A Subjective Model of Temporal Preferences

Haluk Ergin* Todd Sarver†

March 17, 2009

Preliminary Draft
(Please do not circulate)

Abstract

We study preferences for timing of resolution of objective uncertainty in a simple menu choice model with two stages of information arrival. We characterize two general classes of utility representations called linear hidden action representations. The representations can be interpreted as if an unobservable action is taken by the individual (or by the malevolent nature, depending on whether the preference is for early or late resolution of uncertainty) between the two periods.

We illustrate that our general representations allow for a richer class of preferences for timing of resolution of uncertainty than was possible in Kreps and Porteus (1978), and provide a unified framework for studying a variety of well-known preferences in the literature. We show that subjective versions of the class of multi-prior preferences (Gilboa and Schmeidler (1989)) and variational preferences (Maccheroni, Marinacci, and Rustichini (2006)) overlap with the class of linear hidden action preferences exhibiting a preference for late resolution of uncertainty. The costly contemplation model (Ergin and Sarver (2009)) is characterized as a special case of the class of linear hidden action preferences exhibiting a preference for early resolution of uncertainty. A generalization of the Kreps and Porteus (1978) model that allows for preference for flexibility in the sense of Dekel, Lipman, and Rustichini (2001) is also characterized.

*Correspondence address: Washington University in Saint Louis, Department of Economics, Campus Box 1208, Saint Louis, MO 63130. Email: hergin@artsci.wustl.edu.

†Correspondence address: Northwestern University, Department of Economics, 2001 Sheridan Road, Evanston, IL 60208. Email: tsarver@northwestern.edu.

1 Introduction

This paper considers several new classes of dynamic preferences, providing representations for preferences for both early and late resolution of uncertainty. These preferences are examined in a simple menu-choice model with two-stage objective uncertainty. We axiomatize two linear hidden action representations corresponding to preferences for early and late resolution of uncertainty, and we show that, under a minimality condition, these representations are uniquely identified from the preference. In other words, preferences for early and late resolution of uncertainty can be interpreted as arising from the presence of an unobserved (hidden) action that can be taken between the resolution of the first and second period objective uncertainty.

It is well known that an individual may prefer to have uncertainty resolve at an earlier date in order to be able to condition her future actions on the realization of this uncertainty. For example, an individual may prefer to have uncertainty about her future income resolve earlier so that she can smooth her consumption across time. Suppose an individual has the possibility of receiving a promotion with a substantial salary increase several years into the future. If she is able to learn the outcome of that promotion decision now, then even if she will not actually receive the increased income until a later date, she may choose to increase her current consumption by temporarily decreasing her savings or increasing her debt. On the other hand, if she is not told the outcome of the promotion decision, then by increasing her consumption now, she risks having larger debt and hence suboptimally low consumption in the future. In this example, changing the timing of the resolution of uncertainty benefits the individual by increasing her ability to condition her choices on the outcome of that uncertainty.

Kreps and Porteus (1978) considered a broader class of dynamic preferences that allow for a preference for early (or late) resolution of uncertainty even when the individual's ability to condition her actions of this uncertainty is unchanged. For example, suppose the individual described above has no current savings and is unable to take on debt. Then, even if she learns the outcome of the promotion decision now, she is unable to increase her consumption. Even in this case, the preferences considered by Kreps and Porteus (1978) allow the individual to have a strict preference for that uncertainty to resolve earlier (or later). This additional time-preference for the resolution of uncertainty has proved quite useful in applications, in particular, in macroeconomic models of asset pricing and business cycles.

The setting for our model is a simple two-stage version of the model considered by Kreps and Porteus (1978). However, we allow for more general axioms, which permits us to model a richer set of preferences exhibiting a preference for early or late resolution of uncertainty. In particular, we relax the Strategic Rationality Axiom of Kreps and Porteus (1978) (see Axiom 7) to allow for a preference for flexibility as in Kreps (1979) and Dekel, Lipman, and Rustichini (2001, henceforth DLR). We are able to represent this general class of preferences for early and late resolution of uncertainty as if there is an unobserved (hidden) action that can be taken between the resolution of the first and second period objective uncertainty. In the case of a preference for early resolution of uncertainty, this hidden action can be thought of as an action chosen by the individual. Thus, the individual prefers to have objective uncertainty resolve in the first period so that she can choose this action optimally. In the case of a preference for late resolution of objective uncertainty, this hidden action could be thought of as an action chosen by the (malevolent) nature. In this case, the individual prefers to have objective uncertainty resolve in the second period, after this action has been selected by nature, so as to mitigate natures ability to harm her.

This paper not only provides representations for a more general class of preferences for early and late resolution of uncertainty, but also provides new ways to understand and interpret the temporal preferences considered by Kreps and Porteus (1978). Our linear hidden action model is general enough to encompass the subjective versions of a number of well-known representations in the literature: the multiple priors model of Gilboa and Schmeidler (1989), the variational preferences model of Maccheroni, Marinacci, and Rustichini (2006), the costly contemplation model of Ergin and Sarver (2009), and a version of the temporal preferences model of Kreps and Porteus (1978) extended to allow for subjective uncertainty as in DLR (2001). We identify the preference for temporal resolution of uncertainty implied by each of these representations as well as their additional behavioral implications over the linear hidden action model. The general framework in this paper provides a unification of these well-known representations and provides simple axiomatizations.

2 The Model

Let Z be a finite set of alternatives, and let $\triangle(Z)$ denote the set of all probability distributions on Z, endowed with the Euclidean metric d (generic elements $p, q, r \in \triangle(Z)$). Let $\mathcal{A}$ denote the set of all closed subsets of $\triangle(Z)$, endowed with the Hausdorff metric:

$$d_h(A, B) = \max \left\{ \max_{p \in A} \min_{q \in B} d(p, q), \max_{q \in B} \min_{p \in A} d(p, q) \right\}.$$

Elements of $\mathcal{A}$ are called menus (generic menus $A, B, C \in \mathcal{A}$). Let $\triangle(\mathcal{A})$ denote the set of all Borel probability measures on $\mathcal{A}$, endowed with the weak* topology (generic elements $P, Q, R \in \triangle(\mathcal{A})$).[1] The primitive of the model is a binary relation $\succsim$ on $\triangle(\mathcal{A})$, representing the individual's preferences over lotteries over menus.

We interpret $\succsim$ as corresponding to the individual's choices in the first period of a two-period decision problem. In the beginning of period 1, the individual chooses a lottery P over menus. Later in period 1, the uncertainty associated with the lottery P is resolved and P returns a menu A. In the unmodeled period 2, the individual chooses a lottery p out of A. Later in period 2, the lottery p resolves and returns an alternative z. We will refer to the uncertainty associated with the resolution of P as the *first-stage uncertainty*, and will refer to the uncertainty associated with the resolution of p as the *second-stage uncertainty*. Although the period 2 choice is unmodeled, it will be important for the interpretation of the representations.

Our model is a special case of Kreps and Porteus (1978) with only two periods and no consumption in period 1.[2] A lottery $P \in \triangle(\mathcal{A})$ over menus is a *temporal lottery* (Kreps and Porteus (1978)) if P returns a singleton menu with probability one. An individual facing a temporal lottery makes no choice in period 2, between the resolution of first and second stages of the uncertainty. Note that the set of temporal lotteries can be naturally associated with $\triangle(\triangle(Z))$.

For any $A, B \in \mathcal{A}$ and $\alpha \in [0, 1]$, the convex combination of these two menus is defined by $\alpha A + (1 - \alpha)B \equiv \{\alpha p + (1 - \alpha)q : p \in A \text{ and } q \in B\}$. We let $\text{co}(A)$ denote the convex hull of the menu A and let $\delta_A \in \triangle(\mathcal{A})$ denote the degenerate lottery that puts probability 1 on the menu A. Then, $\alpha\delta_A + (1 - \alpha)\delta_B$ denotes the lottery that puts probability α on the menu A and probability $1 - \alpha$ on the menu B. Finally, for any continuous function $V : \mathcal{A} \to \mathbb{R}$ and $P \in \triangle(\mathcal{A})$, we let $\mathbb{E}_P[V]$ denote the expected value of V under the lottery P, i.e., $\mathbb{E}_P[V] = \int_{\mathcal{A}} V(A)\, P(dA)$.

3 General Representations

We will impose the following set of axioms in all the representation results in the paper. Therefore, it will be convenient to refer to them altogether as Axiom 1.

[1] Given a metric space X, the weak* topology on the set of all finite signed Borel measures on X is the topology where a net of signed measures $\{\mu_d\}_{d \in D}$ converges to a signed measure μ if and only if $\int_X f\, \mu_d(dx) \to \int_X f\, \mu(dx)$ for every bounded continuous function $f : X \to \mathbb{R}$.

[2] The same model is also used in Epstein and Seo (2007) and in Section 4 of Epstein, Marinacci and Seo (2007).

Axiom 1

1. (Weak Order): $\succsim$ *is complete and transitive.*

2. (Continuity): *The upper and lower contour sets,* $\{P \in \triangle(\mathcal{A}) : P \succsim Q\}$ *and* $\{P \in \triangle(\mathcal{A}) : P \precsim Q\}$*, are closed in the weak* topology.*

3. (First-Stage Independence): *For any* $P, Q, R \in \triangle(\mathcal{A})$ *and* $\alpha \in (0,1)$*,*

$$P \succ Q \quad \Rightarrow \quad \alpha P + (1-\alpha)R \succ \alpha Q + (1-\alpha)R.$$

4. (L–Continuity): *There exist* $A^*, A_* \in \mathcal{A}$ *and* $M \geq 0$ *such that for every* $A, B \in \mathcal{A}$ *and* $\alpha \in [0,1]$ *with* $\alpha \geq M d_h(A,B)$*,*

$$(1-\alpha)\delta_A + \alpha\delta_{A^*} \succsim (1-\alpha)\delta_B + \alpha\delta_{A_*}.$$

5. (Indifference to Randomization (IR)): *For every* $A \in \mathcal{A}$*,* $\delta_A \sim \delta_{co(A)}$*.*

Axioms 1.1 and 1.2 are standard. Axiom 1.3 is the von Neumann-Morgenstern independence axiom imposed with respect to the first-stage uncertainty. Axioms 1.1–1.3 ensure that there exists a continuous function $V : \mathcal{A} \to \mathbb{R}$ such that $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$. Given Axioms 1.1–1.3, Axiom 1.4 is a technical condition implying the Lipschitz continuity of V.[3] Axiom 1.5 is introduced in DLR (2001). It is justified if the individual choosing from the menu A in period 2 can also randomly select an alternative from the menu, for example, by flipping a coin. In that case, the menus A and co(A) offer the same set of options, and hence they are identical from the perspective of the individual.

The next two axioms from Kreps and Porteus (1978) will be key in our representation results.

Axiom 2 (Preference for Early Resolution of Uncertainty (PERU)) *For any* $A, B \in \mathcal{A}$ *and* $\alpha \in (0,1)$*,*

$$\alpha\delta_A + (1-\alpha)\delta_B \succsim \delta_{\alpha A + (1-\alpha)B}.$$

Axiom 3 (Preference for Late Resolution of Uncertainty (PLRU)) *For any* $A, B \in \mathcal{A}$ *and* $\alpha \in (0,1)$*,*

$$\delta_{\alpha A + (1-\alpha)B} \succsim \alpha\delta_A + (1-\alpha)\delta_B.$$

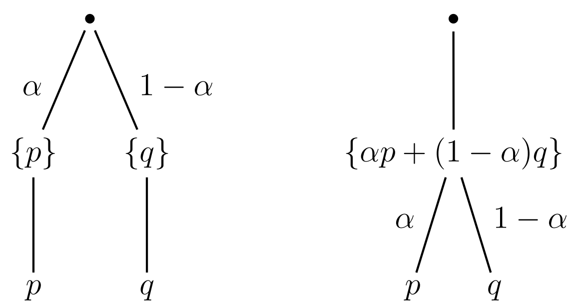

Figure 1: Timing of Resolution of Uncertainty with Temporal Lotteries

To understand Axioms 2 and 3, suppose first that $A = \{p\}$ and $B = \{q\}$ for some $p, q \in \triangle(Z)$. In this case, period 2 choice out of a menu is degenerate, and the two equations above are statements about the preference between the temporal lotteries $\alpha\delta_{\{p\}} + (1-\alpha)\delta_{\{q\}}$ and $\delta_{\alpha\{p\}+(1-\alpha)\{q\}}$. The temporal lottery $\alpha\delta_{\{p\}} + (1-\alpha)\delta_{\{q\}}$ corresponds to the first tree in Figure 1, in which the uncertainty regarding whether lottery p or q is selected resolves in period 1. The temporal lottery $\delta_{\alpha\{p\}+(1-\alpha)\{q\}}$ corresponds to the second tree in Figure 1, in which the same uncertainty resolves in period 2.[4] PERU requires a weak preference the first temporal lottery, whereas PLRU requires a weak preference for the second temporal lottery.

In the general case where A and B need not be singletons, the interpretations of PERU and PLRU are more subtle, since $\delta_{\alpha A+(1-\alpha)B}$ and $\alpha\delta_A + (1-\alpha)\delta_B$ involve non-degenerate period 2 choices. Note that in $\alpha\delta_A + (1-\alpha)\delta_B$, the uncertainty regarding whether the menu A or B is selected resolves in period 1, before the individual makes her choice out of the selected menu. On the other hand, in $\delta_{\alpha A+(1-\alpha)B}$ no uncertainty is resolved in period 1. The period 2 choice of a lottery $\alpha p + (1-\alpha)q$ from the convex combination menu $\alpha A + (1-\alpha)B$ is identical to a pair of choices $p \in A$ and $q \in B$, where after the individual chooses (p, q), p is selected with probability α and q is selected with probability $1-\alpha$. Therefore, the period 2 choice out of the menu $\alpha A + (1-\alpha)B$ can be interpreted as a complete contingent plan out of the menus A and B.

The key distinction between the two lotteries over menus is that in $\delta_{\alpha A+(1-\alpha)B}$ the period 2 choice is made prior to the resolution of the uncertainty regarding whether the choice from A or the choice from B will be implemented, whereas in $\alpha\delta_A + (1-\alpha)\delta_B$,

[3]In models with preferences over menus over lotteries, analogous L–continuity axioms can be found in and Dekel, Lipman, Rustichini, and Sarver (2007, henceforth DLRS), Sarver (2008), and Ergin and Sarver (2009).

[4]In both temporal lotteries, the remaining uncertainty, i.e., the outcome of p conditional on p being selected and the outcome of q conditional on q being selected, is resolved in period 2.

the same uncertainty is resolved in period 1 before the individual makes a choice out of the selected menu. Therefore, PERU can be interpreted as the individual's preference to make a choice out of the menu after learning which menu is selected, whereas PLRU is the individual's preference to make her choice out of the menus before learning which menu is selected.

We will also characterize the special case of our representation where the individual has a weak preference for larger menus.

Axiom 4 (Monotonicity) *For any $A, B \in \mathcal{A}$, $A \subset B$ implies $\delta_B \succsim \delta_A$.*

Since expected-utility functions on $\triangle(Z)$ are equivalent to vectors in $\mathbb{R}^Z$, we will use the notation $u(p)$ and $u \cdot p$ interchangeably for any expected utility function $u \in \mathbb{R}^Z$. We define the set of *normalized (non-constant) expected-utility functions* on $\triangle(Z)$ to be

$$\mathcal{U} = \left\{ u \in \mathbb{R}^Z : \sum_{z \in Z} u_z = 0, \sum_{z \in Z} u_z^2 = 1 \right\}.$$

We are ready to introduce our general representations.

Definition 1 A *Maximum [Minimum] Linear Hidden Action (max-LHA [min-LHA]) representation* is a pair $(\mathcal{M}, c)$ consisting of a compact set of finite signed Borel measures $\mathcal{M}$ on $\mathcal{U}$ and a lower semi-continuous function $c : \mathcal{M} \to \mathbb{R}$ such that:

1. $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$, where $V : \mathcal{A} \to \mathbb{R}$ is defined by Equation (1) [(2)]:
$$V(A) = \max_{\mu \in \mathcal{M}} \left(\int_{\mathcal{U}} \max_{p \in A} u(p)\, \mu(du) - c(\mu) \right) \tag{1}$$
$$V(A) = \min_{\mu \in \mathcal{M}} \left(\int_{\mathcal{U}} \max_{p \in A} u(p)\, \mu(du) + c(\mu) \right). \tag{2}$$

2. The set $\mathcal{M}$ is *minimal*: For any compact proper subset $\mathcal{M}'$ of $\mathcal{M}$, the function V' obtained by replacing $\mathcal{M}$ with $\mathcal{M}'$ in Equation (1) [(2)] is different from V.

The pair $(\mathcal{M}, c)$ is an *LHA representation* if it is a max-LHA or a min-LHA representation. An LHA representation $(\mathcal{M}, c)$ is *monotone* if all measures in $\mathcal{M}$ are positive.

Before we interpret the representations, we will first argue that after appropriately renormalizing the set of ex post utility functions, one can reinterpret the integral term in Equations (1) and (2) as an expectation. More specifically, suppose that μ is a nonnegative measure with $\lambda = \mu(\mathcal{U}) > 0$. Consider the probability measure π on $\mathcal{V} = \lambda\mathcal{U}$ which (heuristically) puts $\mu(u)/\lambda$ weight on each $v = \lambda u \in \mathcal{V}$. Then, by a simple change of variables, the above integral can be rewritten as the expectation $\int_{\mathcal{V}} \max_{p \in A} v(p)\pi(dv)$ which can be interpreted as follows. The individual anticipates that her ex post utility functions will be distributed according to π. Conditional on each realization of her ex post utility $v \in \mathcal{V}$, she will chose a lottery in A that maximizes v. She aggregates across different possible realizations of v by taking expectation with respect to π.[5] Therefore, each measure μ can be interpreted as a reduced-form representation of the individual's subjective uncertainty about her ex post (period 2) utility function over $\triangle(Z)$. We conjecture that analogous interpretations are possible if μ is a signed measure, by introducing regret and temptation to the current framework.[6]

We next interpret Equation (1). In period 1, the individual anticipates that after the first-stage uncertainty is resolved but before she makes her choice in period 2, she will be able to select an action μ from a set $\mathcal{M}$. Each action μ affects the distribution of the individual's ex post utility functions over $\triangle(Z)$, at cost $c(\mu)$. As argued in the above paragraph, the integral in Equation (1) can be interpreted as a reduced-form representation for the value of the action μ when the individual chooses from menu A, which is linear in the menu A. For each menu A, the individual maximizes the value minus cost of her action.

The interpretation of Equation (2) is dual. In period 1, the individual anticipates that after the first-stage uncertainty is resolved but before she makes her choice in period 2, the (malevolent) nature will select an action μ from a set $\mathcal{M}$. The individual anticipates the nature to chose an action which minimizes the value to the individual plus a cost term. The function c can be interpreted as being related to the pessimism attitude of the individual. For constant c, she expects the nature to chose an action that outright minimizes her utility from a menu. Different cost functions put different restrictions on the individual's perception of the malevolent nature's objective.[7]

In the above representations, both the set of available actions and and their costs are subjective in that they are part of the representation. Therefore $\mathcal{M}$ and c are

[5] It is noteworthy that any probability measure π' such that μ and π' are absolutely continuous with respect to each other can be used in the above argument. The reason is that, heuristically, in the LHA-representations, the weight $\mu(u)$ on the ex post utility function u captures both the cardinality of the ex post utility and its probability, and these two effects cannot be separated from behavior as in models with state dependent utility. See Kreps (1988) for an elaborate discussion of the state-dependence issue.

[6] See Sarver (2008), Gul and Pesendorfer (2001), and DLR (2008).

[7] This interpretation is due to Maccheroni, Marinacci, and Rustichini (2006) which we discuss in more detail in the following subsection.

not directly observable to the modeler and need to be identified from the individual's preferences. Note that in both Equations (1) and (2), it is possible to enlarge the set of actions by adding a new action μ to the set $\mathcal{M}$ at a prohibitively high cost $c(\mu)$ without affecting the equations. Therefore, in order to identify $(\mathcal{M}, c)$ from the preference, we also impose an appropriate minimality condition on the set $\mathcal{M}$.

We postpone more concrete interpretations of the set of actions and costs to the discussion of the applications of LHA-representations in the following section. We are now ready to state our general representation result.

Theorem 1 *A. The preference $\succsim$ has a max-LHA [min-LHA] representation if and only if it satisfies Axiom 1 and PERU [PLRU].*

B. The preference $\succsim$ has a monotone max-LHA [min-LHA] representation if and only if it satisfies Axiom 1, PERU [PLRU], and monotonicity.[8]

The special case of LHA representations satisfying indifference to timing of resolution of uncertainty (i.e., both PERU and PLRU) are those where $\mathcal{M}$ is a singleton. In that case, the constant cost can be dropped out from Equations (1) and (2), leading to an analogue of DLR (2001)'s additive representation where the individual is risk neutral with respect to the first-stage uncertainty.

We next give a brief intuition about Theorem 1.A. Axiom 1 guarantees the existence of a Lipschitz continuous function $V : \mathcal{A} \to \mathbb{R}$ such that $V(co(A)) = V(A)$ and $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$. In terms of this expected utility representation, it is easy to see that PERU corresponds to convexity of V and PLRU corresponds to concavity of V. The set $\mathcal{A}^c$ of convex menus can be mapped one-to-one to the set Σ of support functions, preserving the metric and the linear operations. Therefore, by using the property $V(co(A)) = V(A)$ and mimicking the construction in DLR (2001), V can be thought of as a function defined on the subset Σ of the Banach space $C(\mathcal{U})$ of continuous real-valued functions on $\mathcal{U}$. We then apply a variation of the classic duality principle that convex [concave] functions can be written as the supremum [infimum] of affine functions lying below [above] them.[9] Finally, we apply the Riesz representation theorem to write each such continuous affine function as an integral against a measure μ minus [plus] a scalar $c(\mu)$. Theorem 1.B states that additionally imposing monotonicity guarantees that all measures in the LHA representation are positive.

[8]In part B, IR can be dropped for the case of the max-LHA representation, because it is implied by weak order, continuity, first-stage independence, PERU, and monotonicity.

[9]See Rockafellar (1970), Phelps (1993), and Appendix A of the current paper for variations of this duality result.

We show that the uniqueness of the LHA representations follows from the affine uniqueness of V and a result about the uniqueness of the dual representation of a convex function in the theory of conjugate convex functions (see Theorem 9 in Appendix A). A similar application of the duality and uniqueness results can be found in Ergin and Sarver (2009).

Theorem 2 *If $(\mathcal{M}, c)$ and $(\mathcal{M}', c')$ are two max-LHA [min-LHA] representations for $\succsim$, then there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $\mathcal{M}' = \alpha\mathcal{M}$ and $c'(\alpha\mu) = \alpha c(\mu) + \beta$ for all $\mu \in \mathcal{M}$.*

4 Applications

4.1 Multiple Priors and Variational Preferences

A preference for late resolution of uncertainty could arise if an individual would like to delay the resolution of objective lotteries for hedging reasons. In this section, we formalize this intuition by showing that the monotone min-LHA model is equivalent to two representations that have natural interpretations in terms of ambiguity-aversion. The following multiple-priors representation allows for ambiguity regarding the distribution over ex post subjective states and is intuitively similar to the multiple-priors representation proposed by Gilboa and Schmeidler (1989) in the Anscombe-Aumann setting.

Definition 2 A *Subjective-State-Space Multiple-Priors (SSMP) representation* is a quadruple $((\Omega, \mathcal{F}), U, \Pi)$ where Ω is a state space endowed with the σ-algebra $\mathcal{F}$, $U : \Omega \to \mathbb{R}^Z$ is a Z–dimensional, $\mathcal{F}$–measurable, and bounded random vector, and Π is a set of probability measures on $(\Omega, \mathcal{F})$, such that $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$, where $V : \mathcal{A} \to \mathbb{R}$ is defined by

$$V(A) = \min_{\pi \in \Pi} \int_{\Omega} \max_{p \in A} U(\omega) \cdot p \;\pi(d\omega), \tag{3}$$

and the minimization in Equation (3) has a solution for every $A \in \mathcal{A}$.

The next representation is similar in spirit to the variational representation considered by Maccheroni, Marinacci, and Rustichini (2006) in the Anscombe-Aumann setting.

Definition 3 A *Subjective-State-Space Variational (SSV) representation* is a quintuple $((\Omega, \mathcal{F}), U, \Pi, c)$ where Ω is a state space endowed with the σ-algebra $\mathcal{F}$, $U : \Omega \to \mathbb{R}^Z$ is a Z–dimensional, $\mathcal{F}$–measurable, and bounded random vector, Π is a set of probability measures on $(\Omega, \mathcal{F})$, and $c : \Pi \to \mathbb{R}$ is a function, such that $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$, where $V : \mathcal{A} \to \mathbb{R}$ is defined by

$$V(A) = \min_{\pi \in \Pi} \left(\int_\Omega \max_{p \in A} U(\omega) \cdot p \; \pi(d\omega) + c(\pi) \right), \tag{4}$$

and the minimization in Equation (4) has a solution for every $A \in \mathcal{A}$.[10]

The SSV representation generalizes the SSMP representation by allowing a "cost" $c(\pi)$ to be assigned to each measure π in the representation. In the Anscombe-Aumann setting, the class of variational preferences considered by Maccheroni, Marinacci, and Rustichini (2006) is strictly larger than the class of multiple-prior expected-utility preferences considered by Gilboa and Schmeidler (1989). However, we show that in the current setting, the SSMP and SSV representations are equivalent in the sense that the set of preferences that can be represented using an SSMP representation is precisely the set of preferences that can be represented using an SSV representation. The reason for this equivalence in the subjective versions of the representations is the state-dependence of the utility functions in the representations. The following Theorem formalizes this claim and, moreover, states that a preference $\succsim$ can be represented by one of these representations if and only if it has a monotone min-LHA representation.

Theorem 3 *Let $V : \mathcal{A} \to \mathbb{R}$. Then, the following are equivalent:*

1. *There exists a monotone min-LHA representation such that V is given by Equation* (2).
2. *There exists an SSMP representation such that V is given by Equation* (3).
3. *There exists an SSV representation such that V is given by Equation* (4).

[10] Note that for simplicity, we directly assume in the SSMP and SSV representations that the minimization in Equations (3) and (4) have a solution. One alternative approach that does not require this indirect assumption on the parameters would be to replace the minimums in Equations (3) and (4) with infima, in which case Theorem 3 would continue to hold. A second alternative is to impose topological assumptions on the parameters that would guarantee the existence of a minimum, for instance assuming that Ω is a metric space, $\mathcal{F}$ is the Borel σ–algebra on Ω, U is continuous, Π is weak*-compact, and c is lower semi-continuous.

The following immediate corollary provides the axiomatic foundation for the SSMP and SSV representations.

Corollary 1 *A preference $\succsim$ has a SSMP representation if and only if it has a SSV representation if and only if it satisfies Axiom 1, PLRU, and monotonicity.*

We next outline the intuition behind Theorem 3 for the case where in the min-LHA, SSMP, and SSV representations, the sets $\mathcal{M}$, Π, Ω are finite, $\mathcal{F} = 2^{\Omega}$, and the measures in $\mathcal{M}$ have finite support. In this special case, the compactness, lower semi-continuity, measurability, and boundedness properties in the representations are automatically satisfied.

First consider (1) $\Rightarrow$ (3). Fix a monotone min-LHA representation $(\mathcal{M}, c)$, and define V by Equation (2). Take any $\mu \in \mathcal{M}$, and define a measure π_μ on the set $\mu(\mathcal{U})\mathcal{U}$ by $\pi_\mu(v) = \frac{\mu(u)}{\mu(\mathcal{U})}$ for $v = \mu(\mathcal{U})u$, $u \in \mathcal{U}$. This transformation ensures that π_μ is a probability measure. Moreover, since $\pi_\mu(v)v = \mu(u)u$ for $v = \mu(\mathcal{U})u$, for any $A \in \mathcal{A}$,

$$\int_{\mu(\mathcal{U})\mathcal{U}} \max_{p \in A} v(p)\, \pi_\mu(dv) = \int_{\mathcal{U}} \max_{p \in A} u(p)\, \mu(du).$$

Let $\Omega = \bigcup_{\mu \in \mathcal{M}} \mu(\mathcal{U})\mathcal{U}$, and define $U : \Omega \to \mathbb{R}^Z$ by $U(\omega) = \omega$. Let $\Pi = \{\pi_\mu : \mu \in \mathcal{M}\}$ and $\tilde{c}(\pi_\mu) = c(\mu)$. Then, V can be expressed in the following SSV form:

$$V(A) = \min_{\pi \in \Pi} \left(\int_{\Omega} \max_{p \in A} U(\omega) \cdot p \; \pi(d\omega) + \tilde{c}(\pi) \right).$$

The idea of this construction is identical to the interpretation we gave for the integral term in the LHA-representations: Since utility is state-dependent, the weight of the states in the linear aggregator cannot be uniquely pinned down, and integration against the positive measure μ can be reexpressed as integration against a probability measure π_μ after appropriately rescaling the state-dependent utility functions.

To see the intuition for (3) $\Rightarrow$ (2), consider an SSV representation $((\Omega, \mathcal{F}), U, \Pi, c)$, and define V by Equation (4). Let $\tilde{\Omega} = \Omega \times \Pi$ and $\tilde{\mathcal{F}} = 2^{\tilde{\Omega}}$. Let $\mathbf{1} \in \mathbb{R}^Z$ denote the vector whose coordinates are equal to 1, and define $\tilde{U} : \tilde{\Omega} \to \mathbb{R}^Z$ by $\tilde{U}(\omega, \pi) = U(\omega) + c(\pi)\mathbf{1}$ for any $\tilde{\omega} = (\omega, \pi) \in \tilde{\Omega}$. Take any probability measure $\pi \in \Pi$, and define a new measure ρ_π on $\tilde{\Omega}$ by $\rho_\pi(\omega, \pi) = \pi(\omega)$ and $\rho_\pi(\omega, \pi') = 0$ for any $\pi' \neq \pi$. It is immediate that ρ_π is a probability measure on $(\tilde{\Omega}, \tilde{\mathcal{F}})$. Also, for any $A \in \mathcal{A}$,

$$\begin{aligned}\int_{\tilde{\Omega}} \max_{p\in A} \tilde{U}(\tilde{\omega}) \cdot p \; \rho_\pi(d\tilde{\omega}) &= \int_{\Omega\times\Pi} \left[\max_{p\in A} U(\omega)\cdot p + c(\pi')\right] \rho_\pi(d\omega, d\pi') \\ &= \int_{\Omega} \max_{p\in A} U(\omega)\cdot p \; \pi(d\omega) + c(\pi).\end{aligned}$$

Letting $\tilde{\Pi} = \{\rho_\pi : \pi \in \Pi\}$, we see that V can be expressed in the following SSMP form:

$$V(A) = \min_{\rho\in\tilde{\Pi}} \int_{\tilde{\Omega}} \max_{p\in A} \tilde{U}(\tilde{\omega}) \cdot p \; \rho(d\tilde{\omega}).$$

The idea behind this argument is also a consequence of state-dependence of utility which allows any constant to be absorbed in the integral term. Above, integration against the probability measure π plus the constant $c(\pi)$ is reexpressed as integration against the probability measure ρ_π whose support is a subset of states in $\Omega \times \{\pi\}$ where the utility functions are "shifted" by $c(\pi)\mathbf{1}$.

To see the intuition for (2) $\Rightarrow$ (1), consider an SSMP representation $((\Omega, \mathcal{F}), U, \Pi)$, and define V by Equation (3). By the definition of $\mathcal{U}$, since $U(\omega)$ is an expected-utility function for each $\omega \in \Omega$, there exist $u(\omega) \in \mathcal{U}$, $\alpha(\omega) \geq 0$, and $\beta(\omega) \in \mathbb{R}$ such that

$$U(\omega) = \alpha(\omega)u(\omega)\cdot p + \beta(\omega)\mathbf{1}.$$

For each $\pi \in \Pi$, define the measure μ_π on $\mathcal{U}$ by $\mu_\pi(u) = \sum_{\omega\in\Omega: u(\omega)=u} \alpha(\omega)\pi(\omega)$ for each $u \in \mathcal{U}$. Define the function $c : \Pi \to \mathbb{R}$ by $c(\pi) = \sum_{\omega\in\Omega} \beta(\omega)\pi(\omega)$. Then, for any $A \in \mathcal{A}$,

$$\begin{aligned}\sum_{\omega\in\Omega} \max_{p\in A} U(\omega)\cdot p \; \pi(\omega) &= \sum_{\omega\in\Omega} \max_{p\in A} [u(\omega)\cdot p] \, \alpha(\omega)\pi(\omega) + \sum_{\omega\in\Omega} \beta(\omega)\pi(\omega) \\ &= \int_{\mathcal{U}} \max_{p\in A} u(p)\mu_\pi(du) + c(\pi).\end{aligned}$$

Let $\mathcal{M} = \{\mu_\pi : \pi \in \Pi\}$ and let $\tilde{c}(\mu) = \min\{c(\pi) : \pi \in \Pi \text{ and } \mu_\pi = \mu\}$. Then, V can be expressed in the following min-LHA form:

$$V(A) = \min_{\mu\in\mathcal{M}} \left(\int_{\mathcal{U}} \max_{p\in A} u(p) \, \mu(du) + \tilde{c}(\mu) \right). \tag{5}$$

It can also be shown that sequentially removing measures from $\mathcal{M}$ that are not strictly optimal in Equation (5) for some $A \in \mathcal{A}$ leads to a minimal set of measures $\mathcal{M}' \subset \mathcal{M}$.

Our SSMP representation bears some similarity to a representation considered by Epstein, Marinacci and Seo (2007, Theorem 1). One main distinction between our representation and theirs is that they impose a normalization on the state-dependent utility function in their representation. As the above arguments illustrate, the key to the equivalence proposed in Theorem 3 is the state-dependence of the utility functions in the SSMP and SSV representations. If a normalization as in Epstein, Marinacci and Seo (2007) were imposed on the utility functions in the SSV and SSMP representations, then the equivalence of these representations would no longer hold, as it would not be possible to "absorb" the cost function of the SSV representation into the utility function to obtain an SSMP representation. Moreover, although these representations would continue to be special cases of the monotone min-LHA representation, it would not be possible to write every monotone min-LHA representation as an SSV representation since it would not always be possible to "absorb" the magnitude of the measure into the utility function. Theorem 3 illustrates that imposing either a normalization on utility functions (as in the min-LHA representation) or a normalization that measures be probabilities (as in the SSMP and SSV representations) is not restrictive; however, imposing both normalizations simultaneously would place a non-trivial additional restriction on the representations.

4.2 Costly Contemplation

A special case of the max-LHA representation is one of subjective information acquisition/costly contemplation, where the measures in $\mathcal{M}$ can be interpreted as a reduced-form representation of information about one's tastes and the integral term $\int_{\mathcal{U}} \max_{p \in A} u(p)\, \mu(du)$ can be interpreted as the ex ante value of information μ given menu A. In this application, we need to additionally impose the following axiom.

Axiom 5 (Reversibility of Degenerate Decisions (RDD)) *For any $A \in \mathcal{A}$, $p, q \in \triangle(Z)$, and $\alpha \in [0,1]$,*

$$\beta \delta_{\alpha A + (1-\alpha)\{p\}} + (1-\beta)\delta_{\{q\}} \sim \beta \delta_{\alpha A + (1-\alpha)\{q\}} + (1-\beta)\delta_{\{p\}}$$

where $\beta = 1/(2-\alpha)$.

We will call a choice out of a singleton menu as a degenerate decision. To interpret Axiom 5, consider Figure 2. The first tree represents $\beta \delta_{\alpha A + (1-\alpha)\{p\}} + (1-\beta)\delta_{\{q\}}$, where the individual makes a choice out of the menu $\alpha A + (1-\alpha)\{p\}$ with probability β, and makes a degenerate choice out of the menu $\{q\}$ with probability $1-\beta$. A choice out of

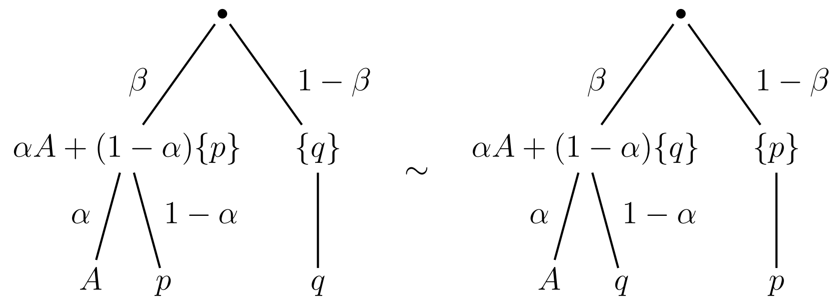

Figure 2: Reversibility of Degenerate Decisions ($\beta = 1/(2-\alpha)$)

the menu $\alpha A + (1-\alpha)\{p\}$ can be interpreted as a contingent plan, where initially in period 2 the individual determines a lottery out of A, and then her choice out of A is executed with probability α and the fixed lottery p is executed with the remaining $1-\alpha$ probability. Similarly, the second tree represents $\beta\delta_{\alpha A+(1-\alpha)\{q\}} + (1-\beta)\delta_{\{p\}}$ where the roles of p and q are reversed.

If one interprets the individual's behavior as one of costly contemplation/subjective information acquisition, then her optimal contemplation strategy might change as the probability α that her choice out of A is executed changes, since her return to contemplation will be higher for higher values of α. However, since the probability that her choice out of A will be executed is the same in both $\alpha A+(1-\alpha)\{p\}$ and $\alpha A+(1-\alpha)\{q\}$, it is reasonable to expect that her contemplation strategy would be the same for both contingent planning problems, although she need not be indifferent between $\delta_{\alpha A+(1-\alpha)\{p\}}$ and $\delta_{\alpha A+(1-\alpha)\{q\}}$ depending on her preference between $\delta_{\{p\}}$ and $\delta_{\{q\}}$. The RDD axiom requires the individual to be indifferent between the two trees in Figure 2 when the probabilities of the paths leading to lotteries p and q are the same, i.e., when $\beta(1-\alpha) = 1-\beta$, or equivalently, $\beta = 1/(2-\alpha)$.

Given a max-LHA representation $(\mathcal{M}, c)$, we show that RDD is equivalent to the following consistency requirement on the set of measures $\mathcal{M}$:

Definition 4 A *Reduced-Form Costly Contemplation (RFCC) representation* is a max-LHA representation $(\mathcal{M}, c)$ where the set $\mathcal{M}$ is *consistent*: For each $\mu, \nu \in \mathcal{M}$ and $p \in \triangle(Z)$,

$$\int_{\mathcal{U}} u(p)\, \mu(du) = \int_{\mathcal{U}} u(p)\, \nu(du).$$

An RFCC representation $(\mathcal{M}, c)$ is *monotone* if all the measures in $\mathcal{M}$ are positive.

We show in Ergin and Sarver (2009) that the consistency condition above is key for the interpretation of the max-LHA representation as a subjective information acquisition problem. More specifically, V satisfies Equation (1) for some monotone RFCC representation $(\mathcal{M}, c)$ if and only if

$$V(A) = \max_{\mathcal{G} \in \mathbf{G}} \left(\mathbb{E}\left[\max_{p \in A} \mathbb{E}\big[U \big| \mathcal{G}\big] \cdot p \right] - c(\mathcal{G}) \right) \tag{6}$$

where $(\Omega, \mathcal{F}, P)$ is a probability space, $U : \Omega \to \mathbb{R}^Z$ is a random vector interpreted as the individual's state-dependent utility, $\mathbf{G}$ is a collection of sub-σ-algebras of $\mathcal{F}$ representing the set of signals that the individual can acquire, and $c(\mathcal{G})$ denotes the cost of subjective signal $\mathcal{G} \in \mathbf{G}$.[11]

As a special case of the max-LHA representation, the RFCC representation satisfies PERU. However, it always satisfies indifference to timing of resolution of uncertainty when restricted to temporal lotteries, i.e., for all $p, q \in \triangle(Z)$ and $\alpha \in (0,1)$:

$$\alpha \delta_{\{p\}} + (1-\alpha)\delta_{\{q\}} \sim \delta_{\{\alpha p + (1-\alpha) q\}}.$$[12]

Therefore, an individual with RFCC preferences never has a strict PERU unless if she has non-degenerate choices in period 2.

We next present an RFCC representation theorem as an application of Theorem 1. Since RFCC representation is a special case of the max-LHA representation, the uniqueness of the RFCC representation is immediately implied by the uniqueness of the max-LHA representation.

Theorem 4 *A. The preference $\succsim$ has a RFCC representation if and only if it satisfies Axiom 1, PERU, and RDD.*

B. The preference $\succsim$ has a monotone RFCC representation if and only if it satisfies

[11] The costly contemplation representation in Equation (3) is similar to the functional form considered by Ergin (2003), whose primitive is a preference over menus taken from a finite set of alternatives.

[12] This property can also be established directly as a consequence of RDD and first-stage independence. Fix any $p, q \in \triangle(Z)$ and $\alpha \in (0,1)$. Letting $\beta = 1/(2-\alpha)$ and $A = \{p\}$, RDD implies

$$\beta \delta_{\{p\}} + (1-\beta)\delta_{\{q\}} \sim \beta \delta_{\{\alpha p + (1-\alpha) q\}} + (1-\beta)\delta_{\{p\}}.$$

Since $\beta = 1/(2-\alpha)$ implies that $\beta = 1 - \beta + \alpha\beta$ and $1 - \beta = (1-\alpha)\beta$, the left side of this expression is equal to $(1-\beta)\delta_{\{p\}} + \alpha\beta\delta_{\{p\}} + (1-\alpha)\beta\delta_{\{q\}}$. Hence,

$$\beta\big[\alpha\delta_{\{p\}} + (1-\alpha)\delta_{\{q\}}\big] + (1-\beta)\delta_{\{p\}} \sim \beta\delta_{\{\alpha p + (1-\alpha)q\}} + (1-\beta)\delta_{\{p\}},$$

which, by first-stage independence, implies $\alpha\delta_{\{p\}} + (1-\alpha)\delta_{\{q\}} \sim \delta_{\{\alpha p + (1-\alpha) q\}}$.

Axiom 1, PERU, RDD, and monotonicity.

4.3 Kreps and Porteus (1978) and DLR (2001)

In this section, we will introduce a generalization of Kreps and Porteus (1978) and DLR (2001) additive representations, and characterize the intersection of this class with LHA preferences. The first axiom we consider is the standard von Neumann-Morgenstern independence axiom imposed on the second-stage uncertainty. It is satisfied by RFCC preferences, but not by SSMP and SSV preferences, since in the latter contexts the individual may benefit from hedging even when she makes no choice in period 2.

Axiom 6 (Second-Stage Independence) *For any $p, q, r \in \triangle(Z)$, and $\alpha \in (0,1)$,*

$$\delta_{\{p\}} \succ \delta_{\{q\}} \quad \Rightarrow \quad \delta_{\{\alpha p+(1-\alpha)r\}} \succ \delta_{\{\alpha q+(1-\alpha)r\}}.$$

Under weak order and continuity, the following axiom from Kreps (1979) guarantees that every menu is indifferent to its best singleton subset. Kreps and Porteus (1978) assume the same relationship between the individual's ranking of menus and alternatives.[13]

Axiom 7 (Strategic Rationality) *For any $A, B \in \mathcal{A}$, $\delta_A \succsim \delta_B$ implies $\delta_A \sim \delta_{A\cup B}$.*

Suppose there exists a continuous function $V : \mathcal{A} \to \mathbb{R}$ such that $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$. Then, the preference $\succsim$ satisfies strategic rationality if and only if

$$V(A) = \max_{p\in A} V(\{p\}), \tag{7}$$

i.e., the restriction of $\succsim$ to temporal lotteries determines the entire preference $\succsim$. If such an individual is indifferent to timing of resolution of uncertainty when choosing among temporal lotteries, then Equation (7) implies that she must always be indifferent to timing of resolution of uncertainty. This is in contrast with RFCC preferences, where

[13] To be precise, Kreps and Porteus (1978) consider both a period 1 preference $\succsim$ over first-stage lotteries in $\triangle(\mathcal{A})$ and a period 2 preference $\succsim_2$ over second-stage lotteries in $\triangle(Z)$. It is easy to show that imposing their temporal consistency axiom (Axiom 3.1 in their paper) on this pair of preferences ($\succsim$,$\succsim_2$) implies that the period 1 preference $\succsim$ satisfies strategic rationality. Conversely, if the period 1 preference $\succsim$ satisfies strategic rationality along with continuity, then there exists *some* period 2 preference $\succsim_2$ such that the pair ($\succsim$,$\succsim_2$) satisfies their temporal consistency axiom. Moreover, in this case, the period 1 preference $\succsim$ satisfies our second-stage independence axiom if and only if this period 2 preference $\succsim_2$ satisfies the substitution axiom of Kreps and Porteus (1978, Axiom 2.3).

the individual is indifferent to timing of resolution of uncertainty when choosing among temporal lotteries, but may exhibit a strict PERU when she faces non-degenerate choices in period 2. Although Kreps and Porteus (1978)'s setup is rich enough to distinguish between attitudes towards timing of resolution of uncertainty over temporal lotteries (without period 2 choice) and over more general lotteries over menus (with period 2 choice), the fact that they implicitly impose Axiom 7 throughout their analysis prevents them from doing so.

A second implication of strategic rationality is that it rules out a strict preference for flexibility (Kreps (1979)), i.e., situations where the union of two menus is strictly better than each menu separately: $\delta_{A \cup B} \succ \delta_A$ and $\delta_{A \cup B} \succ \delta_B$. The reason is that, by Equation (7), the individual behaves as if she has no uncertainty in period 1 about her ex post preference ranking over $\triangle(Z)$.

The following axiom from DLR (2001) is the standard independence requirement applied to convex combinations of menus when there is no first-stage uncertainty.

Axiom 8 (Mixture Independence) *For any $A, B, C \in \mathcal{A}$ and $\alpha \in (0,1)$,*

$$\delta_A \succ \delta_B \quad \Rightarrow \quad \delta_{\alpha A + (1-\alpha) C} \succ \delta_{\alpha B + (1-\alpha) C}.$$

It is easy to see that mixture independence is stronger than second-period independence, but in the presence of Axiom 1, it is weaker than the combination of second-period independence and strategic rationality.

We next consider a class of representations that generalize the DLR (2001) additive representation where there is no objective first-stage uncertainty, and Kreps and Porteus (1978) representation where there is no subjective second-stage uncertainty. Unlike the Kreps and Porteus (1978) representation, the following class of representations are compatible with a strict preference for flexibility.

Definition 5 A *Kreps-Porteus-Dekel-Lipman-Rustichini (KPDLR) representation* is a pair (ϕ, μ) where μ is a finite signed Borel measure on $\mathcal{U}$ and $\phi : [a, b] \to \mathbb{R}$ is a Lipschitz continuous and strictly increasing function where $[a, b] = \{\int_{\mathcal{U}} \max_{p \in A} u(p)\, \mu(du) : A \in \mathcal{A}\}$ and $-\infty < a \leq b < +\infty$, such that $P \succsim Q$ if and only if $\mathbb{E}_P[V] \geq \mathbb{E}_Q[V]$, where $V : \mathcal{A} \to \mathbb{R}$ is defined by:

$$V(A) = \phi\Big(\int_{\mathcal{U}} \max_{p \in A} u(p)\, \mu(du) \Big). \tag{8}$$

A *monotone KPDLR representation* is a KPDLR representation (ϕ, μ) where μ is a positive measure. A *Kreps-Porteus representation* is a monotone KPDLR representation (ϕ, μ) where $\mu(\mathcal{U} \setminus \{u\}) = 0$ for some $u \in \mathbb{R}^Z$.

It is easy to see that two KPDLR representations (ϕ, μ) and (ψ, ν) induce the same preference if and only if there exists $\lambda, \alpha > 0, \beta \in \mathbb{R}$ such that $\mu = \lambda\nu$ and $\phi(t) = \alpha\psi(t/\lambda) + \beta$. Since it is possible to have $\int_{\mathcal{U}} u(p)\, \mu(du) = \lambda \int_{\mathcal{U}} u(p)\, \nu(du)$ for all $p \in \triangle(Z)$ even when $\mu \neq \lambda\nu$, this implies in particular that in KPDLR representations, the preference restricted to temporal lotteries does *not* determine the entire preference. On the other hand, in a Kreps-Porteus representation (ϕ, μ) where $\mu(\mathcal{U} \setminus \{u\}) = 0$, the V in Equation (8) can be rewritten as:

$$V(A) = \phi\Big(\max_{p \in A} v(p)\Big), \tag{9}$$

where $v = \mu(u)u$. Since V in Equation (9) satisfies Equation (7), in this case the preference restricted to temporal lotteries does determine the entire preference.

We are ready to state our KPDLR and Kreps-Porteus representation results. Theorem 5.B is a special case of Theorem 1 in Kreps and Porteus (1978).[14]

Theorem 5 *A. The preference $\succsim$ has a KPDLR representation if and only if it satisfies Axiom 1 and mixture independence.*

B. The preference $\succsim$ has a Kreps-Porteus representation if and only if it satisfies Axiom 1, second-stage independence, and strategic rationality.

C. If the preference $\succsim$ has the KPDLR representation (ϕ, μ), then $\succsim$ satisfies PERU [PLRU] if and only if ϕ is convex [concave].

Since a Kreps-Porteus preference satisfies Axiom 7, the individual has PERU [PLRU] if and only if she has PERU [PLRU] for temporal lotteries. More generally, Theorem 5.C implies that if an individual's preference has a KPDLR representation and has the property that for every $A \in \mathcal{A}$ there exists a $p \in \triangle(Z)$ such that $\delta_A \sim \delta_{\{p\}}$, then the individual has PERU [PLRU] if and only if she has PERU [PLRU] for temporal lotteries. Therefore, under the aforementioned property, the individual must have the same attitudes towards timing of resolution of uncertainty over temporal lotteries and over more general lotteries over menus in a KPDLR representation.

[14] The only difference is that Kreps and Porteus (1978) only require ϕ to be continuous. We additionally require Lipschitz continuity of ϕ, since we impose the L–continuity axiom throughout the paper.

The class of KPDLR preferences need not be a subset of LHA representations, but the subclass satisfying PERU or PLRU do. We next characterize the subclass of KPDLR preferences which satisfies PERU and PLRU within the class of LHA preferences.

Theorem 6 *Let $V : \mathcal{A} \to \mathbb{R}$ and let μ be a finite signed Borel measure on $\mathcal{U}$. Then, there exists a KPDLR representation (ϕ, μ) with convex [concave] ϕ such that V is given by Equation* (8) *if and only if there exists a max-LHA [min-LHA] representation $(\mathcal{M}, c)$ such that V is given by Equation* (1) *[*(2)*] and $\mathcal{M} \subset \{\lambda\mu : \lambda \in \mathbb{R}_+\}$.*

Theorem 6 suggests that if the KPDLR representation satisfies PERU or PLRU, then it is possible to rewrite the function V in Equation (8) as a LHA representation where all measures are multiples of the fixed measure μ. Consider the case of a positive measure μ, and consider again the probability measure π on $\mathcal{V} = \mu(\mathcal{U})\mathcal{U}$ that (heuristically) puts weight $\pi(v) = \mu(u)/\mu(\mathcal{U})$ on each $v = \mu(\mathcal{U})u \in \mathcal{V}$. An interpretation of the LHA representation in Theorem 6 is that all actions lead to the same distribution π over ex post utilities in $\mathcal{V}$, but each action $\lambda\mu \in \mathcal{M}$ changes the magnitude of the ex post utilities by a common scalar multiple λ.[15] In the special case of Kreps-Porteus preferences, all measures are degenerate and put their weight on the same ex post von Neumann-Morgenstern utility function over $\triangle(Z)$. In particular, the individual has no uncertainty about her ex post preference ranking over lotteries in $\triangle(Z)$, and different actions only affect the strength of her ex post preference.

[15]Suppose without loss of generality that $\max\{\lambda : \lambda\mu \in \mathcal{M}\} = 1$. Then, one can also interpret each action $\lambda\mu \in \mathcal{M}$ to lead to the distribution π over $\mathcal{V}$ with probability λ, and to the ex post preference $0 \in \mathbb{R}^Z$ with probability $1 - \lambda$. Under this interpretation, the choice of action affects the probability of the 0 ex post preference, but not the conditional probability distribution π over $\mathcal{V}$.

Appendix

A Mathematical Preliminaries

In this section we present some general mathematical results that will be used to prove our representation and uniqueness theorems. Our main results will center around a classic duality relationship from convex analysis. Throughout this section, let X be a real Banach space, and let X^* denote the space of all continuous linear functionals on X.

We now introduce the standard definition of the subdifferential of a function.

Definition 6 Suppose $C \subset X$ and $f : C \to \mathbb{R}$. For $x \in C$, the *subdifferential* of f at x is defined to be

$$\partial f(x) = \{x^* \in X^* : \langle y - x, x^* \rangle \leq f(y) - f(x) \text{ for all } y \in C\}.$$

The subdifferential is useful for the approximation of convex functions by affine functions. It is straightforward to show that $x^* \in \partial f(x)$ if and only if the affine function $h : X \to \mathbb{R}$ defined by $h(y) = f(x) + \langle y - x, x^* \rangle$ satisfies $h \leq f$ and $h(x) = f(x)$. It should also be noted that when X is infinite-dimensional it is possible to have $\partial f(x) = \emptyset$ for some $x \in C$, even if f is convex. However, it can be shown that if $C \subset X$ is convex and $f : C \to \mathbb{R}$ is Lipschitz continuous and convex, then $\partial f(x) \neq \emptyset$ for all $x \in C$ (see Lemma 2 in Appendix A.1). The formal definition of Lipschitz continuity follows:

Definition 7 Suppose $C \subset X$. A function $f : C \to \mathbb{R}$ is said to be *Lipschitz continuous* if there is some real number K such that $|f(x) - f(y)| \leq K\|x - y\|$ for every $x, y \in C$. The number K is called a *Lipschitz constant* of f. A function f is said to be *locally Lipschitz continuous* if for every $x \in C$, there exists $\varepsilon > 0$ such that f is Lipschitz continuous on $B_\varepsilon(x) \cap C = \{y \in C : \|y - x\| < \varepsilon\}$.

We now introduce the definition of the conjugate of a function.

Definition 8 Suppose $C \subset X$ and $f : C \to \mathbb{R}$. The *conjugate* (or *Fenchel conjugate*) of f is the function $f^* : X^* \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(x^*) = \sup_{x \in C} \big[\langle x, x^* \rangle - f(x)\big].$$

There is an important duality between f and f^*. Lemma 1 summarizes certain properties of f^* that are useful in establishing this duality.[16] We include a proof for completeness.

[16]For a complete discussion of the relationship between f and f^*, see Ekeland and Turnbull (1983) or Holmes (1975). A finite-dimensional treatment can be found in Rockafellar (1970).

Lemma 1 *Suppose $C \subset X$ and $f : C \to \mathbb{R}$. Then,*

1. *f^* is lower semi-continuous in the weak* topology.*
2. *$f(x) \geq \langle x, x^* \rangle - f^*(x^*)$ for all $x \in C$ and $x^* \in X^*$.*
3. *$f(x) = \langle x, x^* \rangle - f^*(x^*)$ if and only if $x^* \in \partial f(x)$.*

Proof: (1): For any $x \in C$, the mapping $x^* \mapsto \langle x, x^* \rangle - f(x)$ is continuous in the weak* topology. Therefore, for all $\alpha \in \mathbb{R}$, $\{x^* \in X^* : \langle x, x^* \rangle - f(x^*) \leq \alpha\}$ is weak* closed. Hence,

$$\{x^* \in X^* : f^*(x^*) \leq \alpha\} = \bigcap_{x \in C} \{x^* \in X^* : \langle x, x^* \rangle - f(x) \leq \alpha\}$$

is closed for all $\alpha \in \mathbb{R}$. Thus, f^* is lower semi-continuous.

(2): For any $x \in C$ and $x^* \in X^*$, we have

$$f^*(x^*) = \sup_{y \in C} \big[\langle y, x^* \rangle - f(y)\big] \geq \langle x, x^* \rangle - f(x),$$

and therefore $f(x) \geq \langle x, x^* \rangle - f^*(x^*)$.

(3): By the definition of the subdifferential, $x^* \in \partial f(x)$ if and only if

$$\langle y, x^* \rangle - f(y) \leq \langle x, x^* \rangle - f(x). \tag{10}$$

for all $y \in C$. By the definition of the conjugate, Equation (10) holds if and only if $f^*(x^*) = \langle x, x^* \rangle - f(x)$, which is equivalent to $f(x) = \langle x, x^* \rangle - f^*(x^*)$. ■

Suppose that $C \subset X$ is convex and $f : C \to \mathbb{R}$ is Lipschitz continuous and convex. As noted above, this implies that $\partial f(x) \neq \emptyset$ for all $x \in C$. Therefore, by parts 2 and 3 of Lemma 1, we have

$$f(x) = \max_{x^* \in X^*} \big[\langle x, x^* \rangle - f^*(x^*)\big] \tag{11}$$

for all $x \in C$.[17] In order to establish the existence of a minimal set of measures in the proof of Theorem 1, it is useful to establish that under certain assumptions, there is a minimal compact subset of X^* for which Equation (11) holds. Let C_f denote the set of all $x \in C$ for which the subdifferential of f at x is a singleton:

$$C_f = \{x \in C : \partial f(x) \text{ is a singleton}\}. \tag{12}$$

Let $\mathcal{N}_f$ denote the set of functionals contained in the subdifferential of f at some $x \in C_f$:

$$\mathcal{N}_f = \{x^* \in X^* : x^* \in \partial f(x) \text{ for some } x \in C_f\}. \tag{13}$$

[17] This is a slight variation of the classic Fenchel-Moreau theorem. The standard version of this theorem states that if $f : X \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous and convex, then $f(x) = f^{**}(x) \equiv \sup_{x^* \in X^*}[\langle x, x^* \rangle - f^*(x^*)]$. See, e.g., Proposition 1 in Ekeland and Turnbull (1983, p97).

Finally, let $\mathcal{M}_f$ denote the closure of $\mathcal{N}_f$ in the weak* topology:

$$\mathcal{M}_f = \overline{\mathcal{N}_f}. \tag{14}$$

Before stating our first main result, recall that the *affine hull* of a set $C \subset X$, denoted $\text{aff}(C)$, is defined to be the smallest affine subspace of X that contains C. Also, a set $C \subset X$ is said to be a *Baire space* if every countable intersection of dense open subsets of C is dense.

Theorem 7 *Suppose (i) X is a separable Banach space, (ii) C is a convex subset of X that is a Baire space (when endowed with the relative topology) such that* $\text{aff}(C)$ *is dense in X,*[18] *and (iii) $f : C \to \mathbb{R}$ is Lipschitz continuous and convex. Then, $\mathcal{M}_f$ is weak* compact, and for any weak* compact $\mathcal{M} \subset X^*$,*

$$\mathcal{M}_f \subset \mathcal{M} \iff f(x) = \max_{x^* \in \mathcal{M}} \left[\langle x, x^* \rangle - f^*(x^*)\right] \quad \forall x \in C.$$

The proof of Theorem 7 is contained in Appendix A.2 and is based on a subdifferentiability result that is established in Appendix A.1. Although the proofs of these results are somewhat involved, the intuition for Theorem 7 is fairly simple. We already know from Lemma 1 that for any $x \in C_f$, $f(x) = \max_{x^* \in \mathcal{N}_f}[\langle x, x^* \rangle - f^*(x^*)]$. The main result of Appendix A.1 shows that under the assumptions of Theorem 7, C_f is dense in C. Therefore, it can be shown that for any $x \in C$,

$$f(x) = \max_{x^* \in \mathcal{M}_f} \left[\langle x, x^* \rangle - f^*(x^*)\right].$$

In addition, if $\mathcal{M}$ is a weak* compact subset of X^* and $\mathcal{M}_f$ is not a subset of $\mathcal{M}$, then there exists $x^* \in \mathcal{N}_f$ such that $x^* \notin \mathcal{M}$. That is, there exists $x \in C_f$ such that $\partial f(x) = \{x^*\}$ and $x^* \notin \mathcal{M}$. Therefore, Lemma 1 implies $f(x) > \max_{x^* \in \mathcal{M}}[\langle x, x^* \rangle - f^*(x^*)]$.

In the proof of Theorem 1, we will construct an LHA representation in which $\mathcal{M}_f$, for a certain function f, is the set of measures. In the proof of part B of Theorem 1, we will use the following result to establish that monotonicity leads to a positive set of measures. For this next result, assume that X is a Banach lattice.[19] Let $X_+ = \{x \in X : x \geq 0\}$ denote the *positive cone* of X. A function $f : C \to \mathbb{R}$ on a subset C of X is *monotone* if $f(x) \geq f(y)$ whenever $x, y \in C$ are such that $x \geq y$. A continuous linear functional $x^* \in X^*$ is *positive* if $\langle x, x^* \rangle \geq 0$ for all $x \in X_+$.

Theorem 8 *Suppose C is a convex subset of a Banach lattice X, such that at least one of the following conditions holds:*

1. *$x \vee x' \in C$ for any $x, x' \in C$, or*
2. *$x \wedge x' \in C$ for any $x, x' \in C$.*

[18] In particular, if C is closed, then by the Baire Category theorem, then C is a Baire space. Also, note that if C contains the origin, then the affine hull of C is equal to the span of C.

[19] See Aliprantis and Border (1999, p302) for a definition of Banach lattices.

Let $f : C \to \mathbb{R}$ be locally Lipschitz continuous, convex, and monotone. Then, the functionals in $\mathcal{M}_f$ are positive.

The proof of Theorem 8 is contained in Appendix A.3.

Finally, the following result will be used in the proof of Theorem 2 to establish the uniqueness of the LHA representation.

Theorem 9 *Suppose X is a Banach space and C is a convex subset of X. Let $\mathcal{M}$ be a weak* compact subset of X^*, and let $c : \mathcal{M} \to \mathbb{R}$ be weak* lower semi-continuous. Define $f : C \to \mathbb{R}$ by*

$$f(x) = \max_{x^* \in \mathcal{M}} \big[\langle x, x^* \rangle - c(x^*)\big]. \tag{15}$$

Then,

1. *The function f is Lipschitz continuous and convex.*
2. *For all $x \in C$, there exists $x^* \in \partial f(x)$ such that $x^* \in \mathcal{M}$ and $f^*(x^*) = c(x^*)$. In particular, this implies $\mathcal{N}_f \subset \mathcal{M}$, $\mathcal{M}_f \subset \mathcal{M}$, and $f^*(x^*) = c(x^*)$ for all $x^* \in \mathcal{N}_f$.*
3. *If C is also compact (in the norm topology), then $f^*(x^*) = c(x^*)$ for all $x^* \in \mathcal{M}_f$.*

The proof of Theorem 9 is contained in Appendix A.4.

A.1 A Subdifferentiability Result

Mazur (1933) showed that if X is a separable Banach space and $f : C \to \mathbb{R}$ is a continuous convex function defined on a convex open subset C of X, then the set of points x where f is Gâteaux differentiable is a dense G_δ set in C (where the notation G_δ indicates a countable intersection of open sets).[20] In this section, we extend Mazur's theorem by replacing the assumption that C is open with the weaker assumptions that C is a Baire space (when endowed with the relative topology) and that the affine hull of C is dense in X.

Theorem 10 *Suppose (i) X is a separable Banach space, (ii) C is a convex subset of X that is a Baire space (when endowed with the relative topology) such that* $\operatorname{aff}(C)$ *is dense in X, and (iii) $f : C \to \mathbb{R}$ is locally Lipschitz continuous and convex. Then, the set of points x where $\partial f(x)$ is a singleton is a dense G_δ set (in the relative topology) in C.*

Note that Mazur's theorem is a special case of Theorem 10. First, if C is an open subset of X, then C is a Baire space and $\operatorname{aff}(C) = X$. Second, any continuous convex function f defined on an open set C is locally Lipschitz continuous (see Phelps (1993, Proposition 1.6)). Therefore, if C is open, then our continuity assumption coincides with that of Mazur's theorem.

[20]For a textbook treatment, see Phelps (1993, Theorem 1.20). An equivalent characterization in terms of closed convex sets and smooth points can be found in Holmes (1975, p171).

Finally, it is a standard result that a continuous convex function f defined on an open set C is Gâteaux differentiable at a point x if and only if the subdifferential $\partial f(x)$ is a singleton set (see Phelps (1993, Proposition 1.8)). Thus, if C is open, then the conclusion of Theorem 10 also coincides with the conclusion of Mazur's theorem.[21]

The remainder of this section is devoted to the proof of Theorem 10. Our proof of Theorem 10 will follow a similar approach to the proof of Mazur's theorem found in Phelps (1993). Theorem 10 also follows from results in Verona and Verona (1990). We choose to provide the direct proof, since it is simpler than showing how the results in Verona and Verona (1990) apply to our context. We begin by establishing that the subdifferential of a Lipschitz continuous and convex function is nonempty at every point.

Lemma 2 *Suppose C is a convex subset of a Banach space X. If $f : C \to \mathbb{R}$ is Lipschitz continuous and convex, then $\partial f(x) \neq \emptyset$ for all $x \in C$. In particular, if $K \geq 0$ is a Lipschitz constant of f, then for all $x \in C$ there exists $x^* \in \partial f(x)$ with $\|x^*\| \leq K$.*

Proof: The epigraph of f is defined by $\text{epi}(f) = \{(x, t) \in C \times \mathbb{R} : t \geq f(x)\}$. Note that $\text{epi}(f) \subset X \times \mathbb{R}$ is a convex set because f is convex with a convex domain C. Now, define

$$H = \{(x, t) \in X \times \mathbb{R} : t < -K\|x\|\}.$$

It is easily seen that H is nonempty and convex. Also, since $\|\cdot\|$ is necessarily continuous, H is open (in the product topology).

Let $x \in C$ be arbitrary. Let $H(x)$ be the translate of H so that its vertex is $(x, f(x))$; that is, $H(x) = (x, f(x)) + H$. We claim that $\text{epi}(f) \cap H(x) = \emptyset$. To see this, note first that

$$H(x) = \{(y, t) \in X \times \mathbb{R} : t < f(x) - K\|y - x\|\}.$$

Now, suppose $(y, t) \in \text{epi}(f)$, so that $t \geq f(y)$. By Lipschitz continuity, we have $f(y) \geq f(x) - K\|y - x\|$. Therefore, $t \geq f(x) - K\|y - x\|$, which implies $(y, t) \notin H(x)$.

Since $H(x)$ is open and nonempty, it has an interior point. We have also shown that $H(x)$ and $\text{epi}(f)$ are disjoint convex sets. Therefore, a version of the separating hyperplane theorem (see Aliprantis and Border (1999, Theorem 5.50)) implies there exists a nonzero continuous linear functional $(x^*, \lambda) \in X^* \times \mathbb{R}$ that separates $H(x)$ and $\text{epi}(f)$. That is, there exists a scalar δ such that

$$\langle y, x^* \rangle + \lambda t \leq \delta \quad \text{if} \quad (y, t) \in \text{epi}(f), \text{ and} \tag{16}$$

$$\langle y, x^* \rangle + \lambda t \geq \delta \quad \text{if} \quad (y, t) \in H(x). \tag{17}$$

[21] The equivalence of assumptions above need not hold if C is not open. For instance, it is possible to construct a set C satisfying the assumptions of Theorem 10 and a function $f : C \to \mathbb{R}$ that is both continuous and linear on C such that f is not locally Lipschitz continuous and $\partial f(x)$ is not a singleton for any $x \in C$. It can also be shown that there exists a set C and a function $f : C \to \mathbb{R}$ satisfying the assumptions of Theorem 10 such that f is not Gâteaux differentiable at any $x \in C$.

Clearly, we cannot have $\lambda > 0$. Also, if $\lambda = 0$, then Equation (17) implies $x^* = 0$. This would contradict (x^*, λ) being a nonzero functional. Therefore, $\lambda < 0$. Without loss of generality, we can take $\lambda = -1$, for otherwise we could renormalize (x^*, λ) and δ by dividing by $|\lambda|$.

Since $(x, f(x)) \in \operatorname{epi}(f)$, we have $\langle x, x^* \rangle - f(x) \leq \delta$. For all $\varepsilon > 0$, we have $(x, f(x) - \varepsilon) \in H(x)$, which implies $\langle x, x^* \rangle - f(x) + \varepsilon \geq \delta$. Therefore, $\langle x, x^* \rangle - f(x) = \delta \geq \langle y, x^* \rangle - f(y)$ for all $y \in C$. Equivalently, we can write $f(y) - f(x) \geq \langle y - x, x^* \rangle$. Thus, $x^* \in \partial f(x)$.

It remains only to show that $\|x^*\| \leq K$. Suppose to the contrary. Then, there exists $y \in X$ such that $\langle y, x^* \rangle < -K\|y\|$, and hence there also exists $\varepsilon > 0$ such that $\langle y, x^* \rangle + \varepsilon < -K\|y\|$. Therefore,

$$\langle y + x, x^* \rangle - f(x) + K\|y\| + \varepsilon < \langle x, x^* \rangle - f(x) = \delta,$$

which, by Equation (17), implies $(y + x, f(x) - K\|y\| - \varepsilon) \notin H(x)$. However, this contradicts the definition of $H(x)$. Thus, $\|x^*\| \leq K$. ■

By definition, for every point in the domain of a locally Lipschitz continuous function, there exists a neighborhood on which the function is Lipschitz continuous. Therefore, the following lemma allows the preceding result to be applied to locally Lipschitz functions:

Lemma 3 *Suppose C is a convex subset of a Banach space X, and fix any $x \in C$ and $\varepsilon > 0$. Then, $\partial f(y) = \partial f|_{B_\varepsilon(x) \cap C}(y)$ for all $y \in B_\varepsilon(x) \cap C$.*

Proof: Fix any $y \in B_\varepsilon(x) \cap C$. It is immediate that $\partial f(y) \subset \partial f|_{B_\varepsilon(x) \cap C}(y)$. To see that $\partial f|_{B_\varepsilon(x) \cap C}(y) \subset \partial f(y)$, fix any $x^* \in \partial f|_{B_\varepsilon(x) \cap C}(y)$. For any $z \in C$, there exists $\lambda \in (0, 1)$ such that $\lambda z + (1 - \lambda)y \in B_\varepsilon(x) \cap C$, so by the definition of the subdifferential and the convexity of f, we have

$$\lambda\langle z - y, x^* \rangle = \langle [\lambda z + (1 - \lambda)y] - y, x^* \rangle \leq f(\lambda z + (1 - \lambda)y) - f(y) \leq \lambda[f(z) - f(y)].$$

Thus, $x^* \in \partial f(y)$. We conclude that $\partial f|_{B_\varepsilon(x) \cap C}(y) = \partial f(y)$ for all $y \in B_\varepsilon(x) \cap C$. ■

For the remainder of this section, assume that X, C, and f are as in Theorem 10. Note that for any $x, y \in C$, we have $\operatorname{span}(C - x) = \operatorname{span}(C - y)$. In addition, since $\operatorname{aff}(C)$ is dense in X, it follows that $\operatorname{span}(C - y) = \operatorname{aff}(C) - y$ is also dense in X. Since any subset of a separable Banach space is separable, $\operatorname{span}(C - y)$ is separable for any $y \in C$. Let $\{x_n\} \subset \operatorname{span}(C - y)$ be a sequence which is dense in $\operatorname{span}(C - y)$ and hence also dense in X. For each $K, m, n \in \mathbb{N}$, let $A_{K,m,n}$ denote the set of all $x \in C$ for which there exist $x^*, y^* \in \partial f(x)$ such that

$$\|x^*\|, \|y^*\| \leq K \text{ and } \langle x_n, x^* - y^* \rangle \geq \tfrac{1}{m}.$$

The following lemmas establish the key properties of $A_{K,m,n}$ that will be needed for our proof of Theorem 10.

Lemma 4 *Let X, C, and f be as in Theorem 10. Then, the set of $x \in C$ for which $\partial f(x)$ is a singleton is $\bigcap_{K,m,n}(C \setminus A_{K,m,n})$.*

Proof: Clearly, if $\partial f(x)$ is a singleton, then $x \in \bigcap_{K,m,n}(C \setminus A_{K,m,n})$. To prove the converse, we will show that if $\partial f(x)$ is not a singleton for $x \in C$, then $x \in A_{K,m,n}$ for some $K, m, n \in \mathbb{N}$. We first show that $\partial f(x) \neq \emptyset$ for all $x \in C$. To see this, fix any $x \in C$. Since f is locally Lipschitz continuous, there exists $\varepsilon > 0$ such that f is Lipschitz continuous on $B_\varepsilon(x) \cap C$. Therefore, by Lemma 2, $\partial f|_{B_\varepsilon(x)\cap C}(x) \neq \emptyset$. By Lemma 3, this implies that $\partial f(x) \neq \emptyset$.

Suppose $\partial f(x)$ is not a singleton. Since $\partial f(x)$ is nonempty, there exist $x^*, y^* \in \partial f(x)$ such that $x^* \neq y^*$. Hence, there exists $y \in X$ such that $\langle y, x^* - y^* \rangle > 0$. Since $\{x_n\}$ is dense in X, by the continuity of $x^* - y^*$, there exists $n \in \mathbb{N}$ such that $\langle x_n, x^* - y^* \rangle > 0$. Thus, there exists $m \in \mathbb{N}$ such that $\langle x_n, x^* - y^* \rangle \geq \frac{1}{m}$. Therefore, taking $K \in \mathbb{N}$ such that $\|x^*\|, \|y^*\| \leq K$, we have $x \in A_{K,m,n}$. ■

Lemma 5 *Let X, C, and f be as in Theorem 10. Then, $A_{K,m,n}$ is a closed subset of C (in the relative topology) for any $K, m, n \in \mathbb{N}$.*

Proof: Consider any net $\{z_d\}_{d\in D} \subset A_{K,m,n}$ such that $z_d \to z$ for some $z \in C$. We will show that $z \in A_{K,m,n}$. For each $d \in D$, choose $x_d^*, y_d^* \in \partial f(z_d)$ such that $\|x_d^*\|, \|y_d^*\| \leq K$ and $\langle x_n, x_d^* - y_d^* \rangle \geq \frac{1}{m}$. Since $\{x^* \in X^* : \|x^*\| \leq K\}$ is weak* compact by Alaoglu's theorem (e.g., see Aliprantis and Border (1999, Theorem 6.25)), any net in this ball has a weak*-convergent subnet. Thus, without loss of generality, we can assume there exist $x^*, y^* \in X^*$ with $\|x^*\|, \|y^*\| \leq K$ such that $x_d^* \xrightarrow{w^*} x^*$ and $y_d^* \xrightarrow{w^*} y^*$. Hence, by the norm-boundedness of the net $\{x_d^*\}_{d\in D}$, the definition of the subdifferential, and the continuity of f, for any $y \in C$,

$$\langle y - z, x^* \rangle = \lim_d \langle y - z_d, x_d^* \rangle \leq \lim_d \big[f(y) - f(z_d)\big] = f(y) - f(z),$$

which implies $x^* \in \partial f(z)$.[22] A similar argument shows $y^* \in \partial f(z)$. Finally, since

$$\langle x_n, x^* - y^* \rangle = \lim_d \langle x_n, x_d^* - y_d^* \rangle \geq \tfrac{1}{m},$$

we have $z \in A_{K,m,n}$, and hence $A_{K,m,n}$ is relatively closed. ■

[22] The first equality follows from a standard result: Let $K \geq 0$, $x \in X$, and $x^* \in X^*$, and let $\{x_d\}_{d\in D} \subset X$ and $\{x_d^*\}_{d\in D} \subset X^*$ be nets such that $x_d \to x$, $x_d^* \xrightarrow{w^*} x^*$, and $\|x_d^*\| \leq K$ for all $d \in D$. Then,

$$\begin{aligned} |\langle x_d, x_d^* \rangle - \langle x, x^* \rangle| &\leq |\langle x_d - x, x_d^* \rangle| + |\langle x, x_d^* - x^* \rangle| \\ &\leq \|x_d - x\| \|x_d^*\| + |\langle x, x_d^* - x^* \rangle| \\ &\leq \|x_d - x\| K + |\langle x, x_d^* - x^* \rangle| \to 0, \end{aligned}$$

which implies $\langle x_d, x_d^* \rangle \to \langle x, x^* \rangle$.

Lemma 6 *Let X, C, and f be as in Theorem 10. Then, $C \setminus A_{K,m,n}$ is dense in C for any $K, m, n \in \mathbb{N}$.*

Proof: Since C is convex, it is straightforward to show that

$$\operatorname{aff}(C) = \{\lambda x + (1-\lambda)y : x, y \in C \text{ and } \lambda \in \mathbb{R}\}. \tag{18}$$

Consider arbitrary $K, m, n \in \mathbb{N}$ and $z \in C$. We will find a sequence $\{z_k\} \subset C \setminus A_{K,m,n}$ such that $z_k \to z$. Recall that $x_n \in \operatorname{span}(C - y)$ for any choice of $y \in C$. Thus, $z + x_n \in z + \operatorname{span}(C - z) = \operatorname{aff}(C)$, so by Equation (18), there exist $x, y \in C$ and $\lambda \in \mathbb{R}$ such that $\lambda x + (1-\lambda)y = z + x_n$. Let us first suppose $\lambda > 1$; we will consider the other cases shortly. Note that $\lambda > 1$ implies $0 < \frac{\lambda-1}{\lambda} < 1$. Consider any sequence $\{a_k\} \subset (0, \frac{\lambda-1}{\lambda})$ such that $a_k \to 0$. Define a sequence $\{y_k\} \subset C$ by $y_k = a_k y + (1-a_k)z$, and note that $y_k \to z$. We claim that for each $k \in \mathbb{N}$, $y_k + \frac{a_k}{\lambda-1}x_n \in C$. To see this, note the following:

$$\begin{aligned} y_k + \tfrac{a_k}{\lambda-1}x_n &= a_k y + (1-a_k)z + \tfrac{a_k}{\lambda-1}\big(\lambda x + (1-\lambda)y - z\big) \\ &= \big(1 - \tfrac{a_k\lambda}{\lambda-1}\big)z + \tfrac{a_k\lambda}{\lambda-1}x. \end{aligned}$$

Since $0 < a_k < \frac{\lambda-1}{\lambda}$, we have $0 < \frac{a_k\lambda}{\lambda-1} < 1$. Thus, $y_k + \frac{a_k}{\lambda-1}x_n$ is a convex combination of z and x, so it is an element of C. This is illustrated in Figure 3.

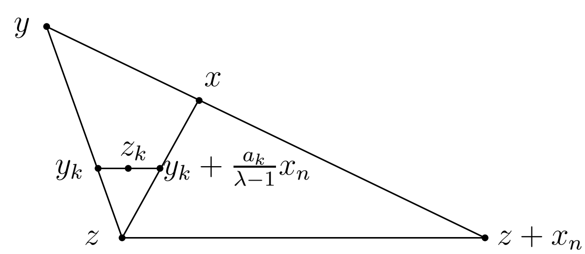

Figure 3: Construction of the sequence $\{z_k\}$

Consider any $k \in \mathbb{N}$. Because C is convex, we have $y_k + tx_n \in C$ for all $t \in (0, \frac{a_k}{\lambda-1})$. Define a function $g : (0, \frac{a_k}{\lambda-1}) \to \mathbb{R}$ by $g(t) = f(y_k + tx_n)$, and note that g is convex. It is a standard result that a convex function on an open interval in $\mathbb{R}$ is differentiable for all but (at most) countably many points of this interval (see Phelps (1993, Theorem 1.16)). Let $t_k \in (0, \frac{a_k}{\lambda-1})$ be such that $g'(t_k)$ exists, and let $z_k = y_k + t_k x_n$. If $x^* \in \partial f(z_k)$, then it is straightforward to verify that the linear mapping $t \mapsto t\langle x_n, x^*\rangle$ is in the subdifferential of g at t_k. Since g is differentiable at t_k, it can only have one element in its subdifferential at that point. Therefore, for any $x^*, y^* \in \partial f(z_k)$, we have $\langle x_n, x^*\rangle = \langle x_n, y^*\rangle$; hence, $z_k \in C \setminus A_{K,m,n}$. Finally, note that since $0 < t_k < \frac{a_k}{\lambda-1}$ and $a_k \to 0$, we have $t_k \to 0$. Therefore, $z_k = y_k + t_k x_n \to z$.

Above, we did restrict attention to the case of $\lambda > 1$. However, if $\lambda < 0$, then let $\lambda' = 1 - \lambda > 1$, $x' = y$, $y' = x$, and the analysis is the same as above. If $\lambda \in [0, 1]$, then note that $z + x_n \in C$. Similar to the preceding paragraph, for any $k \in \mathbb{N}$, define a function $g : (0, \frac{1}{k}) \to \mathbb{R}$ by $g(t) = f(z + tx_n)$. Let $t_k \in (0, \frac{1}{k})$ be such that $g'(t_k)$ exists, and let $z_k = z+t_k x_n$. Then, as argued above, we have $z_k \in C \setminus A_{K,m,n}$ for all $k \in \mathbb{N}$ and $z_k \to z$. ■

Proof of Theorem 10: By Lemma 4, the set of $x \in C$ for which $\partial f(x)$ is a singleton is $\bigcap_{K,m,n}(C \setminus A_{K,m,n})$. By Lemmas 5 and 6, for each $K, m, n \in \mathbb{N}$, $C \setminus A_{K,m,n}$ is open (in the relative topology) and dense in C. Since C is a Baire space, every countable intersection of open dense subsets of C is also dense. This completes the proof. ■

A.2 Proof of Theorem 7

Before proving Theorem 7, we establish a similar result for the case where f is assumed to be locally Lipschitz continuous instead of Lipschitz continuous.

Lemma 7 *Suppose X, C, and f satisfy (i)-(iii) in Theorem 10. Then, for any weak*-closed $\mathcal{M} \subset X^*$, the following are equivalent:*

1. *$\mathcal{M}_f \subset \mathcal{M}$.*
2. *For all $x \in C$, the maximization problem*

$$\max_{x^* \in \mathcal{M}} \big[\langle x, x^* \rangle - f^*(x^*)\big],$$

has a solution and the maximum value is equal to $f(x)$.

Proof: $(1 \Rightarrow 2)$: Let $x \in C$ be arbitrary. By Theorem 10, C_f is dense in C, so there exists a net $\{x_d\}_{d \in D} \subset C_f$ such that $x_d \to x$.[23] Since f is locally Lipschitz continuous, there exists $\varepsilon > 0$ such that f is Lipschitz continuous on $B_\varepsilon(x) \cap C$. Let $K \geq 0$ be a Lipschitz constant of $f|_{B_\varepsilon(x) \cap C}$. Without loss of generality assume that $x_d \in B_\varepsilon(x)$ for all $d \in D$.

For each $d \in D$, by Lemma 2, there exists $x_d^* \in \partial f|_{B_\varepsilon(x) \cap C}(x_d)$ such that $\|x_d^*\| \leq K$. By Lemma 3, $x_d^* \in \partial f|_{B_\varepsilon(x) \cap C}(x_d) = \partial f(x_d)$. Therefore, $\{x_d^*\}_{d \in D} \subset \mathcal{M}_f \cap \{x^* \in X^* : \|x^*\| \leq K\}$, where the intersection is weak* compact since $\mathcal{M}_f$ is weak* closed and $\{x^* \in X^* : \|x^*\| \leq K\}$ is weak* compact by Alaoglu's theorem. Therefore, $\{x_d^*\}_{d \in D}$ has a convergent subnet. Without loss of generality, suppose the net itself converges, so that $x_d^* \xrightarrow{w^*} x^*$ for some $x^* \in \mathcal{M}_f$. By the norm-boundedness of the net $\{x_d^*\}_{d \in D}$, the definition of the subdifferential, and the continuity of f, for any $y \in C$,

$$\langle y - x, x^* \rangle = \lim_d \langle y - x_d, x_d^* \rangle \leq \lim_d \big[f(y) - f(x_d)\big] = f(y) - f(x),$$

[23] If C were also assumed to be open, then one could apply Mazur (1933)'s theorem here instead of Theorem 10. However, in a number of applications, such as the current paper, Epstein, Marinacci and Seo (2007), and Ergin and Sarver (2009), the domain C has an empty interior, yet it satisfies the assumptions of our Theorem 10.

which implies $x^* \in \partial f(x)$.[24] Since $x \in C$ was arbitrary, we conclude that for all $x \in C$, there exists $x^* \in \mathcal{M}_f \subset \mathcal{M}$ such that $x^* \in \partial f(x)$. Then, by parts 2 and 3 of Lemma 1, we conclude that for all $x \in C$,

$$f(x) = \max_{x^* \in \mathcal{M}} \left[\langle x, x^* \rangle - f^*(x^*)\right].$$

($2 \Rightarrow 1$): Fix any $x \in C_f$. Since the maximization problem in condition 2 is assumed to have a solution for every point in C, there exists $x^* \in \mathcal{M}$ such that $f(x) = \langle x, x^* \rangle - f^*(x^*)$, which implies $x^* \in \partial f(x)$ by part 3 of Lemma 1. However, $x \in C_f$ implies $\partial f(x) = \{x^*\}$, and hence $\partial f(x) \subset \mathcal{M}$. Since $x \in C_f$ was arbitrary, we have $\mathcal{N}_f \subset \mathcal{M}$. Because $\mathcal{M}$ is weak* closed, we have $\mathcal{M}_f = \overline{\mathcal{N}_f} \subset \mathcal{M}$. ■

It can be shown that Lemma 7 fails to hold if the maximum in condition 2 is not assumed to exist and maximum is replaced with supremum. We now complete the proof of Theorem 7 by showing that in the special case when f is Lipschitz continuous, $\mathcal{M}_f$ is compact and the existence of the maximum in condition 2 of Lemma 7 can be guaranteed by restricting attention to compact $\mathcal{M}$.

Proof of Theorem 7: If $K \geq 0$ is a Lipschitz constant of f, then Lemma 2 implies that for all $x \in C$ there exists $x^* \in \partial f(x)$ with $\|x^*\| \leq K$. Therefore, if $\partial f(x) = \{x^*\}$, then $\|x^*\| \leq K$. Thus, we have $\|x^*\| \leq K$ for all $x^* \in \mathcal{N}_f$, and hence also for all $x^* \in \mathcal{M}_f$. Since $\mathcal{M}_f$ is a weak*-closed and norm-bounded set in X^*, it is weak* compact by Alaoglu's theorem.

Note that $\mathcal{M} \subset X^*$ is weak* closed since it is weak* compact. Therefore, the ($\Rightarrow$) direction immediately follows from Lemma 7. To see that the ($\Leftarrow$) direction is implied by Lemma 7, it is enough to note that the maximum taken over measures in $\mathcal{M}$ is well-defined. The mapping $x^* \mapsto \langle x, x^* \rangle$ is weak* continuous, and f^* is weak* lower semi-continuous by part 1 of Lemma 1. Therefore, $x^* \mapsto \langle x, x^* \rangle - f^*(x^*)$ is weak* upper semi-continuous and hence attains a maximum on any weak*-compact set. ■

A.3 Proof of Theorem 8

We begin with a lemma for Lipschitz continuous functions.

Lemma 8 *Suppose C is a convex subset of a Banach lattice X. Let $f : C \to \mathbb{R}$ be Lipschitz continuous, convex, and monotone. Let $K \geq 0$ be a Lipschitz constant of f. If $x \in C$ is such that at least one of the following conditions holds:*

1. *for any $x' \in C$, $x \vee x' \in C$, or*

2. *for any $x' \in C$, $x \wedge x' \in C$,*

then there exists a positive $x^* \in \partial f(x)$ with $\|x^*\| \leq K$.

[24] See footnote 22 for an explanation of the first equality.

Proof: Let epi(f), H, and $H(x)$ be as defined in the proof of Lemma 2. Remember that epi(f) and $H(x)$ are non-empty and convex, $H(x)$ is open, and epi$(f) \cap H(x) = \emptyset$ for all $x \in C$. Define $I(x) = H(x) + X_+ \times \{0\}$. Then, $I(x) \subset X \times \mathbb{R}$ is convex as the sum of two convex sets, and it has a non-empty interior since it contains the nonempty open set $H(x)$.

We will show that if $x \in C$ satisfies either of the conditions stated in the lemma, then epi$(f) \cap I(x) = \emptyset$:

Case 1 — $x \vee x' \in C$ for any $x' \in C$: Suppose for a contradiction that $(x', t) \in$ epi$(f) \cap I(x)$. Then, $x' \in C$, and there exist $y \in X$, $z \in X_+$ such that $x' = x + y + z$ and $t < f(x) - K\|y\|$. Let $\bar{x} = x \vee x' \in C$ and $\bar{y} = \bar{x} - x'$. Note that $|\bar{y}| = \bar{y} = (x - x')^+$ and $-y = x - x' + z \geq x - x'$. Hence,

$$|y| = |-y| \geq (-y)^+ \geq (x - x')^+ = |\bar{y}|.$$

Since X is a Banach lattice, the above inequality implies that $\|y\| \geq \|\bar{y}\|$. Monotonicity of f implies that $f(\bar{x}) \geq f(x)$. Thus, $t < f(x) - K\|y\| \leq f(\bar{x}) - K\|\bar{y}\|$. The previous inequality and $\bar{y} = \bar{x} - x'$ imply that $(x', t) \in H(\bar{x})$, a contradiction to epi$(f) \cap H(\bar{x}) = \emptyset$. Thus, epi$(f) \cap I(x) = \emptyset$.

Case 2 — $x \wedge x' \in C$ for any $x' \in C$: Suppose again for a contradiction that $(x', t) \in$ epi$(f) \cap I(x)$. Then, $x' \in C$, and there exist $y \in X$, $z \in X_+$ such that $x' = x + y + z$ and $t < f(x) - K\|y\|$. Let $\bar{x} = x \wedge x' \in C$ and $\bar{y} = x - \bar{x}$. We again have $|\bar{y}| = \bar{y} = (x - x')^+$ and $-y = x - x' + z \geq x - x'$, which implies $|y| \geq |\bar{y}|$ and hence $\|y\| \geq \|\bar{y}\|$. Thus, $t < f(x) - K\|y\| \leq f(x) - K\|\bar{y}\|$. Since $\bar{y} = x - \bar{x}$, this inequality implies $(\bar{x}, t) \in H(x)$. Monotonicity of f implies that $f(\bar{x}) \leq f(x') \leq t$, which implies $(\bar{x}, t) \in$ epi(f), a contradiction to epi$(f) \cap H(x) = \emptyset$. Thus, epi$(f) \cap I(x) = \emptyset$.

We have shown that under either condition 1 or 2, $I(x)$ and epi(f) are disjoint convex sets and $I(x)$ has nonempty interior. Therefore, the same version of the separating hyperplane theorem used in the proof of Lemma 2 implies that there exists a nonzero continuous linear functional $(x^*, \lambda) \in X^* \times \mathbb{R}$ that separates $I(x)$ and epi(f). That is, there exists a scalar δ such that

$$\langle y, x^* \rangle + \lambda t \leq \delta \quad \text{if} \quad (y, t) \in \text{epi}(f), \text{ and} \tag{19}$$

$$\langle y, x^* \rangle + \lambda t \geq \delta \quad \text{if} \quad (y, t) \in I(x). \tag{20}$$

Note that Equation (19) is the same as Equation (16), and Equation (20) implies Equation (17). Therefore, by the exact same arguments as in the proof of Lemma 2, we can without loss of generality let $\lambda = -1$ and conclude that $\delta = \langle x, x^* \rangle - f(x)$, $x^* \in \partial f(x)$, and $\|x^*\| \leq K$.

It only remains to show that x^* is positive. Let $y \in X_+$. Then, for any $\varepsilon > 0$, $(x + y, f(x) - \varepsilon) \in I(x)$. By Equation (20),

$$\langle x + y, x^* \rangle - f(x) + \varepsilon \geq \delta = \langle x, x^* \rangle - f(x),$$

and hence $\langle y, x^* \rangle \geq -\varepsilon$. Since the latter holds for all $\varepsilon > 0$ and $y \in X_+$, we have that $\langle y, x^* \rangle \geq 0$ for all $y \in X_+$. Therefore, x^* is positive. ■

Proof of Theorem 8: First, consider the case where condition 1 holds. Note that $|x \vee x' - x| = (x' - x)^+ \leq |x' - x|$ for any $x, x' \in X$. Since X is a Banach lattice, this implies $\|x \vee x' - x\| \leq \|x' - x\|$. Let $x \in C_f$. Since f is locally Lipschitz continuous, there exists $\varepsilon > 0$ such that f is Lipschitz continuous on $B_\varepsilon(x) \cap C$. Take any $x' \in B_\varepsilon(x) \cap C$. Then, $x \vee x' \in C$, so by the above observations, $x \vee x' \in B_\varepsilon(x) \cap C$. Therefore, by Lemma 8, there exists a positive $x^* \in \partial f|_{B_\varepsilon(x) \cap C}(x)$. By Lemma 3 and $x \in C_f$, we have that $\partial f(x) = \{x^*\}$. Therefore, the functionals in $\mathcal{N}_f$ are positive. Since the set of positive functionals are weak* closed in X^*, every $x^* \in \mathcal{M}_f = \overline{\mathcal{N}_f}$ is also positive.

Next, consider the case where condition 2 holds. Note that $|x - x \wedge x'| = (x - x')^+ \leq |x - x'|$ for any $x, x' \in X$. Since X is a Banach lattice, this implies $\|x - x \wedge x'\| \leq \|x - x'\|$. Let $x \in C_f$. As above, since f is locally Lipschitz continuous, there exists $\varepsilon > 0$ such that f is Lipschitz continuous on $B_\varepsilon(x) \cap C$. Take any $x' \in B_\varepsilon(x) \cap C$. Then, $x \wedge x' \in C$, and hence $x \wedge x' \in B_\varepsilon(x) \cap C$. Therefore, by Lemma 8, there exists a positive $x^* \in \partial f|_{B_\varepsilon(x) \cap C}(x)$. As argued in the previous case, this implies that the functionals in $\mathcal{M}_f$ are all positive. ■

A.4 Proof of Theorem 9

(1): First, note that a solution to Equation (15) must exist for every $x \in C$ since $\mathcal{M}$ is weak* compact, $\langle x, \cdot \rangle$ is weak* continuous, and c is weak* lower semi-continuous. As the maximum of a collection of affine functions, f is obviously convex. To see that f is Lipschitz continuous, note that by the weak* compactness of $\mathcal{M}$ and Alaoglu's Theorem (see Theorem 6.25 in Aliprantis and Border (1999)), there exists $K \geq 0$ such that $\|x^*\| \leq K$ for all $x^* \in \mathcal{M}$. Fix any $x, y \in C$, and let $x^* \in \mathcal{M}$ be a solution to Equation (15) at x. Thus, $f(x) = \langle x, x^* \rangle - c(x^*)$. Since $f(y) \geq \langle y, x^* \rangle - c(x^*)$, we have

$$f(x) - f(y) \leq \langle x - y, x^* \rangle \leq \|x - y\| \|x^*\| \leq \|x - y\| K.$$

A similar argument shows $f(y) - f(x) \leq \|x - y\| K$, and hence $|f(x) - f(y)| \leq \|x - y\| K$. Thus, f is Lipschitz continuous.

(2): Fix any $x \in C$. Let $x^* \in \mathcal{M}$ be a solution to Equation (15), so $f(x) = \langle x, x^* \rangle - c(x^*)$. For any $y \in C$, we have $f(y) \geq \langle y, x^* \rangle - c(x^*)$ and hence $f(y) - f(x) \geq \langle y - x, x^* \rangle$. Therefore, $x^* \in \partial f(x)$. By part 3 of Lemma 1, this implies that

$$\langle x, x^* \rangle - f^*(x^*) = f(x) = \langle x, x^* \rangle - c(x^*),$$

so $f(x^*) = c(x^*)$. To see the other claims, take any $x^* \in \mathcal{N}_f$. Then, there exists $x \in C_f$ such that $\partial f(x) = \{x^*\}$. By the preceding arguments, this implies $x^* \in \mathcal{M}$ and $f^*(x^*) = c(x^*)$.

Since this is true for any $x^* \in \mathcal{N}_f$, we have $\mathcal{N}_f \subset \mathcal{M}$. Since $\mathcal{M}$ is weak* closed, $\mathcal{M}_f = \overline{\mathcal{N}_f} \subset \mathcal{M}$.

(3): We first show that $c(x^*) \geq f^*(x^*)$ for any $x^* \in \mathcal{M}$. Fix any $x^* \in \mathcal{M}$. Then, by the definition of f,

$$\begin{aligned} & f(x) \geq \langle x, x^* \rangle - c(x^*) \quad \forall x \in C \\ \Longrightarrow \; & c(x^*) \geq \langle x, x^* \rangle - f(x) \quad \forall x \in C \\ \Longrightarrow \; & c(x^*) \geq \sup_{x \in C} \big[\langle x, x^* \rangle - f(x) \big] = f^*(x^*). \end{aligned}$$

Since $\mathcal{M}_f \subset \mathcal{M}$ by part 2, this implies $c(x^*) \geq f^*(x^*)$ for all $x^* \in \mathcal{M}_f$. Therefore, it remains only to show that $c(x^*) \leq f^*(x^*)$ for all $x^* \in \mathcal{M}_f$. Fix any $x^* \in \mathcal{M}_f$. Since $\mathcal{M}_f = \overline{\mathcal{N}_f}$, there exists a net $\{x_d^*\}_{d \in D}$ in $\mathcal{N}_f$ that converges to x^* in the weak* topology. Recall from part 2 that $f^*(x^*) = c(x^*)$ for all $x^* \in \mathcal{N}_f$. Therefore,

$$c(x^*) \leq \liminf_d c(x_d^*) = \liminf_d f^*(x_d^*),$$

where the inequality follows from lower semi-continuity of c (see Theorem 2.39 in Aliprantis and Border (1999, p43)). The proof is completed by showing that $\liminf_d f^*(x_d^*) \leq f^*(x^*)$. To see this, first note that by the definition of the limit inferior, there exists a subnet of $\{f^*(x_d^*)\}_{d \in D}$ that converges to $\liminf_d f^*(x_d^*)$. Without loss of generality, assume that the net itself converges to $\liminf_d f^*(x_d^*)$, so $\lim_d f^*(x_d^*) = \liminf_d f^*(x_d^*)$. Since C is compact and f is continuous (by part 1), for each $d \in D$ there exists $x_d \in C$ such that

$$f^*(x_d^*) = \sup_{x \in C} \big[\langle x, x_d^* \rangle - f(x) \big] = \langle x_d, x_d^* \rangle - f(x_d).$$

Since C is compact, the net $\{x_d\}_{d \in D}$ must have a subnet that converges to some limit $x \in C$. Again, without loss of generality, assume that the net itself converges, so $\lim_d x_d = x$.

As in part 1, note that by the compactness of $\mathcal{M}$ and Alaoglu's Theorem, there exists $K \geq 0$ such that $\|x_d^*\| \leq K$ for all $d \in D$. Since $x_d \to x$, $x_d^* \xrightarrow{w^*} x^*$, and $\|x_d^*\| \leq K$ for all $d \in D$, we have $\langle x_d, x_d^* \rangle \to \langle x, x^* \rangle$ (see footnote 22). Given this result and the continuity of f, we have

$$\liminf_d f^*(x_d^*) = \lim_d f^*(x_d^*) = \lim_d \big[\langle x_d, x_d^* \rangle - f(x_d) \big] = \langle x, x^* \rangle - f(x) \leq f^*(x^*),$$

which completes the proof.

B Proof of Theorem 1

Note that $\mathcal{A}$ is a compact metric space since $\triangle(Z)$ is a compact metric space (see, e.g., Munkres (2000, p280–281) or Theorem 1.8.3 in Schneider (1993, p49)). We begin by showing that weak order, continuity, and first-stage independence imply that $\succsim$ has an expected-utility representation.

Lemma 9 *A preference $\succsim$ over $\triangle(\mathcal{A})$ satisfies weak order, continuity, and first-stage independence if and only if there exists a continuous function $V : \mathcal{A} \to \mathbb{R}$ such that $\succsim$ is represented by $\mathbb{E}_P[V]$. Furthermore, if $V : \mathcal{A} \to \mathbb{R}$ and $V' : \mathcal{A} \to \mathbb{R}$ are continuous functions such that $\mathbb{E}_P[V]$ and $\mathbb{E}_P[V']$ represent the same preference over $\triangle(\mathcal{A})$, then there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $V' = \alpha V + \beta$.*

Proof: This is a standard result. For example, it is asserted without proof in Corollary 5.22 of Kreps (1988). Alternatively, it can be verified that $\triangle(\mathcal{A})$ and $\succsim$ satisfy the conditions of Theorem 10.1 of Fishburn (1970), and hence there exists a bounded V such that $\succsim$ is represented by $\mathbb{E}_P[V]$. Since the mapping $P \mapsto \mathbb{E}_P[V]$ is affine, it is straightforward to show that the continuity axiom implies this mapping is weak* continuous. Therefore, for any sequence $\{A_n\} \subset \mathcal{A}$ and any $A \in \mathcal{A}$,

$$A_n \to A \implies \delta_{A_n} \xrightarrow{w^*} \delta_A \implies V(A_n) = \mathbb{E}_{\delta_{A_n}}[V] \to \mathbb{E}_{\delta_A}[V] = V(A),$$

which implies that V is continuous. Uniqueness of V follows from the uniqueness part of the mixture-space theorem (see Kreps (1988, Theorem 5.11) or Fishburn (1970, Theorem 8.4)). ■

Let $\mathcal{A}^c \subset \mathcal{A}$ denote the collection of all convex menus. It is a standard exercise to show that $\mathcal{A}^c$ is a closed subset of $\mathcal{A}$, and hence $\mathcal{A}^c$ is also compact (see Theorem 1.8.5 in Schneider (1993, p50)). Our stategy for proving the sufficiency of the axioms will be to show that the function V described in Lemma 9 satisfies the max-LHA [min-LHA] formula on $\mathcal{A}^c$. Using the IR axiom, it will then be straightforward to show that V satisfies the max-LHA [min-LHA] formula on all of $\mathcal{A}$.

The following lemma shows the implications of our other axioms.

Lemma 10 *Suppose that $V : \mathcal{A} \to \mathbb{R}$ is a continuous function such that $\mathbb{E}_P[V]$ represents the preference $\succsim$ over $\triangle(\mathcal{A})$. Then:*

1. *If $\succsim$ satisfies L–continuity, then V is Lipschitz continuous on $\mathcal{A}^c$, i.e., there exist $K > 0$ such that $|V(A) - V(B)| \leq K d_h(A, B)$ for any $A, B \in \mathcal{A}^c$.*[25]

2. *If V is Lipschitz continuous (on $\mathcal{A}$), then $\succsim$ satisfies L–continuity.*

3. *The preference $\succsim$ satisfies PERU [PLRU] if and only if V is convex [concave].*

4. *The preference $\succsim$ satisfies monotonicity if and only if V is monotone (i.e., $A \subset B$ implies $V(B) \geq V(A)$ for any $A, B \in \mathcal{A}$).*

[25] If $\succsim$ also satisfies IR, then it can be shown that V is Lipschitz continuous on $\mathcal{A}$.

Proof: Claims 3 and 4 follow immediately from the definitions. To prove claim 1, we use the arguments in the proof of Lemma 13 in Ergin and Sarver (2009). Suppose that $\succsim$ satisfies L–continuity. Let $K \equiv 2M[V(A^*) - V(A_*)]$. We first show that for any $A, B \in \mathcal{A}^c$:

$$d_h(A,B) \leq \tfrac{1}{2M} \Longrightarrow |V(A) - V(B)| \leq K d_h(A,B). \tag{21}$$

In particular, this will imply that $K \geq 0$. Suppose that $d_h(A,B) \leq \frac{1}{2M}$ and let $\alpha \equiv M d_h(A,B)$. Then, $\alpha \leq 1/2$ and

$$V(B) - V(A) \leq \tfrac{\alpha}{1-\alpha}[V(A^*) - V(A_*)] \leq 2\alpha[V(A^*) - V(A_*)] = K d_h(A,B),$$

where the first inequality follows from L–continuity, the second inequality follows from $\alpha \leq 1/2$, and the equality follows form the definitions of α and K. Interchanging the roles of A and B above, we also have that $V(A) - V(B) \leq K d_h(A,B)$, proving Equation (21).

Next, we use the argument in the proof of Lemma 8 in the supplementary appendix of DLRS (2007) to show that for any $A, B \in \mathcal{A}^c$:

$$|V(A) - V(B)| \leq K d_h(A,B), \tag{22}$$

i.e., the requirement $d_h(A,B) \leq \frac{1}{2M}$ in Equation (21) is not necessary. To see this, take any sequence $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_n < \lambda_{n+1} = 1$ such that $(\lambda_{i+1} - \lambda_i) d_h(A,B) \leq \frac{1}{2M}$. Let $A_i = \lambda_i A + (1 - \lambda_i)B$. It is straightforward to verify that:[26]

$$d_h(A_{i+1}, A_i) = (\lambda_{i+1} - \lambda_i) d_h(A,B) \leq \tfrac{1}{2M}.$$

Combining this with the triangle inequality and Equation (21), we obtain

$$|V(A) - V(B)| \leq \sum_{i=0}^{n} |V(A_{i+1}) - V(A_i)|$$
$$\leq K \sum_{i=0}^{n} d_h(A_{i+1}, A_i) = K \sum_{i=0}^{n} (\lambda_{i+1} - \lambda_i) d_h(A,B) = K d_h(A,B).$$

Hence, by Equation (22), V is Lipschitz continuous on $\mathcal{A}^c$ with a Lipschitz constant K.

To prove claim 2, suppose that V is Lipschitz continuous with a Lipschitz constant $K > 0$. Let A^* be a maximizer of V on $\mathcal{A}$ and let A_* be a minimizer of V on $\mathcal{A}$. If $V(A^*) = V(A_*)$, then $P \sim Q$ for any $P, Q \in \triangle(\mathcal{A})$, implying that L–continuity holds trivially for A^*, A_*, and $M = 0$. If $V(A^*) > V(A_*)$, then let $M \equiv K/[V(A^*) - V(A_*)] > 0$. For any $A, B \in \mathcal{A}$ and $\alpha \in [0,1]$ with $\alpha \geq M d_h(A,B)$, we have

$$(1-\alpha)[V(B) - V(A)] \leq V(B) - V(A) \leq K d_h(A,B) \leq K\alpha/M = \alpha[V(A^*) - V(A_*)],$$

[26] Note that the convexity of the menus A and B is needed for the first equality.

which implies the conclusion of L–continuity. ■

We now follow a construction similar to the one in DLR (2001) to obtain from V a function W whose domain is the set of support functions. As in the text, let

$$\mathcal{U} = \left\{ u \in \mathbb{R}^Z : \sum_{z \in Z} u_z = 0, \sum_{z \in Z} u_z^2 = 1 \right\}.$$

For any $A \in \mathcal{A}^c$, the support function $\sigma_A : \mathcal{U} \to \mathbb{R}$ of A is defined by $\sigma_A(u) = \max_{p \in A} u \cdot p$. For a more complete introduction to support functions, see Rockafellar (1970) or Schneider (1993). Let $C(\mathcal{U})$ denote the set of continuous real-valued functions on $\mathcal{U}$. When endowed with the supremum norm $\|\cdot\|_\infty$, $C(\mathcal{U})$ is a Banach space. Define an order $\geq$ on $C(\mathcal{U})$ by $f \geq g$ if $f(u) \geq g(u)$ for all $u \in \mathcal{U}$. Let $\Sigma = \{\sigma_A \in C(\mathcal{U}) : A \in \mathcal{A}^c\}$. For any $\sigma \in \Sigma$, let

$$A_\sigma = \bigcap_{u \in \mathcal{U}} \left\{ p \in \triangle(Z) : u \cdot p = \sum_{z \in Z} u_z p_z \leq \sigma(u) \right\}.$$

Lemma 11 *1. For all $A \in \mathcal{A}^c$ and $\sigma \in \Sigma$, $A_{(\sigma_A)} = A$ and $\sigma_{(A_\sigma)} = \sigma$. Hence, σ is a bijection from $\mathcal{A}^c$ to Σ.*

2. For all $A, B \in \mathcal{A}^c$ and any $\lambda \in [0,1]$, $\sigma_{\lambda A + (1-\lambda)B} = \lambda \sigma_A + (1-\lambda)\sigma_B$.

3. For all $A, B \in \mathcal{A}^c$, $d_h(A,B) = \|\sigma_A - \sigma_B\|_\infty$.

4. Σ is convex and compact, and $0 \in \Sigma$.

Proof: Parts 1–3 are standard results that can be found in Rockafellar (1970) or Schneider (1993).[27] For instance, in Schneider (1993), part 1 follows from Theorem 1.7.1, part 2 follows from Theorem 1.7.5, and part 3 follows from Theorem 1.8.11.

For part 4, note that the set Σ is convex by the convexity of $\mathcal{A}^c$ and part 2 of this lemma. As discussed above, the set $\mathcal{A}^c$ is compact, and hence by parts 1 and 3 of this lemma, Σ is a compact subset of the Banach space $C(\mathcal{U})$. Also, if we take $q = (1/|Z|, \ldots, 1/|Z|) \in \triangle(Z)$, then $u \cdot q = 0$ for all $u \in \mathcal{U}$. This implies $\sigma_{\{q\}} = 0$, and hence $0 \in \Sigma$. ■

The following lemma shows that a function defined on $\mathcal{A}^c$ can be transformed into a function on Σ.

Lemma 12 *Suppose $V : \mathcal{A}^c \to \mathbb{R}$, and define a function $W : \Sigma \to \mathbb{R}$ by $W(\sigma) = V(A_\sigma)$. Then:*

1. $V(A) = W(\sigma_A)$ for all $A \in \mathcal{A}^c$.

[27] The standard setting for support functions is the set of nonempty closed and convex subsets of $\mathbb{R}^n$. However, by imposing our normalizations on the domain of the support functions $\mathcal{U}$, the standard results are easily adapted to our setting of nonempty closed and convex subsets of $\triangle(Z)$.

2. V is Lipschitz continuous if and only if W is Lipschitz continuous.

3. If V is convex [concave] if and only if W is convex [concave].

4. V is monotone if and only if W is monotone (i.e., $\sigma \leq \sigma'$ implies $W(\sigma) \leq W(\sigma')$ for any $\sigma, \sigma' \in \Sigma$).

Proof: (1): This follows immediately from part 1 of Lemma 11.

(2): If V is Lipschitz continuous with a Lipschitz constant $K > 0$, then by parts 1 and 3 of Lemma 11, for any $A, B \in \mathcal{A}^c$,

$$|W(\sigma_A) - W(\sigma_B)| = |V(A) - V(B)| \leq K d_h(A, B) = K\|\sigma_A - \sigma_B\|_\infty.$$

A similar argument shows if W Lipschitz continuous, then V is Lipschitz continuous.

(3): If V is convex, then by parts 1 and 2 of Lemma 11, for any $A, B \in \mathcal{A}^c$ and $\lambda \in [0, 1]$,

$$\begin{aligned} W(\lambda\sigma_A + (1-\lambda)\sigma_B) &= W(\sigma_{\lambda A+(1-\lambda)B}) = V(\lambda A + (1-\lambda)B) \\ &\leq \lambda V(A) + (1-\lambda)V(B) = \lambda W(\sigma_A) + (1-\lambda)W(\sigma_B). \end{aligned}$$

Similar arguments can be used to show that convexity of W implies convexity of V and that V is concave if and only if W is concave.

(4): This is an immediate consequence of the following fact, which is easy to see from part 1 of Lemma 11 and the definitions of σ_A and A_σ: For all $A, B \in \mathcal{A}^c$, $A \subset B$ if and only if $\sigma_A \leq \sigma_B$. ■

We denote the set of continuous linear functionals on $C(\mathcal{U})$ (the dual space of $C(\mathcal{U})$) by $C(\mathcal{U})^*$. It is well-known that $C(\mathcal{U})^*$ is the set of finite signed Borel measures on $\mathcal{U}$, where the duality is given by:

$$\langle f, \mu \rangle = \int_{\mathcal{U}} f(u)\mu(du)$$

for any $f \in C(\mathcal{U})$ and $\mu \in C(\mathcal{U})^*$.[28]

For any function $W : \Sigma \to \mathbb{R}$, define the subdifferential ∂W and the conjugate W^* as in Appendix A. Also, define Σ_W, $\mathcal{N}_W$, and $\mathcal{M}_W$ as in Equations (12), (13), and (14), respectively:

$$\begin{aligned} \Sigma_W &= \{\sigma \in \Sigma : \partial W(\sigma) \text{ is a singleton}\}, \\ \mathcal{N}_W &= \{\mu \in C(\mathcal{U})^* : \mu \in \partial W(\sigma) \text{ for some } \sigma \in \Sigma_W\}, \\ \mathcal{M}_W &= \overline{\mathcal{N}_W}, \end{aligned}$$

[28]Since $\mathcal{U}$ is a compact metric space, by the Riesz representation theorem (see Royden (1988, p357)), each continuous linear functional on $C(\mathcal{U})$ corresponds uniquely to a finite signed Baire measure on $\mathcal{U}$. Since $\mathcal{U}$ is a locally compact separable metric space, the Baire sets and the Borel sets of $\mathcal{U}$ coincide (see Royden (1988, p332)). Hence, the set of Baire and Borel finite signed measures also coincide.

where the closure is taken with respect to the weak* topology. We now apply Theorem 7 to the current setting.

Lemma 13 *Suppose $W : \Sigma \to \mathbb{R}$ is Lipschitz continuous and convex. Then, $\mathcal{M}_W$ is weak* compact, and for any weak* compact $\mathcal{M} \subset C(\mathcal{U})^*$,*

$$\mathcal{M}_W \subset \mathcal{M} \iff W(\sigma) = \max_{\mu \in \mathcal{M}} \left[\langle \sigma, \mu \rangle - W^*(\mu) \right] \quad \forall \sigma \in \Sigma.$$

Proof: We simply need to verify that $C(\mathcal{U})$, Σ, and W satisfy the assumptions of Theorem 7. Since $\mathcal{U}$ is a compact metric space, $C(\mathcal{U})$ is separable (see Theorem 8.48 of Aliprantis and Border (1999)). By part 4 of Lemma 11, Σ is a closed and convex subset of $C(\mathcal{U})$ containing the origin. Since Σ is a closed subset of a Banach space, it is a Baire space by the Baire Category theorem. Although the result is stated slightly differently, it is shown in Hörmander (1954) that $\mathrm{span}(\Sigma)$ is dense in $C(\mathcal{U})$. This result is also proved in DLR (2001). Since $0 \in \Sigma$ implies that $\mathrm{aff}(\Sigma) = \mathrm{span}(\Sigma)$, the affine hull of Σ is therefore dense in $C(\mathcal{U})$. Finally, W is Lipschitz continuous and convex by assumption. ■

B.1 Sufficiency of the axioms for the max-LHA representations

To prove the sufficiency of the axioms for the max-LHA representation in part A, suppose that $\succsim$ satisfies Axiom 1 and PERU. By Lemma 9, there exists a continuous function $V : \mathcal{A} \to \mathbb{R}$ such that $\mathbb{E}_P[V]$ represents $\succsim$. Moreover, by Lemma 10, the restriction of V to the set $\mathcal{A}^c$ of convex menus is Lipschitz continuous and convex. With slight abuse of notation, we also denote this restriction by V. By Lemma 12, the function $W : \Sigma \to \mathbb{R}$ defined by $W(\sigma) = V(A_\sigma)$ is Lipschitz continuous and convex. Therefore, by Lemma 13, for all $\sigma \in \Sigma$,

$$W(\sigma) = \max_{\mu \in \mathcal{M}_W} \left[\langle \sigma, \mu \rangle - W^*(\mu) \right].$$

This implies that for all $A \in \mathcal{A}$,

$$\begin{aligned} V(A) = V(\mathrm{co}(A)) &= W(\sigma_{\mathrm{co}(A)}) \\ &= \max_{\mu \in \mathcal{M}_W} \left(\int_{\mathcal{U}} \max_{p \in \mathrm{co}(A)} u(p)\, \mu(du) - W^*(\mu) \right) \\ &= \max_{\mu \in \mathcal{M}_W} \left(\int_{\mathcal{U}} \max_{p \in A} u(p)\, \mu(du) - W^*(\mu) \right), \end{aligned}$$

where the first equality follows from IR and the second equality follows from part 1 of Lemma 12. The function W^* is lower semi-continuous by part 1 of Lemma 1, and $\mathcal{M}_W$ is compact by Lemma 13. It is also immediate from Lemma 13 that $\mathcal{M}_W$ satisfies the minimality condition in Definition 1. Therefore, $(\mathcal{M}_W, W^*)$ is a max-LHA representation for $\succsim$.

To prove the sufficiency of the axioms for the monotone max-LHA representation in part B, suppose that, in addition, $\succsim$ satisfies monotonicity. Then, by Lemmas 10 and 12, the function W is monotone. Also, note that for any $A, B \in \mathcal{A}^c$, $\sigma_A \vee \sigma_B = \sigma_{A \cup B}$. Hence, $\sigma \vee \sigma' \in \Sigma$ for any $\sigma, \sigma' \in \Sigma$. Therefore, by Theorem 8, the measures in $\mathcal{M}_W$ are positive.

B.2 Sufficiency of the axioms for the min-LHA representations

To prove the sufficiency of the axioms for the min-LHA representation in part A, suppose that $\succsim$ satisfies Axiom 1 and PLRU. By Lemma 9, there exists a continuous function $V : \mathcal{A} \to \mathbb{R}$ such that $\mathbb{E}_P[V]$ represents $\succsim$. By Lemmas 10 and 12, the function $W : \Sigma \to \mathbb{R}$ defined by $W(\sigma) = V(A_\sigma)$ is Lipschitz continuous and concave. Define a function $\bar{W} : \Sigma \to \mathbb{R}$ by $\bar{W}(\sigma) = -W(\sigma)$. Then, $\bar{W}$ is Lipschitz continuous and convex, so by Lemma 13, for all $\sigma \in \Sigma$,

$$\bar{W}(\sigma) = \max_{\mu \in \mathcal{M}_{\bar{W}}} \big[\langle \sigma, \mu \rangle - \bar{W}^*(\mu)\big].$$

Let $\mathcal{M} \equiv -\mathcal{M}_{\bar{W}} = \{-\mu : \mu \in \mathcal{M}_{\bar{W}}\}$, and define $c : \mathcal{M} \to \mathbb{R}$ by $c(\mu) = \bar{W}^*(-\mu)$. Then, for any $\sigma \in \Sigma$,

$$\begin{aligned} W(\sigma) = -\bar{W}(\sigma) &= \min_{\mu \in \mathcal{M}_{\bar{W}}} \big[- \langle \sigma, \mu \rangle + \bar{W}^*(\mu)\big] \\ &= \min_{\mu \in \mathcal{M}} \big[- \langle \sigma, -\mu \rangle + \bar{W}^*(-\mu)\big] \\ &= \min_{\mu \in \mathcal{M}} \big[\langle \sigma, \mu \rangle + c(\mu)\big]. \end{aligned}$$

This implies that for all $A \in \mathcal{A}$,

$$\begin{aligned} V(A) = V(\mathrm{co}(A)) &= W(\sigma_{\mathrm{co}(A)}) \\ &= \min_{\mu \in \mathcal{M}} \left(\int_{\mathcal{U}} \max_{p \in \mathrm{co}(A)} u(p)\, \mu(du) + c(\mu) \right) \\ &= \min_{\mu \in \mathcal{M}} \left(\int_{\mathcal{U}} \max_{p \in A} u(p)\, \mu(du) + c(\mu) \right), \end{aligned}$$

where the first equality follows from IR and the second equality follows from part 1 of Lemma 12. The function $\bar{W}^*$ is lower semi-continuous by part 1 of Lemma 1, which implies that c is lower semi-continuous. The compactness of $\mathcal{M}$ follows from the compactness of $\mathcal{M}_{\bar{W}}$, which follows from Lemma 13. Also, by Lemma 13 and the above construction, it is immediate that $\mathcal{M}$ satisfies the minimality condition in Definition 1. Therefore, $(\mathcal{M}, c)$ is a min-LHA representation for $\succsim$.

To prove the sufficiency of the axioms for the monotone min-LHA representation in part B, suppose that, in addition, $\succsim$ satisfies monotonicity. Then, by Lemmas 10 and 12, the function W is monotone. Let $\hat{\Sigma} \equiv -\Sigma = \{-\sigma : \sigma \in \Sigma\}$, and define a function $\hat{W} : \hat{\Sigma} \to \mathbb{R}$ by $\hat{W}(\sigma) \equiv \bar{W}(-\sigma) = -W(-\sigma)$. Notice that $\hat{W}$ is monotone and convex: By the monotonicity of W, for any $\sigma, \sigma' \in \hat{\Sigma}$,

$$\sigma \leq \sigma' \implies -\sigma \geq -\sigma' \implies \hat{W}(\sigma) = -W(-\sigma) \leq -W(-\sigma') = \hat{W}(\sigma).$$

By the concavity of W, for any $\sigma, \sigma' \in \hat{\Sigma}$ and $\lambda \in [0, 1]$,

$$\begin{gathered}\hat{W}(\lambda\sigma + (1-\lambda)\sigma') = -W(\lambda(-\sigma) + (1-\lambda)(-\sigma')) \\ \leq -\lambda W(-\sigma) - (1-\lambda)W(-\sigma') = \lambda\hat{W}(\sigma) + (1-\lambda)\hat{W}(\sigma').\end{gathered}$$

Also, for any $A, B \in \mathcal{A}^c$, $(-\sigma_A) \wedge (-\sigma_B) = -(\sigma_A \vee \sigma_B) = -\sigma_{A \cup B}$. Hence, $\sigma \wedge \sigma' \in \hat{\Sigma}$ for any $\sigma, \sigma' \in \hat{\Sigma}$. Therefore, by Theorem 8, the measures in $\mathcal{M}_{\hat{W}}$ are positive. For any $\mu \in C(\mathcal{U})^*$ and $\sigma, \sigma' \in \hat{\Sigma}$, note that

$$\hat{W}(\sigma') - \hat{W}(\sigma) \geq \langle \sigma' - \sigma, \mu \rangle \iff \bar{W}(-\sigma') - \bar{W}(-\sigma) \geq \langle \sigma' - \sigma, \mu \rangle = \langle -\sigma' + \sigma, -\mu \rangle,$$

and hence $\mu \in \partial\hat{W}(\sigma) \iff -\mu \in \partial\bar{W}(-\sigma)$. In particular, $\hat{\Sigma}_{\hat{W}} = -\Sigma_{\bar{W}}$ and $\mathcal{N}_{\hat{W}} = -\mathcal{N}_{\bar{W}}$. Taking closures, we have $\mathcal{M}_{\hat{W}} = -\mathcal{M}_{\bar{W}} = \mathcal{M}$. Thus, the measures in $\mathcal{M}$ are positive.

B.3 Necessity of the axioms

We begin by demonstrating some of the properties of the function V defined by an LHA representation.

Lemma 14 *Suppose $(\mathcal{M}, c)$ is an LHA representation.*

1. *If $(\mathcal{M}, c)$ is a max-LHA representation and $V : \mathcal{A} \to \mathbb{R}$ is defined by Equation (1), then V is Lipschitz continuous and convex. In addition, defining the function $W : \Sigma \to \mathbb{R}$ by $W(\sigma) = V(A_\sigma)$, we have $\mathcal{M} = \mathcal{M}_W$ and $c(\mu) = W^*(\mu)$ for all $\mu \in \mathcal{M}$.*

2. *If $(\mathcal{M}, c)$ is a min-LHA representation and $V : \mathcal{A} \to \mathbb{R}$ is defined by Equation (2), then V is Lipschitz continuous and concave. In addition, defining the function $W : \Sigma \to \mathbb{R}$ by $W(\sigma) = V(A_\sigma)$, we have $\mathcal{M} = -\mathcal{M}_{-W}$ and $c(\mu) = [-W]^*(-\mu)$ for all $\mu \in \mathcal{M}$.*

Proof: (1): By the definitions of V and W, we have

$$W(\sigma) = \max_{\mu \in \mathcal{M}} \left[\langle \sigma, \mu \rangle - c(\mu)\right], \quad \forall \sigma \in \Sigma.$$

By part 1 of Theorem 9, W is Lipschitz continuous and convex. Therefore, the restriction of V to $\mathcal{A}^c$ is Lipschitz continuous and convex by Lemma 12. Let $K > 0$ be any Lipschitz constant of $V|_{\mathcal{A}^c}$, and take any $A, B \in \mathcal{A}$. It is easily verified that $V(A) = V(\mathrm{co}(A))$, $V(B) = V(\mathrm{co}(B))$, and $d_h(\mathrm{co}(A), \mathrm{co}(B)) \leq d_h(A, B)$. Hence,

$$|V(A) - V(B)| = |V(\mathrm{co}(A)) - V(\mathrm{co}(B))| \leq K d_h(\mathrm{co}(A), \mathrm{co}(B)) \leq K d_h(A, B),$$

which implies that V is Lipschitz continuous on all of $\mathcal{A}$ with the same Lipschitz constant K. Also, for any $A, B \in \mathcal{A}$ and $\lambda \in [0,1]$,

$$\begin{aligned} V(\lambda A + (1-\lambda)B) &= V(\text{co}(\lambda A + (1-\lambda)B)) = V(\lambda \text{co}(A) + (1-\lambda)\text{co}(B)) \\ &\leq \lambda V(\text{co}(A)) + (1-\lambda)V(\text{co}(B)) = \lambda V(A) + (1-\lambda)V(B), \end{aligned}$$

which implies that V is convex on $\mathcal{A}$. Also, by parts 2 and 3 of Theorem 9 and the compactness of Σ, $\mathcal{M}_W \subset \mathcal{M}$ and $W^*(\mu) = c(\mu)$ for all $\mu \in \mathcal{M}_W$. By Lemma 13 and the minimality of $\mathcal{M}$, this implies $\mathcal{M} = \mathcal{M}_W$, and hence $c(\mu) = W^*(\mu)$ for all $\mu \in \mathcal{M}$.

(2): Define a function $\bar{W} : \Sigma \to \mathbb{R}$ by $\bar{W}(\sigma) = -W(\sigma)$. Then, for any $\sigma \in \Sigma$,

$$\begin{aligned} \bar{W}(\sigma) = -W(\sigma) &= -\min_{\mu \in \mathcal{M}} \left[\langle \sigma, \mu \rangle + c(\mu)\right] \\ &= \max_{\mu \in \mathcal{M}} \left[\langle \sigma, -\mu \rangle - c(\mu)\right] \\ &= \max_{\mu \in -\mathcal{M}} \left[\langle \sigma, \mu \rangle - c(-\mu)\right]. \end{aligned}$$

By the same arguments used above, this implies that $\bar{W}$ is Lipschitz continuous and convex, which in turn implies that V is Lipschitz continuous and concave. Moreover, the above arguments imply that $-\mathcal{M} = \mathcal{M}_{\bar{W}}$ and $c(-\mu) = \bar{W}^*(\mu)$ for all $\mu \in -\mathcal{M}$. Thus, $\mathcal{M} = -\mathcal{M}_{\bar{W}} = -\mathcal{M}_{-W}$ and $c(\mu) = \bar{W}^*(-\mu) = [-W]^*(-\mu)$ for all $\mu \in \mathcal{M}$. ■

Suppose that $\succsim$ has a max-LHA [min-LHA] representation $(\mathcal{M}, c)$, and suppose $V : \mathcal{A} \to \mathbb{R}$ is defined by Equation (1) [(2)]. Since $\mathbb{E}_P[V]$ represents $\succsim$ and V is continuous (by Lemma 14), $\succsim$ satisfies weak order, continuity, and first-stage independence by Lemma 9. Since V is Lipschitz continuous and convex [concave] by Lemma 14, $\succsim$ satisfies L–continuity and PERU [PLRU] by Lemma 10. Since $V(A) = V(\text{co}(A))$ for all $A \in \mathcal{A}$, it is immediate that $\succsim$ satisfies IR. Finally, if the measures in $\mathcal{M}$ are positive, then it is obvious that V is monotone, which implies that $\succsim$ satisfies monotonicity.

C Proof of Theorem 2

Throughout this section, we will continue to use notation and results for support functions that were established in Appendix B. Suppose $(\mathcal{M}, c)$ and $(\mathcal{M}', c')$ are two max-LHA representations $\succsim$. Define $V : \mathcal{A} \to \mathbb{R}$ and $V' : \mathcal{A} \to \mathbb{R}$ for these respective representations, and define $W : \Sigma \to \mathbb{R}$ and $W' : \Sigma \to \mathbb{R}$ by $W(\sigma) = V(A_\sigma)$ and $W'(\sigma) = V'(A_\sigma)$. By part 1 of Lemma 14, $\mathcal{M} = \mathcal{M}_W$ and $c(\mu) = W^*(\mu)$ for all $\mu \in \mathcal{M}$. Similarly, $\mathcal{M}' = \mathcal{M}_{W'}$ and $c'(\mu) = W'^*(\mu)$ for all $\mu \in \mathcal{M}'$.

Since V is continuous (by Lemma 14), the uniqueness part of Lemma 9 implies that there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $V' = \alpha V - \beta$. This implies that $W' = \alpha W - \beta$. Therefore, for any $\sigma, \sigma' \in \Sigma$,

$$W(\sigma') - W(\sigma) \geq \langle \sigma' - \sigma, \mu \rangle \iff W'(\sigma') - W'(\sigma) \geq \langle \sigma' - \sigma, \alpha\mu \rangle,$$

and hence $\partial W'(\sigma) = \alpha \partial W(\sigma)$. In particular, $\Sigma_{W'} = \Sigma_W$ and $\mathcal{N}_{W'} = \alpha \mathcal{N}_W$. Taking closures we also have that $\mathcal{M}_{W'} = \alpha \mathcal{M}_W$. Since from our earlier arguments $\mathcal{M}' = \mathcal{M}_{W'}$ and $\mathcal{M} = \mathcal{M}_W$, we conclude that $\mathcal{M}' = \alpha \mathcal{M}$. Finally, let $\mu \in \mathcal{M}$. Then,

$$c'(\alpha\mu) = \sup_{\sigma \in \Sigma} \left[\langle \sigma, \alpha\mu \rangle - W'(\sigma) \right] = \alpha \sup_{\sigma \in \Sigma} \left[\langle \sigma, \mu \rangle - W(\sigma) \right] + \beta = \alpha c(\mu) + \beta,$$

where the first and last equalities follow from our earlier findings that $c' = W'^*|_{\mathcal{M}_{W'}}$ and $c = W^*|_{\mathcal{M}_W}$.

The proof of the uniqueness of the min-LHA representation is similar and involves an application of part 2 of Lemma 14.

D Proof of Theorem 3

($1 \Rightarrow 3$): Fix a monotone min-LHA representation $(\mathcal{M}, c)$, and define V by Equation (2). Since $\mathcal{M}$ is compact, there is $\kappa > 0$ such that $\mu(\mathcal{U}) \leq \kappa$ for all $\mu \in \mathcal{M}$. Let $\Omega = \cup_{\lambda \in [0,\kappa]} \lambda \mathcal{U}$ and let $\mathcal{F}$ be the Borel σ-algebra generated by the relative topology of Ω in $\mathbb{R}^Z$. Define $U : \Omega \to \mathbb{R}^Z$ by $U(\omega) = \omega$.

For each $\mu \in \mathcal{M}$, define the probability measure π_μ on $(\Omega, \mathcal{F})$ as follows. If $\mu(\mathcal{U}) = 0$, let π_μ be the degenerate probability measure that puts probability one on $0 \in \Omega$, i.e., for any $E \in \mathcal{F}$, $\pi_\mu(E) = 1$ if $0 \in E$, and $\pi_\mu(E) = 0$ otherwise. If $\mu(\mathcal{U}) > 0$, then define the probability measure $\tilde{\mu}$ on $\mathcal{U}$ and its Borel σ-algebra by $\tilde{\mu}(E) = \frac{1}{\mu(\mathcal{U})} \mu(E)$ for any measurable $E \subset \mathcal{U}$. Define the function $f_\mu : \mathcal{U} \to \Omega$ by $f_\mu(u) = \mu(\mathcal{U}) u$. Note that f is measurable because it is continuous. Finally, let π_μ be defined by $\pi_\mu = \tilde{\mu} \circ f_\mu^{-1}$. Then,

$$\int_\Omega \max_{p \in A} U(\omega) \cdot p \; \pi_\mu(d\omega) = \int_{\mathcal{U}} \max_{p \in A} u(p) \, \mu(du)$$

for any $A \in \mathcal{A}$.[29] Let $\Pi = \{\pi_\mu : \mu \in \mathcal{M}\}$ and $\tilde{c}(\pi_\mu) = c(\mu)$. Then, V can be expressed in the

[29] This is easy to see if $\mu(\mathcal{U}) = 0$. If $\mu(\mathcal{U}) > 0$, then define the function $g : \Omega \to \mathbb{R}$ by $g(\omega) = \max_{p \in A} U(\omega) \cdot p$. To see that g is $\mathcal{F}$-measurable, let B be a countable dense subset of A. At each $\omega \in \Omega$, $\max_{p \in A} \tilde{U}(\omega) \cdot p$ exists and is equal to $\sup_{p \in B} \tilde{U}(\omega) \cdot p$. For each $p \in B$, $\tilde{U} \cdot p$ is $\mathcal{F}$–measurable as a convex combination of $\mathcal{F}$–measurable random variables. Hence, g an $\mathcal{F}$–measurable as the pointwise supremum of countably many $\mathcal{F}$–measurable random variables (see Billingsley (1995, p184), Theorem 13.4(i)). Then,

$$\int_\Omega \max_{p \in A} U(\omega) \cdot p \; \pi_\mu(d\omega) = \int_{\mathcal{U}} \max_{p \in A} \mu(\mathcal{U}) u(p) \, \tilde{\mu}(du) = \int_{\mathcal{U}} \max_{p \in A} u(p) \, \mu(du),$$

where the first equality follows from the change of variables formula $\int_\Omega g(\omega) \, (\tilde{\mu} \circ f_\mu^{-1})(d\omega) =$

following SSV form:

$$V(A) = \min_{\pi \in \Pi} \left(\int_{\Omega} \max_{p \in A} U(\omega) \cdot p \ \pi(d\omega) + \tilde{c}(\pi) \right).$$

(3 $\Rightarrow$ 2): Let $((\Omega, \mathcal{F}), U, \Pi, c)$ be an SSV representation, and define V by Equation (4). Let the subset $\Pi' \subset \Pi$ stand for the set of $\pi \in \Pi$ such that there exists $A \in \mathcal{A}$ for which π solves the minimization problem in Equation (4). Note that Equation (4) continues to hold when Π is replaced by Π', i.e.,

$$V(A) = \min_{\pi \in \Pi'} \left(\int_{\Omega} \max_{p \in A} U(\omega) \cdot p \ \pi(d\omega) + c(\pi) \right) \tag{23}$$

for all $A \in \mathcal{A}$.

We first show that c is bounded on Π'. Note that since U is bounded, there exists $\kappa > 0$ such that the absolute value of the integral term in Equation (23) is bounded by κ for every menu $A \in \mathcal{A}$ and probability measure in $\pi \in \Pi'$. Take any $\pi, \pi' \in \Pi'$, and suppose that they solve the minimization in Equation (23) for menus A and A', respectively. Then, optimality of π at A implies:

$$c(\pi) - c(\pi') \leq \int_{\Omega} \max_{p \in A} U(\omega) \cdot p \ \pi'(d\omega) - \int_{\Omega} \max_{p \in A} U(\omega) \cdot p \ \pi(d\omega) \leq 2\kappa.$$

Similarly, optimality of π' at A' implies:

$$-2\kappa \leq \int_{\Omega} \max_{p \in A'} U(\omega) \cdot p \ \pi'(d\omega) - \int_{\Omega} \max_{p \in A'} U(\omega) \cdot p \ \pi(d\omega) \leq c(\pi) - c(\pi').$$

Therefore, $|c(\pi) - c(\pi')| \leq 2\kappa$ for any $\pi, \pi' \in \Pi'$, implying that c is bounded on Π'.

Let $\tilde{\Omega} = \Omega \times \Pi'$. Let $\mathcal{G}$ be any σ-algebra on Π' that contains all singletons and such that $c|_{\Pi'} : \Pi' \to \mathbb{R}$ is $\mathcal{G}$-measurable (e.g. $\mathcal{G} = 2^{\Pi'}$). Let $\tilde{\mathcal{F}} = \mathcal{F} \otimes \mathcal{G}$ be the product σ-algebra on $\tilde{\Omega}$. Let $\mathbf{1} \in \mathbb{R}^Z$ denote the vector whose coordinates are equal to 1, and define $\tilde{U} : \tilde{\Omega} \to \mathbb{R}^Z$ by $\tilde{U}(\omega, \pi) = U(\omega) + c(\pi)\mathbf{1}$ for any $\tilde{\omega} = (\omega, \pi) \in \tilde{\Omega}$. Note that $\tilde{U}$ is $\tilde{\mathcal{F}}$-measurable and bounded.[30]

For any $\pi \in \Pi'$, define the function $f_\pi : \Omega \to \tilde{\Omega}$ by $f_\pi(\omega) = (\omega, \pi)$. Note that f_π is

---

$\int_{\mathcal{U}} g(f_\mu(u)) \, \tilde{\mu}(du)$.

[30] $\tilde{U}$ is bounded because U is bounded on Ω and c is bounded on Π'. To see that $\tilde{U}$ is $\tilde{\mathcal{F}}$-measurable, note that since U is $\mathcal{F}$-measurable and $\tilde{\mathcal{F}}$ is the product of the σ-algebras $\mathcal{F}$ and $\mathcal{G}$, the function $f : \tilde{\Omega} \to \mathbb{R}^Z$ defined by $f(\omega, \pi) = U(\omega)$ is $\tilde{\mathcal{F}}$-measurable. Also note that since $c|_{\Pi'}$ is $\mathcal{G}$-measurable, and $\tilde{\mathcal{F}}$ is the product of the σ-algebras $\mathcal{F}$ and $\mathcal{G}$, the function $g : \tilde{\Omega} \to \mathbb{R}^Z$ defined by $g(\omega, \pi) = c(\pi)\mathbf{1}$ is also $\tilde{\mathcal{F}}$-measurable. Therefore, $\tilde{U}$ is $\tilde{\mathcal{F}}$-measurable as the sum of the two $\tilde{\mathcal{F}}$-measurable functions f and g.

measurable.[31] Define the probability measure ρ_π on $(\tilde{\Omega}, \tilde{\mathcal{F}})$ by $\rho_\pi = \pi \circ f_\pi^{-1}$. For any $A \in \mathcal{A}$,

$$\begin{aligned}\int_{\tilde{\Omega}} \max_{p\in A} \tilde{U}(\tilde{\omega}) \cdot p \;\rho_\pi(d\tilde{\omega}) &= \int_\Omega \max_{p\in A} \tilde{U}(f_\pi(\omega)) \cdot p \;\pi(d\omega) \\ &= \int_\Omega \left[\max_{p\in A} U(\omega)\cdot p + c(\pi)\right] \pi(d\omega) \\ &= \int_\Omega \max_{p\in A} U(\omega)\cdot p \;\pi(d\omega) + c(\pi),\end{aligned}$$

where the first equality above follows from the change of variables formula.[32]

Letting $\tilde{\Pi}' = \{\rho_\pi : \pi \in \Pi'\}$, by Equation (23), we see that V can be expressed in the following SSMP form:

$$V(A) = \min_{\rho\in\tilde{\Pi}'} \int_{\tilde{\Omega}} \max_{p\in A} \tilde{U}(\tilde{\omega})\cdot p \;\rho(d\tilde{\omega}).$$

($2 \Rightarrow 1$): Let $((\Omega, \mathcal{F}), U, \Pi)$ be an SSMP representation, and define V by Equation (3). It is easy to see that V is monotone and concave. We next show that V is Lipschitz continuous. For every $\pi \in \Pi$, define $f_\pi : \mathcal{A} \to \mathbb{R}$ by

$$f_\pi(A) = \int_\Omega \max_{p\in A} U(\omega)\cdot p \;\pi(d\omega).$$

Since U is bounded, there exists $\kappa > 0$ such that $\|U(\omega)\| \leq \kappa$ for all $\omega \in \Omega$. Let $A, B \in \mathcal{A}$. Given a state $\omega \in \Omega$, let p^* be a solution of $\max_{p\in A} U(\omega)\cdot p$. By definition of Hausdorff distance, there exists $q^* \in B$ such that $\|p^* - q^*\| \leq d_h(A, B)$. Then,

$$\begin{aligned}&\max_{p\in A} U(\omega)\cdot p - \max_{q\in B} U(\omega)\cdot q = U(\omega)\cdot p^* - \max_{q\in B} U(\omega)\cdot q \\ &\leq U(\omega)\cdot p^* - U(\omega)\cdot q^* \leq \|U(\omega)\|\,\|p^* - q^*\| \leq \kappa\, d_h(A, B).\end{aligned}$$

Taking the expectation of the above inequality with respect to π, we obtain:

$$f_\pi(A) - f_\pi(B) \leq \kappa\, d_h(A, B).$$

Hence f_π is Lipschitz continuous with a Lipschitz constant κ that does not depend on $\pi \in \Pi$. Since V is the pointwise minimum of f_π over $\pi \in \Pi$, it is also Lipschitz continuous with the same Lipschitz constant κ.

Since $V : \mathcal{A} \to \mathbb{R}$ is monotone, concave, Lipschitz continuous, and it satisfies the IR

[31] To see this, note that the collection $\tilde{\mathcal{F}}'$ of sets $E \subset \tilde{\Omega}$ satisfying $\{\omega \in \Omega : (\omega, \pi') \in E\} \in \mathcal{F}$ for every $\pi' \in \Pi'$, is a σ-algebra. Since $\tilde{\mathcal{F}}'$ contains both $F \times \Pi'$ and $\Omega \times G$ for every $F \in \mathcal{F}$ and $G \in \mathcal{G}$, we have that $\tilde{\mathcal{F}} = \mathcal{F} \otimes \mathcal{G} \subset \tilde{\mathcal{F}}'$. It is easy to see that f_π would be measurable if $\tilde{\Omega}$ were endowed with the σ-algebra $\tilde{\mathcal{F}}'$. Therefore, f_π is measurable since $\tilde{\Omega}$ is endowed with the coarser σ-algebra $\tilde{\mathcal{F}}$.

[32] To see this, define the function $g : \tilde{\Omega} \to \mathbb{R}$ by $g(\tilde{\omega}) = \max_{p\in A} \tilde{U}(\tilde{\omega})\cdot p$. By a similar argument as in Footnote 29, g is $\tilde{\mathcal{F}}$-measurable. Then, the change of variables formula is $\int_{\tilde{\Omega}} g(\tilde{\omega})\,(\pi \circ f_\pi^{-1})(d\tilde{\omega}) = \int_\Omega g(f_\pi(\omega))\,\pi(d\omega)$.

condition $V(A) = V(\text{co}(A))$ for all $A \in \mathcal{A}$, the construction in Appendix B.2 implies that there exists a monotone min-LHA representation such that V is given by Equation (2).

E Proof of Theorem 4

We define the set of *translations* to be

$$\Theta \equiv \left\{ \theta \in \mathbb{R}^Z : \sum_{z \in Z} \theta_z = 0 \right\}.$$

For $A \in \mathcal{A}$ and $\theta \in \Theta$, define $A + \theta \equiv \{p + \theta : p \in A\}$. Intuitively, adding θ to A in this sense simply "shifts" A. Also, note that for any $p, q \in \triangle(Z)$, we have $p - q \in \Theta$.

Definition 9 A function $V : \mathcal{A} \to \mathbb{R}$ is *translation linear* if there exists $v \in \mathbb{R}^Z$ such that for all $A \in \mathcal{A}$ and $\theta \in \Theta$ with $A + \theta \in \mathcal{A}$, we have $V(A + \theta) = V(A) + v \cdot \theta$.

Lemma 15 *Suppose that $V : \mathcal{A} \to \mathbb{R}$ is a function such that $\mathbb{E}_P[V]$ represents the preference $\succsim$ over $\triangle(\mathcal{A})$. Then, V is translation linear if and only if $\succsim$ satisfies RDD.*

Proof: Assume that $\mathbb{E}_P[V]$ represents the preference $\succsim$. Then, it is easy to see that $\succsim$ satisfies RDD if and only if

$$V(\alpha A + (1-\alpha)\{p\}) - V(\alpha A + (1-\alpha)\{q\}) = (1-\alpha)[V(\{p\}) - V(\{q\})] \tag{24}$$

for any $\alpha \in [0,1]$, $A \in \mathcal{A}$, and $p, q \in \triangle(Z)$.

If there exists $v \in \mathbb{R}^Z$ such that for all $A \in \mathcal{A}$ and $\theta \in \Theta$ with $A + \theta \in \mathcal{A}$, we have $V(A + \theta) = V(A) + v \cdot \theta$, then both sides of Equation (24) are equal to $(1-\alpha)v \cdot (p - q)$, showing that $\succsim$ satisfies RDD.

If $\succsim$ satisfies RDD, then define the function $f : \triangle(Z) \to \mathbb{R}$ by $f(p) = V(\{p\})$ for all $p \in \triangle(Z)$. Let $\alpha \in [0,1]$ and $p, q \in \triangle(Z)$, then

$$\begin{aligned}
2f(\alpha p + (1-\alpha)q) &= [f(\alpha p + (1-\alpha)q) - f(\alpha p + (1-\alpha)p)] \\
&\quad + [f(\alpha p + (1-\alpha)q) - f(\alpha q + (1-\alpha)q)] + f(p) + f(q) \\
&= (1-\alpha)[f(q) - f(p)] + \alpha[f(p) - f(q)] + f(p) + f(q) \\
&= 2[\alpha f(p) + (1-\alpha)f(q)],
\end{aligned}$$

where the second equality follows from Equation (24) and the definition of f. Therefore, $f(\alpha p + (1-\alpha)q) = \alpha f(p) + (1-\alpha)f(q)$ for any $\alpha \in [0,1]$ and $p, q \in \triangle(Z)$. It is standard to show that this implies that there exists $v \in \mathbb{R}^Z$ such that $f(p) = v \cdot p$ for all $p \in \triangle(Z)$.

To see that V is translation linear, let $A \in \mathcal{A}$ and $\theta \in \Theta$ be such that $A + \theta \in \mathcal{A}$. If $\theta = 0$, then the conclusion of translation linearity follows trivially, so without loss of generality assume that $\theta \neq 0$. Ergin and Sarver (2009) show in the proof of their Lemma 4 that if $A \in \mathcal{A}$ and $A + \theta \in \mathcal{A}$ for some $\theta \in \Theta \setminus \{0\}$, then there exist $A' \in \mathcal{A}$, $p, q \in \triangle(Z)$, and $\alpha \in (0,1]$ such that $A = (1-\alpha)A' + \alpha\{p\}$, $A + \theta = (1-\alpha)A' + \alpha\{q\}$, and $\theta = \alpha(p-q)$. Then

$$\begin{aligned}
V(A+\theta) - V(A) &= V((1-\alpha)A' + \alpha\{p\}) - V((1-\alpha)A' + \alpha\{q\}) \\
&= \alpha[V(\{p\}) - V(\{q\})] \\
&= \alpha[v \cdot p - v \cdot q] \\
&= v \cdot \theta,
\end{aligned}$$

where the second equality follows from Equation (24) and the third equality follows from the expected utility form of f. Therefore, V is translation linear. ■

We are now ready to prove Theorem 4. The necessity of RDD in parts A and B are straightforward and left to the reader. In the rest of this section, we will continue to use the notation and results from Appendix B. For the sufficiency direction of part A, suppose that $\succsim$ satisfies Axiom 1, PERU, and RDD. Then, $(\mathcal{M}_W, W^*)$ constructed in Appendix B is a max-LHA representation for $\succsim$. Since $\succsim$ satisfies RDD, by Lemma 15, V is translation linear. Let $v \in \mathbb{R}^Z$ such that for all $A \in \mathcal{A}$ and $\theta \in \Theta$ with $A+\theta \in \mathcal{A}$, we have $V(A+\theta) = V(A) + v \cdot \theta$. Let $q = (1/|Z|, \ldots, 1/|Z|) \in \triangle(Z)$. By Lemma 22 of Ergin and Sarver (2009), for all $\mu \in \mathcal{M}_W$ and $p \in \triangle(Z)$, $\langle \sigma_{\{p\}}, \mu \rangle = v \cdot (p - q)$. The consistency of $\mathcal{M}_W$ follows immediately from this fact because for any $\mu, \mu' \in \mathcal{M}_W$ and $p \in \triangle(Z)$, we have

$$\int_{\mathcal{U}} u(p)\, \mu(du) = \langle \sigma_{\{p\}}, \mu \rangle = v \cdot (p - q) = \langle \sigma_{\{p\}}, \mu' \rangle = \int_{\mathcal{U}} u(p)\, \mu'(du).$$

If $\succsim$ additionally satisfies monotonicity, then $(\mathcal{M}_W, W^*)$ above is a monotone LHA representation for $\succsim$. Therefore, the sufficiency direction of part B also follows from the consistency of $\mathcal{M}_W$ established above.

References

Aliprantis, C., and K. Border (1999): *Infinite Dimensional Analysis*. Berlin, Germany: Springer-Verlag.

Billingsley, P. (1995): *Probability and Measure*. New York: John Wiley and Sons, Inc.

Dekel, E., B. Lipman, and A. Rustichini (2001): "Representing Preferences with a Unique Subjective State Space," *Econometrica*, 69, 891–934.

Dekel, E., B. Lipman, and A. Rustichini (2008): "Temptation Driven Preferences," *Review of Economic Studies*, forthcoming.

Dekel, E., B. Lipman, A. Rustichini, and T. Sarver (2007): "Representing Preferences with a Unique Subjective State Space: Corrigendum," *Econometrica*, 75, 591–600.

Ekeland, I., and T. Turnbull (1983): *Infinite-Dimensional Optimization and Convexity*. Chicago: The University of Chicago Press.

Epstein, L. G., M. Marinacci, and K. Seo (2007): "Coarse Contingencies and Ambiguity," *Theoretical Economics*, 2, 355–394.

Epstein, L. G. and K. Seo (2007), "Subjective States: A More Robust Model," Mimeo.

Ergin, H. (2003): "Costly Contemplation," Mimeo.

Ergin, H., and T. Sarver (2009): "A Unique Costly Contemplation Representation," Mimeo.

Fishburn, P. C. (1970): *Utility Theory for Decision Making*. New York: John Wiley & Sons.

Gilboa, I., and D. Schmeidler (1989): "Maxmin Expected Utility with Non-Unique Priors," *Journal of Mathematical Economics*, 18, 141–153.

Gul, F., and W. Pesendorfer (2001): "Temptation and Self-Control," *Econometrica*, 69, 1403–1435.

Holmes, R. (1975): *Geometric Functional Analysis and Its Applications*. New York: Springer-Verlag.

Hörmander, L. (1954): "Sur la Fonction d'Appui des Ensembles Convexes dans un Espace Localement Convexe," *Arkiv för Matematik*, 3, 181–186.

Kreps, D. (1979): "A Representation Theorem for Preference for Flexibility," *Econometrica*, 47, 565–578.

Kreps, D. (1988): *Notes on the Theory of Choice*. Colorado: Westview Press.

Kreps, D. and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74, 1447-1498.

Mazur, S. (1933): "Über konvexe Mengen in linearen normierten Räumen," *Studia Mathematica*, 4, 70–84.

Munkres, J. (2000): *Topology*. Upper Saddle River, NJ: Prentice Hall.

Phelps, R. R. (1993): *Convex Functions, Monotone Operators, and Differentiability*. Berlin, Germany: Springer-Verlag.

Rockafellar, R. T. (1970): *Convex Analysis*. Princeton, NJ: Princeton University Press.

Royden, H. L. (1988): *Real Analysis*. Englewood Cliffs, NJ: Prentice Hall.

Sarver, T. (2008), "Anticipating Regret: Why Fewer Options May Be Better," *Econometrica*, 76, 263–305.

Schneider, R. (1993): *Convex Bodies: The Brunn-Minkowski Theory*. Cambridge: Cambridge University Press.

Verona A. and M. E. Verona (1990): "Locally Efficient Monotone Operators," *Proceedings of the American Mathematical Society*, 109, 195–204.