Representing Preferences with a Unique Subjective State Space: Corrigendum: Supplementary Appendix

Eddie Dekel*
Barton L. Lipman†
Aldo Rustichini‡
Todd Sarver§

Current Draft
September 2006

In Section 1, we define our axioms and the additive EU representation, and we state the corrected version of Theorem 4.A of Dekel, Lipman, and Rustichini (2001, henceforth DLR (2001)). In Section 2, we provide a complete and almost entirely self–contained proof of this representation theorem.

1 Axioms and Additive EU Representation

Let $B = \{b_1, \ldots, b_K\}$ denote a set of pure outcomes. Let $\Delta(B)$ denote the set of probability distributions on B. Finally, let $\succ$ denote a preference relation on the set of nonempty subsets of $\Delta(B)$ where this space is endowed with the Hausdorff topology. Let $d_h(x, y)$ denote the Hausdorff distance between x and y.[1]

*Economics Dept., Northwestern University, and School of Economics, Tel Aviv University. E–mail: dekel@northwestern.edu.

†Boston University. E–mail: blipman@bu.edu.

‡University of Minnesota. E–mail: arust@econ.umn.edu.

§Northwestern University. E–mail: tsarver@northwestern.edu

[1]If we let d denote the Euclidean metric on ΔB, then the Hausdorff distance is defined by

$$d_h(x, y) = \max\Big\{ \sup_{\beta\in x} \inf_{\beta'\in y} d(\beta, \beta'), \sup_{\beta\in y} \inf_{\beta'\in x} d(\beta, \beta') \Big\}.$$

DLR (2001) consider the following axioms on $\succ$.

Axiom 1 (Weak Order) $\succ$ *is asymmetric and negatively transitive.*

Axiom 2 (Continuity) *For any x, the strict upper and lower contour sets $\{y \subseteq \Delta(B) \mid y \succ x\}$ and $\{y \subseteq \Delta(B) \mid y \prec x\}$ are open in the Hausdorff topology.*

For any sets x and y and any $\lambda \in [0,1]$, define

$$\lambda x + (1-\lambda)y = \{\lambda\beta + (1-\lambda)\beta' \mid \beta \in x \text{ and } \beta' \in y\}.$$

Axiom 3 (Independence) *If $x \succ y$, then for all z and all $\lambda \in (0,1]$,*

$$\lambda x + (1-\lambda)z \succ \lambda y + (1-\lambda)z.$$

Axiom 4 (Monotonicity) *If $x \subseteq y$, then $y \succeq x$.*

See DLR (2001) for discussion of these axioms.

In the main text of corrigendum, we consider the following additional continuity axiom.

Axiom 5 (L–Continuity) *There exist menus x^* and x_* and $N > 0$ such that for every $\varepsilon \in (0, 1/N)$, for every x and y with $d_h(x,y) \leq \varepsilon$,*

$$(1-N\varepsilon)x + N\varepsilon x^* \succeq (1-N\varepsilon)y + N\varepsilon x_*.$$

See the corrigendum for discussion of this axiom. The following version of continuity is standard in the literature — see, for example, Fishburn (1970) or Kreps (1988).

Axiom 6 (vNM Continuity) *If $x \succ y \succ z$, then there exists $\lambda, \bar{\lambda} \in (0,1)$ such that $\lambda x + (1-\lambda)z \succ y \succ \bar{\lambda}x + (1-\bar{\lambda})z$.*

We say that a function $u : \Delta(B) \to \mathbf{R}$ is an expected–utility function if $u(\beta) = \sum_{b\in B} u(b)\beta(b)$, where we abuse notation by letting b also denote the degenerate distribution with probability 1 on b. Obviously, such a function is completely defined by specifying the vector in $\mathbf{R}^K$ giving the utility of pure outcomes. Hence we also refer to such vectors as expected–utility functions.

Definition 1 *An* additive EU representation *is a measure space S, a measurable state–dependent utility function $U : \Delta(B) \times S \to \mathbf{R}$, and a (countably additive) signed measure μ on S such that (i) $V : 2^{\Delta(B)} \setminus \{\emptyset\} \to \mathbf{R}$ defined by*

$$V(x) = \int_S \sup_{\beta \in x} U(\beta, s)\, \mu(ds) \tag{1}$$

represents $\succ$ and (ii) $U(\cdot, s)$ is an expected–utility function for each $s \in S$.

There are several differences between the definition of an additive EU representation given here and the original definition given in DLR (2001). We now discuss those differences and their implications.

1. The definition in DLR (2001) only requires the measure to be finitely additive. Our necessity arguments do not rely on countable additivity and our sufficiency proof establishes that the measure is countably additive. Hence a representation with a finitely additive measure exists if and only if a representation with a countably additive measure exists.

2. DLR (2001) state explicitly that V is continuous. We show in Lemma 4 that one implication of our definition is the stronger condition that V is Lipschitz continuous.

3. DLR (2001) require that every state $s \in S$ be "relevant," which, loosely speaking, implies that none of the states could be dropped from the representation without altering the underlying preference.[2] Our definition allows for the possibility that some states are not relevant. It is not hard to show that if we choose the state space to be the support of μ, then the relevancy requirement is satisfied.

4. DLR (2001) require that S be nonempty and use a nontriviality axiom to ensure this is possible. Our definition allows for a measure which is identically zero, in which case the support, the set of "relevant" states, is empty. We find this more convenient, but it is easy to show that adding DLR's nontriviality axiom would ensure a nonempty support.

5. DLR (2001) require that if $s, s' \in S$, $s \neq s'$, then $U(\cdot, s)$ and $U(\cdot, s')$ do not represent the same expected–utility preference. In this sense, there are no "redundant" states. It will be obvious from our proof of the representation theorem that this nonredundancy condition could be imposed without affecting the results. We omit the requirement for simplicity.

The following is the corrected statement of Theorem 4.A in DLR (2001).

[2] The interested reader should consult DLR (2001) for a precise definition of a relevant state.

Theorem 1 *The preference $\succ$ has an additive EU representation if and only if it satisfies weak order, vNM continuity, L–continuity, and independence. Furthermore, $\succ$ also satisfies monotonicity if and only if the measure μ is positive.*

This theorem differs from the result claimed by DLR only in the continuity requirements, replacing their continuity axiom with vNM continuity and L–continuity. An implication of the representation theorem is that the assumptions of Theorem 1 imply continuity since the representation constructed is continuous.

The proof of Theorem 1 is contained in Section 2. We now note an interesting relationship between the axioms.

Lemma 1 *If $\succ$ satisfies monotonicity, then it satisfies L–continuity.*

Proof. Let $x^* = \Delta(B)$, $x_* = \{(1/K, \ldots, 1/K)\}$, and $N = K$ where K is the number of pure outcomes. Take any $\varepsilon \in (0, 1/N)$ and $x, y \subseteq \Delta(B)$ such that $d_h(x, y) \leq \varepsilon$. We will show that

$$(1 - N\varepsilon)y + N\varepsilon x_* \subseteq (1 - N\varepsilon)x + N\varepsilon x^*,$$

which, given monotonicity, will yield the desired result:

$$(1 - N\varepsilon)x + N\varepsilon x^* \succeq (1 - N\varepsilon)y + N\varepsilon x_*.$$

To shorten notation, let $\beta^* = (1/K, \ldots, 1/K)$. Take any $\beta \in (1 - N\varepsilon)y + N\varepsilon x_*$, so $\beta = (1 - N\varepsilon)\beta_y + N\varepsilon\beta^*$ for some $\beta_y \in y$. It is clear from the definition of the Hausdorff distance that $d_h(x, y) \leq \varepsilon$ implies $\inf_{\beta' \in x} \|\beta_y - \beta'\|_E \leq \varepsilon$ where $\|\cdot\|_E$ denotes the Euclidean norm. Since $\varepsilon \in (0, 1/N)$, we have $\varepsilon/(1 - N\varepsilon) > \varepsilon$, so there exists $\beta_x \in x$ such that $\|\beta_y - \beta_x\|_E \leq \varepsilon/(1 - N\varepsilon)$.

Define $\hat{\beta} \in \mathbf{R}^K$ as follows:

$$\hat{\beta} = \beta^* + \frac{1 - N\varepsilon}{N\varepsilon}(\beta_y - \beta_x)$$

We claim that $\hat{\beta} \in \Delta(B)$. It is obvious that $\sum_i \hat{\beta}(b_i) = 1$, so we need only verify that $\hat{\beta}(b_i) \geq 0$ for all i. Since $N = K$, we see that for all i,

$$|\hat{\beta}(b_i) - \beta^*(b_i)| \leq \|\hat{\beta} - \beta^*\|_E = \frac{1 - K\varepsilon}{K\varepsilon}\|\beta_y - \beta_x\|_E \leq \frac{1 - K\varepsilon}{K\varepsilon}\;\frac{\varepsilon}{1 - K\varepsilon} = \frac{1}{K}.$$

For all i, $\beta^*(b_i) = 1/K$, so we have $|\hat{\beta}(b_i) - 1/K| \leq 1/K$, which implies $\hat{\beta}(b_i) \geq 0$. Thus $\hat{\beta} \in \Delta(B)$. Finally,

$$\begin{aligned}(1 - N\varepsilon)\beta_x + N\varepsilon\hat{\beta} &= (1 - N\varepsilon)\beta_x + N\varepsilon\beta^* + (1 - N\varepsilon)(\beta_y - \beta_x)\\ &= (1 - N\varepsilon)\beta_y + N\varepsilon\beta^* = \beta,\end{aligned}$$

so $\beta \in (1-N\varepsilon)x + N\varepsilon x^*$. Since $\beta \in (1-N\varepsilon)y + N\varepsilon x_*$ was arbitrary, this completes the proof. ■

In light of Lemma 1, we have the following corollary to Theorem 1.

Theorem 2 *The preference $\succ$ has an additive EU representation with a positive measure μ if and only if it satisfies weak order, vNM continuity, independence, and monotonicity.*

2 Proof of Theorem 1

2.1 Preliminaries: Support Functions

Define X to be the set of all nonempty, closed, and convex subsets of $\Delta(B)$. Then, for all nonempty $x \subseteq \Delta(B)$, we have $\text{conv}(\text{cl}(x)) \in X$, where $\text{cl}(x)$ denotes the closure of x (in the Euclidean topology on $\Delta(B)$) and $\text{conv}(x)$ denotes the convex hull of x. It will sometimes be useful to work with the set X instead of the set of all menus because of a natural relationship that exists between the set of closed and convex sets and a certain class of continuous functions known as the support functions. In this section, we formally define the support functions and discuss some of their properties.

First, let $S^K = \{s \in \mathbf{R}^K : \sum_i s_i = 0 \text{ and } \sum_i s_i^2 = 1\}$ be the set of *normalized (non–constant) expected–utility functions* on $\Delta(B)$. For any $x \in X$, the *support function* $\sigma_x : S^K \to \mathbf{R}$ of x is defined by $\sigma_x(s) = \max_{\beta \in x} \beta \cdot s$. For a more complete introduction to support functions, see Rockafellar (1970) or Schneider (1993). Let $C(S^K)$ denote the set of continuous real–valued functions on S^K. When endowed with the supremum norm $\|\cdot\|$, $C(S^K)$ is a Banach space. Define an order $\geq$ on $C(S^K)$ by $f \geq g$ if $f(s) \geq g(s)$ for all $s \in S^K$. Let $C = \{\sigma_x \in C(S^K) : x \in X\}$. For any $\sigma \in C$, let

$$x_\sigma = \bigcap_{s \in S^K} \left\{ \beta \in \Delta(B) \middle| \beta \cdot s = \sum_i \beta(b_i) s_i \leq \sigma(s) \right\}.$$

The following properties of the support functions will be useful.

Lemma 2 *1. For all $x \in X$ and $\sigma \in C$, $x_{(\sigma_x)} = x$ and $\sigma_{(x_\sigma)} = \sigma$. Hence the mapping $x \mapsto \sigma_x$ is a bijection from X to C.*

2. For all $x, y \in X$ and $\lambda \in (0,1)$, $\sigma_{\lambda x + (1-\lambda) y} = \lambda \sigma_x + (1-\lambda)\sigma_y$.

3. *For all* $x, y \in X$, $d_h(x, y) = \|\sigma_x - \sigma_y\|$.

4. *For all* $x, y \in X$, $x \subseteq y \iff \sigma_x \leq \sigma_y$.

Proof. These are standard results that can be found in Rockafellar (1970) or Schneider (1993).[3] For instance, in Schneider (1993), part 1 follows from Theorem 1.7.1, part 2 follows from Theorem 1.7.5, part 3 follows from Theorem 1.8.11, and part 4 can be found on page 37. ■

We also use the following properties of C.

Lemma 3 *C is convex and $\sigma_{\{(1/K,\ldots,1/K)\}} = \mathbf{0} \in C$ where $\mathbf{0}$ denotes the zero function.*

Proof. The convexity of C follows from part 2 of Lemma 2 and the convexity of X. For any $s \in S^K$, we have $\sum_i s_i = 0$ and hence $\sigma_{\{(1/K,\ldots,1/K)\}}(s) = \sum_i (1/K) s_i = 0$. ■

2.2 Necessity of the Axioms

It is easily verified that a preference $\succ$ with an additive EU representation must satisfy weak order, vNM continuity, and independence. In addition, if the measure μ is positive, then it is easily verified that $\succ$ satisfies monotonicity.

It is also easy to see that if $\succ$ has an additive EU representation, then it must satisfy *indifference to closure (IC)* and *indifference to randomization (IR)* in the sense that for any nonempty $x \subseteq \Delta(B)$, $x \sim \mathrm{cl}(x)$ (IC) and $x \sim \mathrm{conv}(x)$ (IR).[4] To see this, simply note that for any expected–utility function $u : \Delta(B) \to \mathbf{R}$,

$$\sup_{\beta \in x} u(\beta) = \max_{\beta \in \mathrm{cl}(x)} u(\beta) = \sup_{\beta \in \mathrm{conv}(x)} u(\beta).$$

Hence the function V defined in Equation (1) must satisfy $V(x) = V(\mathrm{cl}(x)) = V(\mathrm{conv}(x))$.

Finally, it is easy to see that if $\succ$ satisfies IC and IR and satisfies L–continuity on X, then it satisfies L–continuity on all menus. That is, suppose $\succ$ satisfies L–continuity on

[3] The standard setting for support functions is the set of nonempty, closed, and convex subsets of $\mathbf{R}^n$. However, by imposing our normalizations on the domain of the support functions S^K, the standard results are easily adapted to our setting of nonempty, closed, and convex subsets of $\Delta(B)$.

[4] We show in Lemma 6 that IC and IR are implied by our axioms, so we do not need to add them as separate axioms.

X in the sense that there exists N, x^* and x_* such that for all $\varepsilon \in (0, 1/N)$, for every $x, y \in X$ with $d_h(x, y) \leq \varepsilon$,

$$(1 - N\varepsilon)x + N\varepsilon x^* \succeq (1 - N\varepsilon)y + N\varepsilon x_*.$$

Fix any menus x and y not necessarily in X. Let $\hat{x} = \text{conv}(\text{cl}(x))$ and $\hat{y} = \text{conv}(\text{cl}(y))$. It is not hard to show[5] that $d_h(\hat{x}, \hat{y}) \leq d_h(x, y)$. Hence IC, IR, and independence implies the conclusion of L–continuity for these menus.

In light of this, we can show that L–continuity is necessary by showing that L–continuity on X is necessary. To show the latter, we first prove that the V defined in Equation (1) is Lipschitz continuous.

Definition 2 $V : X \to \mathbf{R}$ *is* Lipschitz continuous *if there exists* $\bar{N} \geq 0$ *such that*

$$V(y) - V(x) \leq \bar{N} d_h(x, y), \quad \forall x, y \in X.$$

Lemma 4 *For any additive EU representation, the V defined by Equation* (1)*, restricted to X, is Lipschitz continuous.*

Proof. Define S^K and σ_x as in Section 2.1. Take an additive EU representation (S, U, μ) and define V as in Equation (1). Since each $U(\cdot, s)$ is an expected–utility function, there exist $\mathbf{s} : S \to S^K$, $f : S \to \mathbf{R}_+$, and $g : S \to \mathbf{R}$ such $U(\beta, s) = (\beta \cdot \mathbf{s}(s))f(s) + g(s)$ for all $\beta \in \Delta(B)$, $s \in S$. Note that for any $x \in X$ and $s \in S^K$,

$$\sup_{\beta \in x} \beta \cdot s = \max_{\beta \in x} \beta \cdot s = \sigma_x(s).$$

Hence, $V(x) = \int_S [(\sigma_x \circ \mathbf{s})f + g]\, \mu(ds)$. We can write μ as $\mu^+ - \mu^-$ where both of these measures are positive. Let $\bar{N} = \int_S f\, \mu^+(ds) + \int_S f\, \mu^-(ds)$. Note that $\bar{N}$ is finite.[6] Take

[5]It follows from the definition of the Hausdorff distance that $d_h(\text{cl}(x), \text{cl}(y)) = d_h(x, y)$. We see that $d_h(\text{conv}(\text{cl}(x)), \text{conv}(\text{cl}(y))) \leq d_h(\text{cl}(x), \text{cl}(y))$ by noting the following two inequalities which we leave to the reader to verify:

$$\sup_{\beta \in \text{cl}(x)} \inf_{\beta' \in \text{cl}(y)} d(\beta, \beta') \geq \sup_{\beta \in \text{conv}(\text{cl}(x))} \inf_{\beta' \in \text{conv}(\text{cl}(y))} d(\beta, \beta'),$$
$$\sup_{\beta \in \text{cl}(y)} \inf_{\beta' \in \text{cl}(x)} d(\beta, \beta') \geq \sup_{\beta \in \text{conv}(\text{cl}(y))} \inf_{\beta' \in \text{conv}(\text{cl}(x))} d(\beta, \beta').$$

[6]By Lemma 3 in Sarver (2005), there exist $x, y \in X$ such that $\sigma_x(s) = 0$ and $\sigma_y(s) = c > 0$ for all $s \in S^K$. Then

$$V(y) - V(x) = c \int_S f\, \mu(ds) = c \left[\int_S f\, \mu^+(ds) - \int_S f\, \mu^-(ds) \right].$$

Since V is real–valued, $V(y) - V(x)$ must be real–valued, so $\int_{S^K} f\, \mu^+(ds)$ and $\int_{S^K} f\, \mu^-(ds)$ are finite.

arbitrary $x, y \in X$. Then,

$$\begin{aligned}
V(y) - V(x) &\le \left| \int_S [(\sigma_y \circ \mathbf{s} - \sigma_x \circ \mathbf{s}) f] \, \mu(ds) \right| \\
&= \left| \int_S [(\sigma_y \circ \mathbf{s} - \sigma_x \circ \mathbf{s}) f] \, \mu^+(ds) - \int_S [(\sigma_y \circ \mathbf{s} - \sigma_x \circ \mathbf{s}) f] \, \mu^-(ds) \right| \\
&\le \left| \int_S [(\sigma_{\hat{y}} \circ \mathbf{s} - \sigma_x \circ \mathbf{s}) f] \, \mu^+(ds) \right| + \left| \int_S [(\sigma_y \circ \mathbf{s} - \sigma_x \circ \mathbf{s}) f] \, \mu^-(ds) \right| \\
&\le \int_S \|\sigma_y - \sigma_x\| f \, \mu^+(ds) + \int_S \|\sigma_y - \sigma_x\| f \, \mu^-(ds) \\
&= \bar{N} \|\sigma_x - \sigma_x\| = \bar{N} d_h(x, y),
\end{aligned}$$

where the last equality follow from Lemma 2. Hence V is Lipschitz continuous. ■

Since V is affine, the following lemma establishes the L–continuity of $\succ$ on X and hence L–continuity.

Lemma 5 *If $\succ$ has a representation V which is affine and Lipschitz continuous on X, then $\succ$ satisfies L–continuity on X.*

Proof. Suppose V is an affine representation of $\succ$ which is Lipschitz continuous on X. L–continuity of $\succ$ on X follows trivially if V is constant, so suppose there exist x^*, x_* such that $V(x^*) > V(x_*)$. Since V is Lipschitz continuous, there exists $\bar{N}$ such that $V(y) - V(x) \le \bar{N} d_h(x, y)$ for all $x, y \in X$. Let $N = \bar{N}/[V(x^*) - V(x_*)]$. So for all x and y in X, we have

$$V(y) - V(x) \le N[V(x^*) - V(x_*)] d_h(x, y).$$

So for all x and y with $d_h(x, y) < 1/N$,

$$V(y) - V(x) \le \frac{N d_h(x, y)}{1 - N d_h(x, y)} [V(x^*) - V(x_*)].$$

So for every $\varepsilon \in [d_h(x, y), 1/N)$,

$$V(y) - V(x) \le \frac{N\varepsilon}{1 - N\varepsilon} [V(x^*) - V(x_*)],$$

or equivalently,

$$(1 - N\varepsilon) V(y) + N\varepsilon V(x_*) \le (1 - N\varepsilon) V(x) + N\varepsilon V(x^*).$$

Since V is affine and represents $\succ$, we see that

$$(1 - N\varepsilon) x + N\varepsilon x^* \succeq (1 - N\varepsilon) y + N\varepsilon x_*.$$

Thus $\succ$ is L–continuous on X. ■

As noted earlier, given IC and IR, if $\succ$ is L–continuous on X, it is L–continuous. Hence we have established necessity.

2.3 Sufficiency of the Axioms

In this section, we establish the sufficiency of the axioms. We first note that our axioms imply indifference to closure and indifference to randomization.

Lemma 6 *If $\succ$ is asymmetric and satisfies independence, then for all $x \subseteq \Delta(B)$, $x \sim \text{cl}(x)$ and $x \sim \text{conv}(x)$.*

DLR (2001) uses continuity to derive these properties, so their proofs do not suffice for our purposes.

Proof. Fix any $x \subseteq \Delta(B)$. First, suppose $x \not\sim \text{cl}(x)$. Then independence implies that for every y and every $\lambda \in (0,1]$, $\lambda x + (1-\lambda)y \not\sim \lambda\text{cl}(x) + (1-\lambda)y$. We contradict this by constructing a set y such that

$$\lambda x + (1-\lambda)y = \lambda\text{cl}(x) + (1-\lambda)y, \ \forall\lambda \in (0,1). \tag{2}$$

By asymmetry of $\succ$, equation (2) establishes the needed contradiction.

To construct y, let $\beta^* = (1/K, \ldots, 1/K)$ and let

$$y = \left\{ \beta \in \mathbf{R}^K \,|\, \sum_{i=1}^{K} \beta(b_i) = 1 \text{ and } \|\beta - \beta^*\|_E < 1/K \right\}.$$

For any $\beta \in y$ and any i, $|\beta(b_i) - \beta^*(b_i)| \leq \|\beta - \beta^*\|_E < 1/K$, so $\beta(b_i) > 0$ for all i. Thus $y \subset \Delta(B)$. Take any $\lambda \in (0,1)$. Clearly $\lambda x + (1-\lambda)y \subseteq \lambda\text{cl}(x) + (1-\lambda)y$. To show the opposite inclusion, take any $\beta \in \lambda\text{cl}(x) + (1-\lambda)y$, so $\beta = \lambda\beta_{\bar{x}} + (1-\lambda)\beta_y$ for some $\beta_{\bar{x}} \in \text{cl}(x)$ and $\beta_y \in y$. Let $\varepsilon = 1/K - \|\beta_y - \beta^*\|_E > 0$. Since $\beta_{\bar{x}} \in \text{cl}(x)$, there exists $\beta_x \in x$ such that $\|\beta_{\bar{x}} - \beta_x\|_E < \frac{1-\lambda}{\lambda}\varepsilon$. Let $\hat{\beta} = \beta_y + \frac{\lambda}{1-\lambda}(\beta_{\bar{x}} - \beta_x)$. Then

$$\|\hat{\beta} - \beta^*\|_E \leq \|\beta_y - \beta^*\|_E + \tfrac{\lambda}{1-\lambda}\|\beta_{\bar{x}} - \beta_x\|_E < \|\beta_y - \beta^*\|_E + \varepsilon = 1/K,$$

so $\hat{\beta} \in y$. Therefore,

$$\lambda\beta_x + (1-\lambda)\hat{\beta} = \lambda\beta_x + (1-\lambda)\beta_y + \lambda(\beta_{\bar{x}} - \beta_x) = \beta,$$

so $\beta \in \lambda x + (1-\lambda)y$. Thus $\lambda\text{cl}(x) + (1-\lambda)y \subseteq \lambda x + (1-\lambda)y$, which implies $\lambda x + (1-\lambda)y = \lambda\text{cl}(x) + (1-\lambda)y$.

Second, suppose $x \not\sim \text{conv}(x)$. Then independence implies that for every $\lambda \in (0,1]$, $\lambda x + (1-\lambda)\text{conv}(x) \not\sim \lambda\text{conv}(x) + (1-\lambda)\text{conv}(x) = \text{conv}(x)$. We contradict this by showing that

$$\lambda x + (1-\lambda)\text{conv}(x) = \text{conv}(x), \ \forall\lambda \in [0, 1/K]. \tag{3}$$

As in the first part of the proof, asymmetry of $\succ$ and equation (3) yield a contradiction.

To show that (3) holds, note that $\lambda x + (1-\lambda)\text{conv}(x) \subseteq \text{conv}(x)$. To show the converse, fix any $\beta \in \text{conv}(x)$. Since x can be viewed as a subset of $\mathbf{R}^{K-1}$, Carathéodory's theorem (see, *e.g.*, Theorem 1.1.4 in Schneider (1993)) implies that β is a convex combination of at most K points in x.

In light of this, fix any $\lambda \in [0, 1/K]$ and write β as $\sum_{i=1}^{K} t_i \beta_i = \beta$ where $\beta_i \in x$, $t_i \geq 0$, and $\sum_{i=1}^{K} t_i = 1$. Clearly, there must be some j such that $t_j \geq 1/K$. Define $\hat{t}_i$ for $i = 1, \ldots, K$ by $\hat{t}_j = (t_j - \lambda)/(1-\lambda)$ and for $i \neq j$, $\hat{t}_i = t_i/(1-\lambda)$. Obviously, $\hat{t}_i \geq 0$ for all $i \neq j$. Also, $t_j \geq 1/K \geq \lambda$ implies $\hat{t}_j \geq 0$. Finally, it is easy to show that $\sum_i \hat{t}_i = 1$. Hence $\hat{\beta} \equiv \sum_i \hat{t}_i \beta_i \in \text{conv}(x)$.

$$\lambda \beta_j + (1-\lambda)\hat{\beta} \in \lambda x + (1-\lambda)\text{conv}(x).$$

But it is easy to see that $\lambda \beta_j + (1-\lambda)\hat{\beta} = \beta$. Hence $\text{conv}(x) \subseteq \lambda x + (1-\lambda)\text{conv}(x)$, implying (3). ■

Note that for any $u \in \mathbf{R}^K$ (i.e., any expected–utility function on $\Delta(B)$) and any $x \subseteq \Delta(B)$,

$$\sup_{\beta \in x} \beta \cdot u = \max_{\beta \in \text{conv}(\text{cl}(x))} \beta \cdot u.$$

Thus if we establish an additive EU representation for $\succ$ on X (the set of all nonempty, closed, and convex subsets of $\Delta(B)$) and apply the same functional form for all $x \subseteq \Delta(B)$, then by Lemma 6 the resulting function represents $\succ$ on the set of all menus. The remainder of this section is devoted to establishing an additive EU representation for $\succ$ on X.

Lemma 7 *If $\succ$ satisfies weak order, vNM continuity, and independence, then there exists an affine $V : X \to \mathbf{R}$ that represents $\succ$ on X. Furthermore, V is unique up to an affine transformation.*

Proof. This result follows from the mixture space theorem. For instance, see Fishburn (1970, Theorem 8.4, page 112) or Kreps (1988, Theorem 5.11, page 54). ■

In the preceding lemma, the restriction to X is needed for the mixture space axioms because $\lambda[\lambda' x + (1-\lambda')y] + (1-\lambda)y$ might not equal $\lambda\lambda' x + (1-\lambda\lambda')y$ if x and y are not convex. If x and y are not convex, the second set may be strictly smaller than the first.

The following lemma combines L–continuity with the assumptions of Lemma 7 to obtain a Lipschitz continuous representation.

Lemma 8 *Assume $\succ$ has an affine representation $V : X \to \mathbf{R}$. Then V is Lipschitz continuous if $\succ$ satisfies L–continuity.*

Proof. Suppose $\succ$ satisfies L–continuity. Fix the x^*, x_*, and N of the axiom, any $D \in (0, 1/N)$, and any x and y with $d_h(x,y) \leq D$. Let $\delta = d_h(x,y)$. If $\delta = 0$, then $x, y \in X$ implies $x = y$ in which case the conclusion is obvious. So suppose $\delta > 0$. Then

$$(1 - N\delta)x + N\delta x^* \succeq (1 - N\delta)y + N\delta x_*.$$

Using the affine representation, this implies[7]

$$V(y) - V(x) \leq \frac{N}{1 - N\delta}[V(x^*) - V(x_*)]d_h(x,y).$$

Since $N\delta = Nd_h(x,y) \leq ND < 1$, we have $N/(1 - N\delta) \leq N/(1 - ND) < \infty$. Let $\bar{N} = [N/(1 - ND)][V(x^*) - V(x_*)]$. Then for any x and y with $d_h(x,y) \leq D$, we have

$$V(y) - V(x) \leq \bar{N} d_h(x,y).$$

To complete the proof, we show the same for arbitrary x and y. Fix any x and y and any sequence $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_M < \lambda_{M+1} = 1$ such that $(\lambda_{m+1} - \lambda_m)d_h(x,y) \leq D$. Let $x_m = \lambda_m x + (1 - \lambda_m)y$. Then

$$\begin{aligned} d_h(x_{m+1}, x_m) &= \|\sigma_{x_{m+1}} - \sigma_{x_m}\| \\ &= (\lambda_{m+1} - \lambda_m)\|\sigma_x - \sigma_y\| \\ &= (\lambda_{m+1} - \lambda_m)d_h(x,y). \end{aligned}$$

Hence from the previous part, we see that

$$V(x_{m+1}) - V(x_m) \leq \bar{N}(\lambda_{m+1} - \lambda_m)d_h(x,y).$$

Summing both sides over m from $m = 0$ to $m = M$ gives $V(y) - V(x) \leq \bar{N}d_h(x,y)$, so V is Lipschitz continuous. ■

In light of Lemmas 7 and 8, there exists a Lipschitz continuous and affine function V that represents $\succ$ on X. It is obvious that if $\succ$ also satisfies monotonicity, then V is monotone in the sense that $x \subseteq y$ implies $V(x) \leq V(y)$. Since V is unique up to an affine transformation, we can normalize V so that $V(\{(1/K, \ldots, 1/K)\}) = 0$. Define the function $W : C \to \mathbf{R}$ by $W(\sigma) = V(x_\sigma)$. Then, by part 1 of Lemma 2, $V(x) = W(\sigma_x)$ for all $x \in X$. We say the function W is monotone if for all $\sigma, \sigma' \in C$, $\sigma \leq \sigma'$ implies $W(\sigma) \leq W(\sigma')$.

[7] We have made the implicit assumption that $x^*, x_* \in X$. This assumption is without loss of generality since by Lemma 6 we can replace these sets with their closed convex hulls.

Lemma 9 *W is Lipschitz continuous and linear (i.e., W is affine and $W(\mathbf{0}) = 0$). If V is monotone, then W is monotone.*

Proof. To see that W is affine, let $x, y \in X$ and $\lambda \in (0,1)$. Then, by parts 1 and 2 of Lemma 2 and the affinity of V,

$$\begin{aligned} W(\lambda\sigma_x + (1-\lambda)\sigma_y) &= W(\sigma_{\lambda x+(1-\lambda)y}) = V(\lambda x + (1-\lambda)y) \\ &= \lambda V(x) + (1-\lambda)V(y) = \lambda W(\sigma_x) + (1-\lambda)W(\sigma_y). \end{aligned}$$

By Lemma 3 and the chosen normalization of V, we see that

$$W(\mathbf{0}) = W(\sigma_{\{(1/K,\ldots,1/K)\}}) = V(\{(1/K,\ldots,1/K)\}) = 0.$$

The Lipschitz continuity of W follows from the Lipschitz continuity of V and parts 1 and 3 of Lemma 2. By part 4 of Lemma 2, W inherits monotonicity from V. ■

We proceed by showing that $W : C \to \mathbf{R}$ has a unique continuous linear extension to $C(S^K)$. First, define H and H^* as follows:[8]

$$\begin{aligned} H &= \bigcup_{r\geq 0} rC = \{r\sigma \in C(S^K) | r \geq 0 \text{ and } \sigma \in C\} \\ H^* &= H - H = \{f \in C(S^K) | f = f_1 - f_2 \text{ for some } f_1, f_2 \in H\} \end{aligned}$$

Therefore, if $f \in H^*$, then there exist $\sigma^1, \sigma^2 \in C$ and $r_1, r_2 \geq 0$ such that $f = r_1\sigma^1 - r_2\sigma^2$. We note some relevant properties of H^*.

Lemma 10 *1. H^* is a linear subspace of $C(S^K)$.*

2. For any $f \in H^*$, there exist $\sigma^1, \sigma^2 \in C$ and $r > 0$ such that $f = r(\sigma^1 - \sigma^2)$.

3. H^* is dense in $C(S^K)$.

Proof. *(1):* It is obvious that $f \in H^*$ implies $rf \in H^*$ for any scalar r. Let $f, g \in H^*$. Then, we can write $f = r_1\sigma^1 - r_2\sigma^2$ and $g = \hat{r}_1\hat{\sigma}^1 - \hat{r}_2\hat{\sigma}^2$ where $r_1, r_2, \hat{r}_1, \hat{r}_2 \geq 0$ and $\sigma^1, \sigma^2, \hat{\sigma}^1, \hat{\sigma}^2 \in C$. Define $\bar{\sigma}^1$ and $\bar{\sigma}^2$ as follows:

$$\bar{\sigma}^1 = \begin{cases} \frac{r_1}{r_1+\hat{r}_1}\sigma^1 + \frac{\hat{r}_1}{r_1+\hat{r}_1}\hat{\sigma}^1 & \text{if } r_1 + \hat{r}_1 \neq 0 \\ 0 & \text{otherwise} \end{cases} \qquad \bar{\sigma}^2 = \begin{cases} \frac{r_2}{r_2+\hat{r}_2}\sigma^2 + \frac{\hat{r}_2}{r_2+\hat{r}_2}\hat{\sigma}^2 & \text{if } r_2 + \hat{r}_2 \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

[8]DLR (2001) define $H = \bigcup_{r\geq 0} rC_+$ where $C_+ = \{\sigma \in C | \sigma \geq 0\}$. However, it can be verified that the resulting H^* is the same under either definition of H.

Since C is convex and $\mathbf{0} \in C$ by Lemma 3, we see that $\bar{\sigma}^1, \bar{\sigma}^2 \in C$. Hence $f + g = (r_1 + \hat{r}_1)\bar{\sigma}^1 - (r_2 + \hat{r}_2)\bar{\sigma}^2 \in H^*$, so H^* is a linear subspace.

(2): Let $f \in H^*$, so there exist $\sigma^1, \sigma^2 \in C$ and $r_1, r_2 \geq 0$ such that $f = r_1\sigma^1 - r_2\sigma^2$. Let $r = \max\{r_1, r_2\}$. If $r = 0$, then $f = r'(\mathbf{0} - \mathbf{0})$ for any $r' > 0$, establishing the desired result. Therefore, suppose $r > 0$. Since $\mathbf{0} \in C$ and C is convex, we have $\hat{\sigma}^i \equiv (r_i/r)\sigma^i \in C$. Then, $f = r(\hat{\sigma}^1 - \hat{\sigma}^2)$.

(3): Although stated slightly differently, a classic proof of this result can be found in Hörmander (1954). A complete proof for the current setting is contained in DLR (2001, Lemma 11). ∎

We use the above properties of H^* to establish the following result.

Lemma 11 *Any Lipschitz continuous linear functional $W : C \to \mathbf{R}$ has a unique continuous linear extension to $C(S^K)$. If W is monotone, then this extension is a positive linear functional.*

Proof. First, extend W to H^* by linearity. Specifically, if $f \in H^*$, then by part 2 of Lemma 10 there exist $\sigma^1, \sigma^2 \in C$ and $r > 0$ such that $f = r(\sigma^1 - \sigma^2)$. Therefore, define $\hat{W} : H^* \to \mathbf{R}$ by $\hat{W}(f) = r[W(\sigma^1) - W(\sigma^2)]$. We verify that $\hat{W}$ is uniquely defined. For suppose

$$f = r(\sigma^1 - \sigma^2) = \hat{r}(\hat{\sigma}^1 - \hat{\sigma}^2).$$

Let $\bar{r} = r + \hat{r}$. The claim obviously holds if $\hat{r} = 0$ so assume $\hat{r} > 0$. Then we have

$$\frac{r}{\bar{r}}\sigma^1 + \frac{\hat{r}}{\bar{r}}\hat{\sigma}^2 = \frac{\hat{r}}{\bar{r}}\hat{\sigma}^1 + \frac{r}{\bar{r}}\sigma^2,$$

which is an element of C since C is convex by Lemma 3. Since W is affine, we see that

$$\frac{r}{\bar{r}}W(\sigma^1) + \frac{\hat{r}}{\bar{r}}W(\hat{\sigma}^2) = \frac{\hat{r}}{\bar{r}}W(\hat{\sigma}^1) + \frac{r}{\bar{r}}W(\sigma^2),$$

or equivalently,

$$rW(\sigma^1) - rW(\sigma^2) = \hat{r}W(\hat{\sigma}^1) - \hat{r}W(\hat{\sigma}^2).$$

It is easily verified that $\hat{W}$ is linear. Also, since $W(\mathbf{0}) = 0$, we have $\hat{W}|_C = W$. Thus $\hat{W}$ is the unique linear extension of W to H^*.

By part 1 of Lemma 10, H^* is a linear subspace of $C(S^K)$. We now prove that $\hat{W}$ is a bounded linear functional on H^*. By the Lipschitz continuity of W on C, there exists $\bar{N}$ such that $W(\sigma^1) - W(\sigma^2) \leq \bar{N}\|\sigma^1 - \sigma^2\|$ for all $\sigma^1, \sigma^2 \in C$. Let $f \in H^*$. By part 2 of Lemma 10, there exist $\sigma^1, \sigma^2 \in C$ and $r > 0$ such that $f = r(\sigma^1 - \sigma^2)$. Therefore,

$$|\hat{W}(f)| = r|W(\sigma^1) - W(\sigma^2)| \leq \bar{N}r\|\sigma^1 - \sigma^2\| = \bar{N}\|f\|,$$

so $\hat{W}$ is bounded on H^*. Therefore, we can apply the Hahn–Banach theorem (see Royden (1988, Theorem 4, page 223)) to conclude that $\hat{W}$ has an extension to a continuous linear functional $\bar{W} : C(S^K) \to \mathbf{R}$. Since H^* is dense in $C(S^K)$ by part 3 of Lemma 10, it is easily verified that $\bar{W}$ is the unique continuous extension of $\hat{W}$ to $C(S^K)$.

It remains only to show that if W is monotone, then $\bar{W}$ is a positive linear functional on $C(S^K)$. We first prove that monotonicity of W implies $\hat{W}$ is a positive linear functional on H^*. For suppose $f \in H^*$, $f \geq 0$. Then, there exist $\sigma^1, \sigma^2 \in C$ and $r > 0$ such that $f = r(\sigma^1 - \sigma^2)$. Clearly, we must have $\sigma^1 \geq \sigma^2$, and hence

$$\hat{W}(f) = r[W(\sigma^1) - W(\sigma^2)] \geq 0$$

by the monotonicity of W. Thus $\hat{W}$ is a positive linear functional. Now, let $f \in C(S^K)$, $f \geq 0$. Since H^* is dense in $C(S^K)$, there exists a sequence $\{f_n\} \subset H^*$ such that $f_n \to f$. Without loss of generality, suppose $f_n \geq 0$ for all n.[9] Since $\bar{W}|_{H^*} = \hat{W}$ and $\hat{W}$ is a positive linear functional, $\bar{W}(f_n) \geq 0$ for all n. Then, by continuity, $\bar{W}(f) \geq 0$, so $\bar{W}$ is a positive linear functional. ■

Notice that the uniqueness of the extension in Lemma 11 is not necessary for proving the existence of an additive EU representation. We include this argument anyway as it can be useful for showing other results. For example, one can use it to show uniqueness of the measure on S^K shown to exist in the next lemma.

We have now established that there exists a continuous linear function $\bar{W} : C(S^K) \to \mathbf{R}$ such that $V(x) = \bar{W}(\sigma_x)$ for all $x \in X$. We have also established that if $\succ$ satisfies monotonicity, then $\bar{W}$ is a positive linear functional. We can now apply the Riesz representation theorem, which states that every continuous linear functional on $C(S^K)$ can be represented as integration against a measure.

Lemma 12 *If S^K is compact metrizable space and $\bar{W}$ is a continuous linear functional on $C(S^K)$, then there exists a finite signed Borel measure μ on S^K such that*

$$\bar{W}(f) = \int_{S^K} f(s)\mu(ds).$$

Furthermore, if $\bar{W}$ is a positive linear functional, then μ is positive.

Proof. See Royden (1988, Theorem 25, page 357) or Aliprantis and Border (1999, Theorem 13.15, page 466). For the case of a positive linear functional, see Royden (1988, Theorem 23, page 352). ■

[9] Otherwise, we could take the sequence $\{f'_n\} \subset H^*$ defined by $f'_n = \max\{f_n, 0\}$. The functions in this sequence are elements of H^* because $\mathbf{0} \in H^*$ and H^* is a vector lattice, which implies it contains the pointwise maximum of any two of its elements. A proof that H^* is a vector lattice can be found in DLR (2001, Lemma 11). Then, since $f \geq 0$, we have $\|f'_n - f\| \leq \|f_n - f\|$, and therefore $f'_n \to f$.

Define $U : \Delta(B) \times S^K \to \mathbf{R}$ by $U(\beta, s) = \beta \cdot s$. Then for all $x \in X$,

$$V(x) = \bar{W}(\sigma_x) = \int_{S^K} \sigma_x(s)\mu(ds) = \int_{S^K} \max_{\beta \in x} U(\beta, s)\mu(ds).$$

Furthermore, if $\succ$ satisfies monotonicity, then μ is positive. This completes the proof.

References

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

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

Fishburn, P. (1970): *Utility Theory for Decision Making*, Publications in Operations Research, No. 18. New York: Wiley.

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

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

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

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

Sarver, T. (2005), "Anticipating Regret: Why Fewer Options May Be Better," Mimeo, Boston University.

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