A Unique Costly Contemplation Representation*

Haluk Ergin† Todd Sarver‡

Current Draft:
November 2, 2007

Abstract

We extend the costly contemplation model to preferences over sets of lotteries, assuming that the state-dependent utilities are von Neumann-Morgenstern. The contemplation costs are uniquely pinned down in a reduced form representation, where the decision-maker selects a subjective measure over expected-utility functions instead of a subjective signal over a subjective state space. We show that in this richer setup, costly contemplation is characterized by Aversion to Contingent Planning, Indifference to Randomization, Independence of Degenerate Decisions, and Strong Continuity.

*We thank Faruk Gul, Bart Lipman, Massimo Marinacci, and numerous seminar participants for helpful comments.

†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

In many problems of individual choice, the decision-maker faces some uncertainty about her preferences over the available alternatives. In many cases, she may be able to improve her decision by first engaging in some form of introspection or contemplation about her preferences. However, if this contemplation is psychologically costly for the individual, then she will not wish to engage in any unnecessary contemplation. This will lead a rational individual to exhibit what we will refer to as an "aversion to contingent planning."

To illustrate, consider a simple example. We will take an individual to one of two restaurants. The first one is a seafood restaurant that serves a tuna (t) and a salmon (s) dish, which we denote by $A = \{t, s\}$. The second one is a steak restaurant that serves a filet mignon (f) and a ribeye (r) dish, which we denote by $B = \{f, r\}$. We will flip a coin to determine to which restaurant to go. If it comes up heads then we will buy the individual the meal of her choice in A, and if it comes up tails then we will buy her the meal of her choice in B. We consider presenting the individual with one of the two following decision problems:

Decision Problem 1
We ask the individual to make a complete contingent plan listing what she would choose conditional on each outcome of the coin flip.

Decision Problem 2
We first flip the coin and let the individual know its outcome. She then selects the dish of her choice from the restaurant determined by the coin flip.

It is conceivable that the individual prefers facing the second decision problem rather than the first one. In this case we say that her preferences (over decision problems) exhibit *Aversion to Contingent Planning (ACP)*. Our explanation of ACP is that the individual finds it psychologically costly to figure out her tastes over meals. Because of this cost, she would rather not contemplate on an inconsequential decision: She would rather not contemplate about her choice out of A were she to know that the coin came up tails and her actual choice set is B. In particular, she prefers to learn which choice set (A or B) is the relevant one before contemplating on her choice.

Our main results are a representation and a uniqueness theorem for preferences over sets of lotteries. We interpret that the preferences arise from a choice situation where, initially the individual chooses from among sets of lotteries (menus, options sets, or decision problems) and subsequently chooses a lottery from that set. The only primitive of the model is the preference over sets of lotteries, which corresponds to the individual's choice behavior in the first period; we do not explicitly model the second-period choice out of the sets. The key axiom in our analysis is ACP, and our representation is a reduced form of the costly contemplation representation introduced in Ergin (2003).[1]

Note that in our restaurant example decision problem 1 corresponds to a choice out of $A \times B = \{(t, f), (t, r), (s, f), (s, r)\}$, where for instance, (s, f) is the plan where the individual indicates that she will have the salmon dish from the seafood restaurant if the coin comes up heads and she will have the filet mignon from the steak restaurant if the coin comes up tails. Also, note that each choice of a contingent plan eventually yields a lottery over meals. For example, if the individual chooses (s, f), then she will face the lottery $\frac{1}{2}s + \frac{1}{2}f$ that yields either salmon or filet mignon, each with one-half probability. Hence decision problem 1 is identical to a choice out of the set of lotteries $\frac{1}{2}A + \frac{1}{2}B = \left\{\frac{1}{2}t + \frac{1}{2}f, \frac{1}{2}t + \frac{1}{2}r, \frac{1}{2}s + \frac{1}{2}f, \frac{1}{2}s + \frac{1}{2}r\right\}$. In general, we can represent the set of contingent plans between any two menus as a mixture of these menus, with the weight on each menu corresponding to the probability that it will be the relevant menu. The individual's preference of decision problem 2 to decision problem 1 is thus equivalent to preferring the half-half lottery over A and B (resolving prior to her choice from the menus) to the mixture of the two menus, $\frac{1}{2}A + \frac{1}{2}B$. Under the mild additional assumption that the individual prefers the better menu, say A, to any lottery over the two menus, we find that the individual will prefer the better of the two menus to the set of contingent plans from the two menus. Our ACP axiom is precisely the formalization of this statement: If $A \succsim B$, then $A \succsim \alpha A + (1 - \alpha)B$ for any $\alpha \in [0, 1]$.

We present our model in detail in Section 2. Along with ACP, we consider four additional axioms. The first three are standard axioms in the setting of preferences over menus:[2] (i) weak order, which states that the preference is complete and transitive, (ii) continuity, and (iii) indifference to randomization, which states that allowing an individual to randomize over the items in a menu has no added value or cost to her. We refer to the fourth axiom as *Independence of Degenerate Decisions (IDD)*. This axiom allows for contemplation, but rules out the possibility that the individual's beliefs themselves are changing.

[1]The primitive of the model of Ergin (2003) is a preference over menus taken from a finite set of alternatives. As we discuss in Section 3, the parameters in his representation are not pinned down by the preference. The richer domain of our preferences, menus of lotteries, combined with the reduced form of our representation enables us to uniquely identify the parameters of our representation. Moreover, our richer domain yields additional behavioral implications of costly contemplation, such as ACP.

[2]See Dekel, Lipman, and Rustichini (2001, henceforth DLR) and Dekel, Lipman, Rustichini, and Sarver (2007, henceforth DLRS).

Our representation and uniqueness theorems are contained in Section 2.2. To motivate our representation, recall that the individual in our model can contemplate in order to obtain information about her preferences. One can think of an individual who chooses her contemplation strategy as selecting a signal, the realization of which gives the individual some information about her tastes for the different alternatives in a menu. The information contained in a realization of the signal results in some ex post ranking of the alternatives, and this ranking can be represented by an ex post utility function. Hence, the distribution of a signal translates into a distribution over ex post utility functions, which is precisely the reduced-form approach we take to modeling costly contemplation.

Letting p denote a lottery and A denote a menu of lotteries, Theorem 1.A shows that any preference over menus satisfying our axioms can be represented by a function V of the form:[3]

$$V(A) = \max_{\mu \in \mathcal{M}} \left[\int_{\mathcal{U}} \max_{p \in A} u(p)\, \mu(du) - c(\mu) \right]. \tag{1}$$

The set $\mathcal{U}$ is a collection of ex post utility functions, with the restriction that each $u \in \mathcal{U}$ is a von Neumann-Morgenstern expected-utility function. The set $\mathcal{M}$ is a collection of signed measures on $\mathcal{U}$, and hence each measure $\mu \in \mathcal{M}$ determines a particular weighting of the ex post utility functions.[4] We interpret the measures in $\mathcal{M}$ as representing the available contemplation strategies and $c(\mu)$ as representing the cost of contemplation. Since we wish to interpret the measures as representing a change in information and not a change in belief, the following consistency condition is imposed on the measures in $\mathcal{M}$ for every lottery p:

$$\int_{\mathcal{U}} u(p)\, \mu(du) = \int_{\mathcal{U}} u(p)\, \nu(du), \quad \forall \mu, \nu \in \mathcal{M}. \tag{2}$$

This condition implies that when the individual faces a menu consisting of a single lottery p, she does not improve her utility by contemplating. Thus, contemplation is only of value to the individual if she has the flexibility to condition her choice of lottery on the information she receives from contemplating.

We refer to the representation given by Equations (1) and (2) as a *Reduced-Form Costly-Contemplation (RFCC) representation.* The important special case of our rep-

[3]This representation bears some similarity to the representation for "variational preferences" considered by Maccheroni, Marinacci, and Rustichini (2006) in the Anscombe-Aumann setting. There is also a technical connection between the two representations since we apply similar results from convex analysis to establish our representation theorems, although the setting of our model requires us to develop a stronger version of these results (see Appendix A).

[4]Note that the measures in our representation are not necessarily probability measures; in fact, they may not even be positive measures.

resentation in which all of the measures in $\mathcal{M}$ are positive is referred to as a *monotone RFCC representation*. Theorem 1.B shows that a preference has a monotone RFCC representation if it satisfies the axioms of Theorem 1.A and the following monotonicity axiom: $A \subset B \implies B \succsim A$. In words, the monotonicity axiom states that adding alternatives to any menu is always (weakly) better for the individual. The final result of Section 2.2 is Theorem 2, which establishes the uniqueness of the RFCC representation.

In Section 3, we further motivate the contemplation interpretation of our representation by relating it to an alternative non-reduced-from representation. In Theorem 3, we show that any monotone RFCC representation is equivalent to the following *Costly Contemplation (CC) representation*:

$$V(A) = \max_{\mathcal{G} \in \mathbf{G}} \left\{ \mathbb{E}\left[\max_{p \in A} \mathbb{E}\big[U \big| \mathcal{G}\big] \cdot p \right] - c(\mathcal{G}) \right\}. \tag{3}$$

This representation more closely resembles a standard costly information acquisition problem and is similar to the functional form considered by Ergin (2003). The CC representation includes a probability space $(\Omega, \mathcal{F}, P)$: The set Ω is a state space, $\mathcal{F}$ is a σ-algebra on Ω, and P is a probability measure on $(\Omega, \mathcal{F})$. The expectations in Equation (3) are taken with respect to P. The individual's state-dependent utility is given by $U(\omega) \cdot p$, where $U : \Omega \to \mathbb{R}^Z$ is Z-dimensional random vector. The individual's contemplation strategies are modeled as signals about the state, or equivalently, as σ-algebras on the state space. The set of contemplation strategies is therefore given by a set $\mathbf{G}$ of sub-σ-algebras of $\mathcal{F}$, with $c(\mathcal{G})$ denoting the cost of the contemplation strategy $\mathcal{G} \in \mathbf{G}$. The parameters $((\Omega, \mathcal{F}, P), \mathbf{G}, U, c)$ of the representation in (3) are subjective in the sense that they are not directly observable, but instead must be elicited from the individual's preferences.

Given the perhaps more standard formulation of the costly contemplation representation in Equation (3), it is natural to ask why our focus is on the RFCC representation. As we discuss in Section 3, there are two main reasons for preferring the reduced-form representation. First, the CC representation is equivalent to the monotone RFCC representation, which is a special case of the RFCC representation. The more general (non-monotone) RFCC representation permits a preference for commitment, which can be used to incorporate other factors, such as temptation or regret, into the model. Second, and more importantly, the parameters in the CC representation are not uniquely identified, and we present an example in Section 3 to illustrate this non-uniqueness.

In Section 4, we introduce a variation of our model in which the individual has limited resources to devote to contemplation. That is, the cost of contemplation does not directly affect the utility of the individual, but instead enters indirectly by being constrained to be below some fixed upper bound. We show that such a model is in fact a special case of our RFCC representation, and we introduce the additional axiom needed to obtain this representation for limited contemplation resources.

Our work relates to several other papers in the literature on preferences over menus. This literature originated with Kreps (1979), who considered preferences over menus taken from a finite set of alternatives. Kreps' framework is also used by Ergin (2003) in his model of costly contemplation. DLR (2001) extend Kreps' analysis to the current setting of preferences over menus of lotteries and use the additional structure of this domain to obtain an essentially unique representation. In Section 5.1, we discuss a version of the independence axiom for preferences over menus of lotteries which is used by DLR (2001) in one of their representation results. We illustrate how our axioms relax the independence axiom and why such a relaxation of independence is necessary in order to model costly contemplation.[5] We conclude in Section 5.2 with a brief overview of the so called *infinite regress issue* for models of costly decision-making, and we explain how our results provide an *as if* solution to the issue.

2 A Model of Costly Contemplation

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.[6] Let $\mathcal{A}$ denote the set of all closed subsets of $\triangle(Z)$, endowed with the Hausdorff metric, which is defined by

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

Elements of $\mathcal{A}$ are called menus or option sets. The primitive of our model is a binary relation $\succsim$ on $\mathcal{A}$, representing the individual's preferences over menus. We maintain the interpretation that, after committing to a particular menu A, the individual chooses a lottery out of A in an unmodeled second stage.

For any $A, B \in \mathcal{A}$ and $\alpha \in [0,1]$, define the convex combination of these two menus by $\alpha A + (1-\alpha)B \equiv \{\alpha p + (1-\alpha)q : p \in A \text{ and } q \in B\}$. Let $co(A)$ denote the convex

[5] A relaxation of the independence axiom in the setting of preferences over menus of lotteries is also considered by Epstein, Marinacci and Seo (2007), who model an individual with an incomplete (or coarse) conception of the future. This coarse conception entails a degree of pessimism on the part of the individual, and their resulting representations are intuitively similar to the maxmin representation of Gilboa and Schmeidler (1989).

[6] Since Z is finite, the topology generated by d is equivalent to the weak* topology on $\triangle(Z)$.

hull of the set A.

2.1 Axioms

We impose the following order and continuity axioms.

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

Axiom 2 (Strong Continuity)

1. (Continuity): *For all* $A \in \mathcal{A}$*, the sets* $\{B \in \mathcal{A} : B \succsim A\}$ *and* $\{B \in \mathcal{A} : B \precsim A\}$ *are closed.*

2. (L–Continuity): *There exist* $p^*, p_* \in \triangle(Z)$ *and* $M > 0$ *such that for every* $A, B \in \mathcal{A}$ *and* $\alpha \in (0,1)$ *with* $d_h(A,B) < \alpha/M$*,*

$$(1-\alpha)A + \alpha\{p^*\} \succ (1-\alpha)B + \alpha\{p_*\}.$$

The weak order axiom is entirely standard, as is the first part of the strong continuity axiom. The added assumption of L–continuity is used to obtain Lipschitz continuity of our representation in much the same way that the continuity axiom is used to obtain continuity.[7] To interpret L–continuity, first note that $\{p^*\} \succ \{p_*\}$.[8] For any $A, B \in \mathcal{A}$, continuity therefore implies that there exists $\alpha \in (0,1)$ such that $(1-\alpha)A+\alpha\{p^*\} \succ (1-\alpha)B+\alpha\{p_*\}$. L–continuity implies that such a preference holds for any $\alpha > Md_h(A,B)$, so as A and B get closer, the minimum required weight on p^* and p_* converges to 0 at a smooth rate. The constant M can be thought of as the sensitivity of this minimum α to the distance between A and B.

The next axiom was introduced in DLR (2001).

Axiom 3 (Indifference to Randomization (IR)) *For every* $A \in \mathcal{A}$*,* $A \sim co(A)$*.*

IR is justified if the individual choosing from the menu A can also randomly select an alternative from the menu, for example, by flipping a coin. In that case, the menus A

[7] Similar L–continuity axioms are used in DLRS (2007) and Sarver (2007). There is also a connection between our L–continuity axiom and the properness condition proposed by Mas-Colell (1986).

[8] Let $\alpha = 1/2$. Applying L–continuity with $A = B = \{p^*\}$ implies $\{p^*\} \succ \{\frac{1}{2}p^* + \frac{1}{2}p_*\}$, and applying L–continuity with $A = B = \{p_*\}$ implies $\{\frac{1}{2}p^* + \frac{1}{2}p_*\} \succ \{p_*\}$.

and $co(A)$ offer the same set of options, and hence they are identical from the perspective of the individual.

The next axiom captures an important aspect of our model of costly contemplation.

Axiom 4 (Aversion to Contingent Planning (ACP)) *For any* $\alpha \in [0, 1]$,

$$A \succsim B \implies A \succsim \alpha A + (1 - \alpha)B.$$

To interpret ACP, suppose we were to extend the individual's preferences to lotteries over menus. Let $\alpha \circ A \oplus (1 - \alpha) \circ B$ denote the lottery that yields the menu A with probability α and the menu B with probability $1 - \alpha$. We interpret that this lottery resolves prior to the individual making her choice of alternative from the menus. If instead the individual is asked to make her decision prior to the resolution of the lottery, then she must make a contingent choice. The situation in which the individual makes a contingent choice, p if A and q if B, prior to the resolution of the lottery over menus is equivalent to choosing the alternative $\alpha p + (1 - \alpha)q \in \alpha A + (1 - \alpha)B$. Thus any contingent choice from A and B corresponds to a unique lottery in $\alpha A + (1 - \alpha)B$.[9] As discussed in the introduction, if contemplation is costly for the individual, then she will prefer that a lottery over menus is resolved prior to her choosing an alternative so that she can avoid contingent planning. Hence,

$$\alpha \circ A \oplus (1 - \alpha) \circ B \succsim \alpha A + (1 - \alpha)B. \tag{4}$$

If in addition this extended preference satisfies stochastic dominance in the sense that $A \succsim B$ implies $A \succsim \alpha \circ A \oplus (1 - \alpha) \circ B$, then Equation (4) implies ACP.[10]

Suppose the individual is asked to make a contingent plan, and she is told that she will be choosing from the menu A with probability α and from the menu $\{p\}$ with probability $1 - \alpha$. We refer to a choice from the singleton menu $\{p\}$ as a *degenerate decision*. When faced with a degenerate decision, there is no benefit to the individual from contemplating. Therefore, if the probability α that her contingent choice from A will be implemented decreases, then her benefit from contemplation decreases. Hence we should expect the individual to choose a less costly level of contemplation as α decreases. However, if α is held fixed, then replacing the degenerate decision $\{p\}$ in

[9] In Appendix F we prove a partial converse of this statement. Namely, for a dense set of menus it is also true that any lottery in $\alpha A + (1 - \alpha)B$ corresponds to a unique contingent plan from A and B.

[10] Let ACP' be the requirement that for all $\alpha \in [0, 1]$: $A \sim B \implies A \succsim \alpha A + (1 - \alpha)B$. All the results in Appendix B rely only on the weaker ACP' property and not the full strength of ACP. Therefore all of our results stated in the main text continue to hold if one replaces ACP by ACP'. The only exceptions are Lemma 33 and Lemma 34 in Appendix F for which we need the original formulation of ACP above.

the mixture $\alpha A + (1-\alpha)\{p\}$ with another degenerate decision $\{q\}$ does not change the probability that the individual's contingent choice from A will be implemented. Therefore, although replacing p with q could affect the individual's utility through its effect on the final composition of lotteries, it will not affect the individual's optimal level of contemplation. The following axiom states that if α is fixed, then there is a type of substitutability of degenerate decisions.

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

$$\alpha A + (1-\alpha)\{p\} \succsim \alpha B + (1-\alpha)\{p\} \implies \alpha A + (1-\alpha)\{q\} \succsim \alpha B + (1-\alpha)\{q\}.$$

Suppose the individual prefers making a contingent plan from $\alpha A + (1-\alpha)\{p\}$ to making one from $\alpha B + (1-\alpha)\{p\}$. As argued above, substituting q for p in both cases will not affect individual's optimal level of contemplation, and hence she will make the same contingent choices from A and B, respectively. IDD states that substituting the degenerate decision $\{q\}$ in the place of $\{p\}$ in both cases will not affect her preferences over these contingent plans. IDD allows for the possibility that the individual contemplates to obtain information about her ex post utility, but it rules out the possibility that the individual changes her beliefs by becoming more optimistic about the utility she will obtain from a given lottery.[11]

Finally, we will also consider the monotonicity axiom of Kreps (1979) in conjunction with our other axioms to obtain a refinement of our representation.

Axiom 6 (Monotonicity) *If $A \subset B$, then $B \succsim A$.*

Note that monotonicity is entirely consistent with costly contemplation and aversion to contingent planning. If additional alternatives are added to a menu A, the individual can always "ignore" these new alternatives and engage in the same contemplation as with the menu A.[12] Therefore, the utility from a menu $B \supset A$ must be at least as great as the utility from the menu A. Although at first glance it may seem that costly contemplation alone could lead to a preference for smaller menus in order to avoid "over-analyzing" the decision, this argument overlooks the fact that the individual chooses her contemplation strategy optimally.

[11]Our IDD axiom is similar in spirit to the weak certainty independence axiom used by Maccheroni, Marinacci, and Rustichini (2006) in the Anscombe-Aumann setting. In their axiom, arbitrary acts play the role of the menus A and B and constant acts play the role of the singleton menus $\{p\}$ and $\{q\}$.

[12]Note that we are assuming it is costless for the individual to "read" the alternatives on the menu. What is costly for the individual is analyzing her tastes for these alternatives.

The possibility of over-analysis could arise if the individual experiences some disutility from not selecting the ex post optimal choice from a menu, for example because of regret. Therefore, regret could lead the individual to sometimes prefer a smaller menu, which we refer to as a *preference for commitment*. Other factors, such as temptation, could also lead to a preference for commitment. Although we do not consider regret and temptation explicitly in this paper, our model without monotonicity is general enough to allow for such factors to be incorporated.[13]

2.2 Representation Result

As we discussed in the introduction, one can think of an individual who chooses her contemplation strategy as selecting a signal, the realization of which gives the individual some information about her tastes for the different alternatives in a menu. The individual's ranking of alternatives conditional on the information contained in a realization of the signal is represented by some ex post utility function. Therefore, the distribution of a signal translates into a distribution over ex post utility functions, which is precisely the reduced-form approach we take to modeling costly contemplation.

Even though contemplation may be costly for the individual in our model, we assume that she is an expected-utility maximizer. Therefore, we impose the condition that all of the ex post utility functions in our representation are von Neumann-Morgenstern. 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. Define the set of *normalized (non-constant) expected-utility functions* on $\triangle(Z)$ to be

$$\mathcal{U} = \left\{ u \in \mathbb{R}^Z : \sum_{z \in Z} u_z = 0, \sum_{z \in Z} u_z^2 = 1 \right\}. \tag{5}$$

For any $\hat{u} \in \mathbb{R}^Z$ (i.e., any expected-utility function), there exist $\alpha \geq 0$, $\beta \in \mathbb{R}$, and $u \in \mathcal{U}$ such that $\hat{u}(p) = \alpha u(p) + \beta$ for all $p \in \triangle(Z)$. Therefore, modulo an affine transformation, $\mathcal{U}$ contains all possible ex post expected-utility functions.

We now define our main representation:[14]

[13]Regret is studied in a related framework by Sarver (2007), and temptation is studied by Gul and Pesendorfer (2001) and Dekel, Lipman, and Rustichini (2007).

[14]Note that we endow the set of all finite signed Borel measures on $\mathcal{U}$ with the weak* topology, that is, the topology where a net $\{\mu_d\}_{d \in D}$ converges to μ if and only if $\int_{\mathcal{U}} f\, \mu_d(du) \to \int_{\mathcal{U}} f\, \mu(du)$ for every continuous function $f : \mathcal{U} \to \mathbb{R}$.

Definition 1 A *Reduced Form Costly Contemplation (RFCC) 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 $V : \mathcal{A} \to \mathbb{R}$ defined by

$$V(A) = \max_{\mu \in \mathcal{M}} \left[\int_{\mathcal{U}} \max_{p \in A} u(p)\, \mu(du) - c(\mu) \right] \tag{6}$$

represents $\succsim$, and:

1. The set $\mathcal{M}$ is *consistent*: For each $\mu, \nu \in \mathcal{M}$ and $p \in \triangle(Z)$,

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

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 (6) no longer represents $\succsim$.

3. There exist $p, q \in \triangle(Z)$ such that $V(\{p\}) > V(\{q\})$.

The consistency condition on the set of measures $\mathcal{M}$ is necessary for our contemplation interpretation of the RFCC representation. Suppose μ and ν represent different contemplation strategies, and suppose $p \in \triangle(Z)$. Since the individual has only one choice when faced with the singleton menu $\{p\}$, she cannot change her choice based on her information. Therefore, the only effect of the individual's contemplation decision on her utility is the effect it has on her contemplation cost c. Thus the first term in the brackets in Equation (6) must be the same for both μ and ν when $A = \{p\}$, which implies the set of measures $\mathcal{M}$ must be consistent.

The minimality condition is needed in order to uniquely identify the parameters in our representation. To see this, note that it is always possible to add a measure $\mu \notin \mathcal{M}$ to the set $\mathcal{M}$ and assign it a cost $c(\mu)$ high enough to guarantee that this measure is never a maximizer in Equation (6). The minimality condition requires that all such unnecessary measures be dropped from the representation. In other words, a minimal set $\mathcal{M}$ in an RFCC representation may not include all possible contemplation strategies available to the individual, but it identifies all of the "relevant" ones.

Condition 3 in the definition is simply a technical requirement relating to the strong continuity axiom. If we take p^* and p_* as in the definition of L–continuity, then $\{p^*\} \succ \{p_*\}$, which gives rise to this third condition. We should also note this "singleton nontriviality" implication of L–continuity is not accidental, as it plays an important role in the proof of our representation theorem.

Finally, note that the normalization of $\mathcal{U}$ in Equation (5) is without loss of generality. Any representation satisfying the conditions of Definition 1 with the modification that $\mathcal{U}$ is allowed to be any (non-normalized) set of ex post expected-utility functions can, after a transformation of the measures and cost function, be written as an RFCC representation with $\mathcal{U}$ defined by Equation (5). Therefore, our normalization of $\mathcal{U}$ imposes no real restrictions on the representation, and our representation theorems would continue to hold without this normalization. However, our uniqueness result would be more difficult to express without this normalization since, intuitively, there would be an extra "degree of freedom" that would need to be taken into account.

An RFCC representation in which all of the measures in $\mathcal{M}$ are positive will be an important special case of our representation which we refer to as the *Monotone RFCC representation.* We now present our main representation theorem:

Theorem 1 *A. The preference $\succsim$ has an RFCC representation if and only if it satisfies weak order, strong continuity, IR, ACP, and IDD.*

B. The preference $\succsim$ has a monotone RFCC representation if and only if it satisfies weak order, strong continuity, monotonicity, ACP, and IDD.[15]

Proof: See Appendix B.

The proof of this result in Appendix B is divided into three parts: In Appendix B.1, we prove some preliminary results for a preference $\succsim$ that satisfies our axioms. Intuition for some of these preliminary results can be found in the discussion in Section 5.1. In Section B.2, we construct a function V that represents $\succsim$ and satisfies certain desirable properties: Lipschitz continuity, convexity, and a type of "translation-linearity" which is closely related to the consistency condition for the measures in our representation. Finally, in Section B.3, we apply duality results from convex analysis to establish that this function V satisfies Equation (6) for some pair $(\mathcal{M}, c)$.

The V defined by Equation (6) for an RFCC representation is a convex function. Although the nonlinearity of this function prevents the use of standard arguments from expected-utility theory, it can still be shown that V is unique up to a positive affine transformation (see Proposition 3 in Appendix B.2). From this it can then be shown that the parameters on an RFCC representation $(\mathcal{M}, c)$ are themselves unique up to a type of positive affine transformation:

[15]Note that the IR axiom is not included in part B of Theorem 1. It is shown in Lemma 5 in Appendix B.1 that IR is implied by the combination of weak order, ACP, monotonicity, and continuity.

Theorem 2 *If $(\mathcal{M}, c)$ and $(\mathcal{M}', c')$ are two RFCC 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 $\mu \in \mathcal{M}$.*

Proof: See Appendix C.

3 Contemplation Interpretation of the Model

In this section, we further motivate the contemplation interpretation of our representation by relating it to the alternative non-reduced-from representation which closely resembles a standard costly information acquisition problem.

Definition 2 A *Costly Contemplation (CC) representation* is a tuple $((\Omega, \mathcal{F}, P), \mathbf{G}, U, c)$ where $(\Omega, \mathcal{F}, P)$ is a probability space, $\mathbf{G}$ is a collection of sub-σ-algebras of $\mathcal{F}$, U is a Z-dimensional, $\mathcal{F}$–measurable, and integrable random vector, and $c : \mathbf{G} \to \mathbb{R}$ is a cost function such that the preference $\succsim$ is represented by

$$V(A) = \max_{\mathcal{G} \in \mathbf{G}} \left\{ \mathbb{E}\left[\max_{p \in A} \mathbb{E}\big[U \big| \mathcal{G}\big] \cdot p \right] - c(\mathcal{G}) \right\} \tag{7}$$

where the outer maximization in (7) has a solution for every $A \in \mathcal{A}$ and there exist $p, q \in \triangle(Z)$ such that $\mathbb{E}[U] \cdot p > \mathbb{E}[U] \cdot q$.[16,17,18]

The costly contemplation representation above is a generalized version of the costly contemplation representation in Ergin (2003).[19] The interpretation of formula (7) is

[16]Let $A \in \mathcal{A}$ and $\mathcal{G} \in \mathbf{G}$. We will show that the term $\mathbb{E}\left[\max_{p \in A} \mathbb{E}\big[U\big|\mathcal{G}\big] \cdot p\right]$ is well defined and finite. This in particular implies that $V(A)$ is finite whenever the outer maximization in (7) has a solution. Let $\tilde{U}$ be an arbitrary version of $\mathbb{E}[U|\mathcal{G}]$. The existence and integrability of $\tilde{U}_z$ follows from integrability of U_z for each $z \in Z$ (see Billingsley (1995, p445)). 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 $\max_{p \in A} \tilde{U} \cdot p = \sup_{p \in B} \tilde{U} \cdot p$ is an $\mathcal{F}$–measurable random variable as the pointwise supremum of countably many $\mathcal{F}$–measurable random variables (see Billingsley (1995, p184) Theorem 13.4 (i)). Note also that for any $p \in \triangle(Z)$, $|\tilde{U} \cdot p| \le \sum_{z \in Z} |\tilde{U}_z|$, hence $\big|\max_{p \in A} \tilde{U} \cdot p\big| \le \sum_{z \in Z} |\tilde{U}_z|$. Therefore integrability of $\max_{p \in A} \tilde{U} \cdot p$ follows from integrability of $\tilde{U}$.

[17]For simplicity, we directly assume that the outer maximization in (7) has a solution, instead of making topological assumptions on $\mathbf{G}$ that may guarantee the existence of a maximum.

[18]The requirement that $\mathbb{E}[U] \cdot p > \mathbb{E}[U] \cdot q$ for some $p, q \in \triangle(Z)$ in the definition of the CC representation is imposed simply to match the singleton-nontriviality condition in the RFCC representation. The main result of the section (Theorem 3) continues to hold if the singleton-nontriviality conditions are dropped from the definitions of both the CC and the RFCC representations.

[19]Ergin (2003) works in the framework introduced by Kreps (1979), where the primitive of the model is a preference over subsets of Z rather than subsets of $\triangle(Z)$. He shows that a preference $\succsim$ over sets

as follows. The individual has a subjective state space Ω representing her tastes over alternatives, endowed with a σ-algebra $\mathcal{F}$. She does not know the realization of the subjective state $\omega \in \Omega$ but has a prior P on $(\Omega, \mathcal{F})$. Her tastes over lotteries in $\triangle(Z)$ are summarized by the random vector U representing her state-dependent vNM function. Her expected utility from a lottery $p \in \triangle(Z)$ conditional on the subjective state $\omega \in \Omega$ is therefore given by $U(\omega) \cdot p = \sum_{z \in Z} p_z U_z(\omega)$.

Before making a choice out of a menu $A \in \mathcal{A}$, the individual may engage in contemplation. A *contemplation strategy* is modeled as a signal about the subjective state which corresponds to a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$. The contemplation strategies available to the individual are given by the collection of σ-algebras $\mathbf{G}$. If the individual carries out the contemplation strategy $\mathcal{G}$, she incurs a psychological cost of contemplation $c(\mathcal{G})$. However, she can then condition her choice out of A on $\mathcal{G}$ and pick an alternative that yields the highest expected utility conditional on the signal realization. Faced with the menu A, the individual chooses an optimal level of contemplation by maximizing the value minus the cost of contemplation. This yields $V(A)$ in Equation (7) as the ex ante value of the option set A. The CC formulation is similar to an optimal information acquisition formula. The difference from a standard information acquisition problem is that the parameters $((\Omega, \mathcal{F}, P), \mathbf{G}, U, c)$ are subjective. Therefore, they are not directly observable, but need to be derived from the individual's preference.

When the individual undertakes the contemplation strategy $\mathcal{G}$, she anticipates that as a result of her contemplation, she will end up with an ex post utility determined by the random variable $\mathbb{E}[U|\mathcal{G}]$. Furthermore, it is enough for her to know the cost and the probability distribution over ex post utility functions associated with each contemplation strategy $\mathcal{G}$ in order to evaluate the ex ante value $V(A)$ in Equation (7). This argument suggests that a CC representation can be rewritten as an RFCC representation, where a contemplation strategy is expressed as a subjective measure over expected-utility functions instead of a subjective signal over a subjective state space. The following equivalence result formalizes this intuition.

Theorem 3 *Let $V : \mathcal{A} \to \mathbb{R}$. Then there exists a monotone RFCC representation such that V is given by Equation* (6) *if and only if there exists a CC representation such that V is given by Equation* (7).

---

of alternatives is monotone ($A \subset B \implies B \succsim A$) if and only if there exist a costly contemplation representation with finite Ω such that $\succsim$ is represented by the ex ante utility function V in (7). The formulation of costly contemplation in this paper allows for infinite subjective state space Ω and extends the formulation to menus of lotteries assuming that state-dependent utility is vNM.

Proof: See Appendix D.

Hence the monotone RFCC representation can be interpreted as a reduced form of the CC representation. An immediate corollary of Theorem 3 and our monotone RFCC representation theorem is the following CC representation theorem.

Corollary 1 *The preference $\succsim$ has a CC representation if and only if it satisfies weak order, strong continuity, monotonicity, ACP, and IDD.*

As argued above, it is not possible to behaviorally distinguish between two sets of CC parameters that induce the same probability distributions over ex post conditional expected-utility functions at the same costs. We next give an example of this non-uniqueness. By the uniqueness and equivalence results established above (Theorems 2 and 3), the non-uniqueness issue associated with CC representations may be overcome by working with equivalent monotone RFCC representations.

Example 1: Let $Z = \{z_1, z_2, z_3\}$, and let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $\hat{\Omega} = \{\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3\}$. Endow both Ω and $\hat{\Omega}$ with their discrete algebras, and let P and $\hat{P}$ be the uniform distributions on Ω and $\hat{\Omega}$, respectively. Consider the following collections of partitions of Ω and $\hat{\Omega}$:

$$\Pi = \Big\{\big\{\{\omega_1, \omega_2\}, \{\omega_3\}\big\}, \big\{\{\omega_1\}, \{\omega_2, \omega_3\}\big\}, \big\{\{\omega_1, \omega_3\}, \{\omega_2\}\big\}\Big\}$$
$$\hat{\Pi} = \Big\{\big\{\{\hat{\omega}_1, \hat{\omega}_2\}, \{\hat{\omega}_3\}\big\}, \big\{\{\hat{\omega}_1\}, \{\hat{\omega}_2, \hat{\omega}_3\}\big\}, \big\{\{\hat{\omega}_1, \hat{\omega}_3\}, \{\hat{\omega}_2\}\big\}\Big\}.$$

Define the collection of contemplation strategies $\mathbf{G}$ to be the collection of algebras of Ω generated by the partitions in Π. Similarly let $\hat{\mathbf{G}}$ consist of algebras of $\hat{\Omega}$ generated by the partitions in $\hat{\Pi}$. Let $c : \mathbf{G} \to \mathbb{R}$ and $\hat{c} : \hat{\mathbf{G}} \to \mathbb{R}$ be the zero functions, i.e., $c(\mathcal{G}) = 0$ for all $\mathcal{G} \in \mathbf{G}$ and $\hat{c}(\hat{\mathcal{G}}) = 0$ for all $\hat{\mathcal{G}} \in \hat{\mathbf{G}}$. Define $U : \Omega \to \mathbb{R}^3$ and $\hat{U} : \hat{\Omega} \to \mathbb{R}^3$ as follows:[20]

$$U(\omega_1) = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}, \qquad U(\omega_2) = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}, \qquad U(\omega_3) = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$$

$$\hat{U}(\hat{\omega}_1) = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}, \qquad \hat{U}(\hat{\omega}_2) = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \qquad \hat{U}(\hat{\omega}_3) = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}.$$

Defining V and $\hat{V}$ as in Equation (7) for each of these respective representations, it is easily verified that $V(A) = \hat{V}(A)$ for any menu $A \in \mathcal{A}$. ■

In the rest of this section we sketch the proof of Theorem 3. We start by constructing an equivalent RFCC representation for a given CC representation $((\Omega, \mathcal{F}, P), \mathbf{G}, U, c)$. Fix $\mathcal{G} \in \mathbf{G}$ and let $\mathbf{1} \in \mathbb{R}^Z$ denote the vector whose coordinates are all equal to 1. Since $\mathbb{E}\big[U\big|\mathcal{G}\big](\omega) \in \mathbb{R}^Z$ for each $\omega \in \Omega$, there exist $\mathcal{G}$–measurable and integrable random variables $\alpha : \Omega \to \mathbb{R}_+$ and $\beta : \Omega \to \mathbb{R}$ and a $\mathcal{G}$–measurable random vector $u : \Omega \to \mathcal{U}$ such that

$$\mathbb{E}\big[U\big|\mathcal{G}\big](\omega) = \alpha(\omega)u(\omega) + \beta(\omega)\mathbf{1}, \ \forall \omega \in \Omega.$$

Moreover, since $\sum_{z\in Z} u_z = 0$ for each $u \in \mathcal{U}$, it follows that $\sum_{z\in Z} \mathbb{E}\big[U_z\big|\mathcal{G}\big](\omega) = |Z|\beta(\omega)$ for all $\omega \in \Omega$. Therefore, by the law of iterated expectations, $\mathbb{E}[\beta] = \frac{1}{|Z|}\sum_{z\in Z}\mathbb{E}[U_z]$. To simplify the exposition, we will assume that $\sum_{z\in Z}\mathbb{E}[U_z] = 0$ and hence $\mathbb{E}[\beta] = 0$.[21]

Define the positive finite measure m on $(\Omega, \mathcal{G})$ via its Radon-Nikodym derivative $\frac{dm}{dP}(\omega) = \alpha(\omega)$ and define the positive measure $\mu_{\mathcal{G}}$ on $\mathcal{U}$ via $\mu_{\mathcal{G}} = m \circ u^{-1}$. Then for any

[20] Under this specification of the random vectors, we have $\mathbb{E}[U] = 0$ and $\mathbb{E}[\hat{U}] = 0$, and hence the singleton nontriviality condition in Definition 2 is not satisfied. However, we allow for this violation purely for expositional simplicity. These representations can be modified to satisfy singleton nontriviality as follows: Add a fourth state to each representation, ω_4 and $\hat{\omega}_4$, respectively, let $U(\omega_4) = \hat{U}(\hat{\omega}_4) \neq 0$, define the sets of partitions to be $\Pi' = \{\pi \cup \{\{\omega_4\}\} : \pi \in \Pi\}$ and $\hat{\Pi}' = \{\hat{\pi} \cup \{\{\hat{\omega}_4\}\} : \hat{\pi} \in \hat{\Pi}\}$, let $\mathbf{G}'$ and $\hat{\mathbf{G}}'$ consist of the algebras generated by the partitions in Π' and $\hat{\Pi}'$ respectively, and let $c : \mathbf{G}' \to \mathbb{R}$ and $\hat{c} : \hat{\mathbf{G}}' \to \mathbb{R}$ be the zero functions.

[21] Without this assumption, we would simply need to subtract the value $\frac{1}{|Z|}\sum_{z\in Z}\mathbb{E}[U_z]$ from the cost function derived below.

menu $A \in \mathcal{A}$:

$$\begin{aligned}
\mathbb{E}\left[\max_{p\in A} \mathbb{E}\big[U\big|\mathcal{G}\big]\cdot p\right] &= \int_{\Omega}\left[\alpha(\omega)\max_{p\in A} u(\omega)\cdot p\right] P(d\omega) + \mathbb{E}\left[\beta\right] \\
&= \int_{\Omega}\left[\max_{p\in A} u(\omega)\cdot p\right] m(d\omega) \\
&= \int_{\mathcal{U}}\left[\max_{p\in A} u\cdot p\right] \mu_{\mathcal{G}}(du)
\end{aligned}$$

where the final equality follows from the change of variables formula. This in particular implies that $\int_{\mathcal{U}} u\, \mu_{\mathcal{G}}(du) = \mathbb{E}[U]$.

Let $\mathcal{M} = \{\mu_{\mathcal{G}} : \mathcal{G} \in \mathbf{G}\}$ and $\tilde{c}(\mu) = \inf\{c(\mathcal{G}) : \mathcal{G} \in \mathbf{G} \text{ and } \mu = \mu_{\mathcal{G}}\}$ for each $\mu \in \mathcal{M}$. By the above paragraph the measures in $\mathcal{M}$ are consistent and for each $A \in \mathcal{A}$, $V(A)$ can be expressed as:

$$V(A) = \max_{\mu\in\mathcal{M}}\left[\int_{\mathcal{U}} \max_{p\in A} u(p)\,\mu(du) - \tilde{c}(\mu)\right].$$

We omit in this proof sketch the remaining arguments that (i) it is possible to obtain a compact subset $\mathcal{M}' \subset \mathcal{M}$ such that the above equation continues to hold if $\mathcal{M}$ is replaced by $\mathcal{M}'$, (ii) the representation with $\mathcal{M}'$ becomes minimal, and (iii) $\tilde{c}$ is lower semi-continuous on $\mathcal{M}'$.

Going back to Example 1, define $u^1, u^2, u^3 \in \mathcal{U}$ by:

$$u^1 = \frac{1}{\sqrt{6}}\begin{pmatrix}2\\-1\\-1\end{pmatrix}, \qquad u^2 = \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\2\\-1\end{pmatrix}, \qquad u^3 = \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\-1\\2\end{pmatrix}.$$

Then it is easy to see that the measures induced by the partitions $\big\{\{\omega_i\},\{\omega_j,\omega_k\}\big\}$ and $\big\{\{\hat{\omega}_i\},\{\hat{\omega}_j,\hat{\omega}_k\}\big\}$ are identical, giving $\frac{\sqrt{6}}{3}$ weight to u^i, $\frac{\sqrt{6}}{3}$ weight to $-u^i$, and zero weight to $\mathcal{U} \setminus \{u^i, -u^i\}$.

For the converse of Theorem 3, suppose that $(\mathcal{M}, c)$ is a monotone RFCC representation. We will construct a canonical information structure that has been used in a number of papers on mechanism design with information acquisition.[22] More specifically let $\Omega = N \times [\triangle(N)]^{\mathcal{M}}$ where $N = \{1, \ldots, n\}$ for $n = |Z|$. The payoff relevant part of the state space is only the first dimension, i.e., $U(i,\lambda) = U(i,\lambda')$ for all $i \in N$ and

[22] For instance, see Bergemann and Valimaki (2002), Bergemann and Valimaki (2006), and Persico (2000).

$\lambda, \lambda' \in [\triangle(N)]^{\mathcal{M}}$. For each measure $\mu \in \mathcal{M}$, consider the random vector $X_\mu : \Omega \to \triangle(N)$ defined by $X_\mu(i, \lambda) = \lambda(\mu)$. Then the signal X_μ reveals the μ^{th} dimension of the state space. We let $\mathcal{G}_\mu$ be the σ-algebra generated by X_μ and set $\mathbf{G} = \{\mathcal{G}_\mu : \mu \in \mathcal{M}\}$.

We will ensure that conditional on acquiring $\mathcal{G}_\mu$, the posterior probabilty of the payoff relevant states almost surely coincide with the actual realization of the signal X_μ in $\triangle(N)$. Since the signal X_μ directly returns the posterior over the payoff relevant dimension of the state space, its distribution is also known as a *standard measure*. (See, e.g., Blackwell (1953).)

Given a measure $\mu \in \mathcal{M}$, define the probability measure $\tilde{\mu}$ on $\mu(\mathcal{U})\mathcal{U}$ by $\tilde{\mu}(E) = \frac{1}{\mu(\mathcal{U})}\mu(\frac{1}{\mu(\mathcal{U})}E)$ for any measurable $E \subset \mu(\mathcal{U})\mathcal{U}$. Then:

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

Hence the integral expression in the RFCC representation can be reinterpreted in a probabilistic sense by rescaling the utility functions in $\mathcal{U}$, where the rescaling coefficient depends on the particular measure.

By compactness of $\mathcal{M}$, there exists $\kappa \geq 0$ such that $\mu(\mathcal{U}) \leq \kappa$ for all $\mu \in \mathcal{M}$. Since $\kappa\mathcal{U}$ is a compact subset of of the $n-1$ dimensional subspace $H = \{u \in R^Z : \sum_z u_z = 0\}$ of $\mathbb{R}^Z$, there exist n affinely independent vectors $v^1, \ldots, v^n \in H$ such that $\cup_{\delta \in [0,\kappa]} \delta\mathcal{U} \subset co(\{v^1, \ldots, v^n\})$. We let these vectors represent the individual's expected utility conditional on the payoff relevant dimension, i.e., we define $U(i, \lambda) = v^i$ for $i \in N$ and $\lambda \in [\triangle(N)]^{\mathcal{M}}$. By affine independence, each point in $co(\{v^1, \ldots, v^n\})$ can be uniquely expressed as a convex combination of the vertices $v^1, \ldots, v^n$. We can therefore interpret each such point as a probability distribution on $N = \{1, \ldots, n\}$ where the probability of $i \in N$ is given by the coefficient of v^i in the unique convex combination. Furthermore for each $\mu \in \mathcal{M}$, the probability measure $\tilde{\mu}$ can be identified with a probability distribution π_μ over $\triangle(N)$. Figure 1 illustrates this construction for the case where $n = 3$.

Consistency of the measures in $\mathcal{M}$ guarantees that the expected value $\alpha \in \triangle(N)$ of π_μ is independent of the particular $\mu \in \mathcal{M}$. We interpret α as the individual's prior beliefs on the payoff-relevant dimension of Ω before she acquires a signal. In Lemma 30 in the appendix, we generalize an observation in Blackwell (1951). An implication of Lemma 30 is that for each $\mu \in \mathcal{M}$, there exist conditional probability measures $(P_\mu(\cdot|1), \ldots, P_\mu(\cdot|n))$ on $\triangle(N)$ such that in the sampling procedure where the individual has prior beliefs $\alpha = (\alpha_1, \ldots, \alpha_n)$ on N and observes a signal drawn with these conditional probabilities: (i) her posterior over N is almost surely the same as the value of her signal, and (ii) the prior distribution over her posteriors is given by π_μ.

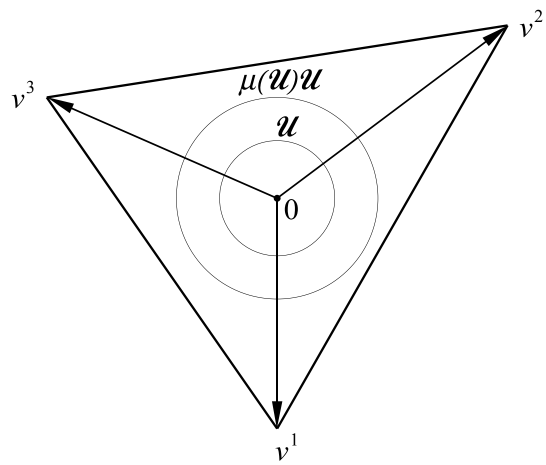

Figure 1: Construction of the CC Representation

We also show in the appendix that the Kolmogorov extension theorem guarantees that there exist a probability measure P on Ω such that the marginal of P on N agrees with α, and the induced probability distribution on the μ^{th} dimension conditional on $\{i\} \times [\triangle(N)]^{\mathcal{M}}$ agrees with $P_\mu(\cdot|i)$ for each $i \in N$. Therefore,

$$\begin{aligned}\mathbb{E}\left[\max_{p\in A}\mathbb{E}\big[U\big|\mathcal{G}_\mu\big]\cdot p\right] &= \mathbb{E}\left[\max_{p\in A}\left[\sum_{i\in N} v^i P(\{i\}\times[\triangle(N)]^{\mathcal{M}}|\mathcal{G}_\mu)\right]\cdot p\right]\\ &= \int_{\triangle(N)}\max_{p\in A}\left[\sum_{i\in N} v^i\beta_i\right]\cdot p\,\pi_\mu(d\beta)\\ &= \int_{\mathcal{U}}\max_{p\in A} u(p)\,\mu(du)\end{aligned}$$

for each $A \in \mathcal{A}$. Hence by defining $\tilde{c}(\mathcal{G}_\mu) = c(\mu)$, V can be expressed as:

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

giving the desired CC representation.

4 Limited Contemplation Resources

In this section we consider an alternative model of costly contemplation in which the cost of contemplation does not directly affect the utility of the individual. Instead, the cost of contemplation enters indirectly when we constrain it to be below some bound k. Such a model may be appropriate in instances where the only cost of contemplation is time and the individual has a limited amount of time to devote to her decision.

Formally, we continue to model contemplation in the reduced from of a compact set of finite signed Borel measures $\mathcal{M}$ over the set of ex post utility functions $\mathcal{U}$, with the requirement $\mathcal{M}$ be consistent and minimal. Let $c : \mathcal{M} \to \mathbb{R}$ be a lower semi-continuous cost function and $k \in \mathbb{R}$ be the maximum allowable contemplation cost. A representation for limited contemplation resources then takes the form of a function $V : \mathcal{A} \to \mathbb{R}$ defined by

$$V(A) = \max_{\mu \in \mathcal{M}} \left[\int_{\mathcal{U}} \max_{p \in A} u(p)\, \mu(du) \right] \quad \text{subject to} \quad c(\mu) \leq k. \tag{8}$$

If we let $\mathcal{M}' = \{\mu \in \mathcal{M} : c(\mu) \leq k\}$, then this representation is equivalent to the following:

$$V(A) = \max_{\mu \in \mathcal{M}'} \left[\int_{\mathcal{U}} \max_{p \in A} u(p)\, \mu(du) \right].$$

Moreover, since c is lower semi-continuous, $\mathcal{M}'$ is also compact. Thus the limited contemplation resources representation in Equation (8) is equivalent to an RFCC representation with a zero cost function, $(\mathcal{M}', 0)$. Since the cost function in an RFCC representation is only unique up to an affine transformation, we see that a preference has a representation as in Equation (8) if and only if it has an RFCC representation $(\mathcal{M}', c')$ where c' is constant.

We now introduce an axiom that characterizes a constant cost of contemplation:

Axiom 7 (Strong IDD) *For any $A, B \in \mathcal{A}$, $p \in \triangle(Z)$, and $\alpha \in (0,1)$,*

$$A \succsim B \iff \alpha A + (1-\alpha)\{p\} \succsim \alpha B + (1-\alpha)\{p\}.$$

As the name suggests, strong IDD is a strengthening of IDD. For suppose

$$\alpha A + (1-\alpha)\{p\} \succsim \alpha B + (1-\alpha)\{p\}$$

for some $A, B \in \mathcal{A}$, $p \in \triangle(Z)$, and $\alpha \in (0,1)$. Strong IDD then implies $A \succsim B$, and applying strong IDD again, we have

$$\beta A + (1-\beta)\{q\} \succsim \beta B + (1-\beta)\{q\}$$

for any $q \in \triangle(Z)$ and $\beta \in (0,1)$. In contrast, IDD only guarantees that the above preference holds for $\beta = \alpha$. Thus, strong IDD implies an independence of degenerate decisions (IDD) and, in addition, independence of the weights on these degenerate decisions.[23]

For intuition, recall that the menu $\alpha A+(1-\alpha)\{p\}$ represents the decision problem in which the individual makes a contingent choice from A, this choice is implemented with probability α, and with probability $1-\alpha$ the individual instead receives p. We argued in Section 2.1 that as α decreases, the individual's benefit from contemplation decreases and hence she will choose a less costly contemplation strategy. However, if the cost of all contemplation strategies is the same, then her optimal contemplation strategy when choosing from the menu A will be the same as her optimal contemplation strategy when choosing from $\alpha A + (1-\alpha)\{p\}$ for any $\alpha \in (0,1)$. Therefore, if $A \succsim B$, then taking the convex combination of these menus with some singleton menu $\{p\}$ could affect the individual's utility through its effect on the final composition of lotteries, but it will not affect her optimal contemplation strategy for each of the respective menus. Hence, her ranking of the menus will not change.

The following theorem formalizes the connection between strong IDD and constant contemplation costs:

Theorem 4 *Suppose the preference $\succsim$ has an RFCC representation $(\mathcal{M}, c)$. Then, $\succsim$ satisfies strong IDD if and only if c is constant.*

Proof: See Appendix E.

Since strong IDD is a strengthening of IDD, the following corollary is immediate:

Corollary 2 *The preference $\succsim$ has an RFCC representation $(\mathcal{M}, c)$ in which c is constant if and only if it satisfies weak order, strong continuity, IR, ACP, and strong IDD.*

[23]Strong IDD is similar in spirit to the certainty independence axiom used by Gilboa and Schmeidler (1989) in the Anscombe-Aumann setting. In their axiom, arbitrary acts play the role of the menus A and B and a constant act plays the role of the singleton menu $\{p\}$. Our discussion of the relationship between strong IDD and IDD parallels the comparison of certainty independence and weak certainty independence found in Section 3.1 of Maccheroni, Marinacci, and Rustichini (2006).

5 Discussion

5.1 Relaxing the Independence Axiom

In this section, we discuss the relationship of our model to the additive EU representation of DLR (2001). We also present graphical intuition for how our axioms relax the independence axiom used in that paper.

Taking $\mathcal{U}$ as in Equation (5), an additive EU representation is a signed Borel measure μ on $\mathcal{U}$ such that $\succsim$ is represented by the following function:[24]

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

Hence, setting aside the singleton nontriviality requirement in the definition of the RFCC representation (see condition 3 in Definition 1), the additive EU representation is a special case of the RFCC representation in which $\mathcal{M} = \{\mu\}$. DLR (2001) and DLRS (2007) show that a preference has an additive EU representation if and only if it satisfies weak order, strong continuity, and the counterpart of the standard independence axiom adapted to sets:

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

$$A \succ B \implies \alpha A + (1-\alpha)C \succ \alpha B + (1-\alpha)C.$$

It is easily verified that under weak order and continuity, independence implies ACP, IR, and IDD. This is not surprising since the additive EU representation is a special case of the RFCC representation. Note also that under weak order and continuity, independence implies a form of indifference to contingent planning: For any $A, B \in \mathcal{A}$ and $\alpha \in [0, 1]$, $A \sim B$ implies $A \sim \alpha A + (1-\alpha)B$. Thus, it is necessary to relax independence in order to allow for costly contemplation with more than one contemplation strategy.

To illustrate how the combination of ACP, IR, and IDD is a relaxation of independence, consider preferences over menus of lotteries over two alternatives. That is, suppose $Z = \{z_1, z_2\}$. In this case, the set of lotteries over Z can be represented as the unit interval $[0, 1]$, with $p \in [0, 1]$ being the probability of alternative z_2. Under IR, we can restrict attention to convex menus since every menu is indifferent to its convex hull. By continuity, we may also restrict attention to closed menus. Closed and convex

[24] This definition of the additive EU representation differs slightly from the one given in DLR (2001). However, the two definitions are easily seen to be equivalent.

menus from $[0,1]$ are simply closed intervals, and hence we are considering preferences over menus of the form $[p,q] \subset [0,1]$ where $p,q \in [0,1]$.

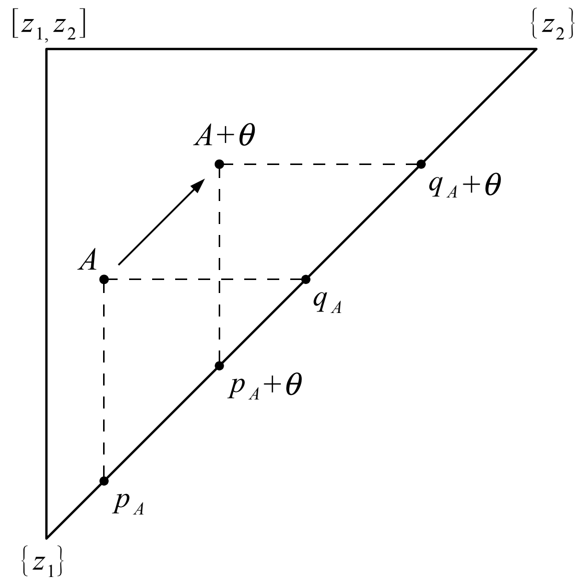

Figure 2: Representing Convex Menus

The set of all menus of this form is illustrated in Figure 2.[25] In this figure, the vertical distance between a point in the triangle and the bottom of the graph indicates the upper bound for the interval represented by that point. The horizontal distance between a point and the left side of the graph indicates the lower bound for the interval represented by that point. Therefore, the set of all singleton menus (i.e., menus of the form $\{p\} = [p,p]$) is represented by the diagonal of the triangle in this figure. The point labeled A in the figure indicates the menu $A = [p_A, q_A]$. Note that we abuse notation slightly and let z_1 denote the lottery the gives z_1 with probability 1, and likewise for z_2. Thus, $\{z_1\}$ corresponds to $\{0\}$, $\{z_2\}$ corresponds to $\{1\}$, and $[z_1, z_2]$ corresponds to $[0,1]$.

When the set of closed and convex menus is represented as in Figure 2, a convex combination of two menus corresponds to the convex combination of the points representing these menus. Therefore, the implication of ACP is simply that the lower contour sets for the preference are convex sets. Before illustrating the implications of IDD, we make a few observations about "translations" of menus. Consider the menu $A = [p_A, q_A]$ indicated in Figure 2, and take some real number θ. Adding the "translation" θ to the

[25] A similar depiction of menus of lotteries can be found in Olszewski (2007).

menu A yields a new menu $A+\theta = [p_A+\theta, q_A+\theta]$. Figure 2 illustrates that translating a menu results in a movement in a direction parallel to the diagonal of the triangle.

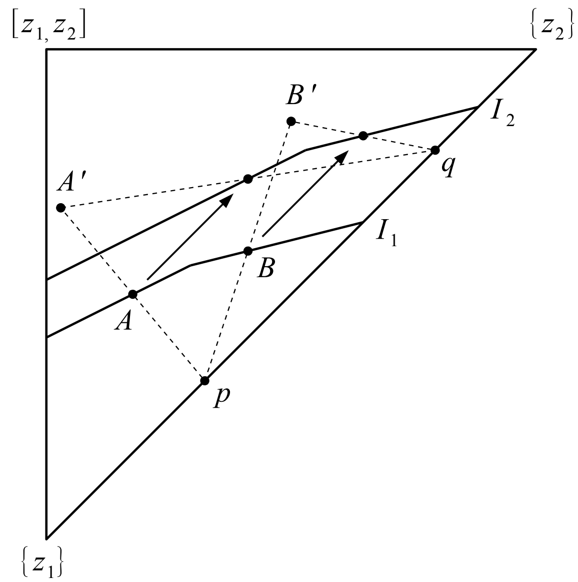

Figure 3: Translation Invariance

Figure 3 builds on these observations to show that IDD implies a type of translation invariance.[26] That is, we will show that if the individual is indifferent between two menus, then she is also indifferent between the new menus obtained by translating them both the same distance in a direction parallel to the diagonal of the triangle. Consider any two menus A and B such that $A \sim B$. Therefore, as illustrated in Figure 3, A and B both lie on the same indifference curve I_1. Note that in order for this preference to satisfy ACP, the lower contour sets of the preference must be convex, and hence the points above I_1 must be preferred to the points below I_1. Figure 3 illustrates that the menus A and B can be written as convex combinations of the singleton menu $\{p\}$ with the menus A' and B', respectively. That is, there exists $\alpha \in (0,1)$ such that $A = \alpha A' + (1-\alpha)\{p\}$ and $B = \alpha B' + (1-\alpha)\{p\}$. Fix any lottery q, and by IDD, we have

$$\alpha A' + (1-\alpha)\{p\} \sim \alpha B' + (1-\alpha)\{p\} \implies \alpha A' + (1-\alpha)\{q\} \sim \alpha B' + (1-\alpha)\{q\}.$$

Thus, the menus $\alpha A' + (1-\alpha)\{q\}$ and $\alpha B' + (1-\alpha)\{q\}$ must also be on the same indifference curve, which is indicated by I_2 in Figure 3. However, letting $\theta = (1-\alpha)(q-$

[26]This property is defined formally in Appendix B.1 and plays an important role in the proof of Theorem 1.

p), it is easily seen that $A + \theta = \alpha A' + (1-\alpha)\{q\}$ and $B + \theta = \alpha B' + (1-\alpha)\{q\}$. In other words, if the menus A and B are both translated by θ, then the individual remains indifferent between them. More generally, it can be show that IDD implies that when the same translation is applied to any two menus, the individual's ranking of these menus is not altered (see Lemma 6).

These figures illustrate that although ACP and IDD allow for "kinks" in indifference curves, these axioms restrict that lower contour sets are convex and indifference curves are translations of each other. Note that the kinks in the indifference curves in Figure 3 indicate a change in the optimal contemplation strategy, and our model allows for a possibly infinite number of kinks. In contrast, the independence axiom requires that indifference curves be linear and hence does not allow for such kinks.

5.2 Infinite Regress

We conclude by making an observation about the infinite regress issue. The infinite regress problem of bounded rationality can be informally explained as follows (see, e.g., Conlisk (1996)): Consider an abstract decision problem D. The standard rational economic individual is typically assumed to solve the problem D optimally without any constraints, no matter how difficult the problem might be. One may be tempted to make the model more realistic by explicitly taking account of the costs of solving it. This leads to a new optimization problem $F(D)$, the problem that incorporates into D the costs of solving D. However, typically $F(D)$ itself is a more difficult problem than D. So if one would like to have an even more realistic model, why not include the cost of solving $F(D)$ explicitly? The latter leads to the new decision problem $F^2(D) = F(F(D))$. This argument can be iterated ad infinitum. The fact that it is not clear at which level $F^n(D)$ one should stop, and how to stop if one stops at any level, corresponds to the infinite regress problem.

The representation result in this paper may be seen as giving an *as if* solution to the infinite regress problem. To the extent that one finds ACP and IDD as convincing behavioral aspects of bounded rationality arising from contemplation costs, there is no loss of generality from restricting attention to $F^1(D)$, the case where the decision maker optimally solves the problem of learning her preferences subject to costs.

Appendix

A Mathematical Preliminaries

In this section we establish 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. After presenting some intermediate results, we describe this duality in Section A.2.

Suppose X is a real Banach space. We now introduce the standard definition of the subdifferential of a function.

Definition 3 Suppose $C \subset X$ is convex 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 important in the approximation of a convex function by affine functions. In fact, it is straightforward to show that $x^* \in \partial f(x)$ if and only if the affine function $h(y) \equiv f(x) + \langle y - x, x^* \rangle$ 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, the following results show that under certain continuity assumptions on f, the subdifferential is always nonempty. For a convex subset C of X, a function $f : C \to \mathbb{R}$ is said to be *Lipschitz continuous* if there is some real number K such that for every $x, y \in C$, $|f(x) - f(y)| \leq K\|x - y\|$. The number K is called a *Lipschitz constant* of f.

Lemma 1 *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: We begin by introducing the standard definition of the epigraph of a function $f : C \to \mathbb{R}$:

$$epi(f) = \{(x, t) \in C \times \mathbb{R} : t \geq f(x)\}.$$

Note that $epi(f) \subset X \times \mathbb{R}$ 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 $epi(f) \cap H(x) = \emptyset$. To see this, note first that

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

Now, suppose $(y,t) \in epi(f)$, 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 $epi(f)$ are disjoint convex sets. Therefore, a version of the Separating Hyperplane Theorem implies there exists a nonzero continuous linear functional $(x^*, \lambda) \in X^* \times \mathbb{R}$ that separates $H(x)$ and $epi(f)$.[27] That is, there exists a scalar δ such that

$$\langle y, x^* \rangle + \lambda t \leq \delta \quad \text{if} \quad (y,t) \in epi(f) \tag{9}$$

and

$$\langle y, x^* \rangle + \lambda t \geq \delta \quad \text{if} \quad (y,t) \in H(x). \tag{10}$$

Clearly, we cannot have $\lambda > 0$. Also, if $\lambda = 0$, then Equation (10) 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) by dividing by $|\lambda|$.

Since $(x, f(x)) \in epi(f)$, we have $\langle x, x^* \rangle - f(x) \leq \delta$. For all $t > 0$, we have $(x, f(x) - t) \in H(x)$, which implies $\langle x, x^* \rangle - f(x) + t \geq \delta$. Therefore, $\langle x, x^* \rangle - f(x) = \delta$, and thus for all $y \in C$,

$$\langle y, x^* \rangle - f(y) \leq \delta = \langle x, x^* \rangle - f(x).$$

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

The following simple lemma will also be useful.

Lemma 2 *Let $K \geq 0$ and let $\{x_d\}_{d \in D} \subset X$ and $\{x_d^*\}_{d \in D} \subset X^*$ be nets such that (i) $\|x_d^*\| \leq K$ for all $d \in D$, and (ii) $x_d \to x$ and $x_d^* \xrightarrow{w^*} x^*$ for some $x \in X$ and $x^* \in X^*$. Then, $\langle x_d, x_d^* \rangle \to \langle x, x^* \rangle$.*

[27] See Aliprantis and Border (1999, Theorem 5.50) or Luenberger (1969, p133).

Proof: We have

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

so that $\langle x_d, x_d^*\rangle \to \langle x, x^*\rangle$. ■

In the next Lemma, we assume that X is a Banach lattice.[28] 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 $x \in X_+$.

Lemma 3 *Suppose C is a convex subset of a Banach lattice X such that for any $x, x' \in C$, $x \vee x' \in C$. If $f : C \to \mathbb{R}$ is Lipschitz continuous, convex, and monotone and if $K \geq 0$ is a Lipschitz constant of f, then for all $x \in C$ there exists a positive $x^* \in \partial f(x)$ with $\|x^*\| \leq K$.*

Proof: Let $epi(f)$, H, and $H(x)$ be as defined in the proof of Lemma 1. 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 non-empty interior since it contains the nonempty open set $H(x)$.

Let $x \in C$ be arbitrary. We claim that $epi(f) \cap I(x) = \emptyset$. Suppose for a contradiction that $(x', s) \in epi(f) \cap I(x)$. Then $x' \in C$, and there exist $y \in X$, $z \in X_+$ such that $x' = x + y + z$, and $s - f(x) < -K\|y\|$. 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 $f(\bar{x}) \geq f(x)$. We therefore have $x' = \bar{x} + \bar{y}$ and $s - f(\bar{x}) \leq s - f(x) < -K\|y\| \leq -K\|\bar{y}\|$. Hence $(x', s) \in H(\bar{x})$, a contradiction to $epi(f) \cap H(\bar{x}) = \emptyset$.

We showed that $I(x)$ and $epi(W)$ 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 1 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

$$\langle y, x^*\rangle + \lambda t \leq \delta \quad \text{if} \quad (y, t) \in epi(f) \tag{11}$$

[28] See Aliprantis and Border (1999, page 302) for a definition of Banach lattices.

and

$$\langle y, x^* \rangle + \lambda t \geq \delta \quad \text{if} \quad (y,t) \in I(x). \tag{12}$$

Note that Equation (11) is the same as Equation (9), and Equation (12) implies Equation (10). Therefore by the exact same arguments as in the proof of Lemma 1, 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 (12),

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

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. ■

A.1 Variation of the Mazur Density Theorem

The Mazur density theorem is a classic result from convex analysis. It states 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 $\partial f(x)$ is a singleton is a dense G_δ set in C.[29] The notation G_δ indicates that a set is the countable intersection of open sets.

We wish to obtain a variation of this theorem by relaxing the assumption that C has a nonempty interior. However, it can be shown that the conclusion of the theorem does not hold for arbitrary convex sets. We will therefore require that the affine hull of C, defined below, is dense in X.

Definition 4 The *affine hull* of a set $C \subset X$, denoted *aff*(C), is defined to be the smallest affine subspace of X that contains C.

That is, the affine hull of C is defined by $x + span(C - C)$ for any fixed $x \in C$. If C is convex, then it is straightforward to show that

$$aff(C) = \{\lambda x + (1 - \lambda) y : x, y \in C \text{ and } \lambda \in \mathbb{R}\}. \tag{13}$$

Intuitively, if we were to draw a line through any two points of C, then that entire line would necessarily be included in any affine subspace that contains C.

We are now ready to state our variation of Mazur's theorem. Essentially, we are able to relax the assumption that C has a nonempty interior and instead assume that *aff*(C) is dense in X if we also replace the continuity assumption with the more restrictive assumption of Lipschitz continuity.

[29] 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).

Proposition 1 *Suppose X is a separable Banach space and C is a closed and convex subset of X containing the origin, and suppose aff(C) is dense in X. If $f : C \to \mathbb{R}$ is Lipschitz continuous and convex, then the set of points x where $\partial f(x)$ is a singleton is a dense G_δ (in the relative topology) set in C.*

Proof: This proof is a variation of the proof of Mazur's theorem found in Phelps (1993). Since any subset of a separable Banach space is separable, *aff*(C) is separable. Let $\{x_n\} \subset$ *aff*(C) be a sequence which is dense in *aff*(C), and hence, by the density of *aff*(C) in X, also dense in X. Let K be a Lipschitz constant of f. For each $m, n \in \mathbb{N}$, let $A_{m,n}$ denote the set of $x \in C$ for which there exist $x^*, y^* \in \partial f(x)$ such that $\|x^*\|, \|y^*\| \leq 2K$ and

$$\langle x_n, x^* - y^* \rangle \geq \tfrac{1}{m}.$$

We claim that if $\partial f(x)$ is not a singleton for $x \in C$, then $x \in A_{m,n}$ for some $m, n \in \mathbb{N}$. By Lemma 1, for all $x \in C$, $\partial f(x) \neq \emptyset$. Therefore, if $\partial f(x)$ is not a singleton, then there exist $x^*, y^* \in \partial f(x)$ such that $x^* \neq y^*$. This does not tell us anything about the norm of x^* or y^*, but by Lemma 1, there exists $z^* \in \partial f(x)$ such that $\|z^*\| \leq K$. Either $z^* \neq x^*$ or $z^* \neq y^*$, so it is without loss of generality that we assume the former. It is straightforward to verify that the subdifferential is convex. Therefore, for all $\lambda \in (0,1)$, $\lambda x^* + (1-\lambda)z^* \in \partial f(x)$, and

$$\|\lambda x^* + (1-\lambda)z^*\| \leq \|z^*\| + \lambda\|x^* - z^*\| \leq 2K$$

for λ sufficiently small. For some such λ, let $w^* = \lambda x^* + (1-\lambda)z^*$. Then, $w^* \neq z^*$ and $\|w^*\| \leq 2K$. Since $w^* \neq z^*$, there exists $y \in X$ such that $\langle y, w^* - z^* \rangle 0$. By the continuity of $w^* - z^*$, there exists a neighborhood N of y such that for all $z \in N$, $\langle z, w^* - z^* \rangle > 0$. Since $\{x_n\}$ is dense in X, there exists $n \in \mathbb{N}$ such that $x_n \in N$. Thus $\langle x_n, w^* - z^* \rangle > 0$, and hence there exists $m \in \mathbb{N}$ such that $\langle x_n, w^* - z^* \rangle > \frac{1}{m}$. Therefore, $x \in A_{m,n}$.

We have just shown that the set of $x \in C$ for which $\partial f(x)$ is a singleton is $\bigcap_{m,n}(C \setminus A_{m,n})$. It remains only show that for each $m, n \in \mathbb{N}$, $C \setminus A_{m,n}$ is open (in the relative topology) and dense in C. Then, we can appeal to the Baire category theorem.

We first show that each $A_{m,n}$ is relatively closed. If $A_{m,n} = \emptyset$, then $A_{m,n}$ is obviously closed, so suppose otherwise. Consider any sequence $\{z_k\} \subset A_{m,n}$ such that $z_k \to z$ for some $z \in C$. We will show that $z \in A_{m,n}$. For each k, choose $x_k^*, y_k^* \in \partial f(z_k)$ such that $\|x_k^*\|, \|y_k^*\| \leq 2K$ and $\langle x_n, x_k^* - y_k^* \rangle \geq \frac{1}{m}$. Since X is separable, the closed unit ball of X^* endowed with the weak* topology is metrizable and compact, which implies any sequence in this ball has a weak*-convergent subsequence.[30] Therefore, the closed ball of radius $2K$ around the origin of X^* has this same property. Thus, without loss of generality, we can assume there exist $x^*, y^* \in X^*$ with $\|x^*\|, \|y^*\| \leq 2K$ such that $x_k^* \xrightarrow{w^*} x^*$ and $y_k^* \xrightarrow{w^*} y^*$. Therefore, for any

[30] For metrizability, see Aliprantis and Border (1999, Theorem 6.34). Compactness follows from Alaoglu's theorem; see Aliprantis and Border (1999, Theorem 6.25). Note that compactness only guarantees that every net has a convergent subnet, but compactness and metrizability together imply that every sequence has a convergent subsequence.

$y \in C$, we have

$$\langle y - z, x^* \rangle = \lim_{k\to\infty} \langle y - z_k, x_k^* \rangle \leq \lim_{k\to\infty} \big[f(y) - f(z_k) \big] = f(y) - f(z).$$

The first equality follows from Lemma 2, the inequality from the definition of the subdifferential, and the last equality from the continuity of f. Therefore, $x^* \in \partial f(z)$. A similar argument shows $y^* \in \partial f(z)$. Finally, since

$$\langle x_n, x^* - y^* \rangle = \lim_{k\to\infty} \langle x_n, x_k^* - y_k^* \rangle \geq \tfrac{1}{m},$$

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

We now need to show that $C \setminus A_{m,n}$ is dense in C for each $m, n \in \mathbb{N}$. Consider arbitrary $m, n \in \mathbb{N}$ and $z \in C$. We will find a sequence $\{z_k\} \subset C \setminus A_{m,n}$ such that $z_k \to z$. Since C contains the origin, $\mathit{aff}(C)$ is a subspace of X. Hence, $z + x_n \in \mathit{aff}(C)$, so Equation (13) implies 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:

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

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.

Consider any $k \in \mathbb{N}$. Because C is convex, we have $y_k + t x_n \in C$ for all $t \in (0, \frac{a_k\lambda}{\lambda-1})$. Define a function $g : (0, \frac{a_k\lambda}{\lambda-1}) \to \mathbb{R}$ by $g(t) = f(y_k + t x_n)$, 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.[31] Let t_k be any $t \in (0, \frac{a_k\lambda}{\lambda-1})$ at which $g'(t)$ 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 a subdifferential to 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 $\langle x_n, x^* \rangle = \langle x_n, y^* \rangle$, and hence $z_k \in C \setminus A_{m,n}$. Finally, note that since $0 < t_k < \frac{a_k\lambda}{\lambda-1}$ and $a_k \to 0$, we have $t_k \to 0$. Therefore, $z_k = y_k + t_k x_n \to z$.

We did restrict attention above 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 in the preceding paragraph, for any $k \in \mathbb{N}$, define a function $g : (0, \frac{1}{k}) \to \mathbb{R}$ by

[31] See Phelps (1993, Theorem 1.16).

$g(t) = f(z + tx_n)$. Let t_k be any $t \in (0, \frac{1}{k})$ at which $g'(t)$ exists, and let $z_k = z + t_k x_n$. Then, as argued above, $z_k \in C \setminus A_{m,n}$ for all $k \in \mathbb{N}$ and $z_k \to z$.

We have now proved that for each $m, n \in \mathbb{N}$, $C \setminus A_{m,n}$ is open (in the relative topology) and dense in C. Since C is a closed subset of a Banach space, it is a Baire space, which implies every countable intersection of (relatively) open dense subsets of C is also dense.[32] This completes the proof. ■

A.2 Fenchel-Moreau Duality

Let X continue to denote a real Banach space. We now introduce the definition of the conjugate of a function.

Definition 5 Suppose $C \subset X$ is convex 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

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

There is an important duality between f and f^*:[33]

Lemma 4 *Suppose $C \subset X$ is convex and $f : C \to \mathbb{R}$. Then,*

1. *f^* is lower semicontinuous 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 semicontinuous.

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

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

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

[32] See Theorems 3.34 and 3.35 of Aliprantis and Border (1999).

[33] For more on this relationship, see Ekeland and Turnbull (1983) or Holmes (1975). A finite-dimensional treatment can be found in Rockafellar (1970).

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

$$\langle y, x^* \rangle - f(y) \leq \langle x, x^* \rangle - f(x). \tag{14}$$

for all $y \in C$. By the definition of the conjugate, Equation (14) holds if and only if $f^*(x^*) = \langle x, x^* \rangle - f(x)$, which is equivalent to $f(x) = \langle x, x^* \rangle - f^*(x^*)$. ■

For the remainder of this section, assume that $C \subset X$ is convex and $f : C \to \mathbb{R}$ is Lipschitz continuous and convex. Then, Lemma 1 implies that $\partial f(x) \neq \emptyset$ for all $x \in C$. Therefore, by parts 2 and 3 of Lemma 4, we have

$$f(x) = \max_{x^* \in X^*} \left[\langle x, x^* \rangle - f^*(x^*) \right] \tag{15}$$

for all $x \in C$. We have just proved a slight variation of the classic Fenchel-Moreau theorem.[34]

We now show that under the assumptions of Proposition 1, there is a minimal compact subset of X^* for which Equation (15) holds. Let C_f denote the set of all $x \in C$ for which the subdifferential of f at x is a singleton:

$$C_f = \{x \in C : \partial f(x) \text{ is a singleton}\}. \tag{16}$$

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

$$\mathcal{N}_f = \{x^* \in X^* : x^* \in \partial f(x), x \in C_f\}. \tag{17}$$

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

$$\mathcal{M}_f = \overline{\mathcal{N}_f}. \tag{18}$$

Proposition 2 *Suppose X, C, and f satisfy the assumptions of Proposition 1. That is, suppose (i) X is a separable Banach space, (ii) C is a closed and convex subset of X containing the origin such that aff(C) is dense in X, and (iii) $f : C \to \mathbb{R}$ is Lipschitz continuous and convex. Then, $\mathcal{M}_f$ is weak* compact, and for any weak* compact $\mathcal{M} \subset X^*$,*

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

Proof: If $K \geq 0$ is a Lipschitz constant of f, then Lemma 1 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 (see Aliprantis and Border, 1999, Theorem 6.25).

[34]The standard version of this theorem states that if $f : X \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous 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).

($\Rightarrow$): Let $x \in C$ be arbitrary. By Proposition 1, C_f is dense in C, so there exists a net $\{x_d\}_{d\in D} \subset C_f$ such that $x_d \to x$. For all $d \in D$, take $x_d^* \in \partial f(x_d)$, and we have $\{x_d^*\}_{d\in D} \subset \mathcal{M}_f$ by the definition of $\mathcal{M}_f$. Since $\mathcal{M}_f$ is weak* compact, every net in $\mathcal{M}_f$ 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 Lemma 2, the definition of the subdifferential, and the continuity of f, for any $y \in C$,

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

which implies $x^* \in \partial f(x)$. 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 4, we conclude that for all $x \in C$,

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

($\Longleftarrow$): First 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 semicontinuous by part 1 of Lemma 4. Therefore, $x^* \mapsto \langle x, x^* \rangle - f^*(x^*)$ is weak* upper semicontinuous and hence attains a maximum on any weak* compact set.

Fix any $x \in C_f$. By the above, 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 4. 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 $\mathcal{M}_f = \overline{\mathcal{N}_f} \subset \mathcal{M}$. ■

B Proof of Theorem 1

The necessity of the axioms in Theorem 1 is straightforward and left to the reader. For the sufficiency direction, let $\mathcal{A}^c \subset \mathcal{A}$ denote the collection of all convex menus. In both parts A and B of Theorem 1, $\succsim$ satisfies IR. In part A, IR is directly assumed whereas in part B it is implied by weak order, continuity, monotonicity, and ACP (see Lemma 5). Therefore for all $A \in \mathcal{A}$, $A \sim co(A) \in \mathcal{A}^c$. Note that for any $u \in \mathcal{U}$, we have

$$\max_{p \in A} u \cdot p = \max_{p \in co(A)} u \cdot p.$$

Thus if we establish the representations in Theorem 1 for convex menus and then apply the same functional form to all of $\mathcal{A}$, then by IR the resulting function represents $\succsim$ on $\mathcal{A}$. Note also that $\mathcal{A}$ is a compact metric space since $\triangle(Z)$ is a compact metric space (see, e.g., Munkres (2000, p279)). 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.

We make some preliminary observations regarding our axioms in Section B.1. We then construct a function V with certain desirable properties in Section B.2. Finally, in Section B.3, we apply the duality results from Appendix A to complete the representation theorem.

B.1 Preliminary Observations

In this section we establish a number of simple implications of the axioms introduced in the text. These results will be useful in subsequent sections.

Lemma 5 *If $\succsim$ satisfies weak order, ACP, monotonicity, and continuity, then it also satisfies IR.*

Proof: Let $A \in \mathcal{A}$. Monotonicity implies that $co(A) \succsim A$, and hence we only need to prove that $A \succsim co(A)$. Let us inductively define a sequence of sets via $A_0 = A$ and $A_k = \frac{1}{2}A_{k-1} + \frac{1}{2}A_{k-1}$ for $k \geq 1$. ACP implies that $A_{k-1} \succsim A_k$, and therefore by transitivity $A \succsim A_k$ for any k. It is straightforward to verify that $d_h(A_k, co(A)) \to 0$, so we have $A \succsim co(A)$ by continuity. ■

For proving our representation theorem, it will be useful to derive an alternative formulation of our IDD axiom. Before introducing this new axiom, we define the set of *translations* to be

$$\Theta \equiv \left\{ \theta \in \mathbb{R}^Z : \sum_{z \in Z} \theta_z = 0 \right\}. \tag{19}$$

Any $\theta \in \Theta$ can be thought of as a signed measure on Z such that $\theta(Z) = 0$. 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 $p - q \in \Theta$. We now give a formulation of IDD in terms of translations.

Axiom 9 (Translation Invariance (TI)) *For any $A, B \in \mathcal{A}$ and $\theta \in \Theta$ such that $A+\theta, B+\theta \in \mathcal{A}$,*

$$A \succsim B \implies A + \theta \succsim B + \theta.$$ [35]

Lemma 6 *The preference $\succsim$ satisfies IDD if and only if it satisfies TI.*

Proof: To see that TI implies IDD, assume that $A, B \in \mathcal{A}$, $p, q \in \triangle(Z)$ are such that $\lambda A + (1-\lambda)\{q\} \succsim \lambda B + (1-\lambda)\{q\}$. Let $A' = \lambda A + (1-\lambda)\{q\}$, $B' = \lambda B + (1-\lambda)\{q\}$ and $\theta = (1-\lambda)(p-q)$. Note that $\theta \in \Theta$, $A' + \theta = \lambda A + (1-\lambda)\{p\} \in \mathcal{A}$, and $B' + \theta = \lambda A + (1-\lambda)\{p\} \in \mathcal{A}$. Hence by TI, $\lambda A + (1-\lambda)\{p\} \succsim \lambda B + (1-\lambda)\{p\}$.

To see that IDD implies TI, assume that $A, B \in \mathcal{A}$ and $\theta \in \Theta$ are such that $A+\theta, B+\theta \in \mathcal{A}$ and $A \succsim B$. If $\theta = 0$, the conclusion of TI holds trivially, so assume that $\theta \neq 0$. Let $Z^- = \{z \in Z : \theta_z < 0\}$. Define θ^+, $\theta^- \in R^Z$ by $\theta^+_z = \max\{0, \theta_z\}$ and $\theta^-_z = \max\{0, -\theta_z\}$ for any $z \in Z$. Then let $\kappa \equiv \sum_{z \in Z} \theta^+_z = \sum_{z \in Z} \theta^-_z > 0$.

[35] Note that TI implies its converse, for suppose $A + \theta \succsim B + \theta$. Then, by TI, $A = (A + \theta) + (-\theta) \succsim (B + \theta) + (-\theta) = B$.

We will first show that for any $r \in A \cup B$,

$$0 \le r_z - \theta_z^- \le 1 - \kappa \;\text{ for all } z \in Z. \tag{20}$$

Note that for any $z \in Z^-$, $r_z - \theta_z^- = r_z + \theta_z \ge 0$ since $r + \theta \in \triangle(Z)$. Note also that if $z \notin Z^-$ then $r_z - \theta_z^- = r_z \ge 0$ since $\theta_z^- = 0$. So for any $z \in Z$,

$$0 \le r_z - \theta_z^- \le \left(1 - \sum_{z' \in Z^- \setminus \{z\}} r_{z'}\right) - \theta_z^- \le \left(1 - \sum_{z' \in Z^- \setminus \{z\}} \theta_{z'}^-\right) - \theta_z^- = 1 - \kappa,$$

establishing Equation (20). Therefore, since $\theta \neq 0$, we have $0 < \kappa \le 1$. Then, $p \equiv \frac{1}{\kappa}\theta^+$, $q \equiv \frac{1}{\kappa}\theta^-$ are in $\triangle(Z)$, and $\theta = \kappa(p - q)$. There are two cases to consider:

First consider the case of $\kappa < 1$. Define subsets A' and B' of $\mathbb{R}^Z$ by

$$A' \equiv \{r' \in \mathbb{R}^Z : r' = \tfrac{1}{1-\kappa}(r - \theta^-) \text{ for some } r \in A\},$$

$$B' \equiv \{r' \in \mathbb{R}^Z : r' = \tfrac{1}{1-\kappa}(r - \theta^-) \text{ for some } r \in B\}.$$

By Equation (20) and the definition of κ, we have that $A', B' \in \mathcal{A}$ and

$$(1-\kappa)A' + \kappa\{q\} = A \succsim B = (1-\kappa)B' + \kappa\{q\}. \tag{21}$$

Next consider the $\kappa = 1$ case. By Equation (20) we have $r = \theta^- = q$ for any $r \in A \cup B$. Therefore $A = B = \{q\}$, and hence Equation (21) holds for any choice of $A', B' \in \mathcal{A}$.

Since Equation (21) holds in each of the two cases above, we conclude by IDD that

$$A + \theta = (1-\kappa)A' + \kappa\{p\} \succsim (1-\kappa)B' + \kappa\{p\} = B + \theta.$$

Therefore TI is satisfied. ■

In light of Lemma 6, we will use IDD and TI interchangeably. We now present one useful consequence of translation invariance.

Lemma 7 *Suppose $\succsim$ satisfies weak order, continuity, and TI. If $A \in \mathcal{A}$, $\theta \in \Theta$, and $\alpha \in (0,1)$ are such that $A + \theta \in \mathcal{A}$, then*

$$A \succsim A + \theta \iff A \succsim A + \alpha\theta \iff A + \alpha\theta \succsim A + \theta. \tag{22}$$

Proof: We will make a simple induction argument. Suppose

$$A + \tfrac{m-1}{n}\theta \succsim A + \tfrac{m}{n}\theta$$

for some $m, n \in \mathbb{N}$ with $m < n$. Then adding $\frac{1}{n}\theta$ to each side of the above and applying TI yields

$$A + \tfrac{m}{n}\theta \succsim A + \tfrac{m+1}{n}\theta.$$

Now suppose that $A \succsim A + \frac{1}{n}\theta$. Then, using induction and the transitivity of $\succsim$, we obtain the following:

$$A \succsim A + \tfrac{1}{n}\theta \succsim \cdots \succsim A + \big(1 - \tfrac{1}{n}\big)\theta \succsim A + \theta. \tag{23}$$

A similar line of reasoning shows that if $A \prec A + \frac{1}{n}\theta$, then we obtain the following:

$$A \prec A + \tfrac{1}{n}\theta \prec \cdots \prec A + \big(1 - \tfrac{1}{n}\big)\theta \prec A + \theta. \tag{24}$$

In sum, Equations (23) and (24) imply that for any $m, n \in \mathbb{N}$, $1 \leq m < n$, we have

$$A \succsim A + \tfrac{1}{n}\theta \iff A \succsim A + \theta \iff A \succsim A + \tfrac{m}{n}\theta \iff A + \tfrac{m}{n}\theta \succsim A + \theta.$$

This establishes Equation (22) for $\alpha \in (0, 1) \cap \mathbb{Q}$. The continuity of $\succsim$ implies that Equation (22) holds for all $\alpha \in (0, 1)$. ■

Although we do not assume that independence holds on $\mathcal{A}$, our other axioms imply that independence does hold for singleton menus.

Axiom 10 (Singleton Independence) *For all $p, q, r \in \triangle(Z)$ and $\lambda \in (0, 1)$,*

$$\{p\} \succsim \{q\} \iff \lambda\{p\} + (1-\lambda)\{r\} \succsim \lambda\{q\} + (1-\lambda)\{r\}.$$

Lemma 8 *If $\succsim$ satisfies weak order, continuity, and TI, then it also satisfies singleton independence.*

Proof: Let $\theta = q - p$ and $\theta' = (1-\lambda)(r - p)$. Then

$$\begin{aligned}\{p\} \succsim \{q\} = \{p\} + \theta &\iff \{p\} \succsim \{p\} + \lambda\theta = (1-\lambda)\{p\} + \lambda\{q\} \\ &\iff \lambda\{p\} + (1-\lambda)\{r\} \succsim \lambda\{q\} + (1-\lambda)\{r\},\end{aligned}$$

where the first equivalence follows from Lemma 7, and the second equivalence follows from TI, $\{p\} + \theta' = \lambda\{p\} + (1-\lambda)\{r\}$, and $(1-\lambda)\{p\} + \lambda\{q\} + \theta' = \lambda\{q\} + (1-\lambda)\{r\}$. Therefore singleton independence is satisfied. ■

Before proceeding, we define the following important subset of $\mathcal{A}^c$:

$$\mathcal{A}^\circ \equiv \{A \in \mathcal{A}^c : \forall \theta \in \Theta \; \exists \alpha > 0 \text{ such that } A + \alpha\theta \in \mathcal{A}^c\}. \tag{25}$$

Thus $\mathcal{A}^\circ$ contains menus that can be translated at least a "little bit" in the direction of any vector in Θ. It is easily verified that $\mathcal{A}^\circ$ is convex. In addition, the following result gives an alternative characterization of $\mathcal{A}^\circ$ along with some other important properties.

Lemma 9 *The set $\mathcal{A}^\circ$ has the following properties:*

1. $\mathcal{A}^\circ = \{A \in \mathcal{A}^c : \exists \varepsilon > 0$ *such that* $\forall p \in A, \forall z \in Z, p_z \geq \varepsilon\}$.

2. *Suppose* $p \in \triangle(Z)$ *is such that* $p_z > 0$ *for all* $z \in Z$*. Then for any* $A \in \mathcal{A}^c$ *and* $\lambda \in [0,1)$, $\lambda A + (1-\lambda)\{p\} \in \mathcal{A}^\circ$.

3. $\mathcal{A}^\circ$ *is dense in* $\mathcal{A}^c$.

Proof: (1): Let $\hat{\mathcal{A}}^\circ \equiv \{A \in \mathcal{A}^c : \exists \varepsilon > 0$ such that $\forall p \in A, \forall z \in Z, p_z \geq \varepsilon\}$. To see that $\hat{\mathcal{A}}^\circ \subset \mathcal{A}^\circ$, take any $A \in \hat{\mathcal{A}}^\circ$ and $\theta \in \Theta$. Let $\varepsilon > 0$ be such that $p_z \geq \varepsilon$ for all $p \in A$ and $z \in Z$. Choose $\alpha > 0$ sufficiently small to ensure that $\alpha \cdot \max_{z\in Z}|\theta_z| \leq \varepsilon$. Then $p_z + \alpha\theta_z \geq p_z - \varepsilon \geq 0$ for all $p \in A$ and $z \in Z$, so $A + \alpha\theta \in \mathcal{A}^c$. Thus $A \in \mathcal{A}^\circ$.

To see that $\mathcal{A}^\circ \subset \hat{\mathcal{A}}^\circ$, take any $A \in \mathcal{A}^\circ$. Fix any $z \in Z$, and take any $\theta \in \Theta$ such that $\theta_z = -1$. Then let $\alpha_z > 0$ be such that $A + \alpha_z\theta \in \mathcal{A}^c$, so for any $p \in A$, $p_z + \alpha_z\theta = p_z - \alpha_z \geq 0$. We obtain such an $\alpha_z > 0$ for every $z \in Z$, so let $\varepsilon \equiv \min_{z\in Z}\alpha_z > 0$. Then for any $p \in A$ and $z \in Z$, $p_z \geq \alpha_z \geq \varepsilon$, so $A \in \hat{\mathcal{A}}^\circ$.

(2): Let $\varepsilon \equiv (1-\lambda)(\min_{z\in Z} p_z) > 0$. Then for any $q \in A$ and $z \in Z$, $\lambda q_z + (1-\lambda)p_z \geq \varepsilon$. Thus $\lambda A + (1-\lambda)\{p\} \in \mathcal{A}^\circ$ by part 1.

(3): It is easily verified that for any $A \in \mathcal{A}^c$, $(1-1/n)A + (1/n)\{p\} \to A$ as $n \to \infty$. Hence $\mathcal{A}^\circ$ is dense in $\mathcal{A}^c$ by part 2. ■

Take p^* and p_* from the L–continuity axiom, and let $\theta^* \equiv p^* - p_*$. We will utilize θ^* a great deal in the construction of our representation, and the following is an important property of θ^*.

Lemma 10 *Suppose* $\succsim$ *satisfies weak order, strong continuity, and TI, and take* $\theta^* = p^* - p_*$*. Let* $A, B \in \mathcal{A}^\circ$ *and* $\alpha, \beta \in \mathbb{R}$ *be such that* $A \sim B$ *and* $A + \alpha\theta^*, B + \beta\theta^* \in \mathcal{A}^c$*. Then,*

$$A + \alpha\theta^* \succsim B + \beta\theta^* \iff \alpha \geq \beta. \tag{26}$$

Proof: We will first show that for any $A \in \mathcal{A}^\circ$, there exists $\gamma > 0$ such that $A + \gamma\theta^* \in \mathcal{A}^c$ and $A + \gamma\theta^* \succ A$. To see this, fix any $A \in \mathcal{A}^\circ$. It follows from part 1 of Lemma 9 that there exist $A' \in \mathcal{A}^c$ and $\gamma \in (0,1)$ such that $A = (1-\gamma)A' + \gamma\{p_*\}$.[36] By L–continuity we have

$$A + \gamma\theta^* = (1-\gamma)A' + \gamma\{p^*\} \succ (1-\gamma)A' + \gamma\{p_*\} = A.$$

Therefore, for any $A \in \mathcal{A}^\circ$ and $\alpha > 0$ such that $A + \alpha\theta^* \in \mathcal{A}^c$, take $\gamma > 0$ such that $A + \gamma\theta^* \in \mathcal{A}^c$ and $A + \gamma\theta^* \succ A$. Applying Lemma 7 to A and $\theta = \max\{\gamma, \alpha\}\theta^*$, we have $A + \alpha\theta^* \succ A$. A similar argument shows that if $A \in \mathcal{A}^\circ$ and $\alpha < 0$ are such that $A + \alpha\theta^* \in \mathcal{A}^c$, then $A \succ A + \alpha\theta^*$.

[36]Take $\varepsilon > 0$ as in Lemma 9, and let $\gamma \equiv \varepsilon$ and $A' \equiv \{q \in \mathbb{R}^Z : q = \frac{1}{1-\gamma}(p - \gamma p_*)$ for some $p \in A\}$. It follows that $A' \in \mathcal{A}^c$ and $A = (1-\gamma)A' + \gamma\{p_*\}$.

Now, let $A, B \in \mathcal{A}^\circ$ and $\alpha, \beta \in \mathbb{R}$ be such that $A \sim B$ and $A + \alpha\theta^*, B + \beta\theta^* \in \mathcal{A}^c$. We prove the equivalence from Equation (26) by considering three cases:

If $\alpha = \beta$ then $\alpha\theta^* = \beta\theta^*$. Hence by TI, $A + \alpha\theta^* \sim B + \beta\theta^*$.

If $\alpha > \beta$, there are three sub-cases to consider. First consider $\alpha > \beta \geq 0$, which implies $0 < \alpha - \beta \leq \alpha$ and hence $A + (\alpha - \beta)\theta^* \in \mathcal{A}^c$. From the above arguments $A + (\alpha - \beta)\theta^* \succ A \sim B$, so by TI, $A + \alpha\theta^* = [A + (\alpha - \beta)\theta^*] + \beta\theta^* \succ B + \beta\theta^*$. Similarly, if $0 \geq \alpha > \beta$, then $\beta \leq \beta - \alpha < 0$ and hence $B + (\beta - \alpha)\theta^* \in \mathcal{A}^c$. From the above arguments $A \sim B \succ B + (\beta - \alpha)\theta^*$, which implies by TI that $A + \alpha\theta^* \succ [B + (\beta - \alpha)\theta^*] + \alpha\theta^* = B + \beta\theta^*$. Finally, $\alpha > 0\beta$ implies $A + \alpha\theta^* \succ A \sim B \succ B + \beta\theta^*$.

If $\beta > \alpha$, then by symmetric arguments $B + \beta\theta^* \succ A + \alpha\theta^*$. ■

B.2 Construction of V

Recall that for any metric space (X, d), $f : X \to \mathbb{R}$ is *Lipschitz continuous* if there is some real number K such that for every $x, y \in X$, $|f(x) - f(y)| \leq K d(x, y)$. The number K is called a *Lipschitz constant* of f. We will construct a function $V : \mathcal{A}^c \to \mathbb{R}$ that represents $\succsim$ on $\mathcal{A}^c$ and has certain desirable properties. We next define the notion of translation-linearity in order to present the main result of this section. Recall that the set of translations, denoted by Θ, is defined in Equation (19).

Definition 6 Suppose that $V : \mathcal{A}^c \to \mathbb{R}$. Then V is *translation-linear* if there exists $v \in \mathbb{R}^Z$ such that for all $A \in \mathcal{A}^c$ and $\theta \in \Theta$ with $A + \theta \in \mathcal{A}^c$, we have $V(A + \theta) = V(A) + v \cdot \theta$.

Proposition 3 *If the preference $\succsim$ satisfies weak order, strong continuity, ACP, and IDD, then there exists a function $V : \mathcal{A}^c \to \mathbb{R}$ with the following properties:*

1. *For any $A, B \in \mathcal{A}^c$, $A \succsim B \iff V(A) \geq V(B)$.*
2. *V is Lipschitz continuous, convex, and translation-linear.*
3. *There exist $p, q \in \triangle(Z)$ such that $V(\{p\}) > V(\{q\})$.*

Moreover, if V and V' are two functions that satisfy 1–3, then there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $V' = \alpha V + \beta$.

The remainder of this section is devoted to the proof of Proposition 3. First note that by taking the p^* and p_* from the L–continuity axiom, it follows that $\{p^*\} \succ \{p_*\}$. Thus part 3 of this proposition follows from part 1.

Let $\mathcal{S} \equiv \{\{p\} : p \in \triangle(Z)\}$ be the set all of singleton sets in $\mathcal{A}^c$. Given the assumptions of Proposition 3 and the results of Lemma 8, $\succsim$ satisfies the von Neumann-Morgenstern axioms on $\mathcal{S}$. Therefore, there exists $v \in \mathbb{R}^Z$ such that for all $p, q \in \triangle(Z)$, $\{p\} \succsim \{q\}$ if and only if $v \cdot p \geq v \cdot q$. We will abuse notation and also treat v as a function $v : \mathcal{S} \to \mathbb{R}$ naturally defined by $v(\{p\}) = v \cdot p$. Note that v is translation-linear since $v(\{p\}+\theta) = v(\{p\}) + v \cdot \theta$ whenever $p \in \triangle(Z)$, $\theta \in \Theta$, and $p + \theta \in \triangle(Z)$.

We want to extend v to a function V on $\mathcal{A}^c$ that represents $\succsim$ and is translation-linear. The outline of the construction of the desired extension is the following: We first restrict attention to menus in $\mathcal{A}^\circ$, as defined in Equation (25). This restriction will allow us to make extensive use of the translation invariance (TI) property defined in the previous section. We will construct a sequence of subsets of $\mathcal{A}^\circ$, starting with $\mathcal{A}^\circ \cap \mathcal{S}$, such that each set is contained in its successor set. We will then extend v sequentially to each of these domains, while still representing $\succsim$ and preserving translation-linearity (with respect to the vector v). The domain will grow to eventually contain all of the sets in $\mathcal{A}^\circ$, and we show how to extend to all of $\mathcal{A}^c$ by continuity. We prove that the resulting function is translation-linear, Lipschitz continuous, and convex.

As above, take p^* and p_* from the L–continuity axiom, and let $\theta^* \equiv p^* - p_*$. Define a sequence $\mathcal{A}_0, \mathcal{A}'_0, \mathcal{A}_1, \mathcal{A}'_1, \ldots$ of subsets of $\mathcal{A}^\circ$ inductively as follows: Let $\mathcal{A}_0 \equiv \mathcal{A}^\circ \cap \mathcal{S}$. By part 1 of Lemma 9, we have that $\mathcal{A}_0 = \{\{p\} : p \in \triangle(Z) \text{ and } \forall z \in Z, p_z > 0\}$. Define $\mathcal{A}'_i$ for all $i \geq 0$ by

$$\mathcal{A}'_i \equiv \{A \in \mathcal{A}^\circ : A \sim B \text{ for some } B \in \mathcal{A}_i\},$$

and define $\mathcal{A}_i$ for all $i \geq 1$ by

$$\mathcal{A}_i \equiv \{A \in \mathcal{A}^\circ : A = B + \alpha\theta^* \text{ for some } \alpha \in \mathbb{R}, B \in \mathcal{A}'_{i-1}\}.$$

Intuitively, we first extend $\mathcal{A}_0$ by including all $A \in \mathcal{A}^\circ$ that are viewed with indifference to some $B \in \mathcal{A}_0$. Then we extend to all translations by multiples of θ^*. We repeat the process, alternating between extension by indifference and extension by translation. Note that $\mathcal{A}_0 \subset \mathcal{A}'_0 \subset \mathcal{A}_1 \subset \mathcal{A}'_1 \subset \cdots$.

We also define a sequence of functions, $V_0, V'_0, V_1, V'_1, \ldots$, from these domains. That is, for all $i \geq 0$, $V_i : \mathcal{A}_i \to \mathbb{R}$ and $V'_i : \mathcal{A}'_i \to \mathbb{R}$. Define these functions recursively as follows:

1. Let $V_0 \equiv v|_{\mathcal{A}_0}$.

2. For $i \geq 0$, if $A \in \mathcal{A}'_i$, then $A \sim B$ for some $B \in \mathcal{A}_i$, so define V'_i by $V'_i(A) \equiv V_i(B)$.

3. For $i \geq 1$, if $A \in \mathcal{A}_i$, then $A = B + \alpha\theta^*$ for some $\alpha \in \mathbb{R}$ and $B \in \mathcal{A}'_{i-1}$, so define V_i by $V_i(A) \equiv V'_{i-1}(B) + \alpha(v \cdot \theta^*)$.

In a series of lemmas, we will show that these are well-defined functions which represent $\succsim$ on their domains and are translation-linear.

First, we present some important properties of $\mathcal{A}_i$ and $\mathcal{A}'_i$ that will be used to prove Lemmas 12 and 13.

Lemma 11 *For any $i \geq 0$:*

1. *If $A \in \mathcal{A}_i$ and $\theta \in \Theta$, then there exists $\bar{\alpha} > 0$ such that:*

$$A + \alpha\theta \in \mathcal{A}_i, \quad \forall \alpha \in [0, \bar{\alpha}].^{37} \tag{27}$$

2. *For all $A, B \in \mathcal{A}'_i$ and $C \in \mathcal{A}^\circ$, $A \succsim C \succsim B$ implies $C \in \mathcal{A}'_i$.*

Proof: (1): First, it follows immediately from part 1 of Lemma 9 that for any $A \in \mathcal{A}^\circ$ and $\theta \in \Theta$, there exists $\bar{\alpha} > 0$ such that

$$A + \alpha\theta \in \mathcal{A}^\circ, \quad \forall \alpha \in [0, \bar{\alpha}]. \tag{28}$$

We now prove by induction. To verify the property on $\mathcal{A}_0 = \mathcal{A}^\circ \cap \mathcal{S}$, take any $A \in \mathcal{A}_0$ and recall that $A = \{p\}$ for some $p \in \triangle(Z)$. Then take $\bar{\alpha} > 0$ such that Equation (28) holds. Then for all $\alpha \in [0, \bar{\alpha}]$, since $p + \alpha\theta \in \triangle(Z)$, we have $A + \alpha\theta \in \mathcal{A}_0$.

We now prove if the property holds for $\mathcal{A}_i$, then it also holds for $\mathcal{A}_{i+1}$. Take any $A \in \mathcal{A}_{i+1}$ and $\theta \in \Theta$. Then $A \sim B$ for some $B \in \mathcal{A}'_i$, and hence $B = C + \beta\theta^*$ for some $C \in \mathcal{A}_i$ and $\beta \in \mathbb{R}$. Choose $\bar{\alpha} > 0$ to be the minimum of that required to satisfy Equation (28) for A and B and to satisfy Equation (27) for C. Fix any $\alpha \in [0, \bar{\alpha}]$. Then $C + \alpha\theta \in \mathcal{A}_i$, and hence $C + \alpha\theta + \beta\theta^* = B + \alpha\theta \in \mathcal{A}'_i$. By TI, $A \sim B$ implies $A + \alpha\theta \sim B + \alpha\theta$, which implies $A + \alpha\theta \in \mathcal{A}_{i+1}$.

(2): We again prove by induction. To prove the result for $\mathcal{A}'_0$, suppose $A, B \in \mathcal{A}'_0$ and $A \succsim C \succsim B$ for some $C \in \mathcal{A}^\circ$. Since $A, B \in \mathcal{A}'_0$, there exist $\{p\}, \{q\} \in \mathcal{A}_0$ such that $\{p\} \sim A \succsim C \succsim B \sim \{q\}$. Continuity implies there exists a $\lambda \in [0,1]$ such that $\{\lambda p + (1-\lambda)q\} \sim C$. By the convexity of $\mathcal{A}_0 = \mathcal{A}^\circ \cap \mathcal{S}$ and the definition of $\mathcal{A}'_0$, this implies that $C \in \mathcal{A}'_0$.

We now show that if $\mathcal{A}'_i$ satisfies the desired condition, then $\mathcal{A}'_{i+1}$ does also. Suppose $A, B \in \mathcal{A}'_{i+1}$ and $A \succsim C \succsim B$ for some $C \in \mathcal{A}^\circ$. If there exist $A', B' \in \mathcal{A}'_i$ such that $A' \succsim C \succsim B'$, then $C \in \mathcal{A}'_i \subset \mathcal{A}'_{i+1}$ by the induction assumption. Thus without loss of generality, suppose $C \succ A'$ for all $A' \in \mathcal{A}'_i$. Since $A \in \mathcal{A}'_{i+1}$, there exists a $A' \in \mathcal{A}_{i+1}$ such that $A' \sim A \succsim C$. Since $A' \in \mathcal{A}_{i+1}$, there exists a $A'' \in \mathcal{A}'_i$ and $\alpha \in \mathbb{R}$ such that $A' = A'' + \alpha\theta^*$. Since $A'' \in \mathcal{A}'_i$ implies $C \succ A''$, this implies $A'' + \alpha\theta^* \succsim C \succ A''$, and therefore $\alpha > 0$ by Lemma 10. By continuity, there exists a $\alpha' \in [0, \alpha]$ such that $A'' + \alpha'\theta^* \sim C$. But $A'' + \alpha'\theta^* \in \mathcal{A}_{i+1}$, so it must be that $C \in \mathcal{A}'_{i+1}$. ■

The following lemmas allow us to prove the desired properties of each V_i and V'_i by induction.

Lemma 12 *For all $i \geq 0$, if V_i is well-defined, translation-linear, and represents $\succsim$ on $\mathcal{A}_i$, then V'_i is also well-defined, translation-linear, and represents $\succsim$ on $\mathcal{A}'_i$.*

Proof: *(Well-defined):* Suppose $A \in \mathcal{A}'_i$ and $B, B' \in \mathcal{A}_i$ are such that $A \sim B$ and $A \sim B'$. Since V_i represents $\succsim$ on $\mathcal{A}_i$ and $\succsim$ is transitive, $V_i(B) = V_i(B')$, and hence $V'_i(A)$ is uniquely

[37] As the proof of this lemma will illustrate, the same property holds for $\mathcal{A}'_i$.

defined.

(Represents $\succsim$): If $A, A' \in \mathcal{A}_i'$ then there exist $B, B' \in \mathcal{A}_i$ such that $A \sim B$ and $A' \sim B'$. Therefore, $V_i'(A) = V_i(B) \geq V_i(B') = V_i'(A')$ if and only if $B \succsim B'$ if and only if $A \succsim A'$, so V_i' represents $\succsim$ on $\mathcal{A}_i'$.

(Translation-linear): Throughout we will use the fact if $\theta \in \Theta$ and $A, A + \theta \in \mathcal{A}_i'$, then $A + \alpha\theta \in \mathcal{A}_i'$ for all $\alpha \in [0, 1]$. This follows by part 2 of Lemma 11 because by Lemma 7, either $A + \theta \succsim A + \alpha\theta \succsim A$ or $A \succsim A + \alpha\theta \succsim A + \theta$.

We first show that V_i' satisfies the following local version of translation-linearity: For all $A \in \mathcal{A}_i'$ and $\theta \in \Theta$ with $A + \theta \in \mathcal{A}_i'$, there exist $\bar{\alpha} > 0$ such that for all $\alpha \in [0, \bar{\alpha}]$,

$$V_i'(A + \alpha\theta) = V_i'(A) + \alpha(v \cdot \theta).$$

To see this property holds, suppose $\theta \in \Theta$ and $A, A + \theta \in \mathcal{A}_i'$. By the definition of $\mathcal{A}_i'$ there exists $B \in \mathcal{A}_i$ such that $A \sim B$. By part 1 of Lemma 11, there exists $\bar{\alpha} \in (0, 1]$ such that $B + \alpha\theta \in \mathcal{A}_i$ for all $\alpha \in [0, \bar{\alpha}]$. Fix any $\alpha \in [0, \bar{\alpha}]$, and $A \sim B$ implies $A + \alpha\theta \sim B + \alpha\theta$ by TI. Therefore, using the translation-linearity of V_i on $\mathcal{A}_i$,

$$V_i'(A + \alpha\theta) = V_i(B + \alpha\theta) = V_i(B) + \alpha(v \cdot \theta) = V_i'(A) + \alpha(v \cdot \theta).$$

We now show that this local version of translation-linearity implies translation-linearity. Fix any $A \in \mathcal{A}_i'$ and $\theta \in \Theta$ with $A + \theta \in \mathcal{A}_i'$, and let

$$\alpha^* \equiv \sup\{\bar{\alpha} \in [0, 1] : V_i'(A + \alpha\theta) = V_i'(A) + \alpha(v \cdot \theta)\ \forall \alpha \in [0, \bar{\alpha}]\}.$$

Note that $V_i'(A + \alpha^*\theta) = V_i'(A) + \alpha^*(v \cdot \theta)$. If $\alpha^* = 0$, this is obvious. If $\alpha^* > 0$, then local translation-linearity applied to $A' = A + \alpha^*\theta \in \mathcal{A}_i'$ and $\theta' = -\alpha^*\theta$ implies there exists $\bar{\alpha} > 0$ such that $V_i'(A + \alpha^*\theta - \bar{\alpha}\theta) = V_i'(A + \alpha^*\theta) - \bar{\alpha}(v \cdot \theta)$. Therefore,

$$\begin{aligned} V_i'(A + \alpha^*\theta) &= V_i'(A + (\alpha^* - \bar{\alpha})\theta) + \bar{\alpha}(v \cdot \theta) \\ &= V_i'(A) + (\alpha^* - \bar{\alpha})(v \cdot \theta) + \bar{\alpha}(v \cdot \theta) \\ &= V_i'(A) + \alpha^*(v \cdot \theta), \end{aligned}$$

where the second equality follows by the definition of α^* since $0 < \alpha^* - \bar{\alpha} < \alpha^*$. It remains only to show that $\alpha^* = 1$. If not, then local translation-linearity applied to $A' = A + \alpha^*\theta \in \mathcal{A}_i'$ and $\theta' = (1 - \alpha^*)\theta$ implies there exists $\bar{\alpha} > 0$ such that for all $\alpha \in [0, \bar{\alpha}]$,

$$\begin{aligned} V_i'(A + \alpha^*\theta + \alpha\theta) &= V_i'(A + \alpha^*\theta) + \alpha(v \cdot \theta) \\ &= V_i'(A) + (\alpha^* + \alpha)(v \cdot \theta). \end{aligned}$$

This would imply $\alpha^* \geq \alpha^* + \bar{\alpha}$, a contradiction. Thus $\alpha^* = 1$. ■

Lemma 13 *For all $i \geq 1$, if V'_{i-1} is well-defined, translation-linear, and represents $\succsim$ on $\mathcal{A}'_{i-1}$, then V_i is also well-defined, translation-linear, and represents $\succsim$ on $\mathcal{A}_i$.*

Proof: *(Well-defined):* Suppose $A \in \mathcal{A}_i$ and $A = B + \alpha\theta^* = B' + \alpha'\theta^*$ for $B, B' \in \mathcal{A}'_{i-1}$ and $\alpha, \alpha' \in \mathbb{R}$. Then $B = B' + (\alpha' - \alpha)\theta^*$, so the translation-linearity of V'_{i-1} implies $V'_{i-1}(B) = V'_{i-1}(B') + (\alpha' - \alpha)(v \cdot \theta^*)$. Therefore, $V'_{i-1}(B) + \alpha(v \cdot \theta^*) = V'_{i-1}(B') + \alpha'(v \cdot \theta^*)$, and hence $V_i(A)$ is uniquely defined.

(Translation-linear): Suppose $\theta \in \Theta$ and $A, A + \theta \in \mathcal{A}_i$. Then there exist $B, B' \in \mathcal{A}'_{i-1}$ and $\alpha, \alpha' \in \mathbb{R}$ such that $A = B + \alpha\theta^*$ and $A + \theta = B' + \alpha'\theta^*$. Then $B' = B + (\alpha - \alpha')\theta^* + \theta$, so the translation-linearity of V'_{i-1} implies $V'_{i-1}(B') = V'_{i-1}(B) + v \cdot [(\alpha - \alpha')\theta^* + \theta]$. By the definition of V_i, we therefore have

$$V_i(A + \theta) = V'_{i-1}(B') + \alpha'(v \cdot \theta^*) = V'_{i-1}(B) + \alpha(v \cdot \theta^*) + v \cdot \theta = V_i(A) + v \cdot \theta.$$

(Represents $\succsim$): Suppose $A, A' \in \mathcal{A}_i$, so that $A = B + \alpha\theta^*$ and $A' = B' + \alpha'\theta^*$ for some $B, B' \in \mathcal{A}'_{i-1}$ and $\alpha, \alpha' \in \mathbb{R}$. There are several different cases to consider, and the interest of brevity we only work through one of them here: $A, A' \succsim B' \succsim B$. Thus $B + \alpha\theta^* \succsim B' \succsim B$, which implies $\alpha \geq 0$ by Lemma 10. Continuity implies there exists $\alpha'' \in [0, \alpha]$ such that $B + \alpha''\theta^* \sim B'$, which therefore implies $B + \alpha''\theta^* \in \mathcal{A}'_{i-1}$. Thus by Lemma 10 and the definition of V_i, we have $A \succsim A'$ if and only if $\alpha - \alpha'' \geq \alpha'$ if and only if

$$\begin{aligned} V_i(A) &= V'_{i-1}(B + \alpha''\theta^*) + (\alpha - \alpha'')(v \cdot \theta^*) \\ &= V'_{i-1}(B') + (\alpha - \alpha'')(v \cdot \theta^*) \\ &\geq V'_{i-1}(B') + \alpha'(v \cdot \theta^*) = V_i(A'). \end{aligned}$$

The other cases are similar.[38] ■

Using induction and the results of Lemmas 12 and 13, we have proved that for all $i \geq 0$, $V_i : \mathcal{A}_i \to \mathbb{R}$ is well-defined, translation-linear, and represents $\succsim$ on $\mathcal{A}_i$. We now define a function $\hat{V} : \bigcup_i \mathcal{A}_i \to \mathbb{R}$ by $\hat{V}(A) \equiv V_i(A)$ if $A \in \mathcal{A}_i$. This is well-defined because if $A \in \mathcal{A}_i$ and $A \in \mathcal{A}_j$, then without loss of generality suppose $j \geq i$, so $\mathcal{A}_i \subset \mathcal{A}_j$. Then $V_j(B) = V_i(B)$ for all $B \in \mathcal{A}_i$, and hence $V_j(A) = V_i(A)$. Note that $\hat{V}$ represents $\succsim$ on $\bigcup_i \mathcal{A}_i$ and is translation-linear.

By the following lemma, we have now established a translation-linear representation for $\succsim$ on all of $\mathcal{A}^\circ$.

Lemma 14 $\mathcal{A}^\circ = \bigcup_i \mathcal{A}_i$.

[38] The only substantively different cases are the variations of the following: $B' \succsim A, A' \succsim B$. However, in this case we can apply Lemma 11, which implies $A, A' \in \mathcal{A}'_{i-1}$, and hence the result is obtained by assumption.

Proof: That $\bigcup_i \mathcal{A}_i \subset \mathcal{A}^\circ$ follows immediately from the definition of $\mathcal{A}_i$, so it remains only to prove that $\mathcal{A}^\circ \subset \bigcup_i \mathcal{A}_i$. Consider any set $A \in \mathcal{A}^\circ$. By part 1 of Lemma 9, there exists some $\alpha > 0$ such that $A + \alpha\theta^*, A - \alpha\theta^* \in \mathcal{A}^\circ$. Fix any $p \in A$, and we therefore have $\{p\} + \alpha\theta^*, \{p\} - \alpha\theta^* \in \mathcal{A}_0 \subset \mathcal{A}^\circ$. For every $\lambda \in [0,1]$, define $A(\lambda) \equiv \lambda A + (1-\lambda)\{p\}$. Note that $A(\lambda) + \alpha\theta^*, A(\lambda) - \alpha\theta^* \in \mathcal{A}^\circ$, which follows from the convexity of $\mathcal{A}^\circ$ since

$$A(\lambda) + \alpha\theta^* = \lambda A + (1-\lambda)\{p\} + \alpha\theta^* = \lambda(A + \alpha\theta^*) + (1-\lambda)(\{p\} + \alpha\theta^*),$$

and similarly for $A(\lambda) - \alpha\theta^*$. By Lemma 10 for all $\lambda \in [0,1]$, $A(\lambda) + \alpha\theta^* \succ A(\lambda) \succ A(\lambda) - \alpha\theta^*$. By continuity, for each λ there exists an open (relative to $[0,1]$) interval $e(\lambda)$ such that $\lambda \in e(\lambda)$ and for all $\lambda' \in e(\lambda)$,

$$A(\lambda) + \alpha\theta^* \succ A(\lambda') \succ A(\lambda) - \alpha\theta^*.$$

Thus $\{e(\lambda) : \lambda \in [0,1]\}$ is an open cover of $[0,1]$. Since $[0,1]$ is compact, there exists a finite subcover, $\{e(\lambda_1), \ldots, e(\lambda_n)\}$. Assume the λ_i's are ordered so that $e(\lambda_i) \cap e(\lambda_{i+1}) \neq \emptyset$, $\{p\} = A(0) \in e(\lambda_1)$, and $A = A(1) \in e(\lambda_n)$. That is, as i increases, $e(\lambda_i)$ moves "farther" from $\{p\}$ and "closer" to A. We can prove that $A(\lambda_1) \in \mathcal{A}_1$ by first observing that

$$A(\lambda_1) + \alpha\theta^* \succ A(0) = \{p\} \succ A(\lambda_1) - \alpha\theta^*,$$

which by continuity implies there exists $\alpha' \in (-\alpha, \alpha)$ such that $A(\lambda_1) + \alpha'\theta^* \sim \{p\}$. This implies $A(\lambda_1) + \alpha'\theta^* \in \mathcal{A}'_0$, which implies that $A(\lambda_1) \in \mathcal{A}_1$. We now show that $A(\lambda_i) \in \mathcal{A}_i$ implies $A(\lambda_{i+1}) \in \mathcal{A}_{i+1}$. If $A(\lambda_i) \in \mathcal{A}_i$, then we also have $A(\lambda_i) + \alpha'\theta^* \in \mathcal{A}_i$ for all $\alpha' \in (-\alpha, \alpha)$. Since $e(\lambda_i) \cap e(\lambda_{i+1}) \neq \emptyset$, choose any $\lambda \in e(\lambda_i) \cap e(\lambda_{i+1})$. Then,

$$A(\lambda_i) + \alpha\theta^* \succ A(\lambda) \succ A(\lambda_i) - \alpha\theta^*$$
$$A(\lambda_{i+1}) + \alpha\theta^* \succ A(\lambda) \succ A(\lambda_{i+1}) - \alpha\theta^*$$

By continuity, there exist $\alpha', \alpha'' \in (-\alpha, \alpha)$ such that $A(\lambda_i) + \alpha'\theta^* \sim A(\lambda) \sim A(\lambda_{i+1}) + \alpha''\theta^*$, which implies $A(\lambda_{i+1}) + \alpha''\theta^* \in \mathcal{A}'_i$. Hence, $A(\lambda_{i+1}) \in \mathcal{A}_{i+1}$. By induction, we conclude that $A(\lambda_i) \in \mathcal{A}_i$ for $i = 1, \ldots, n$, and also that $A \in \mathcal{A}'_n \subset \mathcal{A}_{n+1} \subset \bigcup_i \mathcal{A}_i$. ∎

We have now proved that $\hat{V}$ is translation-linear and represents $\succsim$ on $\mathcal{A}^\circ$. Before extending $\hat{V}$ to $\mathcal{A}^c$, we first establish that $\hat{V}$ is Lipschitz continuous and convex.

Lemma 15 *$\hat{V}$ is Lipschitz continuous.*

Proof: For all $\delta \in (0,1)$, define:

$$\mathcal{A}^\circ_\delta \equiv \{A \in \mathcal{A}^c \ : \ \forall p \in A, \forall z \in Z : p_z \geq \delta\}.$$

We next summarize some straightforward facts about $\mathcal{A}^\circ_\delta$ whose proofs we omit:

1. $\mathcal{A}^\circ_\delta$ is a convex subset of $\mathcal{A}^\circ$.

2. For all $A \in \mathcal{A}_\delta^\circ$ and $\alpha \in (0,\delta)$ there exists a unique menu $A^\alpha \in \mathcal{A}^\circ$ such that $A = (1-\alpha)A^\alpha + \alpha\{p_*\}$.[39]
3. For all $A, B \in \mathcal{A}_\delta^\circ$, $\alpha \in (0,\delta)$: $(1-\alpha)d_h(A^\alpha, B^\alpha) = d_h(A, B)$.
4. For all $A \in \mathcal{A}_\delta^\circ$, $\alpha \in (0,\delta)$: $A + \alpha\theta^* \in \mathcal{A}^\circ$.

Let $K \equiv 2M(v\cdot\theta^*) > 0$ and $\delta \in (0, 1/2)$. We first show that:

$$A, B \in \mathcal{A}_\delta^\circ \;\&\; d_h(A,B) < \tfrac{\delta}{2M} \implies |\hat{V}(A) - \hat{V}(B)| \le K d_h(A,B). \tag{29}$$

Suppose that A, B are as in the left hand side of Equation (29). Let $\alpha \in (2Md_h(A,B), \delta)$. Then

$$d_h(A^\alpha, B^\alpha) = \frac{1}{1-\alpha} d_h(A,B) \le 2d_h(A,B) < \alpha/M < 1/M,$$

where the weak inequality follows from $\alpha < \delta < 1/2$. Applying L–continuity we have:

$$A + \alpha\theta^* = (1-\alpha)A^\alpha + \alpha\{p^*\} \succ (1-\alpha)B^\alpha + \alpha\{p_*\} = B.$$

Since $\hat{V}$ represents $\succsim$ and is translation-linear on $\mathcal{A}^\circ$, we have $\hat{V}(A) + \alpha(v\cdot\theta^*) > \hat{V}(B)$, implying

$$\alpha(v\cdot\theta^*) > \hat{V}(B) - \hat{V}(A).$$

Since the above inequality holds for any $\alpha \in (2Md_h(A,B), \delta)$, we conclude that

$$\hat{V}(B) - \hat{V}(A) \le 2Md_h(A,B)(v\cdot\theta^*) = Kd_h(A,B).$$

Interchanging the roles of A and B above we also have that $\hat{V}(A) - \hat{V}(B) \le Kd_h(A,B)$, proving Equation (29).

Next, we use the argument in the proof of Lemma 8 in the supplementary appendix of DLRS (2007) to show that

$$A, B \in \mathcal{A}_\delta^\circ \implies |\hat{V}(A) - \hat{V}(B)| \le Kd_h(A,B), \tag{30}$$

i.e., the requirement $d_h(A,B) < \frac{\delta}{2M}$ in Equation (29) is not necessary. To see this, take any sequence $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_n < \lambda_{n+1} = 1$ such that $(\lambda_{i+1} - \lambda_i)d_h(A,B) < \frac{\delta}{2M}$. Let $A_i = \lambda_i A + (1-\lambda_i)B$. It is straightforward to verify that

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

[39] The menu A^α is given by $A^\alpha = \left\{q \in \mathbb{R}^Z \;:\; q = \frac{1}{1-\alpha}(p - \alpha p_*) \text{ for some } p \in A\right\}$.

Combining this with the triangular inequality and Equation (29), we obtain

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

To conclude the proof, note that by part 1 of Lemma 9, for any $A, B \in \mathcal{A}^\circ$, there exists a small enough $\delta \in (0, 1/2)$ such that $A, B \in \mathcal{A}^\circ_\delta$. Hence by Equation (30), $\hat{V}$ is Lipschitz continuous on $\mathcal{A}^\circ$ with the Lipschitz constant K. ■

Lemma 16 *$\hat{V}$ is convex.*

Proof: The argument given here is similar to a result contained in a working-paper version of Maccheroni, Marinacci, and Rustichini (2006). We will show that every $A_0 \in \mathcal{A}^\circ$ has a convex and open neighborhood in $\mathcal{A}^\circ$ on which $\hat{V}$ is convex. By a standard result from convex analysis, this implies that $\hat{V}$ is convex on $\mathcal{A}^\circ$.

Let $A_0 \in \mathcal{A}^\circ$. Define $\mathcal{C}$ to be the collection of all closed and bounded non-empty convex subsets of $\{p \in \mathbb{R}^Z : \sum_{z \in Z} p_z = 1\}$, endowed with the Hausdorff metric topology. It follows from part 1 of Lemma 9 that there exists an $\varepsilon > 0$ such that $B_\varepsilon(A_0) \subset \mathcal{A}^\circ$, where we define

$$B_\varepsilon(A_0) \equiv \{A \in \mathcal{C} : d_h(A, A_0) < \varepsilon\}.$$

Note that $d_h(\cdot, \cdot)$ indicates the Hausdorff metric. For any $\theta \in \Theta$ and $A \in \mathcal{C}$, we have $A + \theta \in \mathcal{C}$ and $d_h(A, A + \theta) = \|\theta\|$, where $\|\cdot\|$ indicates the Euclidean norm. There exists $\theta \in \Theta$ such that $\|\theta\| < \varepsilon$ and $v \cdot \theta > 0$.[40] This implies that $A_0 + \theta \in B_\varepsilon(A_0)$ and $A_0 + \theta \succ A_0$. By continuity, there exists $\rho \in (0, \frac{1}{3})$ such that for all $A \in B_{\rho\varepsilon}(A_0)$, $|\hat{V}(A) - \hat{V}(A_0)| < \frac{1}{3}(v \cdot \theta)$. Therefore, if $A, B \in B_{\rho\varepsilon}(A_0)$, then

$$|\hat{V}(A) - \hat{V}(B)| \leq |\hat{V}(A) - \hat{V}(A_0)| + |\hat{V}(A_0) - \hat{V}(B)| < \tfrac{2}{3}(v \cdot \theta).$$

Let $\alpha \equiv \frac{\hat{V}(A) - \hat{V}(B)}{v \cdot \theta}$, which implies $|\alpha| < \frac{2}{3}$. Then, we have

$$\begin{aligned} d_h(A_0, B + \alpha\theta) &\leq d_h(A_0, B) + d_h(B, B + \alpha\theta) \\ &< \rho\varepsilon + \|\alpha\theta\| \\ &< \tfrac{1}{3}\varepsilon + \tfrac{2}{3}\varepsilon = \varepsilon, \end{aligned}$$

[40]For instance $\theta = \alpha\theta^*$ for any $\alpha \in (0, \varepsilon/\|\theta^*\|)$ where $\theta^* = p^* - p_*$.

so $B + \alpha\theta \in B_\varepsilon(A_0) \subset \mathcal{A}^\circ$. Thus $\hat{V}$ is defined at $B + \alpha\theta$. Note that $\alpha(v \cdot \theta) = \hat{V}(A) - \hat{V}(B)$, so that $\hat{V}(B + \alpha\theta) = \hat{V}(B) + \alpha(v \cdot \theta) = \hat{V}(A)$. Since $\succsim$ satisfies ACP, for any $\lambda \in [0, 1]$,

$$\hat{V}(A) \geq \hat{V}(\lambda A + (1 - \lambda)(B + \alpha\theta)).$$

Therefore,

$$\begin{aligned}\hat{V}(A) &\geq \hat{V}(\lambda A + (1 - \lambda)B) + (1 - \lambda)\alpha(v \cdot \theta) \\ &= \hat{V}(\lambda A + (1 - \lambda)B) + (1 - \lambda)(\hat{V}(A) - \hat{V}(B)),\end{aligned}$$

so we have

$$\lambda\hat{V}(A) + (1 - \lambda)\hat{V}(B) \geq \hat{V}(\lambda A + (1 - \lambda)B).$$

Therefore, $\hat{V}$ is convex on the convex and open neighborhood $B_{\rho\varepsilon}(A_0)$ of A_0 in $\mathcal{A}^\circ$. ■

Since $\mathcal{A}^\circ$ is dense in $\mathcal{A}^c$ (see Lemma 9), we can extend $\hat{V}$ to $\mathcal{A}^c$ by continuity. That is, define a function $V : \mathcal{A}^c \to \mathbb{R}$ as follows: For any $A \in \mathcal{A}^c$ there exists a sequence $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}^\circ$ such that $A_n \to A$, so define $V(A) \equiv \lim_{n\to\infty} \hat{V}(A_n)$. Since $\hat{V}$ is Lipschitz continuous, the following lemma establishes that V is well-defined and also Lipschitz continuous. Furthermore, this extension V of $\hat{V}$ represents $\succsim$ on $\mathcal{A}^c$ and preserves the translation-linearity and convexity of $\hat{V}$.

Lemma 17 *The function $V : \mathcal{A}^c \to \mathbb{R}$ is well-defined, and it satisfies properties 1–3 from Proposition 3.*

Proof: By Lemma 9, $\mathcal{A}^\circ$ is dense in $\mathcal{A}^c$. Since $\mathcal{A}^c$ is a compact metric space, it is complete. Since $\hat{V}$ is Lipschitz continuous it is in uniformly continuous (see Aliprantis and Border (1999, page 76)). Therefore by Lemma 3.8 in Aliprantis and Border (1999, page 77), V is well-defined and it is the unique continuous extension of $\hat{V}$ to $\mathcal{A}^c$. To see that V is Lipschitz continuous, let $K > 0$ be a Lipschitz constant for $\hat{V}$ on $\mathcal{A}^\circ$ and let $A, B \in \mathcal{A}^c$. Take sequences $\{A_n\}_{n\in\mathbb{N}}, \{B_n\}_{n\in\mathbb{N}} \subset \mathcal{A}^\circ$ such that $A_n \to A$ and $B_n \to B$. Then,

$$|V(A) - V(B)| = \lim_{n\to\infty} |\hat{V}(A_n) - \hat{V}(B_n)| \leq \lim_{n\to\infty} K d_h(A_n, B_n) = K d_h(A, B).$$

Hence V is Lipschitz continuous with the same constant K.

To see that V is translation-linear, let $A, A+\theta \in \mathcal{A}^c$ for some $\theta \in \Theta$. Fix any $p \in \triangle(Z)$ such that $p_z > 0$ for all $z \in Z$. For all $n \in \mathbb{N}$, define $A_n \equiv (1-1/n)A + (1/n)\{p\}$ and $\theta_n \equiv (1-1/n)\theta$. By Lemma 9, for all $n \in \mathbb{N}$, $A_n \in \mathcal{A}^\circ$ and $A_n + \theta_n = (1-1/n)(A+\theta) + (1/n)\{p\} \in \mathcal{A}^\circ$. Moreover, $A_n \to A$ and $A_n + \theta_n \to A + \theta$ as $n \to \infty$. Therefore,

$$V(A + \theta) - V(A) = \lim_{n\to\infty} \left[\hat{V}(A_n + \theta_n) - \hat{V}(A_n)\right] = \lim_{n\to\infty} v \cdot \theta_n = v \cdot \theta.$$

Thus we see that V is translation-linear on all of $\mathcal{A}^c$. The proof that V is convex is straightforward and follows from a similar line of reasoning; it is therefore omitted.

In order to show that V represents $\succsim$ on $\mathcal{A}^c$, we prove $A \succ B \iff V(A) > V(B)$. Let $\{A_n\}_{n\in\mathbb{N}}, \{B_n\}_{n\in\mathbb{N}} \subset \mathcal{A}^\circ$ be such that $A_n \to A$ and $B_n \to B$ as $n \to \infty$.

To see "$\Rightarrow$", suppose $A \succ B$. By the continuity of $\succsim$, $\{C \in \mathcal{A}^c : A \succ C \succ B\}$ is nonempty and open.[41] Since $\mathcal{A}^\circ$ is dense in $\mathcal{A}^c$, there exists $\bar{A} \in \mathcal{A}^\circ$ such that $A \succ \bar{A} \succ B$. Repeating the same argument for $\bar{A} \succ B$, there exists $\bar{B} \in \mathcal{A}^\circ$ such that $\bar{A} \succ \bar{B} \succ B$. By continuity, $\{C \in \mathcal{A}^c : C \succ \bar{A}\}$ is a neighborhood of A, so there exists $N \in \mathbb{N}$ such that $A_n \succ \bar{A}$ for all $n \geq N$. A similar argument implies there exists $N' \in \mathbb{N}$ such that $\bar{B} \succ B_n$ for all $n \geq N'$. Therefore,

$$V(A) = \lim_{n\to\infty} \hat{V}(A_n) \geq \hat{V}(\bar{A})\hat{V}(\bar{B}) \geq \lim_{n\to\infty} \hat{V}(B_n) = V(B).$$

To show "$\Leftarrow$", we will apply a similar argument using the continuity of V. Suppose $V(A) > V(B)$. By continuity of V, $\{C \in \mathcal{A}^c : V(A) > V(C) > V(B)\}$ is nonempty and open. Since $\mathcal{A}^\circ$ is dense in $\mathcal{A}^c$, there exists $\bar{A} \in \mathcal{A}^\circ$ such that $V(A)V(\bar{A}) > V(B)$. Repeating the same argument for $V(\bar{A})V(B)$, there exists $\bar{B} \in \mathcal{A}^\circ$ such that $V(\bar{A})V(\bar{B}) > V(B)$. By continuity, $\{C \in \mathcal{A}^c : V(C)V(\bar{A})\}$ is a neighborhood of A, so there exists $N \in \mathbb{N}$ such that $\hat{V}(A_n) = V(A_n) > V(\bar{A}) = \hat{V}(\bar{A})$ for all $n \geq N$. A similar argument implies there exists $N' \in \mathbb{N}$ such that $\hat{V}(\bar{B}) > \hat{V}(B_n)$ for all $n \geq N'$. Therefore by continuity of $\succsim$,

$$A = \lim_{n\to\infty} A_n \succsim \bar{A} \succ \bar{B} \succsim \lim_{n\to\infty} B_n = B.$$

Finally, since V represents $\succsim$ on $\mathcal{A}^c$ and $p^* \succ p_*$, we also have that $V(\{p^*\}) > V(\{p_*\})$. ■

The following lemma establishes uniqueness of the representation, completing the proof of Proposition 3.

Lemma 18 *Suppose $\succsim$ satisfies weak order, strong continuity, ACP, and TI. If $V : \mathcal{A}^c \to \mathbb{R}$ and $V' : \mathcal{A}^c \to \mathbb{R}$ are two functions that satisfy 1–3 from Proposition 3, then there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $V' = \alpha V + \beta$.*

Proof: Translation-linearity implies that V and V' are affine on singletons, and therefore the standard vNM uniqueness result implies $V'|_{\mathcal{S}} = \alpha V|_{\mathcal{S}} + \beta$ for some $\alpha > 0$, $\beta \in \mathbb{R}$. By translation-linearity and property 1 from Proposition 3, a simple induction argument shows that $V'|_{\mathcal{A}_i} = \alpha V|_{\mathcal{A}_i} + \beta$ for all i. Hence $V'|_{\mathcal{A}^\circ} = \alpha V|_{\mathcal{A}^\circ} + \beta$. Since $\mathcal{A}^\circ$ is dense in $\mathcal{A}^c$ (see Lemma 9) and the functions V and V' are continuous on $\mathcal{A}^c$, we conclude that $V' = \alpha V + \beta$ on $\mathcal{A}^c$. ■

[41] Note that the sets $\{\lambda \in [0,1] : \lambda A + (1-\lambda)B \succ B\}$ and $\{\lambda \in [0,1] : A \succ \lambda A + (1-\lambda)B\}$ are nonempty and open relative to [0,1] (by continuity of $\succsim$ and continuity of convex combinations), and their union is $[0,1]$. Since $[0,1]$ is connected, their intersection must be non-empty. Hence the set $\{C \in \mathcal{A}^c : A \succ C \succ B\}$ is also nonempty.

B.3 Application of Duality Results

In this section, we apply the duality results from Appendix A to the function V constructed in Section B.2 to obtain the desired RFCC representation. Thus in the remainder of this section assume that V satisfies 1–3 from Proposition 3. Note that if $\succsim$ also satisfies monotonicity, then V is *monotone* in the sense that for all $A, B \in \mathcal{A}^c$ such that $A \subset B$, we have $V(A) \leq V(B)$. We explicitly assume monotonicity of V at the end of this section to prove the stronger representation of Theorem 1.B.

We 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. 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 19 *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 $A, B \in \mathcal{A}^c$, $\sigma_{\lambda A + (1-\lambda) B} = \lambda \sigma_A + (1-\lambda) \sigma_B$.

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

Proof: These are standard results that can be found in Rockafellar (1970) or Schneider (1993). For instance in Schneider (1993), part 1 can be found on p39 (Theorem 1.7.1), part 2 can be found on p37, and part 3 can be found on p53 (Theorem 1.8.11). ■

Lemma 20 *Σ is convex and compact, and $0 \in \Sigma$.*

Proof: The set Σ is convex by the convexity of $\mathcal{A}^c$ and part 2 of Lemma 19. As discussed above, the set $\mathcal{A}^c$ is compact, and hence by parts 1 and 3 of Lemma 19, Σ is a compact subset of the Banach space $C(\mathcal{U})$. Also, if we take $q = (1/|Z|, \dots, 1/|Z|) \in \triangle(Z)$, then $q \cdot u = 0$ for all $u \in \mathcal{U}$. Thus $\sigma_{\{q\}} = 0$, and hence $0 \in \Sigma$. ■

Define the function $W : \Sigma \to \mathbb{R}$ by $W(\sigma) = V(A_\sigma)$. Then, by part 1 of Lemma 19, $V(A) = W(\sigma_A)$ for all $A \in \mathcal{A}^c$. We say the function W is *monotone* if for all $\sigma, \sigma' \in \Sigma$ such that $\sigma \leq \sigma'$ we have $W(\sigma) \leq W(\sigma')$.

Lemma 21 *W is convex and Lipschitz continuous with the same Lipschitz constant as V. If V is monotone, then W is monotone.*

Proof: To see that W is convex, let $A, B \in \mathcal{A}^c$. Then,

$$W(\lambda\sigma_A + (1-\lambda)\sigma_B) = W(\sigma_{\lambda A+(1-\lambda)B}) = V(\lambda A + (1-\lambda)B)$$
$$\leq \lambda V(A) + (1-\lambda)V(B) = \lambda W(\sigma_A) + (1-\lambda)W(\sigma_B)$$

by parts 1 and 2 of Lemma 19 and convexity of V. The function W is Lipschitz continuous with the same Lipschitz constant as V by parts 1 and 3 of Lemma 19. The function W inherits monotonicity from V because of the following fact which is easy to see from part 1 of Lemma 19: for all $A, B \in \mathcal{A}^c$, $A \subset B$ iff $\sigma_A \leq \sigma_B$. ■

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 $\mu \in C(\mathcal{U})^*$.[42]

Define Σ_W, $\mathcal{N}_W$, and $\mathcal{M}_W$ as in Equations (16), (17), and (18), respectively:

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

where the closure is taken with respect to the weak* topology. We now apply Proposition 2 to the current setting.

Lemma 22 *$\mathcal{M}_W$ 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 Proposition 2. Since $\mathcal{U}$ is a compact metric space, $C(\mathcal{U})$ is separable.[43] By Lemma 20, Σ is a closed and convex subset of $C(\mathcal{U})$ containing the origin. Although the result is stated slightly differently, it is shown in Hörmander (1954) that *aff* (Σ) is dense in $C(\mathcal{U})$. This result is also proved in DLR (2001). Finally, W is Lipschitz continuous and convex by Lemma 21. ■

[42]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.

[43]See Theorem 8.48 of Aliprantis and Border (1999).

One consequence of Lemma 22 is that for all $\sigma \in \Sigma$,

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

Therefore, for all $A \in \mathcal{A}^c$,

$$V(A) = \max_{\mu \in \mathcal{M}_W} \left[\int_{\mathcal{U}} \max_{p \in A} (u \cdot p) \mu(du) - W^*(\mu) \right].$$

The function W^* is lower semicontinuous by part 1 of Lemma 4, and $\mathcal{M}_W$ is compact by Lemma 22. It remains only to show that $\mathcal{M}_W$ is consistent and minimal and that monotonicity of W implies each $\mu \in \mathcal{M}_W$ is positive.

Since V is translation-linear, there exists $v \in \mathbb{R}^Z$ such that for all $A \in \mathcal{A}^c$ and $\theta \in \Theta$ with $A + \theta \in \mathcal{A}^c$, we have $V(A + \theta) = V(A) + v \cdot \theta$. The following result shows that a certain subset of $\mathcal{M}_W$ must "agree" with v in a way that will imply the consistency of this subset. In what follows, let $q = (1/|Z|, \ldots, 1/|Z|) \in \triangle(Z)$ and let $\mathcal{A}^\circ \subset \mathcal{A}^c$ be defined as in Equation (25).

Lemma 23 *If $A \in \mathcal{A}^\circ$ and $\mu \in \partial W(\sigma_A)$, then $\langle \sigma_{\{p\}}, \mu \rangle = v \cdot (p - q)$ for all $p \in \triangle(Z)$.*

Proof: Fix any $A \in \mathcal{A}^\circ$ and $\mu \in \partial W(\sigma_A)$. We can apply the definition of the support function to $\theta \in \Theta$, so that $\sigma_{\{\theta\}}(u) = u \cdot \theta$ for $u \in \mathcal{U}$. It is easily verified that for any $A \in \mathcal{A}^c$ and $\theta \in \Theta$, $\sigma_{A+\theta} = \sigma_A + \sigma_{\{\theta\}}$.

We first prove that $\langle \sigma_{\{\theta\}}, \mu \rangle = v \cdot \theta$ for all $\theta \in \Theta$. Fix any $\theta \in \Theta$. Since $A \in \mathcal{A}^\circ$, there exists a $k > 0$ such that $A + k\theta, A - k\theta \in \mathcal{A}^c$. By the translation-linearity of V, we have

$$k(v \cdot \theta) = V(A + k\theta) - V(A) = W(\sigma_{A+k\theta}) - W(\sigma_A).$$

Since $\mu \in \partial W(\sigma_A)$, by part 3 of Lemma 4, $W(\sigma_A) = \langle \sigma_A, \mu \rangle - W^*(\mu)$. Also, by part 2 of the same lemma, $W(\sigma_{A+k\theta}) \geq \langle \sigma_{A+k\theta}, \mu \rangle - W^*(\mu)$. Therefore, we have

$$k(v \cdot \theta) \geq \langle \sigma_{A+k\theta}, \mu \rangle - \langle \sigma_A, \mu \rangle = \langle \sigma_{\{k\theta\}}, \mu \rangle = k \langle \sigma_{\{\theta\}}, \mu \rangle.$$

A similar argument can be used to show that

$$-k(v \cdot \theta) = W(\sigma_{A-k\theta}) - W(\sigma_A) \geq -k \langle \sigma_{\{\theta\}}, \mu \rangle.$$

Hence, we have $k(v \cdot \theta) = k \langle \sigma_{\{\theta\}}, \mu \rangle$, or equivalently, $v \cdot \theta = \langle \sigma_{\{\theta\}}, \mu \rangle$.

We now prove that $\langle \sigma_{\{p\}}, \mu \rangle = v \cdot (p - q)$ for all $p \in \triangle(Z)$. Since $\sum_z u_z = 0$ for $u \in \mathcal{U}$, we have $u \cdot q = 0$ for all $u \in \mathcal{U}$. Clearly, this implies that $\sigma_{\{q\}} = 0$, so that $\langle \sigma_{\{q\}}, \mu \rangle = 0$. For any $p \in \triangle(Z)$, $p - q \in \Theta$, so the above results imply

$$\langle \sigma_{\{p\}}, \mu \rangle = \langle \sigma_{\{p-q\}}, \mu \rangle + \langle \sigma_{\{q\}}, \mu \rangle = \langle \sigma_{\{p-q\}}, \mu \rangle = v \cdot (p - q),$$

which completes the proof. ∎

We showed in Section B.2 that if $q = (1/|Z|, \ldots, 1/|Z|)$, then $\lambda A + (1-\lambda)\{q\} \in \mathcal{A}^\circ$ for any $A \in \mathcal{A}^c$ and $\lambda \in (0,1)$. Therefore, we can use Lemma 23 and the continuity of W to prove the consistency of $\mathcal{M}_W$.

Lemma 24 *If $\mu \in \mathcal{M}_W$, then $\langle \sigma_{\{p\}}, \mu \rangle = v \cdot (p - q)$ for all $p \in \triangle(Z)$.*

Proof: Define $\mathcal{M} \subset \mathcal{M}_W$ by

$$\mathcal{M} \equiv \{\mu \in \mathcal{M}_W : \langle \sigma_{\{p\}}, \mu \rangle = v \cdot (p - q) \text{ for all } p \in \triangle(Z)\}.$$

It is easily verified that $\mathcal{M}$ is a closed subset of $\mathcal{M}_W$ and is therefore compact. We want to show $\mathcal{M}_W \subset \mathcal{M}$, which would imply $\mathcal{M} = \mathcal{M}_W$. By Lemma 22, we only need to verify that $W(\sigma) = \max_{\mu \in \mathcal{M}} \left[\langle \sigma, \mu \rangle - W^*(\mu) \right]$ for all $\sigma \in \Sigma$.

Let $\sigma \in \Sigma$ be arbitrary. For all $\lambda \in (0,1)$, we have $\lambda A_\sigma + (1-\lambda)\{q\} \in \mathcal{A}^\circ$. Note that $\sigma_{\lambda A_\sigma + (1-\lambda)\{q\}} = \lambda \sigma_{(A_\sigma)} + (1-\lambda)\sigma_{\{q\}} = \lambda \sigma$. Therefore, Lemma 23 implies that for all $\lambda \in (0,1)$, $\mathcal{M}_W \cap \partial W(\lambda\sigma) \subset \mathcal{M}$. By Lemma 22, there exists $\mu \in \mathcal{M}_W$ such that $W(\lambda\sigma) = \langle \lambda\sigma, \mu \rangle - W^*(\mu)$, which implies $\mu \in \partial W(\lambda\sigma)$ by part 3 of Lemma 4. Thus $\mathcal{M}_W \cap \partial W(\lambda\sigma) \neq \emptyset$.

Take any net $\{\lambda_d\}_{d \in D}$ such that $\lambda_d \to 1$, and let $\sigma_d \equiv \lambda_d \sigma$, so that $\sigma_d \to \sigma$. From the above, for all $d \in D$ there exists $\mu_d \in \mathcal{M}_W \cap \partial W(\sigma_d) \subset \mathcal{M}$. Since $\mathcal{M}$ is weak* compact, every net in $\mathcal{M}$ has a convergent subnet. Without loss of generality, suppose the net itself converges, so that $\mu_d \xrightarrow{w^*} \mu$ for some $\mu \in \mathcal{M}$. By Lemma 2, the definition of the subdifferential, and the continuity of W, for any $\sigma' \in \Sigma$,

$$\langle \sigma' - \sigma, \mu \rangle = \lim_d \langle \sigma' - \sigma_d, \mu_d \rangle \leq \lim_d \left[W(\sigma') - W(\sigma_d) \right] = W(\sigma') - W(\sigma),$$

which implies $\mu \in \partial W(\sigma)$.[44] Hence, $W(\sigma) = \langle \sigma, \mu \rangle - W^*(\mu)$ by part 3 of Lemma 4. Since $\sigma \in \Sigma$ was arbitrary, this completes the proof. ■

The consistency of $\mathcal{M}_W$ follows immediately from Lemma 24 since for any $\mu, \mu' \in \mathcal{M}_W$ and $p \in \triangle(Z)$, we have

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

Before proving the minimality of $\mathcal{M}_W$, we note the following useful result.

Lemma 25 *For all $\mu \in C(\mathcal{U})^*$ there exists $\sigma \in \Sigma$ such that $W^*(\mu) = \langle \sigma, \mu \rangle - W(\sigma)$.*

Proof: Fix any $\mu \in C(\mathcal{U})^*$. Since W is continuous, the mapping $\sigma \mapsto \langle \sigma, \mu \rangle - W(\sigma)$ is continuous and hence attains a maximum on the compact set Σ. ■

[44] Note that Lemma 2 requires that $\{\mu_d\}_{d \in D}$ be norm bounded, but this follows from the compactness of $\mathcal{M}$ and Alaoglu's Theorem (see Aliprantis and Border, 1999, Theorem 6.25).

We now prove the minimality of $\mathcal{M}_W$.

Lemma 26 *$\mathcal{M}_W$ is minimal.*

Proof: Suppose $\mathcal{M}' \subsetneq \mathcal{M}_W$ is compact and $(\mathcal{M}', W^*|_{\mathcal{M}'})$ is an RFCC representation for $\succsim$. We will show that this is a contradiction.

Define $V' : \mathcal{A}^c \to \mathbb{R}$ as in Equation (6), and define $W' : \Sigma \to \mathbb{R}$ by $W'(\sigma) = V'(A_\sigma)$, so that

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

for all $\sigma \in \Sigma$. By the uniqueness part of Proposition 3, there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $V' = \alpha V - \beta$, which implies $W' = \alpha W - \beta$.

Take any $\bar{\mu} \in \operatorname{argmin}_{\mu \in \mathcal{M}} W^*(\mu)$. Such a $\bar{\mu}$ must exist by the compactness of $\mathcal{M}$ and the lower semi-continuity of W^*. By Lemma 24, for any $p \in \triangle(Z)$,

$$W(\sigma_{\{p\}}) = \max_{\mu \in \mathcal{M}} \left[v \cdot (p - q) - W^*(\mu) \right] = v \cdot (p - q) - W^*(\bar{\mu}).$$

Likewise, by taking $\bar{\mu}' \in \operatorname{argmin}_{\mu \in \mathcal{M}'} W^*(\mu)$, we have that for any $p \in \triangle(Z)$,

$$W'(\sigma_{\{p\}}) = \max_{\mu \in \mathcal{M}'} \left[v \cdot (p - q) - W^*(\mu) \right] = v \cdot (p - q) - W^*(\bar{\mu}').$$

By singleton nontriviality, there exist $p, p' \in \triangle(Z)$ such that $\{p\} \succ \{p'\}$. Thus,

$$W(\sigma_{\{p\}}) - W(\sigma_{\{p'\}}) = v \cdot (p - p') = W'(\sigma_{\{p\}}) - W'(\sigma_{\{p'\}}) > 0,$$

which implies $\alpha = 1$.

Thus $W' = W - \beta$. Since $\mathcal{M}' \subsetneq \mathcal{M}_W$, Lemma 22 requires that there is some $\sigma \in \Sigma$ for which $W(\sigma) \neq W'(\sigma)$. We therefore have $\beta \neq 0$. However, take any $\mu' \in \mathcal{M}'$, and by Lemma 25 there exists $\sigma' \in \Sigma$ such that $W^*(\mu') = \langle \sigma', \mu' \rangle - W(\sigma')$, or equivalently, $W(\sigma') = \langle \sigma', \mu' \rangle - W^*(\mu')$. But then $W'(\sigma') = W(\sigma')$, which requires that $\beta = 0$, a contradiction. ■

We have now completed the proof of Theorem 1.A. The following lemma completes the proof of Theorem 1.B.

Lemma 27 *If W is monotone, then each $\mu \in \mathcal{M}_W$ is positive.*

Proof: $C(\mathcal{U})$ is a Banach lattice (Aliprantis and Border (1999, page 302)) and Σ has the property that if $\sigma, \sigma' \in \Sigma$ then $\sigma \vee \sigma' \in \Sigma$. Therefore by Lemma 3, any $\mu \in \mathcal{N}_W$ must be positive. Since the set of positive measures are weak* closed in $C(\mathcal{U})^*$, we conclude that each measure $\mu \in \mathcal{M}_W = \overline{\mathcal{N}_W}$ is also positive. ■

C Proof of Theorem 2

In the following, let $(\mathcal{M}, c)$ be an RFCC representation of $\succsim$. Let V be as in Equation (6) and define $W : \Sigma \to \mathbb{R}$ by $W(\sigma) \equiv V(A_\sigma)$. Then W is Lipschitz continuous, convex, and it satisfies

$$W(\sigma) = \max_{\mu \in \mathcal{M}} [\langle \sigma, \mu \rangle - c(\mu)] \tag{31}$$

for all $\sigma \in \Sigma$. Since V satisfies 1–3 from Proposition 3, we can use the results of Appendix B.3 as needed.

Lemma 28 *Let $K \geq 0$ and let $\{\mu_d\}_{d \in D}$ be a net in $C(\mathcal{U})^*$ such that (i) $\|\mu_d\| \leq K$ for all $d \in D$, and (ii) $\mu_d \xrightarrow{w^*} \hat{\mu}$ for some $\hat{\mu} \in C(\mathcal{U})^*$. Then $W^*(\mu_d) \to W^*(\hat{\mu})$.*

Proof: By Lemma 25, for each $d \in D$, there exists $\sigma_d \in \Sigma$ such that

$$W^*(\mu_d) = \langle \sigma_d, \mu_d \rangle - W(\sigma_d). \tag{32}$$

Since Σ is compact, there exists a subnet on which $\sigma_d \to \hat{\sigma}$ for some $\hat{\sigma} \in \Sigma$. Without loss of generality, let that subnet be the net itself. By Lemma 2, we then have

$$\langle \sigma_d, \mu_d \rangle \to \langle \hat{\sigma}, \hat{\mu} \rangle. \tag{33}$$

Let $\sigma \in \Sigma$, by the choice of σ_d we have

$$\langle \sigma_d, \mu_d \rangle - W(\sigma_d) \geq \langle \sigma, \mu_d \rangle - W(\sigma).$$

Taking limits above, we obtain

$$\langle \hat{\sigma}, \hat{\mu} \rangle - W(\hat{\sigma}) \geq \langle \sigma, \hat{\mu} \rangle - W(\sigma)$$

by Equation (33) and the continuity of W. Since the above inequality holds for any $\sigma \in \Sigma$, we have that

$$W^*(\hat{\mu}) = \langle \hat{\sigma}, \hat{\mu} \rangle - W(\hat{\sigma}). \tag{34}$$

By Equation (33) and the continuity of W, the limit of the right hand side in Equation (32) is the right hand side in Equation (34). Hence $W^*(\mu_d) \to W^*(\hat{\mu})$. ■

Let $\mu \in \mathcal{M}$. Then, by Equation (31), $W(\sigma) \geq \langle \sigma, \mu \rangle - c(\mu)$, and hence $c(\mu) \geq \langle \sigma, \mu \rangle - W(\sigma)$. Taking the supremum of the right hand side of the latter with respect to $\sigma \in \Sigma$ gives:

$$c(\mu) \geq W^*(\mu) \text{ for all } \mu \in \mathcal{M}. \tag{35}$$

Note also that if $\mu \in \mathcal{M}$ and $\sigma \in \Sigma$ then:

$$W(\sigma) = \langle \sigma, \mu \rangle - c(\mu) \implies \mu \in \partial W(\sigma). \tag{36}$$

To see Equation (36), let $W(\sigma) = \langle \sigma, \mu \rangle - c(\mu)$. For all $\sigma' \in \Sigma$ we have $W(\sigma') \geq \langle \sigma', \mu \rangle - c(\mu)$. Hence $W(\sigma') - W(\sigma) \geq \langle \sigma' - \sigma, \mu \rangle$, which implies $\mu \in \partial W(\sigma)$. We also have

$$\mathcal{M}_W \subset \mathcal{M}. \tag{37}$$

To see Equation (37), let $\mu \in \mathcal{N}_W$. Then there exists $\sigma \in \Sigma$ such that $\partial W(\sigma) = \{\mu\}$. By Equation (36), any maximizer of Equation (31) must be in $\partial W(\sigma) = \{\mu\}$, so μ must be the unique maximizer of Equation (31). In particular $\mu \in \mathcal{M}$, and hence $\mathcal{N}_W \subset \mathcal{M}$. Since $\mathcal{M}$ is closed, $\mathcal{M}_W = \overline{\mathcal{N}_W} \subset \mathcal{M}$.

Lemma 29 *If $\mu \in \mathcal{M}_W$ then $c(\mu) = W^*(\mu)$.*

Proof: First let $\mu \in \mathcal{N}_W$, so there exists $\sigma \in \Sigma$ such that $\partial W(\sigma) = \{\mu\}$. Let $\mu' \in \mathcal{M}$ be a maximizer of Equation (31) for σ. By Equation (36) $\mu' \in \partial W(\sigma)$, so $\mu = \mu'$. Hence μ maximizes Equation (31) for σ, so

$$W(\sigma) = \langle \sigma, \mu \rangle - c(\mu),$$

implying that

$$c(\mu) = \langle \sigma, \mu \rangle - W(\sigma) \leq W^*(\mu).$$

Together with Equations (35) and (37), the above inequality implies that $c(\mu) = W^*(\mu)$.

Now let $\mu \in \mathcal{M}_W$. Then there is a net $\{\mu_d\}_{d \in D}$ in $\mathcal{N}_W$ converging to μ. For each $d \in D$, there is $\sigma_d \in \Sigma$ such that $\partial W(\sigma_d) = \{\mu_d\}$. By Lemma 1 there is an element of $\partial W(\sigma_d)$ with norm less than or equal to K, where $K \geq 0$ denotes a Lipschitz constant of W. We therefore have $\|\mu_d\| \leq K$. Then,

$$W^*(\mu) \leq c(\mu) \leq \liminf_d c(\mu_d) = \liminf_d W^*(\mu_d) = W^*(\mu),$$

where the first inequality follows from Equations (35) and (37), the second inequality follows from lower semi-continuity of c (see Theorem 2.39 in Aliprantis and Border (1999), p43), the third equality follows from the above paragraph, and the final equality follows from Lemma 28. We conclude again that $c(\mu) = W^*(\mu)$. ■

As established in Appendix B.3, $(\mathcal{M}_W, W^*|_{\mathcal{M}_W})$ is an RFCC representation of $\succsim$. By $\mathcal{M}_W \subset \mathcal{M}$ and Lemma 29, the minimality of $(\mathcal{M}, c)$ implies that $\mathcal{M} = \mathcal{M}_W$ and $c = W^*|_{\mathcal{M}_W}$.

To conclude the uniqueness proof, let $(\mathcal{M}, c)$ and $(\mathcal{M}', c')$ be two RFCC representations of $\succsim$. Let V, V', W, and W' be defined accordingly. By the uniqueness part of Proposition 3, there exist $\alpha, \beta \in \mathbb{R}$ with $\alpha > 0$ such that $V' = \alpha V - \beta$. This implies that $W' = \alpha W - \beta$. For any $\mu \in C(\mathcal{U})^*$ and $\sigma, \sigma' \in \Sigma$, note that:

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

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 $c = W^*|_{\mathcal{M}_W}$. This concludes the proof of the theorem.

D Proof of Theorem 3

Given a finite set $N = \{1, \ldots, n\}$, we let $\triangle(N) = \{\alpha \in [0,1]^N : \sum_{i \in N} \alpha_i = 1\}$ denote the simplex over N. In the following, we will always assume without explicit mention that N is endowed with its discrete algebra consisting of all subsets of N, and $\triangle(N)$ is endowed with the Borel σ-algebra $\mathcal{B}$ induced by its Euclidean metric. The integral of an n-dimensional variable is used as a shorthand for the n-tuple of integrals of each dimension of the variable.

Lemma 30 *Let $N = \{1, \ldots, n\}$. Let $\mathcal{F}$ denote the product σ-algebra on $N \times \triangle(N)$ and let $\mathcal{G}$ denote the sub-σ-algebra of $\mathcal{F}$ consisting of events measurable with respect to the second coordinate only, i.e., $\mathcal{G} = \{N \times E \in \mathcal{F} : E \in \mathcal{B}\}$. Let $\alpha \in \triangle(N)$ and π be a probability measure on $(\triangle(N), \mathcal{B})$ that satisfies the* **consistency** *condition:*

$$\int_{\triangle(N)} \lambda \ \pi(d\lambda) = \alpha.$$

Then there exists a probability measure P on $(N \times \triangle(N), \mathcal{F})$ such that:

1. *The marginal of P on N agrees with α, i.e., $P(\{i\} \times \triangle(N)) = \alpha_i$ for all $i \in N$.*

2. *The marginal of P on $\triangle(N)$ agrees with π, i.e., $P(N \times E) = \pi(E)$ for all $E \in \mathcal{B}$.*

3. *The random vector $X : N \times \triangle(N) \to \triangle(N)$ defined by $X(j, \lambda) = \lambda$ for all $(j, \lambda) \in N \times \triangle(N)$ satisfies:*

$$P(\{i\} \times \triangle(N) \,|\, \mathcal{G}) = X_i$$

   P-almost surely for all $i \in N$.[45]

[45] Blackwell (1951) contains a proof of this result for the special case where $\alpha = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$.

Proof: We first define a probability measure $P(\cdot|i)$ on $(\triangle(N), \mathcal{B})$ for each $i \in N$. If $\alpha_i = 0$, fix the probability measure $P(\cdot|i)$ arbitrarily. If $\alpha_i > 0$ then let:

$$P(E|i) = \frac{1}{\alpha_i} \int_E \lambda_i \; \pi(d\lambda) \tag{38}$$

for all $E \in \mathcal{B}$. The consistency condition implies that $P(\cdot|i)$ above is a probability measure. Furthermore

$$\pi(E) = \alpha_1 P(E|1) + \ldots + \alpha_n P(E|n) \tag{39}$$

for all $E \in \mathcal{B}$.

We define the probability measure P on $(N \times \triangle(N), \mathcal{F})$ by:

$$P(F) = \sum_{i \in N} \alpha_i P(\{\lambda \in \triangle(N) : (i, \lambda) \in F\}|i)$$

for all $F \in \mathcal{F}$. The marginal of P on N agrees with α by definition. The marginal of P on $\triangle(N)$ agrees with π by Equation (39).

To verify final claim of the Lemma, fix $i \in N$. Then for any $G = N \times E \in \mathcal{G}$:

$$\int_G X_i(j, \lambda) \; P(dj, d\lambda) = \int_E \lambda_i \; \pi(d\lambda) = P(\{i\} \times E) = P((\{i\} \times \triangle(N)) \cap G),$$

where the first equality follows from the second claim of the Lemma and the second inequality follows from the definition of P. Hence the claim holds by definition of conditional probability.[46] ■

Definition 7 A *CC* representation* is a CC representation $((\Omega, \mathcal{F}, P), \mathbf{G}, U, c)$ such that:

1. $\Omega = N \times \Lambda$ where $N = \{1, \ldots, n\}$ with $n = |Z|$.
2. $U(i, \lambda) = U(i, \lambda')$ for all $i \in N$ and $\lambda, \lambda' \in \Lambda$.
3. $\Lambda = [\triangle(N)]^D$ for some an arbitrary index set D.
4. $\mathcal{F}$ is the product σ-algebra on Ω and $\mathbf{G} = \{\mathcal{G}_d : d \in D\}$ where

   $$\mathcal{G}_d = \{N \times E_d \times [\triangle(N)]^{D \setminus \{d\}} \in \mathcal{F} : E_d \in \mathcal{B}\}$$

   for each $d \in D$.

Proposition 4 *For $V : \mathcal{A} \to \mathbb{R}$, the following are equivalent:*

1. *There exists a monotone RFCC representation such that V is given by Equation* (6)*.*

[46] See for instance Billingsley (1995, p430) for the definition of conditional probability.

2. *There exists a CC representation such that V is given by Equation* (7).

3. *There exists a CC* representation such that V is given by Equation* (7).

Proof: We will prove that "(3) ⇒ (2) ⇒ (1) ⇒ (3)." The implication "(3) ⇒ (2)" holds by definition.

(2) ⇒ (1): Assume that there exists a CC representation $((\Omega, \mathcal{F}, P), \mathbf{G}, U, c)$ where V is given by Equation (7). Then V is monotone and satisfies 1–3 in Proposition 3 in Appendix B.2.[47] Therefore the construction in Appendix B.3 implies that there exists a monotone RFCC representation such that V is given by Equation (6).

(1) ⇒ (3): Assume that there exists a monotone RFCC representation $(\mathcal{M}, c)$ with finite $\mathcal{M}$ such that V is given by Equation (6). Set $D = \mathcal{M}$ and let $n, N, \Lambda, \mathcal{F}$, and $\mathbf{G}$ be as in Definition 7.

Since $\mathcal{M}$ is compact there is $\kappa > 0$ such that $\mu(\mathcal{U}) \leq \kappa$ for all $\mu \in \mathcal{M}$. Define the disc $\mathcal{D} = \cup_{\delta \in [0,\kappa]} \delta\mathcal{U}$ in $\mathbb{R}^Z$. Since $\mathcal{D}$ is compact and $n-1$ dimensional, there exist affinely independent vectors $v^1, \ldots, v^n \in \mathbb{R}^Z$ such that $\mathcal{D} \subset co(\{v^1, \ldots, v^n\})$. Let $U(i, \lambda) = v^i$ for every $i \in N$ and $\lambda \in \Lambda$. By affine independence of $v^1, \ldots, v^n$, for all $u \in \mathcal{D}$, there exist unique coefficients (barycentric coordinates) $\gamma(u) = (\gamma_1(u), \ldots, \gamma_n(u)) \in \triangle(N)$ such that $u = \gamma_1(u)v^1 + \ldots + \gamma_n(u)v^n$. The mapping $\gamma : \mathcal{D} \to \triangle(N)$ is one-to-one and continuous.

[47] It is easy to see that V is monotone, convex, translation-linear, and that there exist $p, q \in \triangle(Z)$ such that $V(\{p\}) > V(\{q\})$. In this footnote, we show that V is Lipschitz continuous. Note that $K = \sum_{z \in Z} \mathbb{E}[|U_z|] > 0$ is finite since U is integrable. Let $\|\cdot\|$ denote the usual Euclidean norm in $\mathbb{R}^Z$. Let $\mathcal{G} \in \mathbf{G}$, define $f_{\mathcal{G}} : \mathcal{A} \to \mathbb{R}$ by

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

Let $A, B \in \mathcal{A}$. Given a state $\omega \in \Omega$, let p^* be a solution of $\max_{p \in A} \mathbb{E}\big[U \big| \mathcal{G}\big](\omega) \cdot p$. By definition of Hausdorff distance there exists $q^* \in B$ such that $\|p^* - q^*\| \leq d_h(A, B)$. Then

$$\begin{aligned} &\max_{p \in A} \mathbb{E}\big[U \big| \mathcal{G}\big](\omega) \cdot p - \max_{q \in B} \mathbb{E}\big[U \big| \mathcal{G}\big](\omega) \cdot q = \mathbb{E}\big[U \big| \mathcal{G}\big](\omega) \cdot p^* - \max_{q \in B} \mathbb{E}\big[U \big| \mathcal{G}\big](\omega) \cdot q \\ &\leq \mathbb{E}\big[U \big| \mathcal{G}\big](\omega) \cdot p^* - \mathbb{E}\big[U \big| \mathcal{G}\big](\omega) \cdot q^* \leq \|\mathbb{E}\big[U \big| \mathcal{G}\big](\omega)\| \times \|p^* - q^*\| \leq \|\mathbb{E}\big[U \big| \mathcal{G}\big](\omega)\| \times d_h(A, B). \end{aligned}$$

Taking the expectation of the above inequality we obtain:

$$f_{\mathcal{G}}(A) - f_{\mathcal{G}}(B) \leq \mathbb{E}\big[\|\mathbb{E}[U \big| \mathcal{G}]\|\big] d_h(A, B).$$

where

$$\mathbb{E}\left[\|\mathbb{E}\big[U \big| \mathcal{G}\big]\|\right] \leq \mathbb{E}\left[\sum_{z \in Z} |\mathbb{E}\big[U_z \big| \mathcal{G}\big]|\right] \leq \mathbb{E}\left[\sum_{z \in Z} \mathbb{E}\big[|U_z| \big| \mathcal{G}\big]\right] = \sum_{z \in Z} \mathbb{E}\left[|U_z|\right] = K.$$

Hence $f_{\mathcal{G}}$ is Lipschitz continuous with a Lipschitz constant K that does not depend on $\mathcal{G}$. Since V is the pointwise maximum of $f_{\mathcal{G}}$ over $\mathcal{G} \in \mathbf{G}$, it is also Lipschitz continuous with the same Lipschitz constant K.

Take an arbitrary $\mu \in \mathcal{M}$. Define the continuous transformation $f_\mu : \mathcal{U} \to \triangle(N)$ by $f_\mu(u) = \gamma(\mu(\mathcal{U})u)$ for all $u \in \mathcal{U}$. Since μ is positive and $\mu(\mathcal{U}) > 0$,

$$\pi_\mu = \frac{1}{\mu(\mathcal{U})} \mu \circ f_\mu^{-1} \tag{40}$$

defines a probability measure on $(\triangle(N), \mathcal{B})$.

Let $\alpha = \int_{\triangle(N)} \beta \, \pi_\mu(d\beta)$, then

$$\sum_{i \in N} \alpha_i v^i = \int_{\mathcal{U}} u \, \mu(du).^{48} \tag{41}$$

In particular, $\alpha = \gamma(\int_{\mathcal{U}} u \, \mu(du))$ is independent of μ by consistency and it is non-zero by singleton non-degeneracy of RFCC representation.

For any $A \in \mathcal{A}$, note that:

$$\int_{\triangle(N)} \max_{p \in A} \left(\sum_{i \in N} \beta_i v^i \right) \cdot p \quad \pi_\mu(d\beta) = \int_{\mathcal{U}} \max_{p \in A} u(p) \, \mu(du).^{49} \tag{42}$$

By Lemma 30 there exists a probability measure P_μ on $N \times \triangle(N)$ such that:

1. The marginal of P_μ on N agrees with α: $P_\mu(\{i\} \times \triangle(N)) = \alpha_i$ for all $i \in N$.
2. The marginal of P_μ on $\triangle(N)$ agrees with π_μ: $P_\mu(N \times E) = \pi_\mu(E)$ for all $E \in \mathcal{B}$.
3. The posterior probability over N conditional on $\mathcal{G} = \{N \times E : E \in \mathcal{B}\}$ is β:

$$P_\mu(\{i\} \times \triangle(N) \,|\, \mathcal{G}) = \beta_i \tag{43}$$

P_μ-almost surely for $(j, \beta) \in N \times \triangle(N)$, for each $i \in N$.

For each $i \in N$, if $\alpha_i > 0$ define the probability measure $P_\mu(\cdot|i)$ on $(\triangle(N), \mathcal{B})$ by

$$P_\mu(E|i) = \frac{1}{\alpha_i} P_\mu(\{i\} \times E)$$

for all $E \in \mathcal{B}$. If $\alpha_i = 0$, fix a probability measure $P_\mu(\cdot|i)$ on $(\triangle(N), \mathcal{B})$ arbitrarily. By Theorem 4.4.6 in Dudley (2002), for each $i \in N$ and nonempty finite subset $\mathcal{M}' \subset \mathcal{M}$, there exists a

[48] To see this, consider the continuous function $g_\mu : \triangle(N) \to \mathbb{R}^Z$ defined by $g_\mu(\beta) = \sum_{i \in N} \beta_i v^i$. Then $\int_{\triangle(N)} g_\mu(\beta) \, [\mu(\mathcal{U})\pi_\mu(d\beta)] = \int_{\mathcal{U}} (g_\mu \circ f_\mu)(u) \, \mu(du)$, by Equation (40) and the change of variables formula. This can be seen to be equivalent to Equation (41) by dividing both sides by $\mu(\mathcal{U})$ and noting that $g_\mu(f_\mu(u)) = \mu(\mathcal{U})u$.

[49] To see this, consider the continuous function $g_\mu : \triangle(N) \to \mathbb{R}$ defined by $g_\mu(\beta) = \max_{p \in A}(\sum_{i \in N} \beta_i v^i) \cdot p$. Then $\int_{\triangle(N)} g_\mu(\beta) \, [\mu(\mathcal{U})\pi_\mu(d\beta)] = \int_{\mathcal{U}} (g_\mu \circ f_\mu)(u) \, \mu(du)$, by Equation (40) and the change of variables formula. This can be seen to be equivalent to Equation (42) by dividing both sides by $\mu(\mathcal{U})$ and noting that $g_\mu(f_\mu(u)) = \max_{p \in A} \mu(\mathcal{U})u \cdot p$.

unique product probability measure $\prod_{\mu\in\mathcal{M}'} P_\mu(\cdot|i)$ on $[\triangle(N)]^{\mathcal{M}'}$ and its associated product σ-algebra. By the Kolmogorov Extension Theorem (see e.g. Corollary 14.27 in Aliprantis and Border (1999)), there exists a unique extension $P(\cdot|i)$ of these finite product probability measures to $\Lambda = [\triangle(N)]^{\mathcal{M}}$ and its associated product σ-algebra.

Define the probability measure P on $(\Omega, \mathcal{F})$ by:

$$P(F) = \sum_{i\in N} \alpha_i P(\{\lambda \in \Lambda : (i,\lambda) \in F\}|i)$$

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

By construction, the marginal of P on the Cartesian product of N and the μ^{th} coordinate of $\Lambda = [\triangle(N)]^{\mathcal{M}}$ is equal to P_μ, i.e.:

$$P(E \times [\triangle(N)]^{\mathcal{M}\setminus\{\mu\}}) = P_\mu(E) \tag{44}$$

for all $E \subset N \times \triangle(N)$ Borel measurable with respect to the product σ-algebra.

By Equations (43) and (44), the posterior probability over N conditional on $\mathcal{G}_\mu$ is $\lambda(\mu)$, i.e., for each $i \in N$:

$$P(\{i\} \times \Lambda \,|\, \mathcal{G}_\mu) = \lambda_i(\mu)$$

P-almost surely for $(j, \lambda) \in \Omega$.

Fix any $\mu \in \mathcal{M}$.

$$\mathbb{E}[U|\mathcal{G}_\mu] = \sum_{i\in N} P(\{i\} \times \Lambda|\mathcal{G}_\mu)\, v^i = \sum_{i\in N} \lambda_i(\mu) v^i \tag{45}$$

for P-almost surely for $(j, \lambda) \in \Omega$.[50]

Since the marginal of P on the Cartesian product of N and the μ^{th} coordinate of $\Lambda = [\triangle(N)]^{\mathcal{M}}$ is P_μ, and the marginal of P_μ on its second coordinate is π_μ, the marginal of P on the μ^{th} coordinate of $\Lambda = [\triangle(N)]^{\mathcal{M}}$ is π_μ. Therefore by Equations (42) and (45), we have that:

$$\mathbb{E}\left[\max_{p\in A} \mathbb{E}\big[U\big|\mathcal{G}_\mu\big] \cdot p\right] = \int_{\mathcal{U}} \max_{p\in A} u(p)\, \mu(du) \tag{46}$$

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

By Equation (46), and defining $\tilde{c}(\mathcal{G}_\mu) = c(\mu)$, V can be expressed as

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

giving the desired CC* representation. ∎

[50]The first equality can be seen by applying Example 34.2 of Billingsley (1999, p446) to each coordinate of U, since U_z is the simple function $\sum_{i\in N} v^i_z I_{\{i\}\times\Lambda}$ for each $z \in Z$.

E Proof of Theorem 4

The necessity of strong IDD is straightforward and left to the reader. For sufficiency, suppose V is defined by Equation (6) for an RFCC representation $(\mathcal{M}, c)$ for the preference $\succsim$. We prove that strong IDD implies c is constant with the following two lemmas:

Lemma 31 *If $\succsim$ satisfies strong IDD, then for any $A \in \mathcal{A}$, $p \in \triangle(Z)$, and $\alpha \in [0,1]$,*

$$V(\alpha A + (1-\alpha)\{p\}) = \alpha V(A) + (1-\alpha)V(\{p\}). \tag{47}$$

Proof: Take $\mathcal{A}^c$, $\mathcal{A}^\circ$, $\mathcal{A}_i$, and $\mathcal{A}'_i$ as defined in Appendix B. It easily verified that for any RFCC representation, the consistency of the measures implies that V is affine on singleton menus:

$$V(\alpha\{q\} + (1-\alpha)\{p\}) = \alpha V(\{q\}) + (1-\alpha)V(\{p\}), \quad \forall p, q \in \triangle(Z).$$

Therefore, Equation (47) holds for all $A \in \mathcal{A}_0$. We prove by induction that Equation (47) holds on $\mathcal{A}_i$ for all $i \geq 0$.

Fix any $i \geq 0$. Our first step is to establish that if Equation (47) holds for all $A \in \mathcal{A}_i$, then it must also hold for all $A \in \mathcal{A}'_i$. For suppose Equation (47) holds on $\mathcal{A}_i$ and $A \in \mathcal{A}'_i$. Then $A \sim B$ for some $B \in \mathcal{A}_i$, and so by stong IDD,

$$\begin{aligned}
V(\alpha A + (1-\alpha)\{p\}) &= V(\alpha B + (1-\alpha)\{p\}) && \text{(by strong IDD)} \\
&= \alpha V(B) + (1-\alpha)V(\{p\}) && \text{(by induction assumption)} \\
&= \alpha V(A) + (1-\alpha)V(\{p\}).
\end{aligned}$$

Next, we establish that if Equation (47) holds for all $A \in \mathcal{A}'_i$, then it must also hold for all $A \in \mathcal{A}_{i+1}$. For suppose Equation (47) holds on $\mathcal{A}'_i$ and $A \in \mathcal{A}_{i+1}$. Then $A = B + \bar{\alpha}\theta^*$ for some $B \in \mathcal{A}'_i$ and $\bar{\alpha} \in \mathbb{R}$. Since the consistency of the measures in an RFCC representation implies that V is translation-linear (as defined in Section B.2), there exists $v \in \mathbb{R}^Z$ such that

$$\begin{aligned}
V(\alpha A + (1-\alpha)\{p\}) &= V\big(\alpha B + (1-\alpha)\{p\} + \alpha\bar{\alpha}\theta^*\big) \\
&= V\big(\alpha B + (1-\alpha)\{p\}\big) + \alpha\bar{\alpha}(v \cdot \theta^*) \\
&= \alpha V(B) + (1-\alpha)V(\{p\}) + \alpha\bar{\alpha}(v \cdot \theta^*) \\
&= \alpha V(B + \bar{\alpha}\theta^*) + (1-\alpha)V(\{p\}) \\
&= \alpha V(A) + (1-\alpha)V(\{p\}).
\end{aligned}$$

By induction, we conclude that Equation (47) holds on $\mathcal{A}_i$ for all $i \geq 0$, and hence by Lemma 14, Equation (47) holds for all $A \in \mathcal{A}^\circ$. Since $\mathcal{A}^\circ$ is dense in $\mathcal{A}^c$ by Lemma 9, the continuity of V implies that the desired property also holds on $\mathcal{A}^c$. Finally, for any $A \in \mathcal{A}$, $co(A) \in \mathcal{A}^c$ and $A \sim co(A)$ by IR. Therefore, by arguments identical to those used above, it follows that Equation (47) holds for all $A \in \mathcal{A}$. ■

Lemma 32 *If V satisfies Equation* (47)*, then c is constant.*

Proof: As in Appendix B.3, define $W : \Sigma \to \mathbb{R}$ by $W(\sigma) = V(A_\sigma)$. By Equation (47) and parts 1 and 2 of Lemma 19, for any $A \in \mathcal{A}$, $p \in \triangle(Z)$, and $\alpha \in [0,1]$,

$$W(\alpha\sigma_A + (1-\alpha)\sigma_{\{p\}}) = \alpha W(\sigma_A) + (1-\alpha)W(\sigma_{\{p\}})$$

It was established in Appendix B.3 that $(\mathcal{M}_W, W^*|_{\mathcal{M}_W})$ is an RFCC representation for $\succsim$. In particular, W satisfies

$$W(\sigma) = \max_{\mu \in \mathcal{M}_W} [\langle \sigma, \mu \rangle - W^*(\mu)]. \tag{48}$$

Since the cost function in an RFCC representation is unique up to an affine transformation, it suffices to show that W^* is constant on $\mathcal{M}_W$. Let $\bar{w} = \min_{\mu \in \mathcal{M}_W} W^*(\mu)$. Note that this minimum is well-defined since W^* is lower semi-continuous and $\mathcal{M}_W$ is compact. Let $\bar{\mu} \in \mathcal{M}_W$ be a minimizing measure, so that $W^*(\bar{\mu}) = \bar{w}$.

We first show that $W^*(\mu) = \bar{w}$ for all $\mu \in \mathcal{N}_W$. Let $\mu \in \mathcal{N}_W$ be arbitrary. By the definition of $\mathcal{N}_W$ and Lemma 4, there exists some $A \in \mathcal{A}$ such that μ is the unique maximizer of Equation (48) at σ_A. That is, $W(\sigma_A) = \langle \sigma_A, \mu \rangle - W^*(\mu) > \langle \sigma_A, \mu' \rangle - W^*(\mu')$ for any $\mu' \in \mathcal{M}_W$, $\mu' \neq \mu$. Now, for any $p \in \triangle(Z)$ and $\alpha \in (0,1)$, choose $\mu' \in \mathcal{M}_W$ that maximizes Equation (48) at $\alpha\sigma_A + (1-\alpha)\sigma_{\{p\}}$. Then,

$$\begin{aligned} \alpha W(\sigma_A) + (1-\alpha)W(\sigma_{\{p\}}) &= W(\alpha\sigma_A + (1-\alpha)\sigma_{\{p\}}) \\ &= \langle \alpha\sigma_A + (1-\alpha)\sigma_{\{p\}}, \mu' \rangle - W^*(\mu') \\ &= \alpha[\langle \sigma_A, \mu' \rangle - W^*(\mu')] + (1-\alpha)[\langle \sigma_{\{p\}}, \mu' \rangle - W^*(\mu')]. \end{aligned}$$

Since $\langle \sigma_A, \mu' \rangle - W^*(\mu') \leq W(\sigma_A)$ and $\langle \sigma_{\{p\}}, \mu' \rangle - W^*(\mu') \leq W(\sigma_{\{p\}})$, the above equation implies that we must in fact have $\langle \sigma_A, \mu' \rangle - W^*(\mu') = W(\sigma_A)$ and $\langle \sigma_{\{p\}}, \mu' \rangle - W^*(\mu') = W(\sigma_{\{p\}})$. By the choice of A, the former implies $\mu' = \mu$. Therefore, the latter implies

$$\langle \sigma_{\{p\}}, \mu \rangle - W^*(\mu) = W(\sigma_{\{p\}}) \geq \langle \sigma_{\{p\}}, \bar{\mu} \rangle - \bar{w}.$$

Consistency implies $\langle \sigma_{\{p\}}, \mu \rangle = \langle \sigma_{\{p\}}, \bar{\mu} \rangle$, and therefore $W^*(\mu) \leq \bar{w}$. Since $\bar{w}$ is the minimum of W^* on $\mathcal{M}_W$, we have $W^*(\mu) = \bar{w}$.

The proof is completed by showing that $W^*(\mu) = \bar{w}$ for all $\mu \in \mathcal{M}_W$. If $\mu \in \mathcal{M}_W$, then there exists a net $\{\mu_d\}_{d \in D}$ in $\mathcal{N}_W$ such that $\mu_d \xrightarrow{w^*} \mu$. Since each μ_d is in $\mathcal{N}_W$, our previous arguments imply that $W^*(\mu_d) = \bar{w}$. Since W^* is lower semi-continuous, it follows that $W^*(\mu) \leq \liminf_d W^*(\mu_d) = \bar{w}$. Since $\bar{w}$ is minimal, we have $W^*(\mu) = \bar{w}$. ■

F Interpretation of ACP

Let $A, B \in \mathcal{A}$ and $\lambda \in (0,1)$. The complete contingent planning problem associated with (λ, A, B) can be described as follows: The individual is asked to chose a pair $(p,q) \in A \times B$ which returns the lottery p with probability λ and the lottery q with probability $1-\lambda$. Therefore each contingent plan $(p,q) \in A \times B$ eventually returns a compound lottery $\lambda p + (1-\lambda)q$ in the menu $\lambda A + (1-\lambda)B$. If, conversely, for each lottery $r \in \lambda A + (1-\lambda)B$ there exists a unique contingent plan that returns r, then the complete contingent planning problem is identical to a choice out of the menu $\lambda A + (1-\lambda)B$.

We say that the triple (λ, A, B) is *invertible* if for all $r \in \lambda A + (1-\lambda)B$ there exist unique $p \in A$ and $q \in B$ for which $r = \lambda p + (1-\lambda)q$. As argued above, the interpretation of the menu $\lambda A + (1-\lambda)B$ as a complete contingent planning problem is justified if the triple (λ, A, B) is invertible. The next two Lemmas show that our results remain unchanged even if we only impose our ACP condition on invertible triples.

Lemma 33 *Let $A, B \in \mathcal{A}$, $\lambda \in (0,1)$, and $\varepsilon > 0$. Then there exist $A', B' \in \mathcal{A}$ such that $d_h(A, A') < \varepsilon$, $d_h(B, B') < \varepsilon$, and (λ, A', B') is invertible.*

Proof: Let $A, B \in \mathcal{A}$, $\lambda \in (0,1)$, and $\varepsilon > 0$. It is easy to see that (λ, A, B) is invertible if $|Z| = 1$. Therefore assume without loss of generality that $|Z| \geq 2$. For any $p \in \triangle(Z)$, let $\|p\|$ denote the Euclidean norm of p in $\mathbb{R}^Z$. Let $N_\varepsilon(p) = \{q \in \triangle(Z) : \|p - q\| < \varepsilon\}$ denote the open ε-ball around p relative to $\triangle(Z)$. Since $\{N_\varepsilon(p) : p \in A\}$ is an open covering of A and A is compact, there is a finite subset A' of A such that $\{N_\varepsilon(p) : p \in A'\}$ covers A. Note that by construction $d_h(A, A') < \varepsilon$.

Similarly there are finitely many lotteries $q_1, \ldots q_n \in B$ such that $N_\varepsilon(q_1), \ldots, N_\varepsilon(q_n)$ cover B. We will construct the desired B', by inductively selecting a $q'_i \in N_\varepsilon(q_i)$ and making sure at each step that $(\lambda, A', \{q'_1, \ldots, q'_i\})$ is invertible. Let $q'_1 = q_1$, then clearly $(\lambda, A', \{q'_1\})$ is invertible. Suppose that $1 \leq i < n$, $(\lambda, A', \{q'_1, \ldots, q'_i\})$ is invertible, and define the sets

$$C = \lambda A' + (1-\lambda)\{q'_1, \ldots, q'_i\} \quad \text{and} \quad D = \left\{ -\frac{\lambda}{1-\lambda} p + \frac{1}{1-\lambda} r \in \mathbb{R}^Z : p \in A', r \in C \right\}.$$

Since $|Z| \geq 2$, $N_\varepsilon(q_{i+1})$ is uncountable. Since D is finite there exists $q'_{i+1} \in N_\varepsilon(q_{i+1}) \setminus D$. We claim that $(\lambda, A', \{q'_1, \ldots, q'_i, q'_{i+1}\})$ is invertible. To see this, it is enough to show that

$$C \cap \left(\lambda A' + (1-\lambda)\{q'_{i+1}\}\right) = \emptyset,$$

since by the inductive assumption $(\lambda, A', \{q'_1, \ldots, q'_i\})$ is invertible. Suppose for a contradiction that $r \in C \cap \left(\lambda A' + (1-\lambda)\{q'_{i+1}\}\right)$, then there exists $p \in A'$ such that $r = \lambda p + (1-\lambda)q'_{i+1}$, which can be rewritten as $q'_{i+1} = -\frac{\lambda}{1-\lambda} p + \frac{1}{1-\lambda} r$, a contradiction to $q'_{i+1} \notin D$. Set $B' = \{q'_1, \ldots, q'_n\}$, then $d_h(B, B') < \varepsilon$ and (λ, A', B') is invertible by induction. ■

Axiom 11 (Weak Aversion to Contingent Planning (WACP)) *For any $A, B \in \mathcal{A}$, if $A \succsim B$, $\lambda \in (0,1)$, and (λ, A, B) is invertible, then $A \succsim \lambda A + (1-\lambda)B$.*

Lemma 34 *If $\succsim$ satisfies weak order, WACP, and continuity, then it also satisfies ACP.*

Proof: Let $A, B \in \mathcal{A}$, $A \succsim B$, and $\lambda \in (0,1)$. By Lemma 33, for each integer n, there exist $A_n, B_n \in \mathcal{A}$ such that $d_h(A_n, A) < \frac{1}{n}$, $d_h(B_n, B) < \frac{1}{n}$, and (λ, A_n, B_n) is invertible. There is a subsequence n_k for which $A_{n_k} \succsim B_{n_k}$ for all k or $B_{n_k} \succsim A_{n_k}$ for all k. Suppose first that the former is true, then by Weak ACP, $A_{n_k} \succsim \lambda A_{n_k} + (1-\lambda)B_{n_k}$ for all k. Continuity implies that in the limit $A \succsim \lambda A + (1-\lambda)B$. Suppose finally that the subsequence is such that $B_{n_k} \succsim A_{n_k}$ for all k. By definition, (λ, A, B) is invertible implies that $(1-\lambda, B, A)$ is invertible, hence $B_{n_k} \succsim (1-\lambda)B_{n_k} + \lambda A_{n_k}$ for all k. Continuity implies that in the limit $B \succsim (1-\lambda)B + \lambda A$, which yields the desired conclusion since $A \succsim B$. ∎

References

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

Bergemann D., and J. Valimaki (2002): "Information Acquisition and Efficient Mechanism Design," *Econometrica*, 70, 1007–1033.

Bergemann D., and J. Valimaki (2006): "Information in Mechanism Design," Cowles foundation working paper, Yale University.

Blackwell, D. (1951): "Comparisons of Experiments," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability,* University of California Press, 93–102.

Blackwell, D. (1953): "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 24, 265–272.

Conlisk, J. (1996), "Why Bounded Rationality," *Journal of Economic Literature*, 34, 669–700.

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 (2007): "Temptation-Driven Preferences," Mimeo, Boston University.

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

Dudley, R. (2002): *Real analysis and Probability.* Cambridge: Cambridge University Press.

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*, forthcoming.

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

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.

Luenberger, D. G. (1969): *Optimization by Vector Space Methods.* New York: John Wiley and Sons, Inc.

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

Mas-Colell, A. (1986): "The Price Equilibrium Existence Problem in Topological Vector Lattices," *Econometrica*, 54, 1039–1054.

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

Olszewski, W. (2007): "Preferences Over Sets of Lotteries," *Review of Economic Studies*, 74, 567–595.

Persico, N. (2000): "Information Acquisition in Auctions," *Econometrica*, 68, 1007–1033.

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. (2007), "Anticipating Regret: Why Fewer Options May Be Better," Mimeo, Northwestern University.

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