

AMBIGUITY AVERSION, ROBUSTNESS, AND THE VARIATIONAL REPRESENTATION OF PREFERENCES

BY FABIO MACCHERONI, MASSIMO MARINACCI, AND ALDO RUSTICHINI¹

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.

KEYWORDS: Ambiguity aversion, model uncertainty, robustness.

1. INTRODUCTION

AMBIGUITY HAS BEEN A CLASSIC ISSUE in decision theory since the seminal work of Ellsberg (1961). 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 (1989) (also known as maxmin expected utility preferences). Agents with these preferences rank payoff profiles f according to the criterion

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

where C is a given convex subset of the set Δ 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.

¹An 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, Ben Polak, 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 is thanked 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-04-52477).

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 because 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 (1994) and Chen and Epstein (2002)). More recently, a different alternative has been explored, starting with the work of Hansen and Sargent (2000, 2001) 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 choice criterion

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

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

Agents who behave according to this choice criterion are considering the possibility that q may not be the appropriate law that governs the phenomenon in which they are interested 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 θ reflects the weight that agents are giving to the possibility that q might not be the correct model. As the parameter θ becomes larger, agents focus more on q as the correct model, giving less importance to possible alternative models p (Proposition 22).

Hansen and Sargent (2000) 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

²Hansen and Sargent (2001) also considered a class of multiple priors preferences with $C = \{p \in \Delta : R(p \parallel q) \leq \eta\}$, which they called *constraint preferences*.

to the motivation that underlies multiple priors preferences. So far, however, this connection has been stated simply 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 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

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

where $\delta_C : \Delta \rightarrow [0, \infty]$ is the indicator function of C (in the sense of convex analysis; see Rockafellar (1970)) given by

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

Like the relative entropy, the indicator function also is a convex function defined on the simplex Δ .

This reformulation clarifies the formal connection with the multiplier preferences in (2) and suggests the general representation

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

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.

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 (1989). 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 $\theta R(p \parallel q)$ 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 (1989), 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 robust preferences that makes 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 into 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).

Although 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 this 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 that deal 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 (1952) and Tobin (1958). Recall that mean–variance preferences are represented by the preference functional

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

In Theorem 24, we show that, on the domain of monotonicity of V , the equality

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

holds, where $G(\cdot \parallel q) : \Delta \rightarrow [0, \infty]$ is the relative Gini concentration index (see Section 4.2.2 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 that deal 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*

In addition to Schmeidler’s (1989) 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: those proposed by Epstein (1999) and by Ghirardato and Marinacci (2002). The key difference in the two approaches lies in the different notion of ambiguity neutrality they use: while Ghirardato and Marinacci (2002) identify ambiguity neutrality with subjective expected utility, Epstein (1999) more generally identifies ambiguity neutrality with probabilistic sophistication. In a nutshell, Ghirardato and Marinacci (2002) claim 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 consider only subjective expected utility preferences as ambiguity neutral preferences.³

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. Although in the paper we adopt the view and terminology of Ghirardato and Marinacci (2002), 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 Epstein (1999), in the same way as it has been done in Epstein (1999) for multiple priors preferences.

To clarify this issue further, in Section 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,

³We refer to Epstein (1999) and Ghirardato and Marinacci (2002) for a detailed presentation and motivation of their approaches. Notice that the notion of ambiguity aversion in Ghirardato and Marinacci (2002) is what provides a foundation for the standard comparative statics exercises in ambiguity for multiple priors preferences that are based on the size of the set of priors.

all examples of our preferences will include special cases of probabilistically sophisticated preferences. For instance, both multiplier and mean–variance preferences are easily seen to be examples of probabilistically sophisticated divergence preferences, although Example 17 shows that this is not the case for general divergence preferences, even for those 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 who has to make choices with only limited information and without a full understanding of what is going on. Some recent psychological literature (e.g., Keren and Gerritsen (1999), Kühberger and Perner (2003)) 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 modeling such psychological trait. In fact, agents who rank payoff profiles according to (4) can be viewed as believing they are 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 (2005) and Rustichini (2005)); for this reason, we do not expatiate further on this interpretation.

The paper is organized as follows. After introducing the setup 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 Schmeidler (1989) 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 [Appendixes](#).

2. SETUP

Consider a set S of *states of the world*, an algebra Σ 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 Σ -measurable. Moreover, we denote by $B_0(\Sigma)$ the set of all real-valued Σ -measurable simple functions, so that $u(f) \in B_0(\Sigma)$ whenever $u: X \rightarrow \mathbb{R}$, and we denote 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 that yields 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 happens in the classic setting of Anscombe and Aumann (1963). 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 :

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

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

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

AXIOM 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.

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

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

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

AXIOM A.6—Nondegeneracy: $f \succ g$ 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 Axioms A.1 and A.4 require preferences to be complete, transitive, and monotone. The latter requirement is basically a state-independence condition, which says that decision makers always (weakly) prefer acts that deliver statewise (weakly) better payoffs, regardless of the state where the better payoffs occur. If a preference relation \succsim satisfies Axioms A.1, A.3, and A.4, then each act $f \in \mathcal{F}$ admits a certainty equivalent $x_f \in X$. (See the proof of Lemma 28 in Appendix B.)

Axiom A.5, due to Schmeidler (1989), is a smoothing axiom that can be interpreted as an ambiguity aversion axiom, as discussed at length in Gilboa and Schmeidler (1989), Schmeidler (1989), Epstein (1999), and Ghirardato and Marinacci (2002).

Axiom A.2 is a weak independence axiom—a variation of an axiom of Gilboa and Schmeidler (1989). It requires independence only with respect to mixing with constant acts, provided the mixing weights are kept constant. Axiom A.2 is weaker than the original axiom of Gilboa and Schmeidler (1989), 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 the rest of this subsection to Axiom A.2.

Consider the following stronger version of Axiom A.2:

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

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

Axiom A.2' is the original axiom of Gilboa and Schmeidler (1989). The next lemma shows how it strengthens Axiom A.2.

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

$$\begin{aligned} \alpha f + (1 - \alpha)x \succsim \alpha g + (1 - \alpha)x \\ \Rightarrow \quad \beta f + (1 - \beta)y \succsim \beta g + (1 - \beta)y. \end{aligned}$$

Axiom A.2 is therefore the special case of Axiom A.2' in which the mixing coefficients α and β are required to be equal. At a conceptual level, Lemma 1 shows that Gilboa and Schmeidler's (1989) certainty independence axiom actually involves two types of independence: independence 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. 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 α and β are very different, say α is close to 1 and β 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 test the plausibility of Axioms A.2 and A.2', is by running experiments. This is possible because both Axioms A.2 and A.2' have clear behavioral implications. For instance, the following (thought) experiment gives a simple testable way to compare Axioms A.2 and A.2', and running this type of experiments will be the subject of future research.

EXAMPLE 2: Consider an urn that contains 90 black and white balls in unknown proportion, and the 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 Axioms A.1 and A.3–A.6, and she displays constant relative risk aversion $\gamma \in (0, 1)$.⁴ If her preferences satisfy Axiom 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 Axiom A.2 and not Axiom 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 Axiom A.2, but it would reveal a violation of Axiom A.2'.

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

⁴Notice that, for all $t > \tau > 0$ and γ , there exist $f, g, x, y, \alpha, \beta$ as discussed in the text such that $\alpha f + (1 - \alpha)x \sim f_t$, $\alpha g + (1 - \alpha)x \sim g_t$, $\beta f + (1 - \beta)y \sim f_\tau$, and $\beta g + (1 - \beta)y \sim g_\tau$.

3.2. Main Result

We can now state our main result, which characterizes preferences that satisfy Axioms A.1–A.6. Here $\Delta = \Delta(\Sigma)$ denotes the set of all finitely additive probabilities on Σ endowed with the weak* topology⁵ 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) *The relation \succsim satisfies Axioms A.1–A.6.*
- (ii) *There exists a nonconstant affine function $u: X \rightarrow \mathbb{R}$ and a grounded, convex, and lower semicontinuous function $c: \Delta \rightarrow [0, \infty]$ such that, for all $f, g \in \mathcal{F}$,*

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

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

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

The representation (6) involves the minimization of a convex lower semicontinuous function, 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 $c^* = \alpha c_0^*$.*

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

⁵That 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 $A \in \Sigma$.

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

As shown in Lemma 29 in Appendix B, the assumption that $u(X)$ is unbounded (above or below) is equivalent to the following axiom (see Kopylov (2001)).

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

We call the variational preferences that satisfy Axiom A.7 *unbounded*.

3.3. Ambiguity Attitudes

We now study the ambiguity attitudes featured by variational preferences. We follow the approach proposed in Ghirardato and Marinacci (2002), 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 $x \in X$,

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

To introduce an absolute notion of ambiguity aversion, as in Ghirardato and Marinacci (2002) 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.*

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

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) *The relation \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 that illustrate 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 because, by Proposition 7, all variational preferences are ambiguity averse. Therefore, the least ambiguity averse functions c^* are those associated with SEU preferences. As 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 denote by c_m^* the maximal ambiguity index of Example 9. Given $\alpha \in (0, 1)$, suppose the right-hand side of (6) is

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

which is the well known ε -contaminated model. In this case,

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

and this is a simple example of an index c^* that does not display 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 ∞ . 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 $\arg \min c_n \subseteq \arg \min c_{n+1}$.⁶

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

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

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

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

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

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 $\bigcup_n \text{dom } c_n$, whereas in (10) it 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 $\bigcup_n \text{dom } c_n$ is very large.

We close with few remarks. First, observe that Lemma 32 in Appendix B shows that the set that Ghirardato and Marinacci (2002) call 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 Rockafellar (1970)), Example 11 can be immediately generalized as follows: the ambiguity index of a convex combination of preference functionals that represent unbounded variational preferences is given by the infimal convolution of their ambiguity indices.

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

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

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

a type of result that can be important in some applications to enable dealing 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 Δ makes it possible to use the powerful de Giorgi–Wijsman theory of Γ convergence (often called epiconvergence; see, e.g., Dal Maso (1993)) to study the behavior of the sequence $\{V_n(f)\}_n$ and, for example, to determine the conditions under which 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 Δ 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 therefore for the multiplier preferences of Hansen and Sargent (2001) and for the mean–variance preferences of Markowitz (1952) and Tobin (1958). For this reason, here we consider this technical extension.

Fortunately, in our setting we can still use the monotone continuity axiom introduced by Arrow (1970) to derive a SEU representation with a countably additive subjective probability (see Chateauneuf, Maccheroni, Marinacci, and Tallon (2005)).

AXIOM 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 $x E_{n_0} f \succ g$.*

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 σ -algebra Σ , while $\Delta^\sigma(q)$ denotes the subset of Δ^σ that consists 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) *The relation \succsim satisfies Axiom A.8.*
- (ii) *For each $t \geq 0$, $\{p \in \Delta : c^*(p) \leq t\}$ is a (weakly) compact subset of Δ^σ .*

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

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

Lemma 30 in Appendix B 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 the variational preferences that satisfy Axiom A.8 *continuous*.

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 (1992) that some authors (notably Epstein (1999)) identify with ambiguity neutrality (or absence of ambiguity altogether).

Our main finding is that what makes a variational preference probabilistically 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 preferences characterized by a suitable symmetry property of the associated ambiguity indices.

Specifically, given a countably additive probability q on the σ -algebra Σ , a preference relation \succsim is *probabilistically sophisticated* (with respect to q) if, given f and g in \mathcal{F} ,

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

For example, if S is finite and q is uniform, then probabilistic sophistication amounts to permutation invariance (i.e., each act is indifferent to all its permutations).

A preference \succsim satisfies (*first order*) *stochastic dominance* (with respect to q) if, given f and g in \mathcal{F} ,

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

Notice that a preference that satisfies stochastic dominance is probabilistically sophisticated.

To characterize probabilistic sophistication, we need to introduce a few well known notions from the theory of stochastic orders (see, e.g., Chong and Rice (1971) and Schmeidler (1979)). Define a partial order \succ_{cx} on $\Delta^\sigma(q)$ by

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

for every convex function ϕ on \mathbb{R} . This is the so-called *convex order* on probability distributions, and by classic results of Rothschild and Stiglitz (1970) and Marshall and Olkin (1979) 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 with respect to q . That is, given p and p' in $\Delta^\sigma(q)$,

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

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* (with respect to q) if $\text{dom } c \subseteq \Delta^\sigma(q)$ and, given p and p' in $\Delta^\sigma(q)$,

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

whereas it is *Shur convex* (with respect to q) if $\text{dom } c \subseteq \Delta^\sigma(q)$ and, given p and p' in $\Delta^\sigma(q)$,

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

A Shur convex function is clearly rearrangement invariant, whereas the converse is, in general, false. A subset C of $\Delta^\sigma(q)$ is *Shur convex* (with respect to q) if its indicator function δ_C is Shur convex.⁷ Finally, we say that q in Δ^σ is *adequate* if either q is nonatomic 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 that characterizes 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 (with respect to q):*

- (i) *The relation \succsim satisfies stochastic dominance.*
- (ii) *The relation \succsim is probabilistically sophisticated.*
- (iii) *The function c^* is rearrangement invariant.*
- (iv) *The function c^* is Shur convex.*

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

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

The proof of Theorem 14 builds on the results of Luxemburg (1967) and Chong and Rice (1971), as well as on some recent elaborations of these results provided by Dana (2005).

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 δ_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 $p \in C$).⁸

To further illustrate Theorem 14, we now introduce a new class of variational preferences that plays an important role in the rest of the paper. As before, assume there is an underlying probability measure $q \in \Delta^\sigma$. Given a Σ -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 ϕ divergence of $p \in \Delta$ with respect to q is given by

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

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 $D_\phi(p \parallel q)$ 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 (1987)). The two most important divergences are the *relative entropy* (or *Kullback–Leibler divergence*) given by $\phi(t) = t \ln t - t + 1$ and the *relative Gini concentration index* (or χ^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 $D_\phi^w(\cdot \parallel q): \Delta \rightarrow [0, \infty]$ is a grounded, convex, and lower semicontinuous function, and the sets*

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

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

⁸This fact can be used to provide an alternative derivation of some of the results of Marinacci (2002), which showed that multiple priors preferences that are probabilistically sophisticated reduce to subjective expected utility when there exists even a single nontrivial unambiguous event. The direct proofs in Marinacci (2002) are, however, shorter and more insightful for the problem with which that article was dealing.

Thanks to the foregoing properties, preferences represented by the functional

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

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

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

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

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

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

Divergence preferences are a very important class of variational preferences and we further study them in the next section. 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 (1985) and Ben-Tal, Ben-Israel, and Teboulle (1991). 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

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

are examples of probabilistically sophisticated variational preferences.⁹

However, the next example shows that even under minimal nonuniformities of the weighting function, divergence preferences are in general not probabilistically sophisticated and they exhibit Ellsberg-type behavior. Therefore,

⁹Analogously, Theorem 14 and Lemma 15 show that all multiple priors preferences with sets of priors $\{p \in \Delta^\sigma(q) : D_\phi(p \parallel q) \leq \eta\}$ are probabilistically sophisticated. These preferences include the constraint preferences of Hansen and Sargent (2001), mentioned in footnote 2.

the probabilistic sophistication of divergence preferences crucially depends on the uniformity of the weight w .

EXAMPLE 17: Consider a standard Ellsberg three color urn, with 30 red balls and 60 balls either green or blue. As usual, consider the 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 one dollar if a red ball is drawn and nothing otherwise, f_G pays one dollar if a green ball is drawn and nothing otherwise, and so on. As is well known, Ellsberg (1961) argued that most subjects rank these acts as

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

Consider a decision maker who has divergence preferences represented by the preference functional V given by (16). Here it is natural to consider a uniform q on the three states. Without loss of generality, 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

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

which is only slightly nonuniform. 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 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} .

To state our results, we need some standard notions of calculus in vector spaces (see Rockafellar (1970) and Phelps (1992)). 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, $\partial V(f)$ consists only 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 $f \in B_0(\Sigma)$,*

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

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

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

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 ϕ is strictly convex. By Theorem 18, they are everywhere differentiable.

Variational preferences that feature 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, although there are plenty of examples of variational preferences that are not everywhere differentiable, since the convex combination of a convex cost function and a strictly convex function is strictly convex, one can approximate arbitrarily well any variational preference with another one that is everywhere differentiable.

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 possible the explicit resolution of optimal problems based on the variational preference functional V . It is worth observing that in Maccheroni, Marinacci, and Rustichini (2006) 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 Epstein and Wang (1994)). By fully characterizing the differentiability of variational preferences, Theorem 18 clarifies the scope of these results. In particular, it shows that although 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 in the preceding text that it is always possible to approximate any variational preference arbitrarily closely with another one that is everywhere differentiable.

¹⁰Here $u'(f)dr$ 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 $\bigcup_{f \in B_0(\Sigma)} \arg \min_{p \in \Delta} (\int u(f) dp + c^*(p))$. Clearly, if c is strictly convex, then a fortiori it is essentially strictly convex.

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 (1989) and the divergence preferences we just introduced. In particular, we show that two important classes of preferences—the multiplier preferences of Hansen and Sargent (2001) and the mean–variance preferences of Markowitz (1952) and Tobin (1958), are special cases of divergence, and therefore variational, preferences.

4.1. Multiple Priors Preferences

Begin with the multiple priors choice model axiomatized by Gilboa and Schmeidler (1989). As we mentioned in Section 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 relationship between Axiom A.2' and the variational formula (6).

In particular, when Axiom A.2' replaces Axiom A.2, the only probabilities in Δ that “matter” in the representation (6) are those to which the decision maker attributes “maximum weight,” that is, those 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) *The relation \succsim satisfies Axiom A.2'.*
- (ii) *For all $f \in \mathcal{F}$,*

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

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

- (iii) *The function c^* takes on only values 0 and ∞ .*

The characterization of the multiple priors model via Axioms A.1, A.2', and A.3–A.6 is due to Gilboa and Schmeidler (1989). Proposition 19 shows how the multiple priors model fits in the representation we established in Theorem 3.

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

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

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

In terms of our representation, by Theorem 3 and Proposition 19 we can make the following statement:

COROLLARY 20: Let \succsim be a variational preference. The following conditions are equivalent:

- (i) The relation \succsim satisfies Axiom A.5'.
- (ii) The relation \succsim is SEU.
- (iii) The relation \succsim satisfies Axiom A.2' and $\{p \in \Delta : c^*(p) = 0\}$ is a singleton. If, in addition, \succsim is unbounded, then (iii) is also equivalent to the statement:
- (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 so as 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

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

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

By Theorem 16, all divergence preferences are examples of continuous variational preferences. As a result, 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 depend only on the parameter θ , which can therefore be interpreted as a coefficient of ambiguity aversion.

COROLLARY 21: Given two (w, ϕ) divergence preferences \succsim_1 and \succsim_2 , the following conditions are equivalent:

- (i) The relation \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 θ becomes closer and closer to 0 (closer and closer to ∞ , resp.).

The limit cases where θ goes either to 0 or to ∞ are described by Proposition 12, which takes an especially stark form for divergence preferences under

¹¹When we want to be specific about w and ϕ , we speak of (w, ϕ) divergence preferences.

some very mild assumptions. To see this, we need a piece of notation: given a simple measurable function $\varphi: S \rightarrow \mathbb{R}$, set

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

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

$$(21) \quad \operatorname{ess\,min}_{s \in S} \varphi(s) = \min_{s \in \operatorname{supp}(q)} \varphi(s).$$

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

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

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

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

When $w \in L^\infty(q)$, divergence preferences therefore tend more and more, as $\theta \rightarrow 0$, to rank acts according to the very cautious criterion given by $\operatorname{ess\,min}_{s \in S} u(f(s))$. In contrast, when ϕ is strictly convex, divergence preferences tend more and more, as $\theta \rightarrow \infty$, 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 ϕ is strictly convex (most examples of divergence D_ϕ^w satisfy this condition; see, e.g., Liese and Vajda (1987)). 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 (1994, 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 ϕ is strictly convex, then $D_\phi^w(\cdot \parallel q) : \Delta(\Sigma) \rightarrow [0, \infty]$ is strictly convex on its effective domain and the variational preference functional $V : B_0(\Sigma) \rightarrow \mathbb{R}$, given by

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

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

$$(22) \quad V'(f; h) = \int h u'(f) dr,$$

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

Consider, for example, multiplier preferences, a special class of divergence preferences in which $c^*(p) = \theta R(p \parallel q)$. By some well known properties of the relative entropy (see Dupuis and Ellis (1997, p. 34)), formula (22) takes, in this case, the neat form

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

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 Axiom A.2'). We hope that this will be achieved in later work and in this regard it is worth observing that the ambiguity index $D_\phi^w(p \parallel q)$ is additively separable, a strong structural property.

After having established the main properties of divergence preferences, we now move on 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

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

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

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

Since the entropy $R^w(p \parallel q)$ is a special case of divergence $D_\phi^w(p \parallel q)$ 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

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

When w is uniform, \succsim is probabilistically sophisticated and it features (a positive multiple of) the standard relative entropy $R(\cdot \parallel q)$ as the index of ambiguity aversion. This is the case considered by Hansen and Sargent (see, e.g., Hansen and Sargent (2000, 2001)), who called this class of entropic preferences in which acts are ranked according to

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

multiplier preferences. Although multiplier preferences are probabilistically sophisticated, Example 17 shows that this is not the case for general entropic preferences that have nonuniform weighting functions w . These “nonuniform” 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 θ as follows: the lower θ is, the more ambiguity averse is the entropic preference. The parameter θ can therefore be interpreted as a coefficient of ambiguity aversion.

We have shown how entropic, and hence 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 Axioms 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 that characterizes 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. Although our model is static, in the follow-up paper (Maccheroni, Marinacci, and Rustichini (2006)), 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 is often the case in choice theory, also here the analysis of the static model is key to paving the way for its dynamic extension.

We close by discussing the related work of Wang (2003), who (in a quite different setting) recently proposed an axiomatization of a class of preferences that include multiplier preferences as special cases. He considered 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 axiomatized the representation

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

His modeling 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 Wang (2003), 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 that consist 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 that have as ambiguity index the sum of δ_C and $\theta R(\cdot \parallel q)$. 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

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

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

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

Like the weighted relative entropy $R^w(p \parallel q)$, the Gini index $G^w(p \parallel q)$ also is a special case of divergence $D_\phi^w(p \parallel q)$ defined in (14), with $\phi(t) =$

$2^{-1}(t-1)^2$.¹² As a result, by Theorem 16, Gini preferences are continuous variational preferences, with ambiguity index $\theta G^w(\cdot \parallel q)$.

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^{mv} ; that is, for all $f, g \in B_0(\Sigma)$,

$$\begin{aligned} f \succsim^{\text{mv}} g &\Leftrightarrow \min_{p \in \Delta^\sigma(q)} \left(\int f \, dp + \theta G(p \parallel q) \right) \\ &\geq \min_{p \in \Delta^\sigma(q)} \left(\int g \, dp + \theta G(p \parallel q) \right). \end{aligned}$$

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. However, 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 (1952) and Tobin (1958) that are defined on $B_0(\Sigma)$ by

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

where Var is the variance with respect to q . These preferences are not monotone unless their domain is suitably restricted to 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 the 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.

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

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

Moreover,

$$\begin{aligned} f \succsim^{\text{mv}} g &\Leftrightarrow \min_{p \in \Delta^\sigma(q)} \left(\int f \, dp + \theta G(p \parallel q) \right) \\ &\geq \min_{p \in \Delta^\sigma(q)} \left(\int g \, dp + \theta G(p \parallel q) \right) \end{aligned}$$

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

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 that satisfy monotonicity. This insight is developed at length in Maccheroni, Marinacci, Rustichini, and Taboga (2004), and we refer the interested reader to that paper for a detailed analysis of this class of variational preferences.

Istituto di Metodi Quantitativi and IGIER, Università Bocconi, Milano, 20100 Italy,

*Dipartimento di Statistica e Matematica Applicata and Collegio Carlo Alberto, Università di Torino, Torino, 10122 Italy; massimo.marinacci@unito.it,
and*

Dept. of Economics, University of Minnesota, Minneapolis, MN 55455, U.S.A.

Manuscript received August, 2004; final revision received April, 2006.

APPENDIX A: NIVELOIDS

The set of all functions in $B_0(\Sigma)$ (resp. $B(\Sigma)$) that take 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., Dunderf and Schwartz (1958, p. 258)).

As is 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 if and only if 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 nonempty subset Φ 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 (1995). 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 if and only if, for all $\varphi, \psi \in B_0(\Sigma, K)$, $k \in K$, and $\alpha \in (0, 1)$,*

- (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 if and only if

- (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 (2005). 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 Dolecki and Greco (1995)). Moreover:

- If Φ is convex and I is concave, then \hat{I} is concave.
- If $\Phi + \mathbb{R} \supseteq B_0(\Sigma)$,¹³ then \hat{I} is the unique niveloid on $B(\Sigma)$ that extends I (see Maccheroni, Marinacci, and Rustichini (2005)).

If Φ is convex and $I: \Phi \rightarrow \mathbb{R}$ is a concave niveloid, direct application of the Fenchel–Moreau theorem (see, e.g., Phelps (1992, p. 42)) to \hat{I} guarantees that

$$\hat{I}(\varphi) = \min_{\mu \in ba(\Sigma)} (\langle \varphi, \mu \rangle - \hat{I}^*(\mu)), \quad \text{where } \hat{I}^*(\mu) = \inf_{\psi \in B(\Sigma)} (\langle \psi, \mu \rangle - \hat{I}(\psi))$$

is the Fenchel conjugate of \hat{I} . If μ 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,

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

see also Föllmer and Schied (2002, Theorem 4.12). Set $I^*(p) = \hat{I}^*(p)$ for each $p \in \Delta(\Sigma)$ and set $\partial_\pi I(\varphi) = \{p \in \Delta(\Sigma) : I(\psi) - I(\varphi) \leq \langle \psi - \varphi, p \rangle \text{ for each } \psi \in \Phi\}$.

The next two results are proved in Maccheroni, Marinacci, and Rustichini (2005).

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

- For each $p \in \Delta(\Sigma)$, $I^*(p) = \inf_{\psi \in \Phi} (\langle \psi, p \rangle - I(\psi))$.
- $I^*: \Delta(\Sigma) \rightarrow [-\infty, 0]$ is concave and weakly* upper semicontinuous.
- 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.
- $\partial_\pi I(k1_S) = \{I^* = 0\} = \arg \max_{p \in \Delta(\Sigma)} I^*(p)$ for all $k \in \mathbb{R}$ such that $k1_S \in \Phi$.

¹³ $\Phi + \mathbb{R}$ 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.

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

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

Moreover, if $\Phi + \mathbb{R} \supseteq B_0(\Sigma)$, I^* is the unique concave and weakly* upper semicontinuous 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

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

PROPOSITION 27: Let $I: B_0(\Sigma, K) \rightarrow \mathbb{R}$ be a normalized concave niveloid, with K unbounded and Σ a σ -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) The set $\{p \in \Delta(\Sigma) : I^*(p) \geq b\}$ 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 $\varphi \in B_0(\Sigma, K)$,

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

APPENDIX 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 Axioms A.1–A.6, then there exists a nonconstant affine function $u: X \rightarrow \mathbb{R}$ and a normalized and concave niveloid $I: B_0(\Sigma, u(X)) \rightarrow \mathbb{R}$ such that $f \succsim g$ if and only if $I(u(f)) \geq I(u(g))$. 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 proof of Lemma 1 is omitted.

LEMMA 28: A binary relation \succsim on \mathcal{F} satisfies Axioms A.1–A.4 and A.6 if and only if there exist a nonconstant affine function $u: X \rightarrow \mathbb{R}$ and a normalized niveloid $I: B_0(\Sigma, u(X)) \rightarrow \mathbb{R}$ such that

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

PROOF: Assume that \succsim on \mathcal{F} satisfies Axioms 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 $\frac{1}{2}x + \frac{1}{2}z \approx \frac{1}{2}y + \frac{1}{2}z$, with-

out loss of generality $\frac{1}{2}x + \frac{1}{2}z \succ \frac{1}{2}y + \frac{1}{2}z$; by Axiom 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 we conclude that $x \succ y$, which is absurd. Then the hypotheses of the mixture space theorem (Hernstein and Milnor (1953)) are satisfied and there exists an affine function $u: X \rightarrow \mathbb{R}$ such that $x \succsim y$ if and only if $u(x) \geq u(y)$. By Axiom 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 Axiom 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 because 1 belongs to the first and 0 belongs to the second; their union is the whole $[0, 1]$. Because $[0, 1]$ 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 $U(f) = u(x_f)$. The function U is well defined because $f \sim x_f$ 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, $f \succsim g$ if and only if $x_f \succsim x_g$ if and only if $u(x_f) \geq u(x_g)$ if and only if $U(f) \geq U(g)$. Therefore, U represents \succsim .

If $f \in \mathcal{F}$, then $u(f) \in B_0(\Sigma, u(X))$. Conversely, if $\varphi \in B_0(\Sigma, u(X))$, then $\varphi(s) = u(x_i)$ 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 Σ . 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, $u(f) = u(g)$ if and only if $u(f(s)) = u(g(s))$ for all $s \in S$ if and only if $f(s) \sim g(s)$ for all $s \in S$, and, by Axiom 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 $\varphi = u(f)$, $\psi = u(g) \in B_0(\Sigma, u(X))$, 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 $k = u(x)$. Then $I(k1_S) = I(u(x)) = U(x) = u(x) = k$. Therefore, I is normalized.

Take $\alpha \in (0, 1)$, $\varphi = u(f) \in B_0(\Sigma, u(X))$, and $k = u(x_k) \in u(X)$; 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 Axiom A.2, $\alpha z + (1 - \alpha)x_k \sim \alpha f + (1 - \alpha)x_k$ and

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

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

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

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

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 niveloid¹⁴ 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 Axiom A.1.

If $f, g \in \mathcal{F}$, $x, y \in X$, and $\alpha \in (0, 1)$, 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 Axiom 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 $\alpha v(f) + (1 - \alpha)v(g) = v(\alpha f + (1 - \alpha)g)$. The inequality $J(v(\alpha_n f + (1 - \alpha_n)g)) \geq J(v(h))$ for all $n \geq 1$ and the continuity of J guarantee $J(v(\alpha f + (1 - \alpha)g)) \geq J(v(h))$. Therefore, $\{\alpha \in [0, 1]: \alpha f +$

¹⁴In 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 $t \in v(X)$, $J(t) = I(t - c) + c = t - c + c = t$.

$(1 - \alpha)g \succsim h\}$ is closed. A similar argument shows that $\{\alpha \in [0, 1]: h \succsim \alpha f + (1 - \alpha)g\}$ too is closed, and Axiom A.3 holds.

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

Finally, because v is not constant and it represents \succsim on X , there exist $x \succ y$ and Axiom A.6 holds too. *Q.E.D.*

PROOFS OF THEOREM 3 AND OF PROPOSITION 6: Assume \succsim satisfies Axioms A.1–A.6. By Lemma 28, there is a nonconstant affine function $u: X \rightarrow \mathbb{R}$ and a normalized niveloid $I: B_0(\Sigma, u(X)) \rightarrow \mathbb{R}$ such that $f \succsim g$ if and only if $I(u(f)) \geq I(u(g))$.

Next we show that Axiom 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 Axiom 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 nonnegative, 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

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

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 $I(\varphi) = \min_{p \in \Delta(\Sigma)} (\int \varphi dp + c(p))$ are, respectively, a nonconstant affine function and a normalized niveloid that represents \succsim . By Lemma 28, \succsim satisfies Axioms A.1–A.4 and A.6. Concavity of I guarantees Axiom A.5.¹⁵ Q.E.D.

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),

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

as desired. The converse is trivial. Q.E.D.

LEMMA 29: Let \succsim be a binary relation on X represented by an affine function $u: X \rightarrow \mathbb{R}$. The set $u(X)$ is unbounded (either below or above) if and only if \succsim satisfies Axiom 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 $I_i(\varphi) = \min_{p \in \Delta(\Sigma)} (\int \varphi dp + c_i^*(p))$ for all $\varphi \in B_0(\Sigma, u_i(X))$.

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

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

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

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

for all $p \in \Delta(\Sigma)$.

Q.E.D.

PROOF OF PROPOSITION 12: Observe that the functions c_n are weakly* lower semicontinuous on Δ 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 $\{\int u(f) \, dp + c_n(p)\}_n$ pointwise converges to

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

Hence, by Dal Maso (1993, Proposition 5.7), this sequence Γ -converges to

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

the weakly* lower semicontinuous envelope of (29). By Dal Maso (1993, Theorem 7.4), this implies

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

(ii) Since $\arg \min c_n = \{p \in \Delta : c_n(p) = 0\}$, the increasing sequence $\{\int u(f) \, dp + c_n(p)\}_n$ pointwise converges (and so, by Dal Maso (1993, p. 47), Γ -converges) to

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

Because Δ is weak-* compact, Dal Maso (1993, Theorem 7.8) implies

$$\begin{aligned} &\liminf_n \min_{p \in \Delta} \left\{ \int u(f) \, dp + c_n(p) \right\} \\ &= \min_{p \in \Delta} \left\{ \int u(f) \, dp + \delta_{\bigcap_n \arg \min c_n}(p) \right\}. \end{aligned} \quad \text{\textit{Q.E.D.}}$$

PROOF OF THEOREM 13: Let (u, c^*) represent \succsim as in Theorem 3 and set, for all $\varphi \in B_0(\Sigma, u(X))$, $I(\varphi) = \min_{p \in \Delta(\Sigma)} (\int \varphi dp + c^*(p))$. It is easy to check that \succsim satisfies Axiom A.8 on \mathcal{F} if and only if I satisfies condition (i) of Proposition 27 on $B_0(\Sigma, u(X))$. Unboundedness of $u(X)$ and the relation $c^* = -I^*$ allow us to apply Proposition 27 and to obtain the desired equivalence. *Q.E.D.*

LEMMA 30: Let \succsim be a variational preference that satisfies Axiom A.8. Then

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

PROOF: Let (u, c^*) represent \succsim as in Theorem 3 and set $I(\varphi) = \min_{p \in \Delta(\Sigma)} (\int \varphi dp + c^*(p))$ 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

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

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 Axiom 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, $\lim_{n \rightarrow \infty} I(\varphi 1_{E_n^c} + (\min \varphi - \varepsilon) 1_{E_n}) > k_j$ because 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

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

and $p(E_n) \rightarrow 0$ (uniformly with respect to $p \in \partial_\pi I(\varphi)$); 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,

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

coincide on $\text{int}(B_0(\Sigma, u(X)))$ and, being continuous, they coincide on $B_0(\Sigma, u(X))$. *Q.E.D.*

The results of Section 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 set $\Gamma_\varphi(b) = \inf\{t \in \mathbb{R} : G_\varphi(t) \leq b\}$ for each $b \in [0, 1]$.

The function G_φ is the *survival function* of φ , and Γ_φ is the *decreasing rearrangement* of φ . Two functions $\varphi, \psi \in L^1(\Sigma, q)$ are *equimeasurable* if $G_\varphi = G_\psi$ (if and only if $\Gamma_\varphi = \Gamma_\psi$). We refer to Chong and Rice (1971) for a comprehensive study of equimeasurability. The preorder \prec_{cx} defined in Section 3.5 can be naturally regarded as a relation on $L^1(\Sigma, q)$ by putting $\varphi \prec_{cx} \psi$ if and only if $\int \phi(\varphi) dq \leq \int \phi(\psi) dq$ for every convex ϕ 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 $\varphi \prec_{<cx} \psi$ if and only if $\int_0^t \Gamma_\varphi(b) db \leq \int_0^t \Gamma_\psi(b) db$ for all $t \in [0, 1]$. Then $\varphi \prec_{<cx} \psi$ if and only if $\varphi \prec_{cx} \psi$ and $\int \varphi dq = \int \psi dq$ (see, e.g., Chong (1976)). Moreover, the equimeasurability relation is the common symmetric part \sim_{cx} of $\prec_{<cx}$ and \prec_{cx} . Simple manipulation of the results of Luxemburg (1967, pp. 125–126) yields the following lemma:

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 $\varphi \prec_{<cx} \varphi'$ implies $H(\varphi) \leq H(\varphi')$.*

PROOF: Let $\varphi \in L^1(\Sigma, q)$. By Luxemburg (1967, Theorem 9.1) for all $\psi \in B_0(\Sigma, K)$,

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

Therefore, $H(\varphi) \leq \sup_{\psi \in B_0(\Sigma, K)} (\int_0^1 \Gamma_\varphi(t) \Gamma_\psi(t) dt - T(\psi))$. 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., Chong and Rice (1971, pp. 57–58)) delivers $H(\varphi) \leq H(\varphi')$. Let $\varphi \prec_{<cx} \varphi'$. By Chong (1976, Theorem 1.1), there is a nonnegative $\varphi'' \in L^1(\Sigma, q)$ such that $\varphi + \varphi'' \prec_{cx} \varphi'$. If H is monotonic, then $H(\varphi) \leq H(\varphi + \varphi'') \leq H(\varphi')$. Q.E.D.

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 $\varphi \prec_{cx} \varphi'$ implies $H(\varphi) \leq H(\varphi')$. (If T is Shur convex and q is not adequate, use Chong and Rice (1971, Theorem 13.8) rather than Luxemburg (1967, Theorem 9.1).) See also Dana (2005).

PROOF OF THEOREM 14: 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 $x \in X$. We prove (i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (i).

(i) \Rightarrow (ii) This step can be proved by a routine argument.

(ii) \Rightarrow (iii) Let \succsim be a continuous unbounded variational preference, which is also probabilistically sophisticated 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, without loss of generality assume $u(X) \supseteq (-\infty, 0]$. Let $x_n \in u^{-1}(-n)$ and $y \in u^{-1}(0)$. Consider the act $f_n = x_n A y$ and the constant act y . Since $q_{f_n} = \delta_y = q_y$, by probabilistic sophistication, $f_n \sim y$ 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, without loss of generality assume $u(X) \supseteq [0, \infty)$. Let $x_n \in u^{-1}(n)$ and $y \in u^{-1}(0)$. 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., Chong and Rice (1971, p. 12)), there exist $x_1, \dots, x_n \in X$, with $u(x_1) > \dots > u(x_n)$, and two partitions $\{A_1, A_2, \dots, A_n\}$ and $\{B_1, B_2, \dots, B_n\}$ of S in Σ , with $q(A_i) = q(B_i)$ for all $i = 1, 2, \dots, n$, such that q -a.s.

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

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 probabilistic sophistication 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 $p \in \Delta^\sigma(\Sigma, q)$,

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

By Lemma 31, $H(\varphi) = \sup_{\psi \in B_0(\Sigma, -u(X))} (\int \varphi \psi \, dq - T(\psi))$ 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

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

where p_ψ 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 $\psi \in P$ if and only if $\psi' \in P$ (see, e.g., Chong and Rice (1971, pp. 15–16)). If $\psi, \psi' \notin P$, then $T(\psi) = \infty = T(\psi')$. If $\psi, \psi' \in P$, then ψ and ψ' are the Radon–Nikodym derivatives of p_ψ 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 Δ^σ for each $t \geq 0$; therefore, $\{\psi \in L^1(\Sigma, q) : T(\psi) \leq t\}$ is a weakly compact subset of $L^1(\Sigma, q)$ and a fortiori T is weakly lower semicontinuous. Since obviously T is convex, Luxemburg (1967, Theorem 13.3) guarantees that T (hence c^*) is Shur convex.¹⁶

(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).¹⁷ Set, for all $\varphi \in$

¹⁶*Different definitions caveat:* Luxemburg (1967, pp. 123–124) defined Shur convexity of T by rearrangement invariance, convexity, and weak lower semicontinuity. His Theorem 13.3 shows that if T has these features, then $\varphi \prec_{cx} \varphi'$ implies $T(\varphi) \leq T(\varphi')$.

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

$B_0(\Sigma)$, $I(\varphi) = \min_{p \in \Delta(\Sigma)} (\int \varphi dp + c(p))$. Then

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

where T is defined as 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., Chong and Rice (1971, 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. Next we show that \succsim satisfies stochastic dominance. Assume that $q(\{s \in S : f(s) \precsim x\}) \leq q(\{s \in S : g(s) \precsim x\})$ 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., Chong and Rice (1971, pp. 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 $f \succsim g$. Q.E.D.

PROOF OF LEMMA 15: Groundedness (in particular, $D_\phi^w(q \parallel q) = 0$) and convexity are trivial, so is Shur convexity if w is uniform. Weak* lower semicontinuity descends from the fact that the sets $\{p \in \Delta : D_\phi^w(p \parallel q) \leq t\}$ 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 Σ -measurable function w with $\min_{s \in S} w(s) > 0$, without requiring $\int w dq = 1$. Set $\underline{w} = \min_{s \in S} w(s)$ and without loss of generality suppose $t \geq 0$. By definition, $\{p \in \Delta : D_\phi^w(p \parallel q) \leq t\} \subseteq \Delta^\sigma(q)$, hence

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

Denote the set on the right-hand side by \mathcal{D} . We show that (a) $\lim_{q(B) \rightarrow 0} p(B) = 0$ uniformly with respect to $p \in \mathcal{D}$ 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., Dunford and Schwartz (1958, Chapter IV)).¹⁸

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

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

¹⁸Part (a) guarantees that \mathcal{D} is relatively sequentially weakly compact; that is, every sequence $\{p_n\}$ in \mathcal{D} admits a weakly convergent subsequence $\{p_{n_j}\}$. Part (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.

Next we show that $\lim_{q(B) \rightarrow 0} p(B) = 0$ uniformly with respect to $p \in \mathcal{C}$ and a fortiori with respect to $p \in \mathcal{D}$. Let $p \in \mathcal{C}$. Clearly,

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

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

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

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

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

This concludes the proof of (a).

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

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

Denote by Π_w the set of all finite partitions π of S in Σ finer than $\pi_w = \{w^{-1}(b) : b \in w(S)\}$. For all $\pi \in \Pi_w$, define on Δ^σ the function

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

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 because π is finer than π_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: We have $D_\phi^w(p \parallel q) = \sup_{\pi \in \Pi_w} D_\phi^w(p_\pi \parallel q_\pi)$ for all $p \in \Delta^\sigma$.

PROOF: Let $p \ll q$. Take an increasingly finer sequence π_n in Π_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 (1995, p. 470)), hence μ -a.e. Continuity of ϕ and the Fatou lemma imply

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

The converse inequality is guaranteed if $\int \phi\left(\frac{dp}{dq}\right) d\mu = \infty$. Else, if $\phi\left(\frac{dp}{dq}\right)$ is μ -integrable, the conditional Jensen 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 $q(B) = 0$, and we can assume it belongs to some $\pi \in \Pi_w$. Therefore,

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

This concludes the proof of the claim.

Q.E.D.

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

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

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

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

This completes the proof when w is simple. Suppose now that w is any Σ -measurable function with $\underline{w} = \inf_{s \in S} w(s) > 0$. Then there exists a sequence of simple Σ -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, $D_\phi^{w_n}(p \parallel q) \uparrow D_\phi^w(p \parallel q)$ for all $p \in \Delta^\sigma(q)$, therefore

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

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

Q.E.D.

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

Q.E.D.

PROOF OF THEOREM 18: Set $I(\varphi) = \min_{p \in \Delta(\Sigma)} (\int \varphi dp + c^*(p))$ for all $\varphi \in B_0(\Sigma)$. Because $-c^*$ 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

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

and I is Gateaux differentiable on $B_0(\Sigma)$ if and only if I^* is strictly concave on line segments in $\bigcup_{\varphi \in B_0(\Sigma)} \partial I(\varphi)$, that is, if and only if c^* is essentially strictly convex.¹⁹

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 Λ 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 additive set function $\int_\cdot u'(f) dp$ in $ba(\Sigma)$. Because this map is linear and $\sigma(ba(\Sigma), B_0(\Sigma))$ - $\sigma(ba(\Sigma), B_0(\Sigma))$ -continuous, and because $\partial I(u(f))$ is weakly* compact and convex, then Λ is $\sigma(ba(\Sigma), B_0(\Sigma))$ -compact and convex.

¹⁹This is proved in Bauschke, Borwein, and Combettes (2001). Notice that our definition of essential strict convexity is weaker than theirs and it guarantees that strict convexity implies essential strict convexity.

By standard results,

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

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

$$\begin{aligned} u(f + th)(s) &= u(f(s) + th(s)) \\ &= u(f(s)) + u'(f(s))th(s) + o(th(s)) \end{aligned}$$

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

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

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

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

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 . Because $u: \mathbb{R} \rightarrow \mathbb{R}$ 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,

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

Because 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 $r_1 \neq r_2$. Q.E.D.

PROOF OF PROPOSITION 19: Let (u, c^*) represent \succsim as in Theorem 3 and without loss of generality assume $[-1, 1] \subseteq u(X)$.

(i) \Rightarrow (ii) By Gilboa and Schmeidler (1989, Theorem 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,

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

Suppose $p \in C$. Then $c^*(p) \leq 0$. Because c^* is nonnegative, 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 $c^*(p) = 0$ if and only if $p \in C$. Therefore, for all $f \in \mathcal{F}$,

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

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

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

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

Suppose, $c^*(p_0) > 0$. There exist a nonnegative (resp. nonpositive), 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 $c^*(p_0) = \infty$.

Q.E.D.

The proof of Corollary 20 is omitted. Just notice that Axiom A.5' can be used to obtain affinity of the functional I that appears in Lemma 28 in the same way in which Axiom 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 let $q \in \Delta(\Sigma)$. The following conditions are equivalent:

- (i) q corresponds to a SEU preference that is 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) \Rightarrow (ii) Suppose \succsim_0 is a SEU preference, with associated subjective probability q and nonconstant affine 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, and $c^* \leq c_0^*$, and by Corollary 20, $c_0^*(q) = 0$; hence, $0 \leq c^*(q) \leq c_0^*(q) = 0$.

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

(iii) \Rightarrow (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 $\varphi \in B_0(\Sigma, u(X))$ and 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 nonempty and $\min_{p \in \Delta(\Sigma)} c^*(p) = 0$. Q.E.D.

PROOF OF PROPOSITION 22: (i) Let $\theta_n \downarrow 0$. By Proposition 12,

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

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

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

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

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

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 Σ . For every $p \in \Delta^\sigma(\Sigma, q)$, define $\psi_p \in B_0(\Sigma)$ by

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

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

It remains to prove that

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

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 ϕ implies that of $D_\phi^w(p \parallel q)$ on its effective domain. Hence, $\arg \min \theta_n D_\phi^w(p \parallel q) = \{q\}$ for each n and so the result follows from Proposition 12. Q.E.D.

PROOF OF PROPOSITION 23: By Theorem 18, it is enough to show that $D_\phi^w(\cdot \parallel q) : \Delta(\Sigma) \rightarrow [0, \infty]$ is strictly convex on its effective domain. Let $p, p' \in \text{dom } D_\phi^w(\cdot \parallel q)$, $p \neq p'$, 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 μ is equivalent to the probability q . By (33), $p, p' \in \text{dom } D_\phi^w(\cdot \parallel q)$ if and only if $\int \phi(\frac{dp}{dq}) d\mu, \int \phi(\frac{dp'}{dq}) d\mu < \infty$ if and only if $\phi(\frac{dp}{dq}), \phi(\frac{dp'}{dq}) \in L^1(\Sigma, \mu)$. Convexity and nonnegativity of ϕ guarantee that

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

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, $\phi(\alpha \frac{dp}{dq} + (1-\alpha) \frac{dp'}{dq}) \in L^1(\Sigma, \mu)$ and

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

that is, $D_\phi^w(\alpha p + (1-\alpha)p' \parallel q) < \alpha D_\phi^w(p \parallel q) + (1-\alpha) D_\phi^w(p' \parallel q)$, as desired. *Q.E.D.*

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 $t \in \mathbb{R}$,

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

so that

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

That is, for all $\varphi \in B_0(\Sigma)$, the Gateaux differential of J at φ 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 $B_0(\Sigma)$ if and only if

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

This relation characterizes the elements of the domain of monotonicity M . In Maccheroni, Marinacci, Rustichini, and Taboga (2004) we show that

$$J(\varphi) = \min_{p \in \Delta^\sigma(q)} \left(\int \varphi dp + \theta G(p \parallel q) \right) \quad \forall \varphi \in M. \quad Q.E.D.$$

REFERENCES

- ANScombe, F. J., and R. J. Aumann (1963): "A Definition of Subjective Probability," *The Annals of Mathematical Statistics*, 34, 199–205. [1453]
- ARROW, K. (1970): *Essays in the Theory of Risk-Bearing*. Amsterdam: North-Holland. [1460]
- BAUSCHKE, H. H., J. M. BORWEIN, and P. L. COMBETTES (2001): "Essential Smoothness, Essential Strict Convexity, and Legendre Functions in Banach Spaces," *Communications in Contemporary Mathematics*, 3, 615–647. [1491]
- BEN-TAL, A. (1985): "The Entropic Penalty Approach to Stochastic Programming," *Mathematics of Operations Research*, 10, 263–279. [1464]
- BEN-TAL, A., A. BEN-ISRAEL, and M. TEBoulLE (1991): "Certainty Equivalents and Information Measures: Duality and Extremal Principles," *Journal of Mathematical Analysis and Applications*, 157, 211–236. [1464]
- BILLINGSLEY, P. (1995): *Probability and Measure* (Third Ed.). New York: Wiley. [1489]
- CHATEAUNEUF, A., F. MACCHERONI, M. MARINACCI, and J.-M. TALLON (2005): "Monotone Continuous Multiple Priors," *Economic Theory*, 26, 973–982. [1460]
- CHEN, Z., and L. G. EPSTEIN (2002): "Ambiguity, Risk, and Asset Returns in Continuous Time," *Econometrica*, 70, 1403–1443. [1448]
- CHONG, K. M. (1976): "Doubly Stochastic Operators and Rearrangement Theorems," *Journal of Mathematical Analysis and Applications*, 56, 309–316. [1484]
- CHONG, K. M., and N. M. RICE (1971): "Equimeasurable Rearrangements of Functions," *Queens Papers in Pure and Applied Mathematics*, 28. [1461,1463,1484–1487]
- DAL MASO, G. (1993): *An Introduction to Γ -Convergence*. Boston: Birkhäuser. [1460,1482]
- DANA, R.-A. (2005): "A Representation Result for Concave Schur Concave Functions," *Mathematical Finance*, 15, 613–634. [1463,1485]
- DOLECKI, S., and G. H. GRECO (1995): "Niveloids," *Topological Methods in Nonlinear Analysis*, 5, 1–22. [1475,1476]
- DUNFORD, N., and J. T. SCHWARTZ (1958): *Linear Operators*, Part I. New York: Wiley. [1475, 1487]
- DUPUIS, P., and R. S. ELLIS (1997): *A Weak Convergence Approach to the Theory of Large Deviations*. New York: Wiley. [1471]
- ELLSBERG, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 643–669. [1447,1465]
- EPSTEIN, L. G. (1999): "A Definition of Uncertainty Aversion," *Review of Economic Studies*, 66, 579–608. [1451,1454,1461]
- EPSTEIN, L. G., and T. WANG (1994): "Intertemporal Asset Pricing under Knightian Uncertainty," *Econometrica*, 62, 283–322. [1448,1467,1470]
- FÖLLMER, H., and A. SCHIED (2002): *Stochastic Finance*. Berlin: de Gruyter. [1476]

- GHIRARDATO, P., AND M. MARINACCI (2002): "Ambiguity Made Precise: A Comparative Foundation," *Journal of Economic Theory*, 102, 251–289. [1451,1454,1457,1459]
- GILBOA, I., AND D. SCHMEIDLER (1989): "Maxmin Expected Utility with a Non-Unique Prior," *Journal of Mathematical Economics*, 18, 141–153. [1447,1449,1452,1454,1468,1493]
- HANSEN, L., AND T. SARGENT (2000): "Wanting Robustness in Macroeconomics," Mimeo, University of Chicago and Stanford University. [1448,1472]
- (2001): "Robust Control and Model Uncertainty," *American Economic Review*, 91, 60–66. [1448,1460,1464,1468,1472]
- HERSTEIN, I. N., AND J. MILNOR (1953): "An Axiomatic Approach to Measurable Utility," *Econometrica*, 21, 291–297. [1478]
- HSU, M., M. BHATT, R. ADOLPHS, D. TRANEL, AND C. F. CAMERER (2005): "Neural Systems Responding to Degrees of Uncertainty in Human Decision-Making," *Science*, 310, 1680–1683. [1452]
- KEREN, G., AND L. E. GERRITSEN (1999): "On the Robustness and Possible Accounts of Ambiguity Aversion," *Acta Psychologica*, 103, 149–172. [1452]
- KOPYLOV, I. (2001): "Procedural Rationality in the Multiple Prior Model," Mimeo, University of Rochester. [1457]
- KÜHBERGER, A., AND J. PERNER (2003): "The Role of Competition and Knowledge in the Ellsberg Task," *Journal of Behavioral Decision Making*, 16, 181–191. [1452]
- LIESE, F., AND I. VAJDA (1987): *Convex Statistical Distances*. Leipzig: Teubner. [1463,1470]
- LUXEMBURG, W. A. J. (1967): "Rearrangement Invariant Banach Function Spaces," *Queens Papers in Pure and Applied Mathematics*, 10, 83–144. [1463,1484–1486]
- MACCHERONI, F., M. MARINACCI, AND A. RUSTICHINI (2005): "Niveloids and Their Extensions," Mimeo. [1476]
- (2006): "Dynamic Variational Preferences," *Journal of Economic Theory*, 128, 4–44. [1467,1472]
- MACCHERONI, F., M. MARINACCI, A. RUSTICHINI, AND M. TABOGA (2004): "Portfolio Selection with Monotone Mean–Variance Preferences," ICER Working Papers, Applied Mathematics Series 27-2004. [1475,1497]
- MACHINA, M. J., AND D. SCHMEIDLER (1992): "A More Robust Definition of Subjective Probability," *Econometrica*, 60, 745–780. [1461]
- MARINACCI, M. (2002): "Probabilistic Sophistication and Multiple Priors," *Econometrica*, 70, 755–764. [1463]
- MARKOWITZ, H. M. (1952): "Portfolio Selection," *Journal of Finance*, 7, 77–91. [1450,1460,1468,1474]
- MARSHALL, A. W., AND I. OLKIN (1979): *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press. [1462]
- PHELPS, R. R. (1992): *Convex Functions, Monotone Operators and Differentiability*. New York: Springer-Verlag. [1466,1476]
- ROCKAFELLAR, R. T. (1970): *Convex Analysis*. Princeton, NJ: Princeton University Press. [1449,1459,1466]
- ROTHSCHILD, M., AND J. E. STIGLITZ (1970): "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 2, 225–243. [1462]
- RUSTICHINI, A. (2005): "Emotion and Reason in Making Decisions," *Science*, 310, 1624–1625. [1452]
- SCHMEIDLER, D. (1979): "A Bibliographical Note on a Theorem of Hardy, Littlewood, and Polya," *Journal of Economic Theory*, 20, 125–128. [1461]
- (1989): "Subjective Probability and Expected Utility Without Additivity," *Econometrica*, 57, 571–587. [1451,1454]
- TOBIN, J. (1958): "Liquidity Preference as Behavior Toward Risk," *Review of Economic Studies*, 25, 65–86. [1450,1460,1468,1474]
- WANG, T. (2003): "A Class of Multi-Prior Preferences," Mimeo, University of British Columbia. [1473]