s \in A$  and  $f(s)$  if  $s \notin A$ .

We assume additionally that  $X$  is a convex subset of a vector space. For instance, this is the case if  $X$  is the set of all the lotteries on a set of prizes, as it happens in the classic setting of Anscombe and Aumann [2]. Using the linear structure of  $X$  we can define as usual for every  $f, g \in \mathcal{F}$  and  $\alpha \in [0, 1]$  the act  $\alpha f + (1 - \alpha)g \in \mathcal{F}$ , which yields  $\alpha f(s) + (1 - \alpha)g(s) \in X$  for every  $s \in S$ .

We model the decision maker's *preferences* on  $\mathcal{F}$  by a binary relation  $\succsim$ . As usual,  $\succ$  and  $\sim$  denote respectively the asymmetric and symmetric parts of  $\succsim$ . If  $f \in \mathcal{F}$ , an element  $x_f \in X$  is a *certainty equivalent* for  $f$  if  $f \sim x_f$ .

## 3 Representation

### 3.1 Axioms

In the sequel we make use of the following properties of  $\succsim$ .

A.1 *Weak Order*. If  $f, g, h \in \mathcal{F}$ : (a) either  $f \succsim g$  or  $g \succsim f$ , (b)  $f \succsim g$  and  $g \succsim h$  imply  $f \succsim h$ .

A.2 *Weak Certainty-Independence*. If  $f, g \in \mathcal{F}$ ,  $x, y \in X$ , and  $\alpha \in (0, 1)$ ,

$$\alpha f + (1 - \alpha)x \succsim \alpha g + (1 - \alpha)x \Rightarrow \alpha f + (1 - \alpha)y \succsim \alpha g + (1 - \alpha)y.$$

A.3 *Continuity*. If  $f, g, h \in \mathcal{F}$ , the sets  $\{\alpha \in [0, 1] : \alpha f + (1 - \alpha)g \succsim h\}$  and  $\{\alpha \in [0, 1] : h \succsim \alpha f + (1 - \alpha)g\}$  are closed.

A.4 *Monotonicity.* If  $f, g \in \mathcal{F}$  and  $f(s) \succsim g(s)$  for all  $s \in S$ , then  $f \succsim g$ .

A.5 *Uncertainty Aversion.* If  $f, g \in \mathcal{F}$  and  $\alpha \in (0, 1)$ ,

$$f \sim g \Rightarrow \alpha f + (1 - \alpha) g \succsim f.$$

A.6 *Non-degeneracy.*  $f \succ g$  for some  $f, g \in \mathcal{F}$ .

Axioms A.1, A.3, A.4 and A.6 are standard assumptions. Axioms A.3 and A.6 are technical assumptions, while A.1 and A.4 require preferences to be complete, transitive, and monotone. The latter requirement is basically a state-independence condition, saying that decision makers always (weakly) prefer acts delivering statewise (weakly) better payoffs, regardless of the state where the better payoffs occur. If a preference relation  $\succsim$  satisfies A.1, A.3, and A.4, then each act  $f \in \mathcal{F}$  admits a certainty equivalent  $x_f \in X$ .<sup>3</sup>

Axiom A.5, due to Schmeidler [48], is a smoothing axiom that can be interpreted as an ambiguity aversion axiom, as discussed at length in [24], [48], [20], and [23].

Axiom A.2 is a weak independence axiom, a variation of an axiom of Gilboa and Schmeidler [24]. It requires independence only with respect to mixing with constant acts, provided the mixing weights are kept constant. A.2 is weaker than the original axiom of Gilboa and Schmeidler [24], and it is this weakening that makes it possible to go beyond the multiple priors model. Because of its importance for our derivation, we devote to A.2 the rest of this subsection.

Consider the following stronger version of A.2.

A.2' *Certainty Independence.* If  $f, g \in \mathcal{F}$ ,  $x \in X$ , and  $\alpha \in (0, 1)$ , then

$$f \succsim g \Leftrightarrow \alpha f + (1 - \alpha)x \succsim \alpha g + (1 - \alpha)x.$$

Axiom A.2' is the original axiom of [24]. The next lemma shows how it strenghtens A.2.

**Lemma 1** *A binary relation  $\succsim$  on  $\mathcal{F}$  satisfies A.2' if and only if, for all  $f, g \in \mathcal{F}$ ,  $x, y \in X$ , and  $\alpha, \beta \in (0, 1]$ , we have:*

$$\alpha f + (1 - \alpha)x \succsim \alpha g + (1 - \alpha)x \Rightarrow \beta f + (1 - \beta)y \succsim \beta g + (1 - \beta)y.$$

Axiom A.2 is therefore the special case of A.2' in which the mixing coefficients  $\alpha$  and  $\beta$  are required to be equal. At a conceptual level, Lemma 1 shows that [24]'s Certainty Independence Axiom actually involves two types of independence: independence

---

<sup>3</sup>See the proof of Lemma 28 in Appendix.



relative to mixing with constants and independence relative to the weights used in such mixing. Our Axiom A.2 retains the first form of independence, but not the second one. In other words, we allow for preference reversals in mixing with constants unless the weights themselves are kept constant.

This is a significant weakening of the Certainty Independence Axiom and its motivation is best seen when the weights  $\alpha$  and  $\beta$  are very different, say  $\alpha$  is close to 1 and  $\beta$  is close to 0. Intuitively, acts  $\alpha f + (1 - \alpha)x$  and  $\alpha g + (1 - \alpha)x$  can then involve far more uncertainty than acts  $\beta f + (1 - \beta)y$  and  $\beta g + (1 - \beta)y$ , which are almost constant acts. As a result, we expect that, at least in some situations, the ranking between the genuinely uncertain acts  $\alpha f + (1 - \alpha)x$  and  $\alpha g + (1 - \alpha)x$  can well differ from that between the almost constant acts  $\beta f + (1 - \beta)y$  and  $\beta g + (1 - \beta)y$ .

Needless to say, even though we believe that such reversals can well occur (from both a positive and normative standpoint), the only way to test them, and so the plausibility of A.2 and A.2', is by running experiments. This is possible since both A.2 and A.2' have clear behavioral implications. For instance, the following (thought) experiment gives a simple testable way to compare A.2 and A.2', and running this type of experiments will be the subject of future research.

**Example 2** Consider an urn containing 90 black and white balls in unknown proportion, and the following bets (payoffs are in dollars):

$t > 0$	Black	White
$f_t$	$t$	$t$
$g_t$	$3t$	$0.01t$

that is,  $f_t$  pays  $t$  dollars whatever happens, while  $g_t$  pays  $3t$  dollars if a black ball is drawn and  $t$  cents otherwise. For example:

10	Black	White		$10^4$	Black	White
$f_{10}$	10	10	and	$f_{10^4}$	10,000	10,000
$g_{10}$	30	0.1		$g_{10^4}$	30,000	100

Assume the decision maker's preferences satisfy A.1, A.3-A.6, and she displays constant relative risk aversion  $\gamma \in (0, 1)$ .<sup>4</sup> If her preferences satisfy A.2', then

$$\text{either } f_t \succsim g_t \text{ for all } t \text{ or } g_t \succsim f_t \text{ for all } t.$$

If, in contrast, her preferences only satisfy A.2 and not A.2', there might exist a threshold  $\bar{t}$  such that

$$f_t \succsim g_t \text{ for all } t \geq \bar{t} \text{ and } g_t \succsim f_t \text{ for all } t \leq \bar{t}.$$

This reversal is compatible with A.2, but it would reveal a violation of A.2'. ▲

---

<sup>4</sup>Notice that, for all  $t > \tau > 0$  and  $\gamma$ , there exist  $f, g, x, y, \alpha, \beta$  as discussed above such that  $\alpha f + (1 - \alpha)x \sim f_t$ ,  $\alpha g + (1 - \alpha)x \sim g_t$ ,  $\beta f + (1 - \beta)y \sim f_\tau$ ,  $\beta g + (1 - \beta)y \sim g_\tau$ .

We close by observing that, in terms of preference functionals, by Theorem 3 and Proposition 19 all preference functionals (4) satisfy A.2 and violate A.2', unless they reduce to the multiple priors form (1).

### 3.2 Main Result

We can now state our main result, which characterizes preferences satisfying axioms A.1-A.6. Here  $\Delta = \Delta(\Sigma)$  denotes the set of all finitely additive probabilities on  $\Sigma$  endowed with the weak\* topology,<sup>5</sup> and  $c : \Delta \rightarrow [0, \infty]$  is said to be *grounded* if its infimum value is zero.

**Theorem 3** *Let  $\succsim$  be a binary relation on  $\mathcal{F}$ . The following conditions are equivalent:*

- (i)  $\succsim$  satisfies conditions A.1-A.6;
- (ii) there exists a non-constant affine function  $u : X \rightarrow \mathbb{R}$  and a grounded, convex, and lower semicontinuous function  $c : \Delta \rightarrow [0, \infty]$  such that, for all  $f, g \in \mathcal{F}$

$$f \succsim g \Leftrightarrow \min_{p \in \Delta} \left( \int u(f) dp + c(p) \right) \geq \min_{p \in \Delta} \left( \int u(g) dp + c(p) \right). \quad (6)$$

For each  $u$  there is a (unique) minimal  $c^* : \Delta \rightarrow [0, \infty]$  satisfying (6), given by

$$c^*(p) = \sup_{f \in \mathcal{F}} \left( u(x_f) - \int u(f) dp \right). \quad (7)$$

The representation (6) involves the minimization of a convex lower semicontinuous functional, which is the most classic variational problem. This motivates the following definition.

**Definition 4** *A preference  $\succsim$  on  $\mathcal{F}$  is called variational if it satisfies axioms A.1-A.6.*

By Theorem 3, variational preferences can be represented by a pair  $(u, c^*)$ . From now on, when we consider a variational preference we will write  $u$  and  $c^*$  to denote the elements of such a pair. Next we give the uniqueness properties of this representation.

**Corollary 5** *Two pairs  $(u_0, c_0^*)$  and  $(u, c^*)$  represent the same variational preference  $\succsim$  as in Theorem 3 if and only if there exist  $\alpha > 0$  and  $\beta \in \mathbb{R}$  such that  $u = \alpha u_0 + \beta$  and  $c^* = \alpha c_0^*$ .*

---

<sup>5</sup>That is, the  $\sigma(\Delta(\Sigma), B_0(\Sigma))$ -topology where a net  $\{p_d\}_{d \in D}$  converges to  $p$  if and only if  $p_d(A) \rightarrow p(A)$  for all  $A \in \Sigma$ .

In Theorem 3 we saw that  $c^*$  is the minimal non-negative function on  $\Delta$  for which the representation (6) holds. More is true when  $u(X) = \{u(x) : x \in X\}$  is unbounded (either below or above):

**Proposition 6** *Let  $\succsim$  be a variational preference with  $u(X)$  unbounded. Then, the function  $c^*$  defined in (7) is the unique non-negative, grounded, convex, and lower semicontinuous function on  $\Delta$  for which (6) holds.*

As shown in Lemma 29 in the Appendix, the assumption that  $u(X)$  is unbounded (above or below) is equivalent to the following axiom (see [32]).

A.7 *Unboundedness.* There exist  $x \succ y$  in  $X$  such that for all  $\alpha \in (0, 1)$  there exists  $z \in X$  satisfying either  $y \succ \alpha z + (1 - \alpha)x$  or  $\alpha z + (1 - \alpha)y \succ x$ .

We call *unbounded* the variational preferences satisfying axiom A.7.

### 3.3 Ambiguity Attitudes

We now study the ambiguity attitudes featured by variational preferences. We follow the approach proposed in [23], to which we refer for a detailed discussion of the notions we use.

Begin with a comparative notion: given two preferences  $\succsim_1$  and  $\succsim_2$ , say that  $\succsim_1$  is *more ambiguity averse than*  $\succsim_2$  if, for all  $f \in \mathcal{F}$  and  $x \in X$ ,

$$f \succsim_1 x \Rightarrow f \succsim_2 x. \quad (8)$$

To introduce an absolute notion of ambiguity aversion, as in [23] we consider Subjective Expected Utility (SEU) preferences as benchmarks for ambiguity neutrality. We then say that a preference relation  $\succsim$  is *ambiguity averse* if it is more ambiguity averse than some SEU preference.

We now apply these notions to our setting. The first thing to observe is that variational preferences are always ambiguity averse.

**Proposition 7** *Each variational preference is ambiguity averse.*

As variational preferences satisfy axiom A.5, and the choice rule resulting from (6) is a maxmin rule, intuitively it is not surprising that variational preferences always display a negative attitude toward ambiguity. Proposition 7 makes precise this intuition.

Next we show that comparative ambiguity attitudes for variational preferences are determined by the function  $c^*$ . Here  $u_1 \approx u_2$  means that there exist  $\alpha > 0$  and  $\beta \in \mathbb{R}$  such that  $u_1 = \alpha u_2 + \beta$ .

**Proposition 8** Given two variational preferences  $\succsim_1$  and  $\succsim_2$ , the following conditions are equivalent:

- (i)  $\succsim_1$  is more ambiguity averse than  $\succsim_2$ ,
- (ii)  $u_1 \approx u_2$  and  $c_1^* \leq c_2^*$  (provided  $u_1 = u_2$ ).

Given that  $u_1 \approx u_2$ , the assumption  $u_1 = u_2$  is just a common normalization of the two utility indices. Therefore, Proposition 8 says that more ambiguity averse preference relations are characterized, up to a normalization, by smaller functions  $c^*$ . Therefore, the function  $c^*$  can be interpreted as an *index of ambiguity aversion*.

We now give few simple examples illustrating this interpretation of the function  $c^*$ .

**Example 9** By Proposition 8, maximal ambiguity aversion is characterized by  $c^*(p) = 0$  for each  $p \in \Delta$ . In this case, (6) becomes:

$$f \succsim g \Leftrightarrow \min_{p \in \Delta} \int u(f) dp \geq \min_{p \in \Delta} \int u(g) dp,$$

that is,

$$f \succsim g \Leftrightarrow \min_{s \in S} u(f(s)) \geq \min_{s \in S} u(g(s)),$$

a form that clearly reflects extreme ambiguity aversion. ▲

**Example 10** Minimal ambiguity aversion corresponds here to ambiguity neutrality as, by Proposition 7, all variational preferences are ambiguity averse. Therefore, the least ambiguity averse functions  $c^*$  are those associated with SEU preferences. As it will be shown in Corollary 20, when a SEU preference is unbounded,  $c^*$  takes the stark form

$$c^*(p) = \begin{cases} 0 & \text{if } p = q, \\ \infty & \text{otherwise,} \end{cases}$$

where  $q$  is the subjective probability associated with the preference. ▲

**Example 11** Denote by  $c_q^*$  the ambiguity neutral index of Example 10, and by  $c_m^*$  the maximal ambiguity index of Example 9. Given  $\alpha \in (0, 1)$ , suppose the r.h.s. of (6) is:

$$(1 - \alpha) \int u(f) dq + \alpha \min_{s \in S} u(f(s)) \geq (1 - \alpha) \int u(g) dq + \alpha \min_{s \in S} u(g(s)),$$

which is the well-known  $\varepsilon$ -contaminated model. In this case,

$$c^*(p) = \min_{p_1, p_2 \in \Delta} \{(1 - \alpha) c_q^*(p_2) + \alpha c_m^*(p_1) : (1 - \alpha) p_2 + \alpha p_1 = p\} = \delta_{(1-\alpha)q+\alpha\Delta}(p),$$

and this is a simple example of an index  $c^*$  not displaying extreme ambiguity attitudes. ▲

According to Proposition 8, variational preferences become more and more (less and less, resp.) ambiguity averse as their ambiguity indices become smaller and smaller (larger and larger, resp.). It is then natural to wonder what happens at the limit, when they go to 0 or to  $\infty$ . The next result answers this question. In reading it recall that  $c_{n+1} \leq c_n$  implies  $\text{dom } c_n \subseteq \text{dom } c_{n+1}$  and  $\arg \min c_n \subseteq \arg \min c_{n+1}$ .<sup>6</sup>

**Proposition 12** (i) *If  $c_n \downarrow$  and  $\lim_n c_n(p) = 0$  whenever this limit is finite, then*

$$\lim_{n \rightarrow \infty} \min_{p \in \Delta} \left\{ \int u(f) dp + c_n(p) \right\} = \min_{p \in \overline{\bigcup_n \text{dom } c_n}} \int u(f) dp \quad (9)$$

for all  $f \in \mathcal{F}$ .

(i) *If  $c_n \uparrow$  and  $\lim_n c_n(p) = \infty$  whenever this limit is not 0, then*

$$\lim_{n \rightarrow \infty} \min_{p \in \Delta} \left\{ \int u(f) dp + c_n(p) \right\} = \min_{p \in \bigcap_n \arg \min c_n} \int u(f) dp \quad (10)$$

for all  $f \in \mathcal{F}$ .

Proposition 12 shows that the limit behavior of variational preferences is described by multiple priors preferences. But, the size of the sets of priors they feature is very different. In fact, in (9) the relevant set of priors is given by  $\overline{\bigcup_n \text{dom } c_n}$ , while in (10) is given by the much smaller set  $\bigcap_n \arg \min c_n$ . For example, Proposition 22 will show that for an important class of variational preferences the set  $\bigcap_n \arg \min c_n$  is just a singleton, so that the limit preference in (10) is actually a SEU preference, while the set  $\overline{\bigcup_n \text{dom } c_n}$  is very large.

We close with few remarks. First, observe that Lemma 32 in the Appendix shows that the set that [23] calls benchmark measures – those probabilities that correspond to SEU preferences less ambiguity averse than  $\succsim$  – is given here by  $\arg \min c^* = \{p \in \Delta : c^*(p) = 0\}$ .

Second, notice that by standard convex analysis results (see [44]), Example 11 can be immediately generalized as follows: the ambiguity index of a convex combination of preference functionals representing unbounded variational preferences is given by the infimal convolution of their ambiguity indices.

Finally, Proposition 12 is an example of a result about the limit behavior of sequences of variational preferences

$$V_n(f) = \min_{p \in \Delta} \left( \int u_n(f) dp + c_n(p) \right),$$

---

<sup>6</sup>  $\text{dom } c$  denotes the effective domain  $\{c < \infty\}$  of  $c$ , while  $\arg \min c_n = \{p \in \Delta : c_n(p) = 0\}$  because  $c_n$  is grounded. Observe that  $\text{dom } c_n$  represents the set of all probabilities that the decision maker considers relevant when ranking acts using the ambiguity index  $c_n$ , while the smaller set  $\arg \min c_n$  contains only the probabilities that are getting the highest weight by this decision maker.

a type of results that can be important in some applications in order to deal with stability issues. These limit results involve the convergence of minima, a classic problem in variational analysis. A noteworthy feature of variational preferences is that the lower semicontinuity of the functions  $\int u_n(f) dp + c_n(p)$  on  $\Delta$  makes it possible to use the powerful de Giorgi-Wijsman theory of  $\Gamma$ -convergence (often called epi-convergence; see, e.g., Dal Maso [12]) to study the behavior of the sequence  $\{V_n(f)\}_n$ , and, for example, to determine under what conditions it converges to some preference functional  $V(f) = \min_{p \in \Delta} (\int u(f) dp + c(p))$  for suitable limit functions  $u$  and  $c$ .

### 3.4 An Extension: Countable Additivity

In Theorem 3 we considered the set  $\Delta$  of all finitely additive probabilities. In applications, however, it is often important to consider countably additive probabilities, which have very convenient analytical properties. For example, we will see momentarily that this is the case for divergence preferences, and so for the multiplier preferences of Hansen and Sargent [27] and for mean-variance preferences of Markowitz [40] and Tobin [49]. For this reason here we consider this technical extension.

Fortunately, in our setting we can still use the Monotone Continuity axiom introduced by Arrow [3] in order to derive a SEU representation with a countably additive subjective probability (see [8]).

**A.8 Monotone Continuity.** If  $f, g \in \mathcal{F}$ ,  $x \in X$ ,  $\{E_n\}_{n \geq 1} \in \Sigma$  with  $E_1 \supseteq E_2 \supseteq \dots$  and  $\bigcap_{n \geq 1} E_n = \emptyset$ , then  $f \succ g$  implies that there exists  $n_0 \geq 1$  such that

$$xE_{n_0}f \succ g.$$

Next we state the countably additive version of Theorem 3. Here  $\Delta^\sigma = \Delta^\sigma(\Sigma)$  denotes the set of all countably additive probabilities defined on a  $\sigma$ -algebra  $\Sigma$ , while  $\Delta^\sigma(q)$  denotes the subset of  $\Delta^\sigma$  consisting of all probabilities that are absolutely continuous with respect to  $q$ ; i.e.,  $\Delta^\sigma(q) = \{p \in \Delta^\sigma : p \ll q\}$ .

**Theorem 13** *Let  $\succsim$  be an unbounded variational preference. The following conditions are equivalent:*

(i)  $\succsim$  satisfies A.8;

(ii)  $\{p \in \Delta : c^*(p) \leq t\}$  is a (weakly) compact subset of  $\Delta^\sigma$  for each  $t \geq 0$ .

*In this case, there exists  $q \in \Delta^\sigma$  such that, for all  $f, g \in \mathcal{F}$ ,*

$$f \succsim g \Leftrightarrow \min_{p \in \Delta^\sigma(q)} \left( \int u(f) dp + c^*(p) \right) \geq \min_{p \in \Delta^\sigma(q)} \left( \int u(g) dp + c^*(p) \right). \quad (11)$$

Lemma 30 of the Appendix shows that even when the preference is not unbounded, axiom A.8 still implies the countable additivity of the probabilities involved in the representation.

In view of these results, we call *continuous* the variational preferences satisfying axiom A.8.

### 3.5 Probabilistic Sophistication

In this section we characterize variational preferences that are probabilistically sophisticated, an important property of preferences introduced by Machina and Schmeidler [39] that some authors (notably Epstein [20]) identify with ambiguity neutrality (or absence of ambiguity altogether).

Our main finding is that what makes a variational preference probabilistic sophisticated is a symmetry property of the ambiguity index. As a result, all classes of variational preferences (for example, multiple priors and divergence preferences) contain a subclass of probabilistically sophisticated ones, characterized by a suitable symmetry property of the associated ambiguity indices.

Specifically, given a countably additive probability  $q$  on the  $\sigma$ -algebra  $\Sigma$ , a preference relation  $\succsim$  is *rearrangement invariant* (w.r.t.  $q$ ) if, given  $f$  and  $g$  in  $\mathcal{F}$ ,

$$q(s \in S : f(s) = x) = q(s \in S : g(s) = x) \text{ for all } x \in X \Rightarrow f \sim g,$$

and it is *probabilistically sophisticated* (w.r.t.  $q$ ) if it is rearrangement invariant and satisfies first order stochastic dominance; that is, given  $f$  and  $g$  in  $\mathcal{F}$ ,

$$q(s \in S : f(s) \lesssim x) \leq q(s \in S : g(s) \lesssim x) \text{ for all } x \in X \Rightarrow f \succsim g.$$

To characterize probabilistic sophistication we need to introduce few well-known notions from the theory of stochastic orders (see, e.g., Chong and Rice [10] and Schmeidler [47]). Define a partial order  $\succ_{cx}$  on  $\Delta^\sigma(q)$  by

$$p \succ_{cx} p' \text{ iff } \int \phi \left( \frac{dp}{dq} \right) dq \geq \int \phi \left( \frac{dp'}{dq} \right) dq$$

for every convex function  $\phi$  on  $\mathbb{R}$ . This is the so-called *convex order* on probability distributions, and by classic results of Rothschild and Stiglitz [45] and Marshall and Olkin [42] we have  $p \succ_{cx} p'$  when the “masses”  $dp(s)$  are more scattered with respect to  $dq(s)$  than the masses  $dp'(s)$ .

An important property of the convex order  $\succ_{cx}$  is that its symmetric part  $\sim_{cx}$  coincides with the identical distribution of the densities w.r.t.  $q$ . That is, given  $p$  and  $p'$  in  $\Delta^\sigma(q)$ ,

$$p \sim_{cx} p' \text{ iff } q \left( s \in S : \frac{dp}{dq}(s) \leq t \right) = q \left( s \in S : \frac{dp'}{dq}(s) \leq t \right) \quad \forall t \in \mathbb{R}. \quad (12)$$

For  $p$  in  $\Delta^\sigma(q)$ , the set  $O(p) = \{p' \in \Delta^\sigma(q) : p' \sim_{cx} p\}$  is called the *orbit* of  $p$ .

A function  $c : \Delta \rightarrow [0, \infty]$  is *rearrangement invariant* (w.r.t.  $q$ ) if  $\text{dom } c \subseteq \Delta^\sigma(q)$  and, given  $p$  and  $p'$  in  $\Delta^\sigma(q)$ ,

$$p \sim_{cx} p' \Rightarrow c(p) = c(p'),$$

while it is *Shur convex* (w.r.t.  $q$ ) if  $\text{dom } c \subseteq \Delta^\sigma(q)$  and, given  $p$  and  $p'$  in  $\Delta^\sigma(q)$ ,

$$p \succ_{cx} p' \Rightarrow c(p) \geq c(p'). \quad (13)$$

A Shur convex function is clearly rearrangement invariant, while the converse is in general false. A subset  $C$  of  $\Delta^\sigma(q)$  is *Shur convex* (w.r.t.  $q$ ) if its indicator function  $\delta_C$  is Shur convex.<sup>7</sup> Finally, we say that  $q$  in  $\Delta^\sigma$  is *adequate* if either  $q$  is non-atomic or  $S$  is finite and  $q$  is uniform.

We can now state our characterization result, which shows that rearrangement invariance is the symmetry property of ambiguity indices characterizing probabilistically sophisticated preferences.

**Theorem 14** *Let  $\succsim$  be a continuous unbounded variational preference. If  $q \in \Delta^\sigma$  is adequate, then the following conditions are equivalent (w.r.t.  $q$ ):*

- (i)  $\succsim$  is probabilistically sophisticated;
- (ii)  $\succsim$  is rearrangement invariant;
- (iii)  $c^*$  is rearrangement invariant;
- (iv)  $c^*$  is Shur convex.

Moreover, (iv) implies (i) for any variational preference even if  $q$  is not adequate.

The proof of Theorem 14 builds on the results of Luxemburg [34] and Chong and Rice [10], as well as on some recent elaborations of these results provided by Dana [13].

As a first straightforward application of Theorem 14, observe that if a set of priors is Shur convex, then the corresponding (continuous) multiple priors preferences are probabilistically sophisticated, and the converse is true in the adequate case. More is true in this case: multiple priors preferences are probabilistically sophisticated if and only if the indicator functions  $\delta_C$  of their sets of priors  $C$  are rearrangement invariant, that is, if and only if the sets  $C$  are orbit-closed ( $O(p) \subseteq C$  for every  $p \in C$ ).<sup>8</sup>

<sup>7</sup>That is,  $\{p' \in \Delta^\sigma(q) : p' \prec_{cx} p\} \subseteq C$  for every  $p \in C$ .

<sup>8</sup>This fact can be used to provide an alternative derivation of some of the results of Marinacci [41], which showed that multiple priors preferences that are probabilistically sophisticated reduce to Subjective Expected Utility when there exists even a single non-trivial unambiguous event. The direct proofs in [41] are, however, shorter and more insightful for the problem which that article was dealing with.



To further illustrate Theorem 14, we now introduce a new class of variational preferences that play an important role in the rest of the paper. As before, assume there is an underlying probability measure  $q \in \Delta^\sigma$ . Given a  $\Sigma$ -measurable function  $w : S \rightarrow \mathbb{R}$  with  $\inf_{s \in S} w(s) > 0$  and  $\int w dq = 1$ , and a convex continuous function  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that  $\phi(1) = 0$  and  $\lim_{t \rightarrow \infty} \phi(t)/t = \infty$ , the  $w$ -weighted  $\phi$ -divergence of  $p \in \Delta$  w.r.t.  $q$  is given by

$$D_\phi^w(p||q) = \begin{cases} \int_S w(s) \phi\left(\frac{dp}{dq}(s)\right) dq(s) & \text{if } p \in \Delta^\sigma(q), \\ \infty & \text{otherwise,} \end{cases} \quad (14)$$

Here  $w$  is the (normalized) weighting function, and in the special case of uniform weighting  $w(s) = 1$  for all  $s \in S$ , we just write  $D_\phi(p||q)$  and we get back to standard divergences, which are a widely used concept of “distance between distributions” in statistics and information theory (see, e.g., Liese and Vajda [33]). The two most important divergences are the *relative entropy* (or *Kullbak-Leibler divergence*) given by  $\phi(t) = t \ln t - t + 1$ , and the *relative Gini concentration index* (or  $\chi^2$ -divergence) given by  $\phi(t) = 2^{-1}(t - 1)^2$ .

The next lemma collects the most important properties of weighted divergences.

**Lemma 15** *A weighted divergence  $D_\phi^w(\cdot||q) : \Delta \rightarrow [0, \infty]$  is a grounded, convex, and lower semicontinuous function, and the sets*

$$\{p \in \Delta : D_\phi^w(p||q) \leq t\} \quad (15)$$

*are (weakly) compact subsets of  $\Delta^\sigma(q)$  for all  $t \in \mathbb{R}$ . Moreover,  $D_\phi(\cdot||q) : \Delta \rightarrow \mathbb{R}$  is Shur convex whenever  $w$  is uniform.*

Thanks to the above properties, preferences represented by the functional

$$V(f) = \min_{p \in \Delta^\sigma(q)} \left\{ \int u(f) dp + \theta D_\phi^w(p||q) \right\}, \quad (16)$$

where  $\theta > 0$  and  $u : X \rightarrow \mathbb{R}$  is an affine function belong to the class of variational preferences. In view of their importance, we call them *divergence preferences*; that is,  $\succsim$  on  $\mathcal{F}$  is a divergence preference if:

$$f \succsim g \Leftrightarrow \min_{p \in \Delta^\sigma(q)} \left\{ \int u(f) dp + \theta D_\phi^w(p||q) \right\} \geq \min_{p \in \Delta^\sigma(q)} \left\{ \int u(g) dp + \theta D_\phi^w(p||q) \right\}.$$

**Theorem 16** *Suppose  $u(X)$  is unbounded. Then, divergence preferences are continuous variational preferences, with index of ambiguity aversion given by*

$$c^*(p) = \theta D_\phi^w(p||q), \quad \forall p \in \Delta.$$

*In particular, these preferences are probabilistically sophisticated whenever  $w$  is uniform.*

Divergence preferences are a very important class of variational preferences, and in the next section we further study them. Unlike multiple priors preferences, they are in general smooth (see Proposition 23), a noteworthy feature for applications.

For a finite state space  $S$ ,  $X = \mathbb{R}$ ,  $u(t) = t$ , and  $w$  uniform, some classes of divergence preferences have been considered by Ben-Tal [5] and Ben-Tal, Ben-Israel, and Teboulle [6]. In the next section we will study two significant examples of divergence preferences, corresponding to the relative entropy and to the relative Gini concentration index. Here it is important to observe that by Theorem 16 all divergence preferences represented by

$$V(f) = \min_{p \in \Delta^\sigma(q)} \left\{ \int u(f) dp + \theta D_\phi(p||q) \right\},$$

are examples of probabilistically sophisticated variational preferences.<sup>9</sup>

However, the next Example shows that, even under minimal non uniformities of the weighting function, divergence preferences are in general not probabilistically sophisticated and they exhibit Ellsberg-type behavior. Therefore, the probabilistic sophistication of divergence preferences crucially depends on the uniformity of the weight  $w$ .

**Example 17** Consider a standard Ellsberg three colors urn, with 30 red balls and 60 balls either green or blue. As usual, consider the following bets:

	Red	Green	Blue
$f_R$	1	0	0
$f_G$	0	1	0
$f_{R \cup B}$	1	0	1
$f_{G \cup B}$	0	1	1

where  $f_R$  pays 1 dollar if a red ball is drawn and nothing otherwise,  $f_G$  pays 1 dollar if a green ball is drawn and nothing otherwise, and so on. As well-known, Ellsberg [19] argued that most subjects rank these acts as follows:

$$f_R \succ f_G \quad \text{and} \quad f_{R \cup B} \prec f_{G \cup B}. \quad (17)$$

Consider a decision maker having divergence preferences represented by the preference functional  $V$  given by (16). Here it is natural to consider a uniform  $q$  on the three states. W.l.o.g., set  $u(0) = 0$  and  $u(1) = 1$ . By Theorem 16, when  $w$  is uniform it cannot be the case that  $V(f_R) > V(f_G)$  and  $V(f_{R \cup B}) < V(f_{G \cup B})$ . However, consider the weighting function  $w : \{R, G, B\} \rightarrow \mathbb{R}$  given by

$$w(R) = 1.01, \quad w(G) = 0.99, \quad \text{and} \quad w(B) = 1,$$

---

<sup>9</sup>Analogously, Theorem 14 and Lemma 15 show that all multiple priors preferences with sets of priors  $\{p \in \Delta^\sigma(q) : D_\phi(p||q) \leq \eta\}$  are probabilistically sophisticated. These preferences include the constraint preferences of Hansen and Sargent [27], mentioned in footnote 1.

which is only slightly non uniform. If we set  $\theta = 1$  and take either the weighted relative entropy  $\phi(t) = t \ln t - t + 1$  or the weighted Gini relative index  $\phi(t) = 2^{-1}(t - 1)^2$ , then some simple computations (available upon request) show that

$$V(f_R) > V(f_G) \quad \text{and} \quad V(f_{R \cup B}) < V(f_{G \cup B}),$$

thus delivering the Ellsberg pattern (17). ▲

### 3.6 Smoothness

Most economic models are based on the optimization of an objective function. When this function is differentiable solving the optimization problem is easier, and the solution has appealing properties. There is a well established set of techniques (first order necessary conditions, envelope theorems, implicit function theorems, and so on) that are extremely useful in both finding a solution of an optimization problem and in characterizing its properties. For example, they make it possible to carry out comparative statics exercises, a key feature in most economic models.

In view of all this, it is important to study whether our variational preference functionals are differentiable. In this section (see Theorem 18) we fully characterize the differentiability properties of variational preference functionals, and show that they are adequate for economic applications.

Throughout this section we assume that  $X$  is the set of all monetary lotteries, that is, the set of all finitely supported probability measures on  $\mathbb{R}$ . An act  $f$  of  $\mathcal{F}$  is *monetary* if  $f(s)$  is a degenerate lottery for every  $s \in S$ , that is, if  $f(s) \in \mathbb{R}$  (with the usual identification of  $z \in \mathbb{R}$  with the degenerate lottery  $d_z \in X$ ). The set of all monetary acts can thus be identified with  $B_0(\Sigma)$ .

We consider a variational preference functional

$$V(f) = \min_{p \in \Delta} \left\{ \int u(f) dp + c^*(p) \right\}$$

restricted to  $B_0(\Sigma)$ , that is, restricted to monetary acts. We also make the standard assumption that the associated utility function  $u$  is concave (thus reflecting risk aversion), strictly increasing, and differentiable on  $\mathbb{R}$ .

In order to state our results we need some standard notions of calculus in vector spaces (see [44] and [43]). Given  $f \in B_0(\Sigma)$ , the *directional derivative* of  $V : B_0(\Sigma) \rightarrow \mathbb{R}$  at  $f$  is the functional  $V'(f; \cdot) : B_0(\Sigma) \rightarrow \mathbb{R}$  defined by

$$V'(f; h) = \lim_{t \downarrow 0} \frac{V(f + th) - V(f)}{t} \quad \forall h \in B_0(\Sigma).$$

The functional  $V$  is (*Gateaux*) *differentiable* at  $f$  if  $V'(f; \cdot)$  is linear and supnorm continuous on  $B_0(\Sigma)$ . In this case,  $V'(f; \cdot)$  is the (*Gateaux*) *differential* of  $V$  at  $f$ .

The *superdifferential* of  $V$  at  $f$  is the set  $\partial V(f)$  of all linear and supnorm continuous functionals  $L : B_0(\Sigma) \rightarrow \mathbb{R}$  such that

$$V'(f; h) \leq L(h) \quad \forall h \in B_0(\Sigma).$$

In particular,  $\partial V(f)$  is a singleton if and only if  $V$  is differentiable at  $f$ . In this case,  $\partial V(f)$  only consists of the differential  $V'(f; \cdot)$ ; i.e.,  $\partial V(f) = \{V'(f; \cdot)\}$ .

We are now ready to state our result. It provides an explicit formula for the superdifferential  $\partial V(f)$  at every  $f \in B_0(\Sigma)$  and a full characterization of differentiability, along with an explicit formula for the differential.

**Theorem 18** *For all  $f \in \mathcal{F}$ ,*

$$\partial V(f) = \left\{ u'(f) dr : r \in \arg \min_{p \in \Delta} \left( \int u(f) dp + c^*(p) \right) \right\}. \quad (18)$$

*In particular,  $V$  is everywhere differentiable on  $B_0(\Sigma)$  if and only if  $c^*$  is essentially strictly convex.<sup>10</sup> In this case,*

$$V'(f; h) = \int hu'(f) dr, \quad (19)$$

where  $\{r\} = \arg \min_{p \in \Delta} (\int u(f) dp + c^*(p))$ .

The strict convexity of the ambiguity index thus characterizes everywhere differentiable variational preference functionals. For example, Proposition 23 will show that divergence preferences have a strictly convex ambiguity index, provided  $\phi$  is strictly convex. By Theorem 18 they are everywhere differentiable.

Variational preferences featuring an index  $c$  that is not strictly convex are, by Theorem 18, not everywhere differentiable in general. This is a large class of preferences, which includes but is much larger than that of multiple priors preferences.

However, though there are plenty of examples of variational preferences that are not everywhere differentiable, they form a small subset of the class of all variational preferences. In fact, the convex combination of a convex cost function and a strictly convex function is strictly convex, and so one can approximate arbitrarily well any variational preference with another one that is everywhere differentiable.

---

<sup>10</sup>Here  $u'(f) dr$  denotes the functional on  $B_0(\Sigma)$  that associates  $\int hu'(f) dr$  to every  $h \in B_0(\Sigma)$ , and  $c$  is *essentially strictly convex* if it is strictly convex on line segments in  $\bigcup_{f \in B_0(\Sigma)} \partial V(f)$ . Clearly, if  $c$  is strictly convex, then *a fortiori* it is essentially strictly convex.

In view of all this, the interest of Theorem 18 is both theoretical and practical. In applications the explicit formulas (18) and (19) are very important because they make it possible the explicit resolution of optimal problems based on the variational preference functional  $V$ . It is worth observing that in [36] we show that a version of these formulas holds also for dynamic variational preferences, the intertemporal version of the variational preferences we are introducing in this paper.

On the theoretical side, Theorem 18 is interesting because in some economic applications of ambiguity aversion the lack of smoothness (and in particular the existence of “kinks” in the indifference curves) has played a key role. For example, it has been used to justify nonparticipation in asset markets (see [21]). By fully characterizing the differentiability of variational preferences, Theorem 18 clarifies the scope of these results. In particular, it shows that though kinks are featured by some important classes of ambiguity averse preferences, they are far from being a property of ambiguity aversion *per se*. Indeed, we just observed above that it is always possible to approximate arbitrarily close any variational preference with another one that is everywhere differentiable.

## 4 Special Cases

In this section we study in some more detail two important classes of variational preferences, the multiple priors preferences of Gilboa and Schmeidler [24] and the divergence preferences we just introduced. In particular we show that two important classes of preferences, the multiplier preferences of Hansen and Sargent [27] and the mean-variance preferences of Markowitz [40] and Tobin [49], are special cases of divergence, and so variational, preferences.

### 4.1 Multiple Priors Preferences

Begin with the multiple priors choice model axiomatized by Gilboa and Schmeidler [24]. As we have mentioned in Subsection 3.1, the multiple priors model is characterized by axiom A.2', a stronger version of our independence axiom A.2. Next we show in greater detail the relation between A.2' and the variational formula (6).

In particular, when A.2' replaces A.2, the only probabilities in  $\Delta$  that “matter” in the representation (6) are the ones to which the decision maker attributes “maximum weight”, that is, the ones in  $\arg \min c^*$ . The set of priors  $C$  used in the multiple priors model is then given by  $\{p \in \Delta : c^*(p) = 0\}$ .

**Proposition 19** *Let  $\succsim$  be a variational preference. The following conditions are equivalent:*

- (i)  $\succsim$  satisfies A.2';

(ii) for all  $f \in \mathcal{F}$ ,

$$\min_{p \in \Delta} \left( \int u(f) dp + c^*(p) \right) = \min_{\{p \in \Delta : c^*(p) = 0\}} \int u(f) dp. \quad (20)$$

If, in addition,  $\succsim$  is unbounded, then (ii) is also equivalent to:

(iii)  $c^*$  only takes on values 0 and  $\infty$ .

The characterization of the multiple priors model via axioms A.1, A.2', and A.3-A.6 is due to Gilboa and Schmeidler [24]. Proposition 19 shows how the multiple priors model fits in the representation we established in Theorem 3.

As well-known, the standard SEU model is the special case of the multiple priors model characterized by the following stronger version of A.5.

A.5' *Uncertainty Neutrality*. If  $f, g \in \mathcal{F}$  and  $\alpha \in (0, 1)$ ,

$$f \sim g \Rightarrow \alpha f + (1 - \alpha) g \sim f.$$

In terms of our representation, by Theorem 3 and Proposition 19 we have:

**Corollary 20** *Let  $\succsim$  be a variational preference. The following conditions are equivalent:*

(i)  $\succsim$  satisfies A.5';

(ii)  $\succsim$  is SEU;

(iii)  $\succsim$  satisfies A.2' and  $\{p \in \Delta : c^*(p) = 0\}$  is a singleton.

If, in addition,  $\succsim$  is unbounded, then (iii) is also equivalent to:

(iv) there exists  $q \in \Delta$  such that  $c^*(q) = 0$  and  $c^*(p) = \infty$  for every  $p \neq q$ .

## 4.2 Divergence Preferences

In the previous section we introduced divergence preferences in order to illustrate our results on probabilistic sophistication. Here we discuss their ambiguity attitudes.

Recall that a preference  $\succsim$  on  $\mathcal{F}$  is a divergence preference if

$$f \succsim g \Leftrightarrow \min_{p \in \Delta^\sigma(q)} \left\{ \int u(f) dp + \theta D_\phi^w(p||q) \right\} \geq \min_{p \in \Delta^\sigma(q)} \left\{ \int u(g) dp + \theta D_\phi^w(p||q) \right\},$$

where  $\theta > 0$ ,  $q$  is a countably additive probability on the  $\sigma$ -algebra  $\Sigma$ ,  $u : X \rightarrow \mathbb{R}$  is an affine function, and  $D_\phi^w(\cdot||q) : \Delta \rightarrow [0, \infty]$  is the  $w$ -weighted  $\phi$ -divergence given by (14).<sup>11</sup>

By Theorem 16, all divergence preferences are examples of continuous variational preferences. As a result, in order to determine their ambiguity attitudes we can invoke Propositions 7 and 8. By the former result, divergence preferences are ambiguity averse. As to comparative attitudes, the next simple consequence of Proposition 8 shows that they only depend on the parameter  $\theta$ , which can therefore be interpreted as a coefficient of ambiguity aversion.

**Corollary 21** *Given two  $(w, \phi)$ -divergence preferences  $\succsim_1$  and  $\succsim_2$ , the following conditions are equivalent:*

- (i)  $\succsim_1$  is more ambiguity averse than  $\succsim_2$ ,
- (ii)  $u_1 \approx u_2$  and  $\theta_1 \leq \theta_2$  (provided  $u_1 = u_2$ ).

According to Corollary 21, divergence preferences become more and more (less and less, resp.) ambiguity averse as the parameter  $\theta$  becomes closer and closer to 0 (closer and closer to  $\infty$ , resp.).

The limit cases where  $\theta$  goes either to 0 or to  $\infty$  are described by Proposition 12, which takes an especially stark form for divergence preferences under some very mild assumptions. To see it, we need a piece of notation: given a simple measurable function  $\varphi : S \rightarrow \mathbb{R}$ , set

$$\text{ess min}_{s \in S} \varphi(s) = \max \{t \in \mathbb{R} : q(\{s \in S : \varphi(s) \geq t\}) = 1\}.$$

For example, when  $q$  has a finite support  $\text{supp}(q)$ , we have

$$\text{ess min}_{s \in S} \varphi(s) = \min_{s \in \text{supp}(q)} \varphi(s). \tag{21}$$

**Proposition 22** (i) *If  $w \in L^\infty(q)$  then, for all  $f \in \mathcal{F}$ ,*

$$\lim_{\theta \downarrow 0} \min_{p \in \Delta^\sigma(q)} \left\{ \int u(f) dp + \theta D_\phi^w(p||q) \right\} = \text{ess min}_{s \in S} u(f(s)).$$

(ii) *If  $\phi$  is strictly convex, then, for all  $f \in \mathcal{F}$ ,*

$$\lim_{\theta \uparrow \infty} \min_{p \in \Delta^\sigma(q)} \left\{ \int u(f) dp + \theta D_\phi^w(p||q) \right\} = \int u(f) dq.$$

---

<sup>11</sup>When we want to be specific about  $w$  and  $\phi$  we speak of  $(w, \phi)$ -divergence preferences.

When  $w \in L^\infty(q)$ , divergence preferences therefore tend more and more, as  $\theta$  goes to 0, to rank acts according to the very cautious criterion given by  $\text{ess min}_{s \in S} u(f(s))$ . In contrast, when  $\phi$  is strictly convex, divergence preferences tend more and more, as  $\theta$  goes to  $\infty$ , to rank acts according to the SEU criterion given by  $\int u(f) dq$ .

The next result, a simple consequence of Theorem 18, shows that divergence preferences are smooth under the assumption that  $\phi$  is strictly convex (most examples of divergence  $D_\phi^w$  satisfy this condition; see, e.g., [33]). This is an important feature of divergence preferences that makes them especially suited for optimization problems, and that also differentiates them from multiple priors preferences, which are not everywhere differentiable (see, e.g., Epstein and Wang [21, p. 295]).

As we did for Theorem 18, we assume that  $X$  is the set of all monetary lotteries, and we regard  $B_0(\Sigma)$  as the collection of all monetary acts. We also assume that the utility function  $u$  is concave, strictly increasing, and differentiable on  $\mathbb{R}$ .

**Proposition 23** *If  $\phi$  is strictly convex, then  $D_\phi^w(\cdot||q) : \Delta(\Sigma) \rightarrow [0, \infty]$  is strictly convex on its effective domain and the variational preference functional  $V : B_0(\Sigma) \rightarrow \mathbb{R}$  given by*

$$V(f) = \min_{p \in \Delta(\Sigma)} \left\{ \int u(f) dp + \theta D_\phi^w(p||q) \right\}, \quad \forall f \in B_0(\Sigma),$$

is everywhere differentiable for all  $\theta > 0$ . In this case,

$$V'(f; h) = \int hu'(f) dr, \tag{22}$$

where  $\{r\} = \arg \min_{p \in \Delta} (\int u(f) dp + \theta D_\phi^w(p||q))$ .

Consider for example multiplier preferences, a special class of divergence preferences in which  $c^*(p) = \theta R(p||q)$ . By some well known properties of the relative entropy (see [18, p. 34]), formula (22) takes in this case the following neat form:

$$V'(f; h) = \frac{\int hu'(f) \exp\left(-\frac{u(f)}{\theta}\right) dq}{\int \exp\left(-\frac{u(f)}{\theta}\right) dq}$$

for all  $f, h \in B_0(\Sigma)$ .

Unlike the multiple priors case, for divergence preferences we do not have an additional axiom that on top of axioms A.1-A.6 would deliver them (for multiple priors preferences the needed extra axiom was A.2'). We hope that this will be achieved in later work and in this regard it is worth observing that the ambiguity index  $D_\phi^w(p||q)$  is additively separable, a strong structural property.

After having established the main properties of divergence preferences, we now move to discuss two fundamental examples of this class of variational preferences.



### 4.2.1 Entropic and Multiplier Preferences

We say that a preference  $\succsim$  on  $\mathcal{F}$  is an *entropic preference* if

$$f \succsim g \Leftrightarrow \min_{p \in \Delta^\sigma(q)} \left( \int u(f) dp + \theta R^w(p||q) \right) \geq \min_{p \in \Delta^\sigma(q)} \left( \int u(g) dp + \theta R^w(p||q) \right),$$

where  $\theta > 0$ ,  $q \in \Delta^\sigma$ ,  $u : X \rightarrow \mathbb{R}$  is an affine function, and  $R^w(\cdot||q) : \Delta \rightarrow [0, \infty]$  is the weighted relative entropy given by:

$$R^w(p||q) = \begin{cases} \int_S w(s) \left( \frac{dp}{dq}(s) \log \frac{dp}{dq}(s) - \frac{dp}{dq}(s) + 1 \right) dq(s) & \text{if } p \in \Delta^\sigma(q), \\ \infty & \text{otherwise.} \end{cases}$$

Since the entropy  $R^w(p||q)$  is a special case of divergence  $D_\phi^w(p||q)$  defined in (14), where  $\phi(t) = t \log t - t + 1$ , entropic preferences are an example of divergence preferences. Hence, by Theorem 16 they are continuous variational preferences, with index of ambiguity aversion given by

$$c^*(p) = \theta R^w(p||q), \quad \forall p \in \Delta.$$

When  $w$  is uniform,  $\succsim$  is probabilistically sophisticated and it features (a positive multiple of) the standard relative entropy  $R(\cdot||q)$  as index of ambiguity aversion. This is the case considered by Hansen and Sargent (see, e.g., [26] and [27]), and they call *multiplier preferences* this class of entropic preferences, in which acts are ranked according to:

$$V(f) = \min_{p \in \Delta^\sigma(q)} \left( \int u(f) dp + \theta R(p||q) \right).$$

Though multiplier preferences are probabilistically sophisticated, Example 17 shows that this not the case for general entropic preferences having non uniform weighting functions  $w$ . These “non uniform” entropic preferences thus provide a specification of preferences that can in general produce Ellsberg-type behavior – and so are ambiguity averse according to all notions of ambiguity available in literature – but that also retain the good analytical tractability of multiplier preferences.

As to the ambiguity attitudes featured by entropic preferences, by Corollary 21 they are characterized by the parameter  $\theta$  as follows: the lower  $\theta$  is, the more ambiguity averse is the entropic preference. The parameter  $\theta$  can therefore be interpreted as a coefficient of ambiguity aversion.

We have shown how entropic, and so multiplier, preferences are a special case of divergence preferences. As we already observed, for divergence preferences we do not have an additional axiom that on top of A.1-A.6 would deliver them (even though we

have been able to point out some strong structural properties that their ambiguity indices satisfy). On the other hand, we view entropic preferences as essentially an analytically convenient specification of variational preferences, much in the same way as, for example, Cobb-Douglas preferences are an analytically convenient specification of homothetic preferences. As a result, in our setting there might not exist behaviorally significant axioms that would characterize entropic preferences (as we are not aware of any behaviorally significant axiom characterizing Cobb-Douglas preferences). Similar considerations apply to the Gini preferences that we will introduce momentarily.

For their macroeconomic applications, Hansen and Sargent are mostly interested in dynamic choice problems. Though our model is static, in the follow-up paper Maccheroni, Marinacci and Rustichini [36] we provide a dynamic version of it and, *inter alia*, we are able to provide some dynamic specifications of multiplier preferences that are time consistent. As it is often the case in choice theory, also here the analysis of the static model is key in paving the way for its dynamic extension.

We close by discussing the related work of Wang [50]. In a quite different setting [50] recently proposed an axiomatization of a class of preferences that include multiplier preferences as special cases. He considers preferences over triplets  $(f, C, q)$ , where  $f$  is a payoff profile,  $q$  is a reference probability, and  $C \subseteq \Delta$  is a confidence region. For such preferences he axiomatizes the following representation:

$$V(f, C, q) = \min_{p \in C} \left\{ \int u(f) dp + \theta R(p||q) \right\}. \quad (23)$$

His modelling is very different from ours: in our setup preferences are defined only on acts, and we derive simultaneously both the utility index  $u$  and the ambiguity index  $c$ ; that is, uncertainty is subjective. In [50], both  $C$  and  $q$  are exogenous, and so uncertainty is objective; moreover, agents' preferences are defined on the significantly larger set of all possible triplets consisting of payoff profile, confidence region, and reference model.

In any case, observe that when (23) is viewed as a preference functional on  $\mathcal{F}$ , then it actually represents variational preferences having as ambiguity index the sum of  $\delta_C$  and  $\theta R(\cdot||q)$ . As a result, Wang's preferences are a special case of variational preferences once they are interpreted in our setting.

### 4.2.2 Gini and Mean-Variance Preferences

We say that a preference  $\succsim$  on  $\mathcal{F}$  is a *Gini preference* if

$$f \succsim g \Leftrightarrow \min_{p \in \Delta^\sigma(q)} \left( \int u(f) dp + \theta G^w(p||q) \right) \geq \min_{p \in \Delta^\sigma(q)} \left( \int u(g) dp + \theta G^w(p||q) \right),$$

where  $\theta > 0$ ,  $q \in \Delta^\sigma$ ,  $u : X \rightarrow \mathbb{R}$  is an affine function, and  $G^w(\cdot||q) : \Delta \rightarrow [0, \infty]$  is the weighted relative Gini index given by:

$$G^w(p||q) = \begin{cases} \int_S w(s) \frac{1}{2} \left( \frac{dp}{dq}(s) - 1 \right)^2 dq(s) & \text{if } p \in \Delta^\sigma(q), \\ \infty & \text{otherwise.} \end{cases}$$

Like the weighted relative entropy  $R^w(p||q)$ , also the Gini index  $G^w(p||q)$  is a special case of divergence  $D_\phi^w(p||q)$  defined in (14), with  $\phi(t) = 2^{-1}(t-1)^2$ .<sup>12</sup> As a result, by Theorem 16, Gini preferences are continuous variational preferences, with ambiguity index  $\theta G^w(\cdot||q)$ .

In particular,  $\succsim$  is probabilistically sophisticated when  $w$  is uniform. In this case, when  $X$  is the set of all monetary lotteries and  $u(t) = t$  for all  $t \in \mathbb{R}$ , we call the restriction of these preferences to the collection  $B_0(\Sigma)$  of all monetary acts *monotone mean-variance preferences*, written  $\succsim^{mmv}$ ; that is, for all  $f, g \in B_0(\Sigma)$ ,

$$f \succsim^{mmv} g \Leftrightarrow \min_{p \in \Delta^\sigma(q)} \left( \int f dp + \theta G(p||q) \right) \geq \min_{p \in \Delta^\sigma(q)} \left( \int g dp + \theta G(p||q) \right).$$

Since the Gini index is, along with Shannon's entropy, a classic concentration index, monotone mean-variance preferences are a natural example of divergence preferences. But, we are not considering them just for this: their main interest lies in the close connection they have with mean-variance preferences.

In fact, consider the classic mean-variance preferences of Markowitz [40] and Tobin [49] defined on  $B_0(\Sigma)$  by:

$$f \succsim^{mv} g \Leftrightarrow \int f dq - \frac{1}{2\theta} \text{Var}(f) \geq \int g dq - \frac{1}{2\theta} \text{Var}(g), \quad (24)$$

where  $\text{Var}$  is the variance with respect to  $q$ . These preferences are not monotone, unless their domain is suitably restricted the set  $M$  on which the (Gateaux) differential of the mean-variance functional  $f \mapsto \int f dq - (1/2\theta) \text{Var}(f)$  is positive (as a linear functional). The convex set  $M$ , called domain of monotonicity of  $\succsim^{mv}$ , is where these preferences do not violate the monotonicity axiom A.4, that is where they are economically meaningful.

---

<sup>12</sup>The classic Gini concentration index can be obtained by normalization from the relative one (with uniform  $w$ ) in the same way Shannon's entropy can be obtained from relative entropy.

**Theorem 24** *The domain of monotonicity  $M$  of  $\succsim^{mv}$  is the set*

$$\left\{ f \in B_0(\Sigma) : f - \int f dq \leq \theta q - a.s. \right\}.$$

Moreover,

$$f \succsim^{mv} g \Leftrightarrow \min_{p \in \Delta^\sigma(q)} \left( \int f dp + \theta G(p||q) \right) \geq \min_{p \in \Delta^\sigma(q)} \left( \int g dp + \theta G(p||q) \right)$$

for all  $f, g \in M$ .

By Theorem 24, mean-variance preferences coincide with monotone mean-variance preferences once they are restricted on their domain of monotonicity  $M$ , which is where they are meaningful.

*Inter alia*, this result suggests that monotone mean-variance preferences are the natural adjusted version of mean-variance preferences satisfying monotonicity. This insight is developed at length in [38], and we refer the interested reader to that paper for a detailed analysis of this class of variational preferences.

## A Niveloids

The set of all functions in  $B_0(\Sigma)$  (resp.  $B(\Sigma)$ ) taking values in the interval  $K \subseteq \mathbb{R}$  is denoted by  $B_0(\Sigma, K)$  (resp.  $B(\Sigma, K)$ ).

When endowed with the supnorm,  $B_0(\Sigma)$  is a normed vector space and  $B(\Sigma)$  is a Banach space. The norm dual of  $B_0(\Sigma)$  (resp.  $B(\Sigma)$ ) is the space  $ba(\Sigma)$  of all bounded and finitely additive set functions  $\mu : \Sigma \rightarrow \mathbb{R}$  endowed with the total variation norm, the duality being  $\langle \varphi, \mu \rangle = \int \varphi d\mu$  for all  $\varphi \in B_0(\Sigma)$  (resp.  $B(\Sigma)$ ) and all  $\mu \in ba(\Sigma)$  (see, e.g., [17, p. 258]).

As well known, the weak\* topologies  $\sigma(ba(\Sigma), B_0(\Sigma))$  and  $\sigma(ba(\Sigma), B(\Sigma))$  coincide on  $\Delta(\Sigma)$ ; moreover, a subset of  $\Delta^\sigma(\Sigma)$  is weakly\* compact iff it is weakly compact (i.e., compact in the weak topology of the Banach space  $ba(\Sigma)$ ). For  $\varphi, \psi \in B(\Sigma)$ , we write  $\varphi \geq \psi$  if  $\varphi(s) \geq \psi(s)$  for all  $s \in S$ .

A functional  $I : \Phi \rightarrow \mathbb{R}$ , defined on a non-empty subset  $\Phi$  of  $B(\Sigma)$ , is a *niveloid* if

$$I(\varphi) - I(\psi) \leq \sup(\varphi - \psi)$$

for all  $\varphi, \psi \in \Phi$ , see Dolecki and Greco [15]. Clearly a niveloid is Lipschitz continuous in the supnorm. A niveloid  $I$  is *normalized* if  $I(k1_S) = k$  for all  $k \in \mathbb{R}$  such that  $k1_S \in \Phi$ ; with a little abuse we sometimes write  $k$  instead of  $k1_S$ .

**Lemma 25** *Let  $0 \in \text{int}(K)$ . A functional  $I : B_0(\Sigma, K) \rightarrow \mathbb{R}$  is a niveloid iff for all  $\varphi, \psi \in B_0(\Sigma, K)$ ,  $k \in K$ ,  $\alpha \in (0, 1)$ ,*

(i)  $\varphi \geq \psi$  implies  $I(\varphi) \geq I(\psi)$ , and

(ii)  $I(\alpha\varphi + (1 - \alpha)k) = I(\alpha\varphi) + (1 - \alpha)k$ .

In this case,  $I$  is concave iff

(iii)  $I(\psi) = I(\varphi)$  implies  $I(\alpha\psi + (1 - \alpha)\varphi) \geq I(\varphi)$ .

Properties (i) and (ii) are called *monotonicity* and *vertical invariance*, respectively. The proof of Lemma 25 can be found in Maccheroni, Marinacci, and Rustichini [35]. If  $I : \Phi \rightarrow \mathbb{R}$  is a niveloid, then

$$\hat{I}(\psi) = \sup_{\varphi \in \Phi} \left[ I(\varphi) + \inf_{s \in S} (\psi(s) - \varphi(s)) \right], \quad \forall \psi \in B(\Sigma)$$

is the least niveloid on  $B(\Sigma)$  that extends  $I$  (see [15]). Moreover,

- if  $\Phi$  is convex and  $I$  is concave, then  $\hat{I}$  is concave;
- if  $\Phi + \mathbb{R} \supseteq B_0(\Sigma)$ ,<sup>13</sup> then  $\hat{I}$  is the unique niveloid on  $B(\Sigma)$  that extends  $I$ ; (see [35]).

If  $I : \Phi \rightarrow \mathbb{R}$  is a concave niveloid, direct application of the Fenchel-Moreau Theorem (see, e.g., [43, p. 42]) to  $\hat{I}$  guarantees that  $\hat{I}(\varphi) = \min_{\mu \in ba(\Sigma)} (\langle \varphi, \mu \rangle - \hat{I}^*(\mu))$ , where  $\hat{I}^*(\mu) = \inf_{\psi \in B(\Sigma)} (\langle \psi, \mu \rangle - \hat{I}(\psi))$  is the *Fenchel conjugate* of  $\hat{I}$ . If  $\mu$  is not positive, there exists  $\varphi \geq 0$  such that  $\langle \varphi, \mu \rangle < 0$ , then  $\langle \alpha\varphi, \mu \rangle - \hat{I}(\alpha\varphi) \leq \alpha \langle \varphi, \mu \rangle - \hat{I}(0)$  for all  $\alpha \geq 0$ , whence  $\hat{I}^*(\mu) = -\infty$ . If  $\mu(S) \neq 1$ , choose  $\psi \in B(\Sigma)$ , then  $\langle \psi + b, \mu \rangle - \hat{I}(\psi + b) = \langle \psi, \mu \rangle - \hat{I}(\psi) + b(\mu(S) - 1)$  for all  $b \in \mathbb{R}$ , and so  $\hat{I}^*(\mu) = -\infty$ . That is,

$$\hat{I}(\varphi) = \min_{p \in \Delta(\Sigma)} (\langle \varphi, p \rangle - \hat{I}^*(p)); \quad (25)$$

see also, Föllmer and Schied [22, Thm. 4.12]. Set  $I^*(p) = \hat{I}^*(p)$  for each  $p \in \Delta(\Sigma)$ , and  $\partial_\pi I(\varphi) = \{p \in \Delta(\Sigma) : I(\psi) - I(\varphi) \leq \langle \psi - \varphi, p \rangle \text{ for each } \psi \in \Phi\}$ .

The next two results are proved in [35].

**Lemma 26** *Let  $\Phi$  be a convex subset of  $B(\Sigma)$  containing at least one constant function, and  $I : \Phi \rightarrow \mathbb{R}$  be a concave and normalized niveloid. Then:*

(i) For each  $p \in \Delta(\Sigma)$ ,  $I^*(p) = \inf_{\psi \in \Phi} (\langle \psi, p \rangle - I(\psi))$ .

(ii)  $I^* : \Delta(\Sigma) \rightarrow [-\infty, 0]$  is concave and weakly\* upper semicontinuous.

---

<sup>13</sup> $\Phi + \mathbb{R}$  is the set  $\{\varphi + b : \varphi \in \Phi, b \in \mathbb{R}\}$ . An important special case in which  $\Phi + \mathbb{R} = B_0(\Sigma)$  is when  $\Phi = B_0(\Sigma, K)$  and  $K$  is unbounded.

(iii) For each  $\varphi \in \Phi$ ,  $\partial_\pi I(\varphi) = \arg \min_{p \in \Delta(\Sigma)} (\langle \varphi, p \rangle - I^*(p))$  and it is not empty.

(iv)  $\partial_\pi I(k1_S) = \{I^* = 0\} = \arg \max_{p \in \Delta(\Sigma)} I^*(p)$  for all  $k \in \mathbb{R}$  such that  $k1_S \in \Phi$ .

(v)  $I^*$  is the maximal functional  $R : \Delta(\Sigma) \rightarrow [-\infty, 0]$  such that

$$I(\varphi) = \inf_{p \in \Delta(\Sigma)} (\langle \varphi, p \rangle - R(p)) \quad \forall \varphi \in \Phi. \quad (26)$$

Moreover, if  $\Phi + \mathbb{R} \supseteq B_0(\Sigma)$ ,  $I^*$  is the unique concave and weakly\* upper semi-continuous function  $R : \Delta(\Sigma) \rightarrow [-\infty, 0]$  such that (26) holds.

(vi) If (26) holds, and  $\Psi \subseteq \Phi$  is such that  $\sup_{s \in S} \psi(s) - \inf_{s \in S} \psi(s) < b$  for all  $\psi \in \Psi$ , then

$$I(\psi) = \inf_{\{p \in \Delta(\Sigma) : R(p) \geq -b\}} (\langle \psi, p \rangle - R(p)) \quad \forall \psi \in \Psi. \quad (27)$$

**Proposition 27** Let  $I : B_0(\Sigma, K) \rightarrow \mathbb{R}$  be a normalized concave niveloid, with  $K$  unbounded and  $\Sigma$  a  $\sigma$ -algebra. Then, the following conditions are equivalent:

(i) If  $\varphi, \psi \in B_0(\Sigma, K)$ ,  $k \in K$ ,  $\{E_n\}_{n \geq 1} \in \Sigma$  with  $E_1 \supseteq E_2 \supseteq \dots$  and  $\bigcap_{n \geq 1} E_n = \emptyset$ , then  $I(\varphi) > I(\psi)$  implies that there exists  $n_0 \geq 1$  such that

$$I(k1_{E_{n_0}} + \varphi 1_{E_{n_0}^c}) > I(\psi).$$

(ii)  $\{p \in \Delta(\Sigma) : I^*(p) \geq b\}$  is a weakly compact subset of  $\Delta^\sigma(\Sigma)$  for each  $b \leq 0$ .

(iii) There exists  $q \in \Delta^\sigma(\Sigma)$  such that  $\{p \in \Delta(\Sigma) : I^*(p) \geq b\}$  is a weakly compact subset of  $\Delta^\sigma(\Sigma, q)$  for each  $b \leq 0$ , and for each  $\varphi \in B_0(\Sigma, K)$ ,

$$I(\varphi) = \min_{p \in \Delta^\sigma(\Sigma, q)} (\langle \varphi, p \rangle - I^*(p)). \quad (28)$$

## B Proofs of the Results in the Main Text

The main result proved in this appendix is Theorem 3. Its proof proceeds as follows: using Lemma 28, we first show that if  $\succsim$  satisfies A.1-A.6, then there exists a non-constant affine function  $u : X \rightarrow \mathbb{R}$  and a normalized and concave niveloid  $I : B_0(\Sigma, u(X)) \rightarrow \mathbb{R}$  such that  $f \succsim g$  iff  $I(u(f)) \geq I(u(g))$ . Then, (25) obtained in Appendix A delivers the desired variational representation  $I(u(f)) = \min_{p \in \Delta(\Sigma)} (\int u(f) dp + c(p))$ .

We now move to the proofs. The standard one of Lemma 1 is omitted.

**Lemma 28** A binary relation  $\succsim$  on  $\mathcal{F}$  satisfies A.1-A.4 and A.6 iff there exist a non-constant affine function  $u : X \rightarrow \mathbb{R}$  and a normalized niveloid  $I : B_0(\Sigma, u(X)) \rightarrow \mathbb{R}$  such that

$$f \succsim g \Leftrightarrow I(u(f)) \geq I(u(g)).$$

**Proof.** Assume  $\succsim$  on  $\mathcal{F}$  satisfies A.1-A.4 and A.6. Let  $x, y \in X$  be such that  $x \sim y$ . If there exists  $z \in X$  such that  $\frac{1}{2}x + \frac{1}{2}z \approx \frac{1}{2}y + \frac{1}{2}z$ , w.l.o.g.  $\frac{1}{2}x + \frac{1}{2}z \succ \frac{1}{2}y + \frac{1}{2}z$ ; by A.2 (we can replace  $z$  with  $x$  to obtain)  $\frac{1}{2}x + \frac{1}{2}x \succ \frac{1}{2}y + \frac{1}{2}x$  and (we can replace  $z$  with  $y$  to obtain)  $\frac{1}{2}x + \frac{1}{2}y \succ \frac{1}{2}y + \frac{1}{2}y$ , and conclude  $x \succ y$ , which is absurd. Then the hypotheses of the Mixture Space Theorem (Hernstein and Milnor [28]) are satisfied, and there exists an affine function  $u : X \rightarrow \mathbb{R}$  such that  $x \succsim y$  iff  $u(x) \geq u(y)$ . By A.6 there exist  $f, g \in \mathcal{F}$  such that  $f \succ g$ . Let  $x, y \in X$  be such that  $x \succsim f(s)$  and  $g(s) \succsim y$  for all  $s \in S$ , then  $x \succsim f \succ g \succsim y$  implies  $x \succ y$ , and  $u$  cannot be constant. Moreover  $u$  is unique up to positive affine transformations and we can assume  $0 \in \text{int}(u(X))$ .

For all  $f \in \mathcal{F}$ , let  $x, y \in X$  be such that  $x \succsim f(s) \succsim y$  for all  $s \in S$ , then  $x \succsim f \succsim y$ . By A.3 the sets  $\{\alpha \in [0, 1] : \alpha x + (1 - \alpha)y \succsim f\}$  and  $\{\alpha \in [0, 1] : f \succsim \alpha x + (1 - \alpha)y\}$  are closed; they are nonempty since 1 belongs to the first and 0 to the second; their union is the whole  $[0, 1]$ . Since  $[0, 1]$  is connected, their intersection is not empty, hence there exists  $\beta \in [0, 1]$  such that  $\beta x + (1 - \beta)y \sim f$ . In particular, any act  $f$  admits a certainty equivalent  $x_f \in X$ .

If  $f \sim x_f$ , set  $U(f) = u(x_f)$ .  $U$  is well defined since  $f \sim x_f$  and  $f \sim y_f$  with  $x_f, y_f \in X$  implies  $x_f \sim y_f$  and  $u(x_f) = u(y_f)$ . Clearly,  $f \succsim g$  iff  $x_f \succsim x_g$  iff  $u(x_f) \geq u(x_g)$  iff  $U(f) \geq U(g)$ . Therefore  $U$  represents  $\succsim$ .

If  $f \in \mathcal{F}$  then  $u(f) \in B_0(\Sigma, u(X))$ . Conversely, if  $\varphi \in B_0(\Sigma, u(X))$ ,  $\varphi(s) = u(x_i)$  if  $s \in A_i$  for suitable  $x_1, \dots, x_N \in X$  and a partition  $\{A_1, A_2, \dots, A_N\}$  of  $S$  in  $\Sigma$ . Therefore, setting  $f(s) = x_i$  if  $s \in A_i$  we have  $\varphi = u(f)$ . We can conclude that  $B_0(\Sigma, u(X)) = \{u(f) : f \in \mathcal{F}\}$ . Moreover,  $u(f) = u(g)$  iff  $u(f(s)) = u(g(s))$  for all  $s \in S$  iff  $f(s) \sim g(s)$  for all  $s \in S$ , and by A.4,  $f \sim g$  or equivalently  $U(f) = U(g)$ .

Define  $I(\varphi) = U(f)$  if  $\varphi = u(f)$ . By what we have just observed,  $I : B_0(\Sigma, u(X)) \rightarrow \mathbb{R}$  is well defined. If  $\varphi = u(f)$  and  $\psi = u(g) \in B_0(\Sigma, u(X))$  and  $\varphi \geq \psi$ , then  $u(f(s)) \geq u(g(s))$  for all  $s \in S$ , and  $f(s) \succsim g(s)$  for all  $s \in S$ , so  $f \succsim g$ ,  $U(f) \geq U(g)$ , and  $I(\varphi) = I(u(f)) = U(f) \geq U(g) = I(u(g)) = I(\psi)$ . Therefore,  $I$  is monotonic. Take  $k \in u(X)$ , say  $k = u(x)$ ,  $I(k1_S) = I(u(x)) = U(x) = u(x) = k$ . Therefore,  $I$  is normalized.

Take  $\alpha \in (0, 1)$ ,  $\varphi = u(f) \in B_0(\Sigma, u(X))$ ,  $k = u(x_k) \in u(X)$ ; denote by  $x_0$  an element in  $X$  such that  $u(x_0) = 0$ . Choose  $x, y \in X$  such that  $x \succsim f(s) \succsim y$  for all  $s \in S$ , then  $\alpha x + (1 - \alpha)x_0 \succsim \alpha f(s) + (1 - \alpha)x_0 \succsim \alpha y + (1 - \alpha)x_0$  for all  $s \in S$ . The technique used in the second paragraph of this proof yields the existence of  $\beta \in [0, 1]$  such that  $\beta(\alpha x + (1 - \alpha)x_0) + (1 - \beta)(\alpha y + (1 - \alpha)x_0) \sim \alpha f + (1 - \alpha)x_0$ , i.e.,  $\alpha z + (1 - \alpha)x_0 \sim \alpha f + (1 - \alpha)x_0$ , where  $z = \beta x + (1 - \beta)y \in X$ . Then, by A.2,

$\alpha z + (1 - \alpha) x_k \sim \alpha f + (1 - \alpha) x_k$ , and

$$\begin{aligned}
I(\alpha\varphi + (1 - \alpha)k) &= I(u(\alpha f + (1 - \alpha)x_k)) = u(\alpha z + (1 - \alpha)x_k) \\
&= \alpha u(z) + (1 - \alpha)k = \alpha u(z) + (1 - \alpha)0 + (1 - \alpha)k \\
&= u(\alpha z + (1 - \alpha)x_0) + (1 - \alpha)k \\
&= I(u(\alpha f + (1 - \alpha)x_0)) + (1 - \alpha)k = I(\alpha\varphi) + (1 - \alpha)k.
\end{aligned}$$

By Lemma 25,  $I$  is a niveloid, and we already proved that it is normalized.

Conversely, assume there exist a non-constant affine function  $u : X \rightarrow \mathbb{R}$  and a normalized niveloid  $I : B_0(\Sigma, u(X)) \rightarrow \mathbb{R}$  such that  $f \succsim g$  iff  $I(u(f)) \geq I(u(g))$ .

Choose  $c \in \mathbb{R}$  such that  $0 \in \text{int}(u(X) + c)$  and set  $v = u + c$ . Define  $J : B_0(\Sigma, v(X)) \rightarrow \mathbb{R}$  by  $J(\varphi) = I(\varphi - c) + c$ . Notice that  $J$  is a normalized niveloid,<sup>14</sup> and

$$\begin{aligned}
f \succsim g &\Leftrightarrow I(u(f)) \geq I(u(g)) \Leftrightarrow I(u(f) + c - c) + c \geq I(u(g) + c - c) + c \\
&\Leftrightarrow I(v(f) - c) + c \geq I(v(g) - c) + c \Leftrightarrow J(v(f)) \geq J(v(g)).
\end{aligned}$$

Clearly,  $\succsim$  satisfies A.1.

If  $f, g \in \mathcal{F}$ ,  $x, y \in X$ ,  $\alpha \in (0, 1)$ , then  $\alpha v(h)$ ,  $(1 - \alpha)v(z)$ ,  $\alpha v(h) + (1 - \alpha)v(z) \in B_0(\Sigma, v(X))$  for  $h = f, g$  and  $z = x, y$ ; moreover

$$\begin{aligned}
&\alpha f + (1 - \alpha)x \succsim \alpha g + (1 - \alpha)x \Rightarrow \\
&J(\alpha v(f) + (1 - \alpha)v(x)) \geq J(\alpha v(g) + (1 - \alpha)v(x)) \Rightarrow \\
&J(\alpha v(f)) + (1 - \alpha)v(x) \geq J(\alpha v(g)) + (1 - \alpha)v(x) \Rightarrow \\
&J(\alpha v(f)) + (1 - \alpha)v(y) \geq J(\alpha v(g)) + (1 - \alpha)v(y) \Rightarrow \\
&J(\alpha v(f) + (1 - \alpha)v(y)) \geq J(\alpha v(g) + (1 - \alpha)v(y)) \Rightarrow \\
&\alpha f + (1 - \alpha)y \succsim \alpha g + (1 - \alpha)y
\end{aligned}$$

and A.2 holds.

If  $f, g, h \in \mathcal{F}$ ,  $\alpha \in [0, 1]$ , and there exists  $\alpha_n \in [0, 1]$  such that  $\alpha_n \rightarrow \alpha$  and  $\alpha_n f + (1 - \alpha_n)g \succsim h$  for all  $n \geq 1$ ; then  $v(\alpha_n f + (1 - \alpha_n)g) = \alpha_n v(f) + (1 - \alpha_n)v(g)$  converges uniformly to  $\alpha v(f) + (1 - \alpha)v(g) = v(\alpha f + (1 - \alpha)g)$ .  $J(v(\alpha_n f + (1 - \alpha_n)g)) \geq J(v(h))$  for all  $n \geq 1$ , and the continuity of  $J$  guarantee  $J(v(\alpha f + (1 - \alpha)g)) \geq J(v(h))$ . Therefore  $\{\alpha \in [0, 1] : \alpha f + (1 - \alpha)g \succsim h\}$  is closed. A similar argument shows that  $\{\alpha \in [0, 1] : h \succsim \alpha f + (1 - \alpha)g\}$  is closed too, and A.3 holds.

<sup>14</sup>In fact, for all  $\varphi, \psi \in B_0(\Sigma, v(X))$ ,

$$J(\varphi) - J(\psi) = I(\varphi - c) + c - I(\psi - c) - c \leq \sup((\varphi - c) - (\psi - c)) = \sup(\varphi - \psi).$$

Moreover, for all  $t \in v(X)$ ,  $J(t) = I(t - c) + c = t - c + c = t$ .



Given  $f, g \in \mathcal{F}$ ,  $f(s) \succsim g(s)$  for all  $s \in S$  iff  $J(v(f(s))) \geq J(v(g(s)))$  for all  $s$  iff  $v(f(s)) \geq v(g(s))$  for all  $s$ , then monotonicity of  $J$  yields  $J(v(f)) \geq J(v(g))$ . This shows A.4.

Finally, since  $v$  is not constant and it represents  $\succsim$  on  $X$ , there exist  $x \succ y$ , and A.6 holds too.  $\blacksquare$

**Proof of Theorem 3 (and Proposition 6).** Assume  $\succsim$  satisfies A.1-A.6. By Lemma 28, there is a non-constant affine function  $u : X \rightarrow \mathbb{R}$  and a normalized niveloid  $I : B_0(\Sigma, u(X)) \rightarrow \mathbb{R}$  such that  $f \succsim g$  iff  $I(u(f)) \geq I(u(g))$ .

Next we show that A.5 implies that  $I : B_0(\Sigma, u(X)) \rightarrow \mathbb{R}$  is concave. Let  $\varphi, \psi \in B_0(\Sigma, u(X))$  be such that  $I(\varphi) = I(\psi)$  and  $\alpha \in (0, 1)$ . If  $f, g \in \mathcal{F}$  are such that  $\varphi = u(f)$  and  $\psi = u(g)$ , then  $f \sim g$  and, by A.5,  $\alpha f + (1 - \alpha)g \succsim f$ , that is,

$$\begin{aligned} I(\alpha\varphi + (1 - \alpha)\psi) &= I(\alpha u(f) + (1 - \alpha)u(g)) = I(u(\alpha f + (1 - \alpha)g)) \\ &\geq I(u(f)) = I(\varphi). \end{aligned}$$

Lemma 25 guarantees the concavity of  $I$ .

The functional  $I : B_0(\Sigma, u(X)) \rightarrow \mathbb{R}$  is, therefore, a concave and normalized niveloid. For all  $p \in \Delta(\Sigma)$ , set  $c^*(p) = -I^*(p)$ . Lemma 26 guarantees that

$$c^*(p) = - \inf_{\psi \in B_0(\Sigma, u(X))} (\langle \psi, p \rangle - I(\psi)) = \sup_{f \in \mathcal{F}} \left( u(x_f) - \int u(f) dp \right)$$

for all  $p \in \Delta(\Sigma)$  (where  $x_f$  is a certainty equivalent for  $f$ ), that  $c^*$  is non-negative, grounded, convex, and weakly\* lower semicontinuous, and that

$$I(\varphi) = \min_{p \in \Delta(\Sigma)} \left( \int \varphi dp + c^*(p) \right) \quad \forall \varphi \in B_0(\Sigma, u(X));$$

see also (25).

Let  $c : \Delta(\Sigma) \rightarrow [0, \infty]$  be a grounded, convex, and weakly\* lower semicontinuous function such that

$$f \succsim g \Leftrightarrow \min_{p \in \Delta(\Sigma)} \left( \int u(f) dp + c(p) \right) \geq \min_{p \in \Delta(\Sigma)} \left( \int u(g) dp + c(p) \right).$$

For all  $\varphi = u(f) \in B_0(\Sigma, u(X))$ ,

$$\begin{aligned} I(\varphi) &= I(u(f)) = u(x_f) = \min_{p \in \Delta(\Sigma)} \left( \int u(x_f) dp + c(p) \right) \\ &= \min_{p \in \Delta(\Sigma)} \left( \int u(f) dp + c(p) \right) = \min_{p \in \Delta(\Sigma)} \left( \int \varphi dp + c(p) \right). \end{aligned}$$

Lemma 26 guarantees  $c^* \leq c$  (this concludes the proof that (i) implies (ii)). Moreover, if  $u(X)$  is unbounded, again Lemma 26 guarantees  $c = c^*$  (this proves Proposition 6).

For the converse, notice that  $u$  and  $I(\varphi) = \min_{p \in \Delta(\Sigma)} \left( \int \varphi dp + c(p) \right)$  are a non-constant affine function and a normalized niveloid representing  $\succsim$ . By Lemma 28,  $\succsim$  satisfies A.1-A.4 and A.6. Concavity of  $I$  guarantees A.5.<sup>15</sup> ■

**Proof of Corollary 5.** Let  $(u_0, c_0^*)$  represent  $\succsim$  as in Theorem 3. If  $(u, c^*)$  is another representation of  $\succsim$  (as in Theorem 3), by (6)  $u$  and  $u_0$  are affine representations of the restriction of  $\succsim$  to  $X$ . Hence, by standard uniqueness results there exist  $\alpha > 0$  and  $\beta \in \mathbb{R}$  such that  $u = \alpha u_0 + \beta$ . By (7),

$$c^*(p) = \sup_{f \in \mathcal{F}} \left( u(x_f) - \int u(f) dp \right) = \sup_{f \in \mathcal{F}} \left( \alpha u_0(x_f) + \beta - \int (\alpha u_0(f) + \beta) dp \right) = \alpha c_0^*(p),$$

as desired. The converse is trivial. ■

**Lemma 29** *Let  $\succsim$  be a binary relation on  $X$  represented by an affine function  $u : X \rightarrow \mathbb{R}$ .  $u(X)$  is unbounded (either below or above) iff  $\succsim$  satisfies A.7.*

The standard proof is omitted. Proposition 7 is a consequence of Lemma 32.

**Proof of Proposition 8.** (ii) trivially implies (i). As to the converse, let  $(u_i, c_i^*)$  represent  $\succsim_i$  as in Theorem 3,  $i = 1, 2$ , and set  $I_i(\varphi) = \min_{p \in \Delta(\Sigma)} \left( \int \varphi dp + c_i^*(p) \right)$  for all  $\varphi \in B_0(\Sigma, u_i(X))$ .

By (8) and the fact that  $u_1$  and  $u_2$  are not constant, we can choose  $u_1 = u_2 = u$ . For all  $f \in \mathcal{F}$ , if  $f \sim_1 x$ , then  $f \succsim_2 x$ ; therefore,  $I_1(u(f)) = u(x) \leq I_2(u(f))$ . This implies  $I_1 \leq I_2$ , and

$$c_1^*(p) = \sup_{\varphi \in B_0(\Sigma, u(X))} \left( I_1(\varphi) - \int \varphi dp \right) \leq \sup_{\varphi \in B_0(\Sigma, u(X))} \left( I_2(\varphi) - \int \varphi dp \right) = c_2^*(p)$$

for all  $p \in \Delta(\Sigma)$ . ■

**Proof of Proposition 12.** Observe that the functions  $c_n$  are weak\* lower semicontinuous on  $\Delta$ , and so is  $\int u(f) dp + c_n(p)$  for each  $n$ . Using this observation we now prove (i) and (ii).

(i) The decreasing sequence  $\left\{ \int u(f) dp + c_n(p) \right\}_n$  pointwise converges to

$$\int u(f) dp + \delta_{\cup_n \text{dom } c_n}(p). \quad (29)$$

---

<sup>15</sup>If  $f \sim g$  and  $\alpha \in (0, 1)$ , then

$$I(u(\alpha f + (1 - \alpha)g)) = I(\alpha u(f) + (1 - \alpha)u(g)) \geq \alpha I(u(f)) + (1 - \alpha)I(u(g)) = I(u(f)).$$

Hence, by [12, Prop. 5.7] this sequence  $\Gamma$ -converges to

$$\int u(f) dp + \delta_{\overline{\cup_n \text{dom } c_n}}(p),$$

the weak\* lower semicontinuous envelope of (29). By [12, Thm 7.4], this implies

$$\lim \min_n \left\{ \int u(f) dp + c_n(p) \right\} = \min_{p \in \Delta} \left\{ \int u(f) dp + \delta_{\overline{\cup_n \text{dom } c_n}}(p) \right\}.$$

(ii) Since  $\arg \min c_n = \{p \in \Delta : c_n(p) = 0\}$ , the increasing sequence  $\left\{ \int u(f) dp + c_n(p) \right\}_n$  pointwise converges (and so, by [12, p. 47],  $\Gamma$ -converges) to

$$\int u(f) dp + \delta_{\cap_n \arg \min c_n}(p).$$

Since  $\Delta$  is weak\* compact, Dal Maso [12, Thm. 7.8] implies

$$\lim \min_n \left\{ \int u(f) dp + c_n(p) \right\} = \min_{p \in \Delta} \left\{ \int u(f) dp + \delta_{\cap_n \arg \min c_n}(p) \right\}.$$

■

**Proof of Theorem 13.** Let  $(u, c^*)$  represent  $\succsim$  as in Theorem 3, and set, for all  $\varphi \in B_0(\Sigma, u(X))$ ,  $I(\varphi) = \min_{p \in \Delta(\Sigma)} \left( \int \varphi dp + c^*(p) \right)$ . It is easy to check that  $\succsim$  satisfies A.8 on  $\mathcal{F}$  iff  $I$  satisfies the condition (i) of Proposition 27 on  $B_0(\Sigma, u(X))$ . Unboundedness of  $u(X)$  and the relation  $c^* = -I^*$  allow to apply Proposition 27 and to obtain the desired equivalence. ■

**Lemma 30** *Let  $\succsim$  be a variational preference that satisfies axiom A.8. Then,*

$$f \succsim g \Leftrightarrow \inf_{p \in \Delta^\sigma(\Sigma)} \left( \int u(f) dp + c^*(p) \right) \geq \inf_{p \in \Delta^\sigma(\Sigma)} \left( \int u(g) dp + c^*(p) \right), \quad \forall f, g \in \mathcal{F}.$$

**Proof.** Let  $(u, c^*)$  represent  $\succsim$  as in Theorem 3, and set  $I(\varphi) = \min_{p \in \Delta(\Sigma)} \left( \int \varphi dp + c^*(p) \right)$  for all  $\varphi \in B_0(\Sigma, u(X))$ . Let  $\varphi \in \text{int}(B_0(\Sigma, u(X)))$ ,  $E_n \downarrow \emptyset$ , and  $\varepsilon > 0$  such that  $\varphi - \varepsilon \in B_0(\Sigma, u(X))$ . Then,

$$I(\varphi 1_{E_n^c} + (\min \varphi - \varepsilon) 1_{E_n}) - I(\varphi) \leq I(\varphi - \varepsilon 1_{E_n}) - I(\varphi) \leq -\varepsilon p(E_n) \leq 0$$

for all  $p \in \partial_\pi I(\varphi)$ . Consider a sequence  $\{k_j\}_{j \geq 1}$  in  $u(X)$  such  $k_j < I(\varphi)$  and  $k_j \uparrow I(\varphi)$ . By A.8,  $I$  satisfies (i) of Proposition 27 on  $B_0(\Sigma, u(X))$ . Then, for all  $j \geq 1$  there is  $n_0 \geq 1$  such that  $k_j < I(\varphi 1_{E_{n_0}^c} + (\min \varphi - \varepsilon) 1_{E_{n_0}})$ . Hence,  $\lim_{n \rightarrow \infty} I(\varphi 1_{E_n^c} + (\min \varphi - \varepsilon) 1_{E_n}) > k_j$  as the sequence  $I(\varphi 1_{E_n^c} + (\min \varphi - \varepsilon) 1_{E_n})$  is increasing. Passing to the limit for  $j \rightarrow \infty$ , we obtain

$$\lim_{n \rightarrow \infty} I(\varphi 1_{E_n^c} + (\min \varphi - \varepsilon) 1_{E_n}) = I(\varphi), \quad (30)$$

and  $p(E_n) \rightarrow 0$  (uniformly w.r.t.  $p \in \partial_\pi I(\varphi)$ ), that is,  $\partial_\pi I(\varphi)$  is a (weakly compact) subset of  $\Delta^\sigma(\Sigma)$ . By Lemma 26,  $\min_{p \in \Delta(\Sigma)} (\int \varphi dp + c^*(p))$  is attained for all  $\varphi \in \text{int}(B_0(\Sigma, u(X)))$  in  $(\partial_\pi I(\varphi))$  and hence in  $\Delta^\sigma(\Sigma)$ . Hence,

$$I(\varphi) = \min_{p \in \Delta(\Sigma)} \left( \int \varphi dp + c^*(p) \right) \text{ and } J(\varphi) = \inf_{p \in \Delta^\sigma(\Sigma)} \left( \int \varphi dp + c^*(p) \right)$$

coincide on  $\text{int}(B_0(\Sigma, u(X)))$ , being continuous, they coincide on  $B_0(\Sigma, u(X))$ .  $\blacksquare$

The results of Subsection 3.5 require some notation and preliminaries. For all  $\varphi \in L^1(\Sigma, q)$ , set  $G_\varphi(t) = q(\{\varphi > t\})$  for each  $t \in \mathbb{R}$  and  $\Gamma_\varphi(b) = \inf\{t \in \mathbb{R} : G_\varphi(t) \leq b\}$  for each  $b \in [0, 1]$ .

$G_\varphi$  is the *survival function* of  $\varphi$ ,  $\Gamma_\varphi$  is the *decreasing rearrangement* of  $\varphi$ . Two functions  $\varphi, \psi \in L^1(\Sigma, q)$  are *equimeasurable* if  $G_\varphi = G_\psi$  (iff  $\Gamma_\varphi = \Gamma_\psi$ ). We refer to Chong and Rice [10] for a comprehensive study of equimeasurability. The preorder  $\prec_{cx}$  defined in Subsection 3.5 can be naturally regarded as a relation on  $L^1(\Sigma, q)$  by putting  $\varphi \prec_{cx} \psi$  iff  $\int \phi(\varphi) dq \leq \int \phi(\psi) dq$  for every convex  $\phi$  on  $\mathbb{R}$ .

Analogously, the definition of Shur convexity (resp. rearrangement invariance) can be spelled for functions  $T : L^1(\Sigma, q) \rightarrow (-\infty, \infty]$  by requiring that  $\varphi \prec_{cx} \psi$  implies  $T(\varphi) \leq T(\psi)$  (resp.  $\varphi \sim_{cx} \psi$  implies  $T(\varphi) = T(\psi)$ ).

Put  $\varphi \prec\prec_{cx} \psi$  iff  $\int_0^t \Gamma_\varphi(b) db \leq \int_0^t \Gamma_\psi(b) db$  for all  $t \in [0, 1]$ . Then  $\varphi \prec_{cx} \psi$  iff  $\varphi \prec\prec_{cx} \psi$  and  $\int \varphi dq = \int \psi dq$  (see, e.g., [11]). Moreover, the equimeasurability relation is the common symmetric part  $\sim_{cx}$  of  $\prec\prec_{cx}$  and  $\prec_{cx}$ . Simple manipulation of the results of Luxemburg [34, p. 125-126] yields the following:

**Lemma 31 (Luxemburg)** *Let  $q$  be adequate. If  $T : B_0(\Sigma, K) \rightarrow (-\infty, \infty]$  is rearrangement invariant, then the function defined for all  $\varphi \in L^1(\Sigma, q)$  by  $H(\varphi) = \sup_{\psi \in B_0(\Sigma, K)} (\int \varphi \psi dq - T(\psi))$  is Shur convex. If  $H$  is monotonic, then  $\varphi \prec\prec_{cx} \varphi'$  implies  $H(\varphi) \leq H(\varphi')$ .*

**Proof.** Let  $\varphi \in L^1(\Sigma, q)$ . By [34, Thm 9.1], for all  $\psi \in B_0(\Sigma, K)$ ,

$$\max \left\{ \int \varphi \psi' dq : \psi' \sim_{cx} \psi, \psi' \in B_0(\Sigma, K) \right\} = \int_0^1 \Gamma_\varphi(t) \Gamma_\psi(t) dt. \quad (31)$$

Therefore  $H(\varphi) \leq \sup_{\psi \in B_0(\Sigma, K)} \left( \int_0^1 \Gamma_\varphi(t) \Gamma_\psi(t) dt - T(\psi) \right)$ . Again by (31), for all  $\psi \in B_0(\Sigma, K)$  there exists  $\psi' \in B_0(\Sigma, K)$  with  $\psi' \sim_{cx} \psi$  such that  $\int_0^1 \Gamma_\varphi(t) \Gamma_\psi(t) dt = \int \varphi \psi' dq$ . Since  $T(\psi') = T(\psi)$ , then

$$\int_0^1 \Gamma_\varphi(t) \Gamma_\psi(t) dt - T(\psi) = \int \varphi \psi' dq - T(\psi) = \int \varphi \psi' dq - T(\psi'),$$

and

$$H(\varphi) = \sup_{\psi \in B_0(\Sigma, K)} \left( \int_0^1 \Gamma_\varphi(t) \Gamma_\psi(t) dt - T(\psi) \right).$$

If  $\varphi \prec_{cx} \varphi'$ , an inequality of Hardy (see, e.g., [10, p. 57-58]) delivers  $H(\varphi) \leq H(\varphi')$ . Let  $\varphi \prec_{\prec_{cx}} \varphi'$ . By [11, Thm 1.1] there is a non-negative  $\varphi'' \in L^1(\Sigma, q)$  such that  $\varphi + \varphi'' \prec_{cx} \varphi'$ . If  $H$  is monotonic, then  $H(\varphi) \leq H(\varphi + \varphi'') \leq H(\varphi')$ . ■

**Remark 1** Analogously, if  $q$  is adequate and  $T : L^1(\Sigma, q) \rightarrow (-\infty, \infty]$  is rearrangement invariant (resp.  $q \in \Delta^\sigma(\Sigma)$  and  $T$  is Shur convex), then the function defined for all  $\varphi \in B_0(\Sigma)$  by  $H(\varphi) = \sup_{\psi \in L^1(\Sigma, q)} (\int \varphi \psi dq - T(\psi))$  is Shur convex. If  $H$  is monotonic, then  $\varphi \prec_{\prec_{cx}} \varphi'$  implies  $H(\varphi) \leq H(\varphi')$ . (If  $T$  is Shur convex and  $q$  is not adequate use [10, Thm 13.8] rather than [34, Thm 9.1]). See also Dana [13].

**Proof of Theorem 14.** We prove (i)  $\Rightarrow$  (ii)  $\Rightarrow$  (iii)  $\Rightarrow$  (iv)  $\Rightarrow$  (i). For  $f \in \mathcal{F}$ ,  $q_f$  denotes the finite support probability on  $X$  defined by  $q_f(x) = q(f^{-1}(x))$  for all  $x \in X$ .

(i)  $\Rightarrow$  (ii) is trivial. (ii)  $\Rightarrow$  (iii) Let  $\succsim$  be a continuous unbounded variational preference, which is also rearrangement invariant with respect to an adequate  $q \in \Delta^\sigma(\Sigma)$ . Let  $(u, c^*)$  represent  $\succsim$  as in Theorem 3, and set, for all  $\varphi \in B_0(\Sigma, u(X))$ ,  $I(\varphi) = \min_{p \in \Delta(\Sigma)} (\int \varphi dp + c^*(p))$ . For all  $p \in \Delta$

$$c^*(p) = \sup_{f \in \mathcal{F}} \left( u(x_f) - \int u(f) dp \right) = \sup_{\psi \in B_0(\Sigma, u(X))} \left( I(\psi) - \int \psi dp \right).$$

Theorem 13 guarantees that  $\{c^* < \infty\} \subseteq \Delta^\sigma(\Sigma)$ . If  $p \in \Delta^\sigma(\Sigma) \setminus \Delta^\sigma(\Sigma, q)$ , there exists  $A \in \Sigma$  such that  $p(A) > 0$  and  $q(A) = 0$ . If  $u(X)$  is unbounded below, w.l.o.g. assume  $u(X) \supseteq (-\infty, 0]$ . Let  $x_n \in u^{-1}(-n)$ ,  $y \in u^{-1}(0)$ . Consider the act  $f_n = x_n A y$ , and the constant act  $y$ . Since  $q_{f_n} = \delta_y = q_y$ , by rearrangement invariance,  $f_n \sim y$  and  $I(-n1_A) = I(u(f_n)) = 0$  for all  $n \geq 1$ ; therefore

$$c^*(p) \geq \sup_{n \geq 1} \left( I(-n1_A) - \int -n1_A dp \right) = \infty.$$

If  $u(X)$  is unbounded above, w.l.o.g. assume  $u(X) \supseteq [0, \infty)$ . Let  $x_n \in u^{-1}(n)$ ,  $y \in u^{-1}(0)$ . Consider the act  $f_n = y A x_n$ . Since  $q_{f_n} = \delta_{x_n} = q_{x_n}$ , then  $f_n \sim x_n$  and  $I(n1_{A^c}) = I(u(f_n)) = n$  for all  $n \geq 1$ ; therefore

$$c^*(p) \geq \sup_{n \geq 1} \left( I(n1_{A^c}) - \int n1_{A^c} dp \right) = \sup_{n \geq 1} (n - n(1 - p(A))) = \infty.$$

We conclude that  $\{c^* < \infty\} \subseteq \Delta^\sigma(\Sigma, q)$ . In particular,  $\varphi = \varphi'$   $q$ -a.s. implies  $I(\varphi) = I(\varphi')$ .

Assume  $\varphi, \psi \in B_0(\Sigma, u(X))$  are equimeasurable. Therefore (see, e.g., [10, p. 12]) there exist  $x_1, \dots, x_n \in X$  and two partitions  $\{A_1, A_2, \dots, A_n\}$  and  $\{B_1, B_2, \dots, B_n\}$  of  $S$  in  $\Sigma$ , with  $q(A_i) = q(B_i)$  for all  $i = 1, 2, \dots, n$ , such that  $q$ -a.s.

$$\varphi = \sum_{i=1}^n u(x_i) 1_{A_i} \text{ and } \psi = \sum_{i=1}^n u(x_i) 1_{B_i}.$$

If  $f(s) = x_i$  for all  $s \in A_i$  and  $g(s) = x_i$  for all  $s \in B_i$ , then  $f, g \in \mathcal{F}$  and  $q_f = q_g$ . By rearrangement invariance of  $\succsim$ , we obtain  $f \sim g$ , whence  $I(\varphi) = I(u(f)) = I(u(g)) = I(\psi)$ , and  $I$  is rearrangement invariant. Setting  $T(\psi) = -I(-\psi)$ ,  $T : B_0(\Sigma, -u(X)) \rightarrow \mathbb{R}$  is rearrangement invariant too. For all  $p \in \Delta^\sigma(\Sigma, q)$ ,

$$c^*(p) = \sup_{\psi \in B_0(\Sigma, u(X))} \left( I(\psi) - \int \psi dp \right) = \sup_{\psi \in B_0(\Sigma, -u(X))} \left( \int \psi \frac{dp}{dq} dq - T(\psi) \right).$$

By Lemma 31,  $H(\varphi) = \sup_{\psi \in B_0(\Sigma, -u(X))} \left( \int \varphi \psi dq - T(\psi) \right)$  for all  $\varphi \in L^1(\Sigma, q)$ , is Shur convex, therefore  $c^*(p)$  is Shur convex, and *a fortiori* rearrangement invariant.

(iii)  $\Rightarrow$  (iv) Consider  $T : L^1(\Sigma, q) \rightarrow [0, \infty]$  defined as follows:

$$T(\psi) = \begin{cases} c^*(p_\psi) & \text{if } \psi \geq 0 \text{ } q\text{-a.s., } \int \psi dq = 1, \\ \infty & \text{otherwise,} \end{cases} \quad (32)$$

where  $p_\psi$  is the element of  $\Delta^\sigma(\Sigma, q)$  such that  $dp_\psi/dq = \psi$ . Denote by  $P$  the set

$$\left\{ \psi \in L^1(\Sigma, q) : \psi \geq 0 \text{ } q\text{-a.s. and } \int \psi dq = 1 \right\}.$$

Let  $\psi' \sim_{cx} \psi$ , then  $\psi \in P$  iff  $\psi' \in P$  (see, e.g., [10, p. 15-16]). If  $\psi, \psi' \notin P$ , then  $T(\psi) = \infty = T(\psi')$ . If  $\psi, \psi' \in P$ , then  $\psi$  and  $\psi'$  are the Radon-Nikodym derivatives of  $p_\psi$  and  $p_{\psi'}$ . By the rearrangement invariance of  $c^*$ ,  $T(\psi') = c^*(p_{\psi'}) = c^*(p_\psi) = T(\psi)$ , and so  $T$  is rearrangement invariant. By Theorem 13,  $\{p \in \Delta : c^*(p) \leq t\}$  is a weakly compact subset of  $\Delta^\sigma$  for each  $t \geq 0$ , therefore  $\{\psi \in L^1(\Sigma, q) : T(\psi) \leq t\}$  is a weakly compact subset of  $L^1(\Sigma, q)$ , *a fortiori*  $T$  is weakly lower semicontinuous. Since obviously  $T$  is convex, [34, Thm 13.3] guarantees that  $T$  (hence  $c^*$ ) is Shur convex.<sup>16</sup>

(iv)  $\Rightarrow$  (i) Let  $\succsim$  be a variational preference. Let  $(u, c)$  represent  $\succsim$  as in Theorem 3, and assume  $c$  is Shur convex (with respect to  $q$ ).<sup>17</sup> Set, for all  $\varphi \in B_0(\Sigma)$ ,  $I(\varphi) = \min_{p \in \Delta(\Sigma)} \left( \int \varphi dp + c(p) \right)$ . Then

$$-I(-\varphi) = - \inf_{p \in \Delta^\sigma(\Sigma, q)} \left( \int -\varphi dp + c(p) \right) = \sup_{\psi \in L^1(\Sigma, q)} \left( \int \varphi \psi dq - T(\psi) \right)$$

<sup>16</sup> *Different definitions caveat:* [34, p. 123-124] define Shur convexity of  $T$  by rearrangement invariance, convexity, and weak lower semicontinuity, Theorem 13.3 shows that if  $T$  has these features then  $\varphi \prec_{cx} \varphi'$  implies  $T(\varphi) \leq T(\varphi')$ .

<sup>17</sup> Notice that we are not assuming that  $\succsim$  is unbounded, or continuous, or that  $c = c^*$ , or that  $q$  is adequate.

where  $T$  is defined like in (32) replacing  $c^*$  with  $c$ . Let  $\psi' \prec_{cx} \psi$ . If  $\psi \notin P$ , then  $T(\psi) = \infty \geq T(\psi')$ . Else if  $\psi \in P$ ,  $\int \psi' dq = \int \psi dq = 1$ , and  $\psi' \geq 0$   $q$ -a.s. (see, e.g., [10, p. 62]), that is,  $\psi' \in P$ . The Shur convexity of  $c$  ensures that  $T(\psi') = c(p_{\psi'}) \leq c(p_\psi) = T(\psi)$ . Thus,  $T$  is Shur convex. By Remark 1,  $H(\varphi) = -I(-\varphi)$  is Shur convex; in particular,  $I(\varphi) = -H(-\varphi)$  is rearrangement invariant.

Assume  $f, g \in \mathcal{F}$  are such that  $q_f = q_g$ , then  $q_{u(f)} = (q_f)_u = (q_g)_u = q_{u(g)}$ . It follows that  $u(f) \sim_{cx} u(g)$ ,  $I(u(f)) = I(u(g))$ , and  $f \sim g$ .

Assume that  $q(\{s \in S : f(s) \lesssim x\}) \leq q(\{s \in S : g(s) \lesssim x\})$  for all  $x \in X$ . For all  $t \in u(X)$ ,  $q(\{s \in S : u(f(s)) > t\}) \geq q(\{s \in S : u(g(s)) > t\})$ , and this is *a fortiori* true if  $t \notin u(X)$ . We conclude that  $G_{u(f)} \geq G_{u(g)}$ ,  $\Gamma_{u(f)} \geq \Gamma_{u(g)}$ ,  $\Gamma_{-u(f)} \leq \Gamma_{-u(g)}$  (see, e.g., [10, p. 30-31]), and  $-u(f) \prec_{cx} -u(g)$ . Since  $H$  is monotonic, Remark 1 yields  $-H(-u(f)) \geq -H(-u(g))$ , so that  $I(u(f)) \geq I(u(g))$  and  $f \succsim g$ . ■

**Proof of Lemma 15.** Groundedness (in particular  $D_\phi^w(q||q) = 0$ ) and convexity are trivial, so is Shur convexity if  $w$  is uniform. Weak\* lower semicontinuity descends from the fact that the sets  $\{p \in \Delta : D_\phi^w(p||q) \leq t\}$  are weakly compact in  $\Delta^\sigma(q)$  for all  $t \in \mathbb{R}$ . The remaining part of the proof is devoted to show this compactness feature.

We first consider a simple  $\Sigma$ -measurable function  $w$  with  $\min_{s \in S} w(s) > 0$ . Set  $\underline{w} = \min_{s \in S} w(s)$ , and w.l.o.g. suppose  $t \geq 0$ . By definition,  $\{p \in \Delta : D_\phi^w(p||q) \leq t\} \subseteq \Delta^\sigma(q)$ , hence

$$\{p \in \Delta : D_\phi^w(p||q) \leq t\} = \{p \in \Delta^\sigma : D_\phi^w(p||q) \leq t\},$$

denote the set on the r.h.s. by  $\mathcal{D}$ . We show that (a)  $\lim_{q(B) \rightarrow 0} p(B) = 0$  uniformly w.r.t.  $p \in \mathcal{D}$ , and (b) if  $\{p_n\}_{n \geq 1} \subseteq \mathcal{D}$  and  $p_n(B) \rightarrow p(B)$  for all  $B \in \Sigma$ , then  $p \in \mathcal{D}$ . Then, a classical result of Bartle, Dunford, and Schwartz guarantees that  $\mathcal{D}$  is weakly compact (see, e.g., [17, Ch. IV]).<sup>18</sup>

(a) Notice that  $D_\phi^{\underline{w}}(p||q) \leq D_\phi^w(p||q)$  for all  $p \in \Delta^\sigma$ , so that

$$\mathcal{D} = \{p \in \Delta^\sigma : D_\phi^{\underline{w}}(p||q) \leq t\} \subseteq \{p \in \Delta^\sigma : D_\phi^w(p||q) \leq t\} = \mathcal{C}.$$

Next we show that  $\lim_{q(B) \rightarrow 0} p(B) = 0$  uniformly w.r.t.  $p \in \mathcal{C}$ , a fortiori w.r.t.  $p \in \mathcal{D}$ . Let  $p \in \mathcal{C}$ , clearly,

$$\int_A \underline{w} \phi \left( \frac{dp}{dq} \right) dq \leq \int \underline{w} \phi \left( \frac{dp}{dq} \right) dq < t + 1 \text{ for all } A \in \Sigma.$$

---

<sup>18</sup>(a) guarantees that  $\mathcal{D}$  is relatively sequentially weakly compact, that is, every sequence  $\{p_n\}$  in  $\mathcal{D}$  admits a weakly convergent subsequence  $\{p_{n_j}\}$ , (b) guarantees that the limit of  $p_{n_j}$  belongs to  $\mathcal{D}$ . That is  $\mathcal{D}$  is sequentially weakly compact, the Eberlein-Smulian Theorem guarantees that  $\mathcal{D}$  is weakly compact.

For all  $\varepsilon > 0$ , since  $\lim_{t \rightarrow \infty} \frac{t}{w\phi(t)} = 0^+$ , there exists  $\delta > 0$  such that

$$0 < \frac{b}{w\phi(b)} < \frac{\varepsilon}{2(t+1)} \text{ if } b > \frac{\varepsilon}{2\delta},$$

in particular  $\phi(b) > 0$ . For all  $B$  such that  $q(B) < \delta$ , and all  $p \in \mathcal{C}$

$$\begin{aligned} p(B) &= \int_B \frac{dp}{dq} dq = \int_{B \cap \left\{ \frac{dp}{dq} \leq \frac{\varepsilon}{2\delta} \right\}} \frac{dp}{dq} dq + \int_{B \cap \left\{ \frac{dp}{dq} > \frac{\varepsilon}{2\delta} \right\}} \frac{dp}{dq} dq \\ &\leq \frac{1}{2} \frac{\varepsilon}{\delta} q(B) + \int_{B \cap \left\{ \frac{dp}{dq} > \frac{\varepsilon}{2\delta} \right\}} \frac{\frac{dp}{dq}}{w\phi\left(\frac{dp}{dq}\right)} w\phi\left(\frac{dp}{dq}\right) dq \\ &< \frac{1}{2} \varepsilon + \frac{\varepsilon}{2(t+1)} \int_{B \cap \left\{ \frac{dp}{dq} > \frac{\varepsilon}{2\delta} \right\}} w\phi\left(\frac{dp}{dq}\right) dq \leq \varepsilon. \end{aligned}$$

This concludes the proof of (a).

Let  $\mu(B) = \int_B w dq$  for all  $B \in \Sigma$ . Notice that

$$D_\phi^w(p||q) = \begin{cases} \int \phi\left(\frac{dp}{dq}\right) d\mu & \text{if } p \in \Delta^\sigma(q), \\ \infty & \text{otherwise.} \end{cases} \quad (33)$$

Denote by  $\Pi_w$  the set of all finite partitions  $\pi$  of  $S$  in  $\Sigma$  finer than  $\pi_w = \{w^{-1}(b) : b \in w(S)\}$ .

For all  $\pi \in \Pi_w$ , define on  $\Delta^\sigma$  the function

$$D_\phi^w(p_\pi||q_\pi) = \sum_{A \in \pi} \phi\left(\frac{p(A)}{q(A)}\right) \mu(A),$$

with the conventions  $\frac{p(A)}{q(A)} = 0$  if  $p(A) = 0$  and  $\phi\left(\frac{p(A)}{q(A)}\right) \mu(A) = \infty$  if  $p(A) > 0$ , when  $q(A) = 0$ . Notice that since  $\pi$  is finer than  $\pi_w$ , then for all  $A \in \pi$ ,  $w(A)$  is a singleton (improperly denoted by  $w(A)$ ), and

$$\mu(A) = \int_A w dq = w(A) q(A).$$

*Claim 1*  $D_\phi^w(p||q) = \sup_{\pi \in \Pi_w} D_\phi^w(p_\pi||q_\pi)$  for all  $p \in \Delta^\sigma$ .

*Proof.* Let  $p \ll q$ . Take an increasingly finer sequence  $\pi_n$  in  $\Pi_w$  such that  $dp/dq$  is  $\sigma(\bigcup_{n=1}^\infty \pi_n)$ -measurable, then

$$\sum_{A \in \pi_n} \frac{p(A)}{q(A)} 1_A \rightarrow \frac{dp}{dq} \quad q\text{-a.e.}$$

(see, e.g., Billingsley [7, p. 470]), hence  $\mu$ -a.e.. Continuity of  $\phi$  and the Fatou Lemma imply

$$\int \phi\left(\frac{dp}{dq}\right) d\mu \leq \liminf \sum_{A \in \pi_n} \phi\left(\frac{p(A)}{q(A)}\right) \mu(A) \leq \sup_{\pi \in \Pi_w} \sum_{A \in \pi} \phi\left(\frac{p(A)}{q(A)}\right) \mu(A).$$



The converse inequality is guaranteed if  $\int \phi \left( \frac{dp}{dq} \right) d\mu = \infty$ . While, if  $\phi \left( \frac{dp}{dq} \right)$  is  $\mu$ -integrable, the conditional Jensen's Inequality guarantees that for all  $\pi \in \Pi_w$ ,

$$\mathbb{E}^m \left( \phi \left( \frac{dp}{dq} \right) \middle| \pi \right) \geq \phi \left( \mathbb{E}^m \left( \frac{dp}{dq} \middle| \pi \right) \right) \quad m\text{-a.e.}$$

where  $m = \mu/\mu(S)$  and  $\mathbb{E}$  is the expectation. Hence,

$$\mathbb{E}^m \left( \phi \left( \frac{dp}{dq} \right) \right) = \mathbb{E}^m \left( \mathbb{E}^m \left( \phi \left( \frac{dp}{dq} \right) \middle| \pi \right) \right) \geq \mathbb{E}^m \left( \phi \left( \mathbb{E}^m \left( \frac{dp}{dq} \middle| \pi \right) \right) \right),$$

for all  $\pi \in \Pi_w$ , but

$$\begin{aligned} \mathbb{E}^m \left( \frac{dp}{dq} \middle| \pi \right) &= \sum_{A \in \pi: m(A) \neq 0} 1_A \left( \frac{1}{m(A)} \int_A \frac{dp}{dq} dm \right) \\ &= \sum_{A \in \pi: \mu(A) \neq 0} 1_A \left( \frac{\mu(S)}{\mu(A)} \frac{1}{\mu(S)} \int_A \frac{dp}{dq} d\mu \right) \\ &= \sum_{A \in \pi: q(A) \neq 0} 1_A \left( \frac{1}{w(A)q(A)} \int_A \frac{dp}{dq} w dq \right) \\ &= \sum_{A \in \pi: q(A) \neq 0} 1_A \left( \frac{1}{w(A)q(A)} w(A)p(A) \right) \\ &= \sum_{A \in \pi} 1_A \left( \frac{p(A)}{q(A)} \right). \end{aligned}$$

Therefore,

$$\begin{aligned} \int \phi \left( \frac{dp}{dq} \right) d\mu &= \mu(S) \mathbb{E}^m \left( \phi \left( \frac{dp}{dq} \right) \right) \geq \mu(S) \mathbb{E}^m \left( \phi \left( \mathbb{E}^m \left( \frac{dp}{dq} \middle| \pi \right) \right) \right) \\ &\geq \mu(S) \mathbb{E}^m \left( \phi \left( \sum_{A \in \pi} 1_A \left( \frac{p(A)}{q(A)} \right) \right) \right) = \sum_{A \in \pi} \phi \left( \frac{p(A)}{q(A)} \right) \mu(A), \end{aligned}$$

so that

$$\int \phi \left( \frac{dp}{dq} \right) d\mu \geq \sup_{\pi \in \Pi_w} \sum_{A \in \pi} \phi \left( \frac{p(A)}{q(A)} \right) \mu(A).$$

If  $p \in \Delta^\sigma \setminus \Delta^\sigma(q)$ , there exists  $B \in \Sigma$  such that  $p(B) > 0$  and  $q(B) = 0$ . Therefore,

$$\sup_{\pi \in \Pi_w} \sum_{A \in \pi} \phi \left( \frac{p(A)}{q(A)} \right) \mu(A) \geq \phi \left( \frac{p(B)}{q(B)} \right) \mu(B) = \infty = D_\phi^w(p||q).$$

This concludes the proof of the Claim.  $\square$

(b) For all  $\pi \in \Pi_w$ , set  $\mathcal{D}_\pi = \{p \in \Delta^\sigma : D_\phi^w(p_\pi||q_\pi) \leq t\}$ . We show that  $\{p_n\}_{n \geq 1} \subseteq \mathcal{D}_\pi$  and  $p_n(B) \rightarrow p(B)$  for all  $B \in \Sigma$ , then  $p \in \mathcal{D}_\pi$ . First notice that  $p \in \Delta^\sigma$  (for the

Vitali-Hahn-Saks Theorem). For all  $A \in \pi$  such that  $q(A) = 0$ , then  $p_n(A) = 0$  for all  $n \geq 1$  (else  $D_\phi^w((p_n)_\pi || q_\pi) = \infty$ ), and hence

$$\phi\left(\frac{p_n(A)}{q(A)}\right) = \phi(0) = \phi\left(\frac{p(A)}{q(A)}\right).$$

For all  $A \in \pi$  such that  $q(A) > 0$ , clearly  $\frac{p_n(A)}{q(A)} \rightarrow \frac{p(A)}{q(A)}$ . We conclude that

$$\sum_{A \in \pi} \phi\left(\frac{p(A)}{q(A)}\right) q(A) = \lim_{n \rightarrow \infty} \sum_{A \in \pi} \phi\left(\frac{p_n(A)}{q(A)}\right) q(A) \leq t,$$

as wanted. Now (b) descends from Claim 1 and the observation that

$$\mathcal{D} = \{p \in \Delta^\sigma : D_\phi^w(p||q) \leq t\} = \left\{p \in \Delta^\sigma : \sup_{\pi \in \Pi_w} D_\phi^w(p_\pi || q_\pi) \leq t\right\} = \bigcap_{\pi \in \Pi_w} \mathcal{D}_\pi.$$

This completes the proof when  $w$  is simple. Suppose now that  $w$  is any  $\Sigma$ -measurable function with  $\underline{w} = \inf_{s \in S} w(s) > 0$ . Then, there exists a sequence of simple  $\Sigma$ -measurable functions  $w_n$  such that  $w_n \uparrow w$ , and  $\min_{s \in S} w_n(s) \geq \underline{w}$  for all  $n \geq 1$ . By Levi's Monotone Convergence Theorem,  $D_\phi^{w_n}(p||q) \uparrow D_\phi^w(p||q)$  for all  $p \in \Delta^\sigma(q)$ , therefore

$$\mathcal{D} = \{p \in \Delta^\sigma(q) : D_\phi^w(p||q) \leq t\} = \left\{p \in \Delta^\sigma(q) : \sup_{n \geq 1} D_\phi^{w_n}(p||q) \leq t\right\} = \bigcap_{n \geq 1} \mathcal{D}_n.$$

We conclude that  $\mathcal{D}$  as well is weakly compact. ■

**Proof of Theorem 16.** Lemma 15 and Theorem 3 guarantees that divergence preferences are variational. Unboundedness of  $u$  together with Proposition 6 guarantee that  $c^*(\cdot) = \theta D_\phi^w(\cdot||q)$ . Finally, Theorem 13 and Lemma 15 imply that  $\succsim$  satisfies A.8. ■

**Proof of Theorem 18.** Set  $I(\varphi) = \min_{p \in \Delta(\Sigma)} (\int \varphi dp + c^*(p))$  for all  $\varphi \in B_0(\Sigma)$ . Since  $-c^*$  coincides with the Fenchel conjugate of  $I$  (see Lemma 26) on  $\Delta(\Sigma)$  and  $I^*(\mu) = -\infty$  for all  $\mu \in ba(\Sigma) \setminus \Delta(\Sigma)$ , then

$$\partial I(\varphi) = \arg \min_{p \in \Delta(\Sigma)} \left( \int \varphi dp + c^*(p) \right), \quad \forall \varphi \in B_0(\Sigma),$$

and  $I$  is Gateaux differentiable on  $B_0(\Sigma)$  iff  $c^*$  is essentially strictly convex.<sup>19</sup>

Let  $f \in B_0(\Sigma)$ . Set  $\Lambda = \{u'(f) dr \mid r \in \arg \min_{p \in \Delta} (\int u(f) dp + c^*(p))\}$ , and notice that  $\Lambda$  is the image of  $\arg \min_{p \in \Delta} (\int u(f) dp + c^*(p)) = \partial I(u(f))$  through the map from  $\Delta(\Sigma)$  to  $ba(\Sigma)$  that associates to  $p(\cdot)$  in  $\Delta(\Sigma)$  the bounded and finitely

---

<sup>19</sup>This is proved in [4]. Notice that our definition of essential strict convexity is weaker than theirs, and it guarantees that strict convexity implies essential strict convexity.

additive set function  $\int_{(\cdot)} u'(f) dp$  in  $ba(\Sigma)$ . Since this map is linear,  $\sigma(ba(\Sigma), B_0(\Sigma))$ - $\sigma(ba(\Sigma), B_0(\Sigma))$ -continuous, and  $\partial I(u(f))$  is weak\* compact and convex, then  $\Lambda$  is  $\sigma(ba(\Sigma), B_0(\Sigma))$ -compact and convex.

By standard results,

$$I'(u(f); u'(f)h) = \min_{p \in \partial I(u(f))} \int u'(f) h dp = \min_{\mu \in \Lambda} \int h d\mu, \quad \forall h \in B_0(\Sigma). \quad (34)$$

Let  $h \in B_0(\Sigma)$  and  $s \in S$ . If  $h(s) \neq 0$ , then

$$u(f + th)(s) = u(f(s) + th(s)) = u(f(s)) + u'(f(s))th(s) + o(th(s))$$

for  $th(s) \rightarrow 0$ , that is,

$$\lim_{th(s) \rightarrow 0} \frac{u(f(s) + th(s)) - u(f(s)) - u'(f(s))th(s)}{th(s)} = 0,$$

and so

$$\lim_{t \downarrow 0} \frac{u(f(s) + th(s)) - u(f(s)) - u'(f(s))h(s)t}{t} = 0. \quad (35)$$

Clearly (35) holds also if  $h(s) = 0$ . Since  $f$  and  $h$  are simple, the above limit is uniform with respect to  $s \in S$ . Therefore, for  $t \downarrow 0$ ,

$$\begin{aligned} 0 &\leq \left| \frac{I(u(f + th)) - I(u(f))}{t} - \min_{\mu \in \Lambda} \int h d\mu \right| \\ &\leq \left| \frac{I(u(f + th)) - I(u(f) + tu'(f)h) + I(u(f) + tu'(f)h) - I(u(f))}{t} - \min_{\mu \in \Lambda} \int h d\mu \right| \\ &\leq \left| \frac{I(u(f + th)) - I(u(f) + tu'(f)h)}{t} \right| + \left| \frac{I(u(f) + tu'(f)h) - I(u(f))}{t} - \min_{\mu \in \Lambda} \int h d\mu \right| \\ &\leq \frac{\|u(f + th) - u(f) - tu'(f)h\|}{t} + o(1), \end{aligned}$$

where the last inequality descends from the Lipschitz continuity of  $I$  and (34). Uniformity of limit (35), then delivers

$$\lim_{t \downarrow 0} \frac{I(u(f + th)) - I(u(f))}{t} = \min_{\mu \in \Lambda} \int h d\mu$$

for all  $h \in B_0(\Sigma)$ , or  $V'(f; \cdot) = \min_{\mu \in \Lambda} \int (\cdot) d\mu$ , that is,  $\partial V(f) = \Lambda$ .

Now, if  $c^*$  is essentially strictly convex, then  $\partial I(u(f))$  is a singleton for every  $f$  in  $B_0(\Sigma)$  and  $\partial V(f) = \{u'(f) dr \mid r \in \partial I(u(f))\}$  is a singleton too. Conversely, assume that  $V$  is Gateaux differentiable on  $B_0(\Sigma)$ , and *per contra* that  $c^*$  is not essentially strictly convex. Then there exists  $\varphi \in B_0(\Sigma)$  such that  $\partial I(\varphi)$  contains two distinct elements  $r_1$  and  $r_2$ . Since  $u : \mathbb{R} \rightarrow \mathbb{R}$  is concave, then it is unbounded below and there

is  $b \in \mathbb{R}$  such that  $\varphi + b \in B_0(\Sigma, u(\mathbb{R}))$ . In particular, there exists  $f \in B_0(\Sigma)$  such that  $u(f) = \varphi + b$  and

$$\begin{aligned} r_1, r_2 \in \partial I(\varphi) &= \arg \min_{p \in \Delta(\Sigma)} \left( \int \varphi dp + c^*(p) \right) \\ &= \arg \min_{p \in \Delta(\Sigma)} \left( \int (\varphi + b) dp + c^*(p) \right) = \partial I(u(f)). \end{aligned}$$

It follows that  $u'(f) dr_1, u'(f) dr_2 \in \partial V(f)$ . Since  $V$  is Gateaux differentiable on  $B_0(\Sigma)$ , then  $u'(f) dr_1 = u'(f) dr_2$ , that is,

$$\int hu'(f) dr_1 = \int hu'(f) dr_2 \quad \forall h \in B_0(\Sigma). \quad (36)$$

Since  $u$  is strictly monotonic, concave, and differentiable, then  $u'(z) \neq 0$  for all  $z \in \mathbb{R}$ , and (36) implies  $\int \psi dr_1 = \int \psi dr_2$  for all  $\psi \in B_0(\Sigma)$ , contradicting  $r_1 \neq r_2$ .  $\blacksquare$

**Proof of Proposition 19.** Let  $(u, c^*)$  represent  $\succsim$  as in Theorem 3, w.l.o.g., assume  $[-1, 1] \subseteq u(X)$ .

(i)  $\Rightarrow$  (ii) By [24, Thm 1], there is a weakly\* compact and convex set  $C \subseteq \Delta(\Sigma)$  such that  $u(x_f) = \min_{p \in C} \int u(f) dp$  for all  $f \in \mathcal{F}$  and each  $x_f \sim f$ . By Theorem 3,

$$c^*(p) = \sup_{f \in \mathcal{F}} \left( \min_{q \in C} \int u(f) dq - \int u(f) dp \right) \quad \forall p \in \Delta(\Sigma). \quad (37)$$

Suppose  $p \in C$ , then  $c^*(p) \leq 0$ , since  $c^*$  is non-negative, we have  $c^*(p) = 0$ . Next, suppose  $p_0 \notin C$ . By the Separating Hyperplane Theorem, there is a simple measurable function  $\varphi : S \rightarrow u(X)$  such that  $\int \varphi dp > \int \varphi dp_0$  for each  $p \in C$ . Hence, taking  $f \in \mathcal{F}$  such that  $\varphi = u(f)$ ,  $\min_{p \in C} \int u(f) dp - \int u(f) dp_0 > 0$ , which in turn implies  $c^*(p_0) > 0$ . We conclude that  $c^*(p) = 0$  iff  $p \in C$ . Therefore, for all  $f \in \mathcal{F}$

$$\min_{p \in \Delta(\Sigma)} \left( \int u(f) dp + c^*(p) \right) = u(x_f) = \min_{p \in C} \int u(f) dp = \min_{p \in \{c^*=0\}} \int u(f) dp.$$

(ii)  $\Rightarrow$  (i) and (iii)  $\Rightarrow$  (ii) are trivial. Now assume  $\succsim$  is unbounded above (resp. below).

(ii)  $\Rightarrow$  (iii) For all  $p \in \Delta(\Sigma)$

$$c^*(p) = \sup_{f \in \mathcal{F}} \left( \min_{q \in \{c^*=0\}} \int u(f) dq - \int u(f) dp \right) = \sup_{\varphi \in B_0(\Sigma, u(X))} \left( \min_{q \in \{c^*=0\}} \int \varphi dq - \int \varphi dp \right)$$

Suppose,  $c^*(p_0) > 0$ . There exist a non-negative (resp. non-positive), simple, measurable function  $\varphi : S \rightarrow u(X)$  and  $\varepsilon > 0$  such that  $\min_{p \in \{c^*=0\}} \int \varphi dp - \int \varphi dp_0 > \varepsilon$ , but  $n\varphi \in B_0(\Sigma, u(X))$  for all  $n \in \mathbb{N}$ , and

$$c^*(p_0) \geq \min_{p \in \{c^*=0\}} \int n\varphi dp - \int n\varphi dp_0 \geq n\varepsilon$$

for all  $n \in \mathbb{N}$ . We conclude  $c^*(p_0) = \infty$ . ■

The proof of Corollary 20 is omitted. Just notice that A.5' can be used to obtain affinity of the functional  $I$  appearing in Lemma 28 in the same way in which A.5 is used to obtain its concavity at the beginning of the proof of Theorem 3.

**Lemma 32** *Let  $\succsim$  be a variational preference represented by  $(u, c^*)$  as in Theorem 3 and  $q \in \Delta(\Sigma)$ . The following conditions are equivalent:*

- (i)  $q$  corresponds to a SEU preference less ambiguity averse than  $\succsim$ ;
- (ii)  $c^*(q) = 0$ ;
- (iii)  $q \in \partial I(k)$  for some (all)  $k \in u(X)$ , where  $I(\varphi) = \min_{p \in \Delta(\Sigma)} (\int \varphi dp + c^*(p))$  for all  $\varphi \in B_0(\Sigma, u(X))$ .

*In particular, any variational preference is ambiguity averse.*

**Proof.** (i) implies (ii). Suppose  $\succsim_0$  is a SEU preference, with associated subjective probability  $q$  and utility index  $u_0$  such that  $\succsim$  is more ambiguity averse than  $\succsim_0$ . By (8), we can assume  $u_0 = u$ . By Proposition 8,  $c^* \leq c_0^*$ , by Corollary 20  $c_0^*(q) = 0$ , and hence  $0 \leq c^*(q) \leq c_0^*(q) = 0$ .

(ii) implies (iii).  $c^*(q) = 0$  iff  $\sup_{\varphi \in B_0(\Sigma, u(X))} (I(\varphi) - \int \varphi dq) = 0$  iff  $I(\varphi) \leq \int \varphi dq$  for all  $\varphi \in B_0(\Sigma, u(X))$  iff  $I(\varphi) - I(k) \leq \int \varphi dq - \int k dq$  for all  $\varphi \in B_0(\Sigma, u(X))$  and all  $k \in u(X)$  iff  $q \in \partial I(k)$  for all  $k \in u(X)$ .

(iii) implies (i). If  $q \in \partial I(k)$  for some  $k \in u(X)$ , then  $I(\varphi) - I(k) \leq \int \varphi dq - \int k dq$  for all  $\varphi \in B_0(\Sigma, u(X))$ , then  $I(\varphi) \leq \int \varphi dq$  for all  $\varphi \in B_0(\Sigma, u(X))$ . Denote by  $\succsim_0$  the SEU preference, with associated subjective probability  $q$  and utility index  $u$ . Notice that for all  $f \in \mathcal{F}$  and  $x \in X$ :  $f \succsim x$  implies  $I(u(f)) \geq u(x)$ , a fortiori  $\int u(f) dq \geq u(x)$  and  $f \succsim_0 x$ .

Ambiguity aversion of  $\succsim$  now follows from the observation that  $\arg \min_{p \in \Delta(\Sigma)} c^*(p)$  is non-empty and  $\min_{p \in \Delta(\Sigma)} c^*(p) = 0$ . ■

**Proof of Proposition 22.** (i) Let  $\theta_n \downarrow 0$ . By Proposition 12,

$$\lim_n \min_{p \in \Delta} \left\{ \int u(f) dp + \theta_n D_\phi^w(p||q) \right\} = \min_{p \in \Delta} \left\{ \int u(f) dp + \delta_{\{D_\phi^w(\cdot||q) < \infty\}}(p) \right\}.$$

If  $w \in L^\infty(q)$  and  $\underline{w} = \inf_{s \in S} w(s)$ , then  $\underline{w} D_\phi(p||q) \leq D_\phi^w(p||q) \leq \|w\|_\infty D_\phi(p||q)$ , which implies  $\text{dom } D_\phi(\cdot||q) = \text{dom } D_\phi^w(\cdot||q)$ . In turn, this implies

$$\lim_n \min_{p \in \Delta} \left\{ \int u(f) dp + \theta_n D_\phi^w(p||q) \right\} = \inf_{p \in \text{dom } D_\phi(\cdot||q)} \int u(f) dp.$$

Next we show that, for all  $\varphi \in B_0(\Sigma)$ ,

$$\inf_{p \in \text{dom } D_\phi(\cdot||q)} \int \varphi dp = \inf_{p \in \Delta^\sigma(\Sigma, q)} \int \varphi dp. \quad (38)$$

Let  $\varphi = \sum_{j=1}^n \alpha_j 1_{A_j}$ , with  $\alpha_1 > \dots > \alpha_n$  and  $\{A_i\}_{i=1}^n$  a partition of  $S$  in  $\Sigma$ . For every  $p \in \Delta^\sigma(\Sigma, q)$ , define  $\psi_p \in B_0(\Sigma)$  by

$$\psi_p(s) = \sum_{i:q(A_i)>0} \frac{p(A_i)}{q(A_i)} 1_{A_i} + \sum_{i:q(A_i)=0} 1_{A_i}$$

It is easy to see that  $\psi_p$  is non-negative and  $\int \psi_p dq = 1$ . Call  $p'$  the element of  $\Delta^\sigma(\Sigma, q)$  such that  $dp'/dq = \psi_p$ . Since  $\psi_p$  is a simple function, then  $D_\phi(p'||q) = \int \phi(\psi_p) dq \in \mathbb{R}$ , so that  $p' \in \text{dom } D_\phi(\cdot||q)$ . Since  $\int \varphi dp' = \int \varphi dp$ , we have  $\{\int \varphi dp : p \in \Delta^\sigma(\Sigma, q)\} \subseteq \{\int \varphi dp : p \in \text{dom } D(\cdot||q)\}$ , which yields (38) as the converse inclusion is trivial.

It remains to prove that

$$\inf_{p \in \Delta^\sigma(q)} \int \varphi dp = \text{ess min}_{s \in S} \varphi(s). \quad (39)$$

Here  $\text{ess min}_{s \in S} \varphi(s) = \min \{\alpha_i : q(A_i) > 0\}$ . Let  $i^* \in \{1, \dots, n\}$  be such that  $\alpha_{i^*} = \min \{\alpha_i : q(A_i) > 0\}$ , and let  $q_{A_{i^*}}$  be the conditional distribution of  $q$  on  $A_{i^*}$ . Then,  $q_{A_{i^*}} \in \Delta^\sigma(q)$ , and  $\int \varphi dq_{A_{i^*}} = \text{ess min}_{s \in S} \varphi(s)$ . This proves (39).

(ii) Let  $\theta_n \uparrow \infty$ . By Proposition 23, strict convexity of  $\phi$  implies that of  $D_\phi^w(p||q)$  on its effective domain. Hence,  $\arg \min \theta_n D_\phi^w(p||q) = \{q\}$  for each  $n$ , and so the result follows from Proposition 12.  $\blacksquare$

**Proof of Proposition 23.** By Theorem 18, it is enough to show that  $D_\phi^w(\cdot||q) : \Delta(\Sigma) \rightarrow [0, \infty]$  is strictly convex on its effective domain. Let  $p, p' \in \text{dom } D_\phi^w(\cdot||q)$ ,  $p \neq p'$ , and  $\alpha \in (0, 1)$ . Let  $\mu(B) = \int_B w dq$  for all  $B \in \Sigma$ . The assumption  $\inf_{s \in S} w(s) > 0$  guarantees that the measure  $\mu$  is equivalent to the probability  $q$ . By (33),  $p, p' \in \text{dom } D_\phi^w(\cdot||q)$  iff  $\int \phi\left(\frac{dp}{dq}\right) d\mu, \int \phi\left(\frac{dp'}{dq}\right) d\mu < \infty$  iff  $\phi\left(\frac{dp}{dq}\right), \phi\left(\frac{dp'}{dq}\right) \in L^1(\Sigma, \mu)$ . Convexity and non-negativity of  $\phi$  guarantee that

$$0 \leq \phi\left(\alpha \frac{dp}{dq} + (1-\alpha) \frac{dp'}{dq}\right) \leq \alpha \phi\left(\frac{dp}{dq}(s)\right) + (1-\alpha) \phi\left(\frac{dp'}{dq}(s)\right) \quad \forall s \in S;$$

strict convexity,  $p \neq p'$ , and  $\alpha \in (0, 1)$  imply that the second inequality is strict on a set  $E \in \Sigma$  with  $q(E) > 0$ , and hence  $\mu(E) > 0$ . Therefore,  $\phi\left(\alpha \frac{dp}{dq} + (1-\alpha) \frac{dp'}{dq}\right) \in L^1(\Sigma, \mu)$  and

$$\int \left[ \alpha \phi\left(\frac{dp}{dq}(s)\right) + (1-\alpha) \phi\left(\frac{dp'}{dq}(s)\right) - \phi\left(\alpha \frac{dp}{dq} + (1-\alpha) \frac{dp'}{dq}\right) \right] d\mu > 0,$$

that is,  $D_\phi^w(\alpha p + (1 - \alpha)p' || q) < \alpha D_\phi^w(p || q) + (1 - \alpha) D_\phi^w(p' || q)$ , as desired.  $\blacksquare$

**Proof of Theorem 24.** The functional  $J : B_0(\Sigma) \rightarrow \mathbb{R}$  defined by  $J(\varphi) = \int \varphi dq - (2\theta)^{-1} \text{Var}_q(\varphi)$  is concave and Gateaux differentiable. Concavity is trivial. Moreover, for all  $\varphi, \psi \in B_0(\Sigma)$  and  $t \in \mathbb{R}$

$$\begin{aligned} J(\varphi + t\psi) &= \int (\varphi + t\psi) dq - \frac{1}{2\theta} \text{Var}_q(\varphi + t\psi) \\ &= -\frac{1}{2\theta} \text{Var}_q(\psi) t^2 + \left[ \int \psi \left( 1 - \frac{1}{\theta} \left( \varphi - \int \varphi dq \right) \right) dq \right] t + J(\varphi), \end{aligned}$$

so that

$$J'(\varphi; \psi) = \lim_{t \rightarrow 0} \frac{J(\varphi + t\psi) - J(\varphi)}{t} = \int \psi \left( 1 - \frac{1}{\theta} \left( \varphi - \int \varphi dq \right) \right) dq. \quad (40)$$

That is, for all  $\varphi \in B_0(\Sigma)$  the Gateaux differential of  $J$  at  $\varphi$  is represented by a measure with Radon-Nikodym derivative with respect to  $q$  given by  $1 - \frac{1}{\theta}(\varphi - \int \varphi dq)$ . Therefore,  $J'(\varphi; \cdot)$  is positive as a linear functional on  $B_0(\Sigma)$  iff

$$q \left( \left\{ s \in S : \varphi(s) - \int \varphi dq \leq \theta \right\} \right) = 1.$$

This relation characterizes the elements of the domain of monotonicity  $M$ . In [38] we show that

$$J(\varphi) = \min_{p \in \Delta^\sigma(q)} \left( \int \varphi dp + \theta G(p || q) \right), \quad \forall \varphi \in M. \quad \blacksquare$$

## References

- [1] C.D. Aliprantis and K.C. Border (1999), *Infinite dimensional analysis*, Springer, New York.
- [2] F. J. Anscombe and R. J. Aumann, A definition of subjective probability, *Annals of Mathematical Statistics*, 34, 199–205, 1963.
- [3] K. Arrow, *Essays in the theory of risk-bearing*, North-Holland, Amsterdam, 1970.
- [4] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces, *Communications in Contemporary Mathematics*, 3, 615–647, 2001.
- [5] A. Ben-Tal, The entropic penalty approach to stochastic programming, *Mathematics of Operations Research*, 10, 263–279, 1985.

- [6] A. Ben-Tal, A. Ben-Israel, and M. Teboulle, Certainty equivalents and information measures: duality and extremal principles, *Journal of Mathematical Analysis and Applications*, 157, 211–236, 1991.
- [7] P. Billingsley, *Probability and Measure, 3rd Edition*, Wiley, New York, 1995.
- [8] A. Chateauneuf, F. Maccheroni, M. Marinacci and J.-M. Tallon, Monotone continuous multiple priors, *Economic Theory*, 26, 973–982, 2005.
- [9] Z. Chen and L.G. Epstein, Ambiguity, risk, and asset returns in continuous time, *Econometrica*, 70, 1403–1443, 2002.
- [10] K.M. Chong and N.M. Rice, Equimeasurable rearrangements of functions, *Queens Papers in Pure and Applied Mathematics*, 28, 1971.
- [11] K.M. Chong, Doubly stochastic operators and rearrangement theorems, *Journal of Mathematical Analysis and Applications*, 56, 309–316, 1976.
- [12] G. Dal Maso, *An introduction to  $\Gamma$ -convergence*, Birkhäuser, Boston, 1993.
- [13] R-A. Dana, A representation result for concave Schur concave functions, mimeo, 2005.
- [14] C. Dellacherie and P-A. Meyer, *Probability and potential C*, North Holland, Amsterdam, 1988.
- [15] Dolecki, S. and G.H. Greco (1995). Niveloids, *Topological Methods in Nonlinear Analysis*, 5, 1–22.
- [16] J. Dow and S. Werlang, Uncertainty aversion, risk aversion, and the optimal choice of portfolio, *Econometrica*, 60, 197-204, 1992.
- [17] N. Dunford and J.T. Schwartz, *Linear Operators, Part I*, Wiley, New York, 1958.
- [18] P. Dupuis and R. S. Ellis, *A weak convergence approach to the theory of large deviations*, Wiley, New York, 1997.
- [19] D. Ellsberg, Risk, ambiguity, and the Savage axioms, *Quarterly Journal of Economics*, 75, 643–669, 1961.
- [20] L.G. Epstein, A definition of uncertainty aversion, *Review of Economic Studies*, 66, 579–608, 1999.
- [21] L.G. Epstein and T. Wang, Intertemporal asset pricing under Knightian uncertainty, *Econometrica*, 62, 283–322, 1994.



- [22] H. Föllmer and A. Schied, *Stochastic finance*, de Gruyter, Berlin, 2002.
- [23] P. Ghirardato and M. Marinacci, Ambiguity made precise: a comparative foundation, *Journal of Economic Theory*, 102, 251–289, 2002.
- [24] I. Gilboa and D. Schmeidler, Maxmin expected utility with a non-unique prior, *Journal of Mathematical Economics*, 18, 141–153, 1989.
- [25] P. Groeneboom, J. Oosterhoff, and F. H. Ruymgaart, Large deviation theorems for empirical probability measures, *Annals of Probability*, 7, 553–586, 1979.
- [26] L. Hansen and T. Sargent, Wanting robustness in macroeconomics, mimeo, 2000.
- [27] L. Hansen and T. Sargent, Robust control and model uncertainty, *American Economic Review*, 91, 60–66, 2001.
- [28] I.N. Herstein and J. Milnor, An axiomatic approach to measurable utility, *Econometrica*, 21, 291–297, 1953.
- [29] M. Hsu, M. Bhatt, R. Adolphs, D. Tranel and C. F. Camerer, Neural systems responding to degrees of uncertainty in human decision-making, *Science*, 310, 1680–1683, 2005.
- [30] G. Keren and L. E. Gerritsen, On the robustness and possible accounts of ambiguity aversion, *Acta Psychologica*, 103, 149–172, 1999.
- [31] A. Kühberger and J. Perner, The role of competition and knowledge in the Ellsberg task, *Journal of Behavioral Decision Making*, 16, 181–191, 2003.
- [32] I. Kopylov, Procedural rationality in the multiple prior model, mimeo, 2001.
- [33] F. Liese and I. Vajda, *Convex Statistical Distances*, Teubner, Leipzig, 1987.
- [34] W.A.J. Luxemburg, Rearrangement invariant Banach function spaces, *Queens Papers in Pure and Applied Mathematics*, 10, 83–144, 1967.
- [35] F. Maccheroni, M. Marinacci and A. Rustichini, Niveloids and their extensions, mimeo, 2005.
- [36] F. Maccheroni, M. Marinacci and A. Rustichini, Dynamic variational preferences, *Journal of Economic Theory*, forthcoming.
- [37] F. Maccheroni, M. Marinacci and A. Rustichini, Asset pricing with variational preferences, in progress, 2006.

- [38] F. Maccheroni, M. Marinacci, A. Rustichini, and M. Taboga, Portfolio selection with monotone Mean-Variance preferences, mimeo, 2004.
- [39] M.J. Machina and D. Schmeidler, A more robust definition of subjective probability, *Econometrica*, 60, 745–780, 1992.
- [40] H. M. Markowitz, Portfolio selection, *Journal of Finance*, 7, 77–91, 1952.
- [41] M. Marinacci, Probabilistic sophistication and multiple priors, *Econometrica* 70, 755–764, 2002.
- [42] A.W. Marshall and I. Olkin, *Inequalities: theory of majorization and its applications*, Academic Press, New York, 1979.
- [43] R. R. Phelps, *Convex functions, monotone operators and differentiability*, Springer-Verlag, New York, 1992.
- [44] R. T. Rockafellar, *Convex analysis*, Princeton Univ. Press, Princeton, 1970.
- [45] M. Rothschild and J.E. Stiglitz, Increasing risk: I. A definition, *Journal of Economic Theory*, 2, 225–243, 1970.
- [46] A. Rustichini, Emotion and reason in making decisions, *Science*, 310, 1624–1625, 2005.
- [47] D. Schmeidler, A bibliographical note on a theorem of Hardy, Littlewood, and Polya, *Journal of Economic Theory*, 20, 125–128, 1979.
- [48] D. Schmeidler, Subjective probability and expected utility without additivity, *Econometrica*, 57, 571–587, 1989.
- [49] J. Tobin, Liquidity preference as behavior toward risk, *Review of Economic Studies*, 25, 65–86, 1958.
- [50] T. Wang, A class of multi-prior preferences, mimeo, 2003.