Disentangling Strict and Weak Choice in Random Expected Utility Models

Evan Piermont* and Roee Teper†

September 14, 2020

Abstract

We ask when a random choice rule without any axiomatic restrictions can be the result of random expected utility maximization in conjunction with a tie breaking rule. To answer this question we introduce and axiomatize the notion of a *choice capacities* (CCs), the (not-necessarily observable) family of sub-additive set functions representing the frequency of choice by *strict* maximization. The set of choice capacities is in bijection with the set of random utility models, and, critically, the set of CCs that dominate a random choice rule characterizes the set of random utility models consistent with it under some tie-breaking rule. We provide conditions on the random choice rule so as to constructively identify the consistent random utility model that produces indifference with the least frequency.

*Royal Holloway, University of London, Department of Economics, `evan.piermont@rhul.ac.uk`

†The authors thank Jay Lu for his thoughtful comments, which improved the manuscript and its exposition.

1 Introduction

Often, the characteristics of agents are observed only in coarse manner; a population of observationally identical decision makers might take distinct actions. From the analyst's perspective, choice appears to be random. The observable data is embodied by a *Random Choice Rule* (RCR): a measure μ_D for each decision problem D; $\mu_D(x)$ represents the frequency that x was chosen from D.

When the set of consumption objects has a linear structure, *Random Expected Utility Models* (REUs)—a set of expected utility functions and a probability measure thereover—are a powerful and tractable tool in organizing and interpreting the observed data. We say μ maximizes the REU if $\mu_D(x)$ is the probability of a utility function u such that $x = \arg\max_{z \in D} u(z)$.

If indifference is permitted, however, a REU no longer determines choice frequencies—if with positive probability $u(x) = u(y)$ then probability x is chosen from $D = \{x, y\}$ is undefined by the REU, as it depends on how ties are broken. Gul and Pesendorfer (2006a) (GP), who introduced the modern decision theoretic machinery of random choice, dealt with this limitation by axiomatizing the maximization of *regular* REUs, those REUs such that ties occur with probability 0. The elimination of ties is achieved via an extremeness axiom asserting that only the extreme points of a decision problem are ever chosen. Lu (2016a) extends this methodology to allow ties with probability 0 or probability 1; Even in models where ties are part of the *interpretation* of the model, the tie breaking procedure is often assumed to correspond to a regular REU. The resulting choice data (i.e., the RCR primitive of the model) corresponds to a regular REU (Ahn and Sarver, 2013; Gul and Pesendorfer, 2006b) and is observationally equivalent to a 1 stage model without any ties. Frick et al. (2017) and Lu and Saito (2019) entertain dynamic environments but handle tie-breaking in ways analogous to the aforementioned static models.

In this paper, we consider RCRs without any axiomatic restrictions, extremeness or otherwise, and (i) characterize the set of REUs that could have generated them under some tie-breaking procedure, (ii) outline the relation between properties of tie-breaking rules and properties of the RCRs there generated, and (iii) study the REUs that rely on indifference the least in order to explain the observed data.

Example. Let $\{a, b\}$ be a set of prizes, with P the set of lotteries thereon. Consider the following observations for decision problems $D = \{a, b\}$ and $D' = \{a, b, c = \frac{1}{2}a + \frac{1}{2}b\}$:

$$\mu_D : \begin{cases} a & \mapsto \frac{2}{3} \\ b & \mapsto \frac{1}{3} \end{cases} \qquad \mu_{D'} : \begin{cases} a & \mapsto \frac{1}{2} \\ b & \mapsto \frac{1}{4} \\ c & \mapsto \frac{1}{4} \end{cases}$$

This random choice rule does not satisfy GP's extremeness axiom and therefore does not maximize any *regular* random expected utility model.

Once we entertain the possibility of arbitrary tie breaking, the data can be explained: Letting $[\alpha, \beta]$ denotes a utility of α for a and β for b, set $u_1 = [1, 0]$. $u_2 = [-1, 0]$ and $u_3 = [0, 0]$. Then let ξ be the measure given utility indices given by $\xi(u_1) = \frac{1}{2}$, $\xi(u_2) = \frac{1}{4}$, and $\xi(u_3) = \frac{1}{4}$. Let τ be a tie breaking rule—itself a random choice rule—such that $\tau_{\{a,b\}}(a) = \frac{2}{3}$ and $\tau_{\{a,b,c\}}(c) = 1$. Then ξ together with the tie-breaking rule τ generate the observed choice data μ. ∎

Characterizing the set of REUs consistent with the observed data is facilitated by the theory of *choice capacities* (CCs). A choice capacity is family of non-additive set functions defined over the elements of each decision problem. CC's generalize random choice rules, dispensing with the additivity requirement.

Every REU, ξ, defines a unique CC that determines the maximal frequency of choice consistent with ξ. Specifically, for an REU ξ, $\rho^\xi_D : 2^D \to [0,1]$ is defined by

$$\rho^\xi_D(A) = \xi(\{u \mid A \cap \arg\max_D u \neq \varnothing\}). \qquad \text{(MAX)}$$

In this paper, we extended the GP model, providing necessary and sufficient conditions for a CC such to maximizes a REU, according to (MAX). In particular, the (implicit) additivity axiom of GP is replaced by *convex modularity* stating the difference between $\rho_D(A) + \rho_D(B)$ and $\rho_D(A \cup B)$ is exactly $\rho_D(\alpha A + (1-\alpha)B)$, whenever A, B and $\alpha A + (1-\alpha)B$ are all subsets of D. In words, the magnitude of the non-additivity is controlled by the value of convex combinations. Notice, if ρ was additive, then when A and B are disjoint, it must be that $\rho_D(\alpha A + (1-\alpha)B) = 0$—interior points can never be chosen, recovering the extremeness of the GP model.

Example (continued). Taking ξ from earlier in the example, we can construct ρ^ξ, according to (MAX). We have $\arg\max_{D'}(u_1) = \{a\}$, $\arg\max_{D'}(u_2) = \{b\}$ and $\arg\max_{D'}(u_3) = \{a, b, c\}$. So $\rho^\xi_{D'}(a) = \xi(\{u_1, u_3\}) = \frac{3}{4}$. Likewise, we have $\rho^\xi_{D'}(\{a, c\}) = \frac{3}{4}$, $\rho^\xi_{D'}(b) = \rho^\xi_{D'}(\{b, c\}) = \frac{1}{2}$, $\rho^\xi_{D'}(c) = \frac{1}{4}$, $\rho^\xi_{D'}(\{a, b\}) = \rho^\xi_{D'}(\{a, b, c\}) = 1$. It is easy to see that for any $A \subseteq D$, we have $\rho^\xi_D(A) = \rho^\xi_{D'}(A)$.

We can verify that ρ^ξ satisfies convex modularity: $\rho^\xi_{D'}(a) + \rho^\xi_{D'}(b) - \rho^\xi_{D'}(\{a, b\}) = \frac{3}{4} + \frac{1}{2} - 1 = \frac{1}{4} = \rho^\xi_{D'}(c)$. Notice also that $\rho^\xi_{D''}(A) \geqslant \mu_{D''}(A)$, for all $D'' \in \{D, D'\}$ and all $A \subseteq D''$. ∎

Generalizing the last observation from the example, if μ is generated by ξ under the tie-breaking rule τ then $\rho^\xi_D(A) \geqslant \mu_D(A)$ for all decision problems D and all $A \subseteq D$. This must be: ρ^ξ represents the maximal choice frequency consistent with ξ, and is therefore point-wise larger than μ which represents one particular consistent choice frequency (that associated with breaking ties according to τ).

The converse also holds: if $\rho^\xi_D(A) \geqslant \mu_D(A)$ for all decision problems D and all $A \subseteq D$, then there a exists tie-breaking rule, τ, such that (ξ, τ) generate μ. This relationship, and the bijection between REUs and the class of CCs axiomatized in this paper, fully characterize the set of REUs consistent with μ. An REU ξ is consistent with the observed choice frequencies μ exactly when ρ^{ξ} point-wise dominates μ.

The set of REUs consistent with a given μ is generally quite large: intuitively, if ξ is consistent with μ, one could construct some ξ' which moves ξ-probability to indifference, then appropriately alter the tie-breaking rule to reconstruct the choice data. Because of this, we turn our attention to *minimal* REUs, meaning that there is no other consistent REU that induces indifference less often. In the Example, μ is also trivially generated by $\xi'([0,0]) = 1$ and $\tau' = \mu$. But (ξ, τ) (as defined in the example) is minimal, inducing indifference with probability $\frac{1}{4}$; under ξ' indifference occurs with as probability 1.

Identifying minimal REUs is important for economic decision making. A modeler who is trying to maximize social welfare in an allocation problem is always better off using an REU that minimizes tie-breaking. If the modeler assumes that the agents have a strict preference for x over y when they are in fact indifferent, then the provision of x rather than y is not harmful. But if she assumes they are indifferent between x and y when in fact y is strictly preferred, allocating x is costly to the agents' welfare. Thus, choosing the model which ascribes strict preference whenever possible is the most robust model when making economic decisions that rely on the data.

If $y \in D$ is in the relative interior of D, then it will be a maximizer from D if and only if all elements of D are indifferent to one another.[1] This indicates that the probability that ξ realizes a utility function indifferent to all elements of D is exactly $\rho^{\xi}_D(y)$ where $y \in D$ is in the relative interior of D. Therefore, if ξ realizes indifference across all elements of D more often than ξ' we have $\rho^{\xi}_D(y) > \rho^{\xi'}_D(y)$. All of this is to say, the identification of a minimal REU consistent with μ can be recast as the

[1]Recall, the relative interior are the set of strict convex combinations of the extreme points of A.

search for a CC that (i) satisfies our axiomatization, (ii) dominates μ and (iii) does not point-wise dominate any other CC that satisfies the first two conditions.

A simple compactness argument ensures that a minimal REU exists, but may not be unique. In this paper, we provide a necessary and sufficient condition, *lexicographic preference for hedging*, on tie-breaking rules such that a unique minimal REU exists and can be identified constructively: by first constructing a CC from μ and then appealing to our (constructive) representation result for CCs. Lexicographic preference for hedging postulates, loosely speaking, if with positive ξ-probability A is the set of maximizers for a decision problem, then there exists some (possibly different) decision problem such that indifference is broken in favor of the relatively interior points of A. That is, the tie-breaking procedure hedges between the maximizing elements; it is as if the decision maker fears she might have misspecified her own utility, but that this concern is of lexicographic importance.

Under lexicographic preference for hedging, we can construct a CC that can 'see' indifference, since indifference shows up through the choice of interior points. Making this precise requires a notational investment, but the logic is clear by returning to our example:

Example (continued again). Taking μ from earlier in the example, we can consider the CC, ρ, that assigns to each element the minimal probability it was a maximizer consistent with all the observed data. For example, what is the minimal probability that a *could be* chosen from $D = \{a, b\}$? It is $\frac{3}{4}$, despite the fact that it is not chosen this frequently.

This is because, by nature of the linearity of expected utility, whenever $c = \frac{1}{2}a + \frac{1}{2}b$ was chosen both a and b were maximizers. Thus, in D', a was a maximizer with probability at least $\frac{3}{4}$. Also, if a is a maximizer of D', it is also a maximizer of D.

So we set $\rho_D(a) = \rho_{D'}(a) = \mu_{D'}(a) + \mu_{D'}(c) = \frac{3}{4}$. Furthering this logic, we have $\rho_D(b) = \rho_{D'}(b) = \mu_{D'}(b) + \mu_{D'}(c) = \frac{1}{2}$, $\rho_{D'}(c) = \mu_{D'}(c) = \frac{1}{4}$, and $\rho_D(\{a,b\}) = \rho_{D'}(\{a,b\}) = \mu_D(\{a,b\}) = 1$. Notice that ρ constructed in this manner coincides with ρ^ξ constructed via (MAX): in other words, ξ given by $\xi([1,0]) = \frac{1}{2}$, $\xi([-1,0]) = \frac{1}{4}$, and $\xi([0,0]) = \frac{1}{4}$ is maximized by this constructed ρ. Moreover, ξ, along with the τ given by $\tau_{\{a,b\}}(a) = \frac{2}{3}$ and $\tau_{\{a,b,c\}}(c) = 1$ generate the observed choice data μ.

Notice that this also shows that ξ is minimal, since $\rho^\xi_{D'}(c) = \mu_{D'}(c)$, so any ξ which yields ties between a and b less often (there is only one kind of tie in this simple 2-dimensional case) would necessarily not dominate μ and therefore fail to be consistent with the observed data. ∎

Discussion and Related Literature

The general problem of understanding if a random choice rule is represented by a random utility function is difficult, see for example McFadden (2005) and the references therein. GP dramatically simplified the problem by considering only random *linear* utilities.

In multi-dimensional spaces, necessary for the linear/expected utility formulation, indifferences reflect how an agent is willing to trade off across different dimensions and cannot be relinquished without also dispensing of *continuity* (Nishimura and Ok, 2014). The current methods of dealing with the issue—namely by allowing individual utility realizations to entertain ties, but assuming this happens in any given decision problem with trivial probability—adds complexity to the representation and limits its economic applicability. For example, private information acquisition is a natural example for random choice: given a common prior, randomness enters because different agents observe different signals. However, conditional on a private signal, an agent will necessarily be indifferent between some alternatives, and may be forced to break ties. This excludes many natural and commonly employed signal structures, such as a common prior and a finite number of private signals. Signal structures with a finite number of signals are natural in that they may arise endogenously under optimal information acquisition. By allowing indifference to obtain with arbitrary probability, our model allows for random choice based on any information structure.

Both (the supplement to) GP and Ahn and Sarver (2013) have the feature that they consider indifference and a tie-breaking rule that satisfies all of the GP axioms. This has the implication that observed choices themselves conform to the GP model. In GP this is to the detriment of identification; a non-extreme choice rule followed by a tie-breaking cannot be distinguished from the extreme choice rule it induces. The observational hurdle is overcome by Ahn and Sarver (2013) by observing, in addition to the random choice rule, data from an ex-ante stage of choices over decision problems themselves.

Lin (2018) studies an environment in which the modeler can observe the frequency with which *subsets* of the decision problem are chosen. The interpretation is that the set of maximizers is chosen, obviating any need for tie breaking. There is a formal equivalence between this model and our model of choice capacities, which is shown in Appendix 5.1. Like this paper, Gul and Pesendorfer (2013) consider non-additive 'choice data,' where non-additivity is an indication of indifference, but consider a finite set of alternatives and take the non-additive object a primitive.

Outline

The following section introduces the formal environment and notations. It also provides a recapitulation of the GP model and axioms. Section 3 discusses our general notion of tie-breaking and the consistency between arbitrary random choice rules and choice capacities. It provides the notion of minimal choice capacities and delineates a method for their construction. Then, in Section 4 we furnish a representation result for choice capacities with respect to random expected utility models. All proofs are contained in the Appendix. In Appendix 5.1, we discuss how choice capacities could be indirectly observed from other types of choice data: random choice functions in which decision makers choose subsets of alternatives and choice with status-quo bias or in the presence of a default option.

2 Preliminaries

Choice Capacities and Random Choice Rules

A finite, non-empty subset of $\mathbb{R}^n$ is referred to as a *menu* or *decision problem*. Let $\mathcal{D}$ denote the set of all decision problems with D as a generic element. For a set $A \subset R^n$, let $\text{conv}(A)$ and $\text{int}(A)$ denote the convex hull and the interior of A, respectively. Moreover, if A is convex then let $\text{ext}(A)$ collect the extreme points of A and $\text{ri}(A)$ denote the relative interior of A. When it is not confusing to do so, we will write $\text{ri}(A)$ and $\text{ext}(A)$ to mean $\text{ri}(\text{conv}(A))$ and $\text{ext}(\text{conv}(A))$ for non-convex A. For a decision problem D let $\text{cv}(D)$ denote the set off all decision problems with the same convex hull: $\text{cv}(D) = \{D' \in \mathcal{D} \mid \text{conv}(D') = \text{conv}(D)\}$.

The objects of interest are choice capacities (CCs): $\rho = \{\rho_D\}_{D\in\mathcal{D}}$ where for each D, ρ_D is a *capacity* over $\mathbb{R}^n$. Specifically ρ_D is a *grounded, normalized, and monotone set function*: i.e., $\rho_D : 2^{\mathbb{R}^n} \to [0,1]$ such that $\rho_D(\varnothing) = 0$, $\rho_D(D) = 1$, $\rho_D(A \cup B) \geqslant \rho_D(A)$. Because we define CCs over $2^{\mathbb{R}^n}$ rather than $\mathcal{D}$ we add the requirement that $\rho_D(A) = \rho_D(A \cap D)$ for all $A \subseteq \mathbb{R}^n$. Call a choice capacity, μ, a *random choice rule* (RCR) if μ_D is additive for all $D \in \mathcal{D}$. (N.B., we will use ρ as a generic not-necessarily additive CC, and μ as a generic RCR, when additivity is assumed). Endow the set of CCs with the topology of weak convergence.[2]

[2]I.e., the appropriate generalization of weak convergence for non-additive measures. In this case, we say that $\rho_n \to \rho$ if $\int_{\mathbb{R}^n} f d\rho_n \to \int_{\mathbb{R}^n} f d\rho$ for every bounded continuous $f : \mathbb{R}^n \to \mathbb{R}$ and where integration refers to *Choquet* integration. When restricted to the set of RCRs, the induced topology

GP show (in their Appendix B) that it is without loss of generality to consider only choice problems that are elements of the n dimensional simplex. This lends the interpretation that there is a set of $n+1$ consumption prizes, and decision problems are sets of lotteries thereover—the resulting representation is interpreted as a probability distribution on vNM indices. The advantage of the more general framework is that it allows other interpretations without any change to the primitive. Indeed, we could interpret each dimension as a 'state-of-the-world,' and a decision problem as a collection of Anscombe-Aumann acts (whose outcomes are in utils). Here, the resulting representation is interpreted as an information representation, a la Lu (2016a), a probability distribution over beliefs regarding the state space.

Random Linear Representations

While a RCR (or constructed CC) corresponds to the observable behavior of a population of agents, we interpret the choices as resulting from the maximization of preference. Here, we take a preference to be a linear function over the n dimensions, which, of course, can be represented by a vector in R^n. When interpreting our primitive as choices over lotteries, the linear function corresponds to a utility index over the $n+1$ prizes.[3] When considering our primitive to be choices over Anscombe-Aumann acts, the linear function corresponds to the relative likelihood of each of the n states.[4]

For $(u,x) \in R^n \times R^n$, we write $u(x)$ to denote the inner product of the vectors u and x. For a given decision problem D, let $M(D,u)$ denote the set of vectors

is exactly the usual topology of weak convergence. See Kawabe (2012); Feng and Nguyen (2007); Girotto et al. (2000).

[3]Per usual, we can normalize the utility of the $(n+1)^{th}$ prize to 0, so that the set of utility functions considered is representable within R^n.

[4]Notice, to make sense of this interpretation, we need to ensure that beliefs can be normalized, hence, the linear function must be a strictly positive vector. This requires additional axioms; see Lu (2016b).

that maximize u over the domain D: $M(D, u) = \arg\max_{x \in D} u(x)$. In dual fashion, for $A \subset R^n$, define $N(D, A)$ to be the set of utilities such that something in A is maximal over D according to u: $N(D, A) = \{u \in \mathbb{R}^n \mid A \cap M(D, u) \neq \varnothing\}$. When $A = \{x\}$ is a singleton, then $N(D, \{x\})$ is the normal cone to D at x. The idea being that if an agent entertains preference u when facing problem D, her selection will be in the set $M(D, u)$. Taking this as given, if we observe the agent choose $x \in A$ from decision problem D, it must be that her preference was in $N(D, A)$.

Let Ω be the smallest algebra on $\mathbb{R}^n$ that contains $N(D, A)$ for all (D, A) (where we set $N(D, A) = \varnothing$ if $A \cap D = \varnothing$). Then a *Random Expected Utility Model* (REU) is a finitely additive probability measure over $(\mathbb{R}^n, \Omega)$. Call a REU *regular* if $\xi(N(D, A)) = \xi(\mathrm{ri}(N(D, A)))$ for all (D, A). A regular REU realizes ties with zero probability. Endow the set of REUs with the topology of weak convergence.

Definition. *Let ξ be a REU. Say that ρ maximizes ξ if $\rho_D(A) = \xi(N(D, A))$ for all (D, A).*

GP define the maximization by a RCR in an analogy to the definition above, but impose consistency between μ and ξ only over singleton sets—without additivity, we must impose consistency directly over all subsets of the decision problem.

The set of CCs satisfying our axioms and the set of REUs are in bijection via the map taking an REU to its maximizer.

Proposition 2.1. *Every REU has a unique maximizer and every ρ maximizes at most one REU.*

In other words, every (finitely additive) measure over linear utilities corresponds to a unique CC, without in any way qualifying the set of permissible measures. Although the definition of a CC requires its value be specified on all subsets, if a CC maximizes an REU than it is completely identified over its value on singletons (this follows from Lemma 5). In light of Proposition 2.1 we can denote the unique maximizer of ξ by ρ^{ξ}.

The GP Model

The GP model imposes four axioms over RCRs to ensure that they maximize a *regular* REU, restated here for reference. We write the axioms in terms of our more general model of capacities as they will play will play a role in the representation Theorem for CCs.

Axiom 1—Monotonicity. *Let $D \subseteq D'$, and let $A \subseteq D$. Then*

$$\rho_D(A) \geqslant \rho_{D'}(A).$$

Axiom 2—Extremeness. *If $A \subseteq D \cap D'$ with* $\operatorname{ext}(D) = \operatorname{ext}(D')$, *then*

$$\rho_D(A) = \rho_{D'}(A).$$

Axiom 3—Linearity. *Let $A \subseteq D$. Then*

$$\rho_{\lambda D+z}(\lambda A + z) = \rho_D(A).$$

for all $\lambda > 0$ and $z \in \mathbb{R}^n$.

Axiom 4—Mixture Continuity. *For $D, D' \in \mathcal{D}$, $\rho_{\lambda D+\lambda' D'}$ is continuous in λ, λ' for $\lambda, \lambda' \geqslant 0$.*

These four axioms are necessary and sufficient for the maximization of a random expected utility model. To ensure that the REU is countably-additive we can add the following additional axiom.

Axiom 5—U-Continuity. *For $\{D_n\}_{n\in\mathbb{N}}$ converging to D,*

$$\limsup \rho_{D_n}(C) \leqslant \rho_D(C)$$

for all closed $C \subseteq \mathbb{R}^n$*.*

Extremeness and U-Continuity are written differently then their counterparts in GP. While they are not in general equivalent, they are equivalent under the assumption that ρ is additive (i.e., over the class of RCRs). To see that our extremeness implies GP's extremeness axiom (i.e., that $\rho_D(\text{ext}(D)) = 1$) notice that $\rho_{\text{ext}(D)}(\text{ext}(D)) = 1$ definitionally and that $\text{ext}(\cdot)$ is idempotent. To see that U-Continuity implies GP's continuity axiom, notice that for measures, $\limsup \rho_{D_n}(C) \leqslant \rho_D(C)$ for all closed C implies weak convergence (Theorem 29.1 in Billingsley (1995)).

Theorem 2.2 (Gul and Pesendorfer, 2006)**.** *The RCR* μ *satisfies Monotonicity, Extremeness, Linearity, and Mixture-Continuity if and only if it maximizes a finitely additive regular REU* ξ*. Moreover,* μ *additionally satisfies U-Continuity if and only if* ξ *is countably-additive.*

3 Random Choice Rules and Tie-Breaking

Despite arising from the maximization of an REU, observed choice data, in the form of an RCR, may not satisfy the GP axioms listed above. Specifically, if the REU in question is not regular, then the RCR depends not only on the REU but also on how indifferences are broken by decision makers. When the ambient menu is D, and a utility is drawn such that some subset $D' \subseteq D$ are all valued maximally, then a tie breaking rule specifies how often the elements of D' get selected. Formally:

Definition. *A* tie-breaking rule *is a set of measures* $\tau = \{\tau^D_{D'}\}_{D \in \mathcal{D}, D' \subseteq D}$ *where each* $\tau^D_{D'} \in \Delta(\mathbb{R}^n)$ *with* $\text{supp}(\tau^D_{D'}) = D'$*. Call* τ *a* tie-breaking RCR *if* $\tau^D_{D''} = \tau^{D'}_{D''}$ *for all* $D, D', D'' \in \mathcal{D}, D'' \subseteq D \cap D'$*.*

In the most general case, a tie-breaking rule can be menu dependent, so that τ might make different selections out of D'' when D'' is the realized set of maximizers in D from when D'' is the realized set of maximizers in D'. When τ depends only on the set of maximizers realized, then it is an RCR, and we suppress the super-script notation when not confusing. Note, even when a tie-breaking rule is an RCR, we impose no additional restrictions on the choices—τ need not be monotone, linear, etc.

Now consider an RCR, μ, thought of as the modelers observable information. Since ties are allowed and tie-breaking might occur in a menu-dependent, non-linear manner, we do not impose any of the GP axioms on μ. We say μ is *consistent* with (ξ, τ) where ξ is an REU and τ is a tie-breaking rule, if

$$\mu_D(A) = \int_{R^n} \tau^D_{M(D,u)}(A)\ \xi(du)$$

Say that μ is consistent with ξ if there exists a τ such that μ is consistent with (ξ, τ). In general, given that μ is consistent with ξ, the set of tie-breaking rules that witness this will not be unique.

The interpretation is that μ is generated by the following behavior: first a utility, u, is drawn according to ξ, then ties are broken over the set of maximizers, $M(D, u) = \arg\max_D u$, according to the rule τ. In this sense, we can decompose a RCR into two components: ξ reflects strict preference, ensuring that a utility maximizer is chosen. Then, if the resulting set of elements strictly preferred to everything else is not a singleton, τ reflects the tie-breaking procedure.

Proposition 3.1. *An RCR μ satisfies any subset of $\boldsymbol{AX} = \{$monotone, linear, mixture continuous, extreme$\}$ if and only if it is consistent with some (ξ, τ) where ξ is an REU and τ is a tie-breaking RCR that satisfies the same subset of $\boldsymbol{AX}$.*

Notice, by considering empty subset of $\boldsymbol{AX}$, Proposition 3.1 states that every RCR can be decomposed into the maximization of an REU followed by a tie-breaking rule. Of course, this is trivial, in that we can take the REU to be the trivial preference and let $\tau = \mu$. By considering the whole of $\boldsymbol{AX}$, we see that if τ is a GP RCR, then the resulting observable choice is also a GP RCR, regardless of ξ. This is, of course, the content of Theorem S1 of the material supplemental to Gul and Pesendorfer (2006a).

RCRs and CCs

Proposition 3.1 can be seen as a negative result; once indifferences are considered there are no observable implications of maximizing an REU, unless restrictions are placed on the tie-breaking procedure. Because of this, understanding RCRs becomes a question not of when an RCR is consistent with an REU, but rather, which is the set of REUs that could have possibly generated the observed data. By the above result, this set is non-empty, and, the relationship between RCRs and more general choice capacities characterizes it completely.

Theorem 3.2. *Let μ be a RCR. Then (i) μ is consistent with ξ if and only if $\rho^{\xi}_{D}(A) \geqslant \mu_D(A)$ for every (D, A) and (ii) the set of REUs consistent with every μ is non-empty, convex, and compact.*

So μ is consistent with ξ if and only if it is pointwise dominated by the associated CC. Thus, given a set of observable data, the modeler can check consistency with a given REU by simply constructing the relevant CC and checking dominance. For example, take a modeler who has already estimated a random expected utility model and then observes additional choice data which does not conform to the exact empirical frequencies of the original dataset. The modeler can easily check if the additional data is consistent with the estimated REU (but under a different tie breaking rule) or if choices are fundamentally inconsistent and thus must have arise from a different maximization problem.

Further, if the modeler has two datasets, μ and μ', and has already estimated which REUs are consistent with this data, then Theorem 3.2 states that those REUs jointly consistent with both is again non-empty, convex, and compact. Implicit in Theorem 3.2 is the fact that while the modeler can find an REU to jointly explain both μ and μ', such a model will generally realize indifference more often than necessary to explain either dataset individually. To see this, let P denote the set of CCs dominating μ and P' those dominating μ'. Then $P \cap P'$ are the CCs jointly dominating both datasets. Of course, because P and P' are both *upwards closed*, their intersection will be pointwise higher than the minimal elements in either and therefore will realize ties more frequently.[5] Thus, the modeler faces a tradeoff between finding a unifying model that jointly explains disparate datasets and minimizing the reliance on indifference to explain the each dataset.

Minimal REUs

Generically, the set of REUs consistent with μ is not a singleton. In an effort to minimize the use of indifference, the modeler can select a representative REU that is minimal in the following sense that no other consistent REU realizes ties less often:

Definition. *For an RCR μ, call ξ μ-*minimal* if (i) ξ is consistent with μ and (ii) ρ^ξ does not strictly point-wise dominate ρ^ζ for any ζ consistent with μ.*

There are many different kinds of ties that can occur (i.e., different sets of elements that could be tied), so our notion of minimality states that no consistent REU realizes all kinds less often.

Proposition 3.3. *The set of μ-minimal REUs is non-empty. Moreover, if ξ is the unique μ-minimal REU and ζ is consistent with μ, then ρ^ζ dominates ρ^ξ.*

[5] As discussed in the introduction, if ρ dominates ρ' then ρ realizes ties more often than ρ'.

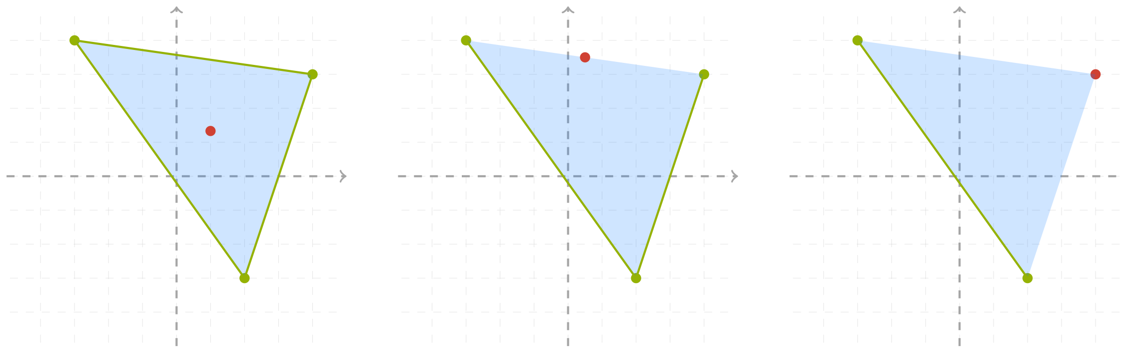

Figure 1: $\mathrm{pi}(D, A)$ for different sets A and the same D. The set A is the red point, $\mathrm{pi}(D, A)$ is in blue, and $D \backslash \mathrm{pi}(D, A)$ is in green.

The first part of the above remark follows directly from compactness arguments. Because this notion of minimality is rather weak, the set of minimal REUs need not be a singleton. However, we will show that for a large class of RCRs there exists a unique minimal REU, which can be found constructively! The second part of the above remark indicates that when ξ is the unique μ-minimal REU, then ξ explains the observed choices and any other explanation must realize *every* kind of indifference weakly more often. Thus, for a modeler who wishes to make recommendations on the basis of strict preference, ξ is the unambiguously best explanation for the for the observed data.

Towards potentially constructing a unique minimal REU, notice first that each RCR μ implicitly defines a capacity over D by associating to each $A \subseteq D$ the minimal probability with which some element of A must have been a maximizer.

Definition. *Let* $\mathrm{pi}(D, A) = \bigcup_{\alpha \in (0,1)} \{\alpha A + (1-\alpha)\mathrm{conv}(D)\}$ *denote the* projective interior *of A in D.*

When an object in $\mathrm{pi}(D, A)$ is chosen, it must be that some object in A maximized the realized preference.[6] Hence, by setting

$$\rho_D(A) = \sup_{D' \in \mathrm{cv}(D)} \mu_{D'}(\mathrm{pi}(D, A)), \tag{$\star$}$$

the constructed capacity ρ represents the minimal probability that A is a maximizer while still being consistent with the observed choices across $\mathrm{cv}(D)$. If this constructed ρ maximizes some ξ then it is easy to see that ξ is uniquely minimal, since it represents the least upper-bound on choice frequencies.

Of course, because tie-breaking behavior might change across different decision problems, the resulting capacities over each decision problem will not in general constitute a CC which maximizes any REU. However, for a large class of tie-breaking rules, this is the case.

Definition. *Say that τ displays a* ***lexicographic preference for hedging*** *if for all $\epsilon > 0$, $D \in \mathcal{D}$ and $A \subseteq D$, there exists a $D' \in \mathrm{cv}(D)$ such that*

$$\tau_{D''}^{D'}(\mathrm{pi}(D, A)) \geqslant 1 - \epsilon,$$

for all $D'' \subseteq D'$ such that $D'' \cap A \neq \varnothing$.

A tie-breaking rule with a lexicographic preference for hedging will, for some decision problem, break ties by choosing strict mixtures of the indifferent elements with arbitrarily high probability. To see how this is captured by the definition, consider $A = \{\alpha x + (1 - \alpha) y\}$, for $x, y \in \mathrm{ext}(D)$. Now, assume that for some u with ξ-positive probability, $x, y \in M(D, u)$. Hence, for some D', when u is realized then A is a maximizer, and the tie-breaking rule places arbitrary probability on $\mathrm{pi}(D, A)$. But notice that all elements of $\mathrm{pi}(D, A)$ are convex combinations that place positive weight on both x and y.

When interior points are chosen, the CC constructed via ($\star$) will be able to "see" indifference, in the sense that for every (D, A), there is is a decision problem in $\mathrm{cv}(D)$

[6] Lemma 2 shows that $\mathrm{pi}(D, A)$ is the union of the interiors of all faces intersecting A.

where $\mathrm{pi}(D, A)$ is chosen with probability arbitrarily close to the probability that some element of A is a maximizer. This latter property is a characterization of maximizing an REU.

Theorem 3.4. *An RCR μ is consistent with ξ and some tie breaking rule displaying a lexicographic preference for hedging, τ, if and only if*

$$\rho_D^\xi(A) = \sup_{D' \in \mathrm{cv}(D)} \mu_{D'}(\mathrm{pi}(D, A)). \tag{$\star$}$$

Moreover, in such cases ξ is the unique μ-minimal REU.

The *only if* direction of Theorem 3.4 indicates that, whenever tie-breaking is sufficiently well-behaved, then the unique minimal REU can be found constructively. Starting with the observable data, μ, the modeler can first constructing a CC via ($\star$). Then, given a CC, the modeler can construct the representing REU by appealing to Theorem 4.1, provided in Section 4 (a representation theorem for CCs which is also constructive).

The *if* direction of Theorem 3.4 can be seen as an behavioral characterization of REU under a lexicographic preference for hedging. The equation ($\star$) transforms any observable data, μ, into a CC. Then, to test the hypothesis of lexicographic preference for hedging, the modeler needs only to check if this CC satisfies the axioms provided in Section 4. Although this procedure requires an additional step of applying ($\star$), it requires no additional data beyond the RCR μ.

While lexicographic preference for hedging is an admittedly abstract property, it is met by many classes of more concrete tie-breaking rules. For example a preference for hedging or uniform randomization.

Definition. *Say τ displays a* ***strong lexicographic preference for hedging*** *if for all $D \in \mathcal{D}$ and $D' \subseteq D$,*

$$\tau_{D'}^D(\mathrm{ri}(D')) = 1,$$

whenever $\mathrm{ri}(D') \cap D \neq 0$.

Such a tie-breaking rule hedges between indifferent options whenever possible.

Definition. *Call* τ ***non-atomic*** *if for all sequences* D^n, E^n *such that* $E^n \subseteq D^n$ *and* $|E^n| \to \infty$, *we have*

$$\tau_{E^n}^{D^n}(x) \to 0$$

for all $x \in \mathbb{R}^n$.

A non-atomic tie breaking rule is one where as the set indifferent alternatives increases without bound, the tie-breaking rule spreads out so that the choice of any single alternative vanishes. A special case of a non-atomic tie-breaking rule is the uniform measure, where τ uniformly chooses across all indifferent options.

Remark 1. *If* τ *is displays a strong lexicographic preference for hedging or is non-atomic it displays a lexicographical preference for hedging.*

4 Axiomatics

Given an RCR μ, Theorem 3.4 provides a method by which the modeler can construct a CC, which, if it maximizes ξ, implies that ξ is the unique μ-minimal REU. But how does the modeler know when the constructed CC maximizes an REU, and if it does, which one? In this section we provide the answer via an axiomatization on CCs. Here, we provide both a test for whether a CC maximizes an REU (if it satisfies the axioms) and if so, a method for constructing it (the constructive representation theorem).

The GP axioms are all still obviously necessary. But, we now need an additional axiom to control how non-additivity can enter ρ.

Axiom 6—CONVEX-MODULARITY. *Let $A, B \subseteq D$ be such that $\alpha A + (1-\alpha)B \subseteq D$ for $\alpha \in (0,1)$. Then*

$$\rho_D(\alpha A + (1-\alpha)B) = \rho_D(A) + \rho_D(B) - \rho_D(A \cup B).$$

Convex-Modularity indicates that the gap between $\rho_D(A \cup B)$ and $\rho_D(A) + \rho_D(B)$ is determined by the convex combinations of the menus. Given our interest in linear utilities, the choice of $\alpha A + (1-\alpha)B$ indicates indifference between A and B; hence any 'non-additivity' of ρ stems directly from indifferences.

Theorem 4.1. *The CC ρ satisfies the GP axioms and Convex-Modularity if and only if it maximizes a finitely additive REU ξ. Moreover, ρ additionally satisfies U-Continuity if and only if ξ is countably-additive.*

The proof of Theorem 4.1 explicitly constructs the measure ξ. As a preliminary, we show two key facts. The first is that ρ is completely determined by its value over singletons. Convex-Modularity places strict limits on the flexibility gained by allowing ρ to be non-additive; if $\rho_D(x)$, $\rho_D(y)$ and $\rho_D(\alpha x + (1-\alpha)y)$ are identified, then so too is $\rho_D(\{x, y\})$; Monotonicity and Extremeness allow us to add the necessary mixtures. Inductively, this determines all choice probabilities. The second fact, replicating results from GP, is that whenever $N(D, \{x\}) = N(D', \{x'\})$ then $\rho_D(x) = \rho_{D'}(x')$.

Armed with these two observations, we construct the measure ξ. For technical reasons, we first identify the measure of the *relative interior* of each $N(D, \{x\})$, then appeal to extension theorems to complete the construction. We proceed inductively on the dimension of the relative interior. To illustrate this we will consider the menu shown in Figure 2. There is a single normal cone of dimension 0, namely $\mathbf{0}$. Since $x \in \text{int}(\text{conv}(D))$ we have $N(D, \{x\}) = \mathbf{0}$, so we can set $\xi(\mathbf{0}) = \rho_D(x)$.

Then, since $N(D', \{y\})$ is 1 dimensional, its boundary is a 0 dimensional convex cone (hence $\mathbf{0}$). As such we can set $\xi(\text{ri}(N(D, \{y\}))) = \rho_D(y) - \rho_D(x)$. Moving up a

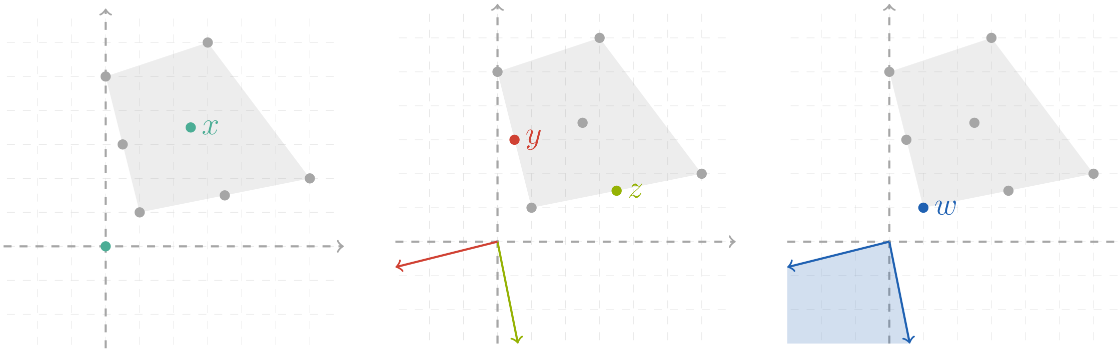

Figure 2: Each panel regards the same fixed menu D (in gray). The first panel $N(D,\{x\})$; the second $N(D,\{y\})$ (in red) and $N(D,\{z\})$ (in green); the third $N(D,\{w\})$.

level, we see that $N(D,\{w\})$ is 2 dimensional and its boundary consists is the union of all three previously identified sets. Therefore $\xi(\mathrm{ri}(N(D,\{w\}))) = \rho_D(w) - \rho_D(y) - \rho_D(z) + \rho_D(x)$. Notice we must add back $\rho_D(x)$ as $N(D,\{x\}) = N(D,\{y\}) \cap N(D,\{z\})$ and was therefore subtracted off twice in prior steps. That this process is well defined and results in a measure representing all choice frequencies is a direct consequence of the above two observations.

The proof equating U-Continuity to countably-additivity is a corollary of Theorem 3 of Lin (2018) and Proposition 5.1.

5 Discussion

5.1 Other constructions of ρ

In this section we describe two other data generating processes that lead to (the identification of) a choice capacity. In each environment, we show how the observed data can be used to construct a choice capacity and extend the Example from the Introduction to fit the data generating process.

Set Valued Choice. In some environments, a modeler might directly observe the entire set of maximizers associated with a decision problem. In other words, the data available to the modeler is the frequency with which each subset of D is chosen—a measure m_D over 2^D. Taking the observed measures $\{m_D\}_{D\in\mathcal{D}}$ as our primitive, we say that $\{m_D\}_{D\in\mathcal{D}}$ *maximizes* a REU, ξ, if for all (D, A),

$$m_D(A) = \xi\big(\{u \mid \arg\max_{z\in D} u(z) = A\}\big).$$

Lin (2018) studies this environment directly, axiomatizing maximization of a REU by placing axioms over $\{m_D\}_{D\in\mathcal{D}}$. Here, we show that this is equivalent to observing a CC. Construct $\{\rho^m_D\}_{D\in\mathcal{D}}$ as follows:

$$\rho^m_D(A) = \sum_{\substack{B\in 2^D,\\ B\cap A\neq\varnothing}} m_D(B). \tag{5.1}$$

To see what is happening, take the following example:

Example. Let ξ and $D = \{a, b, c = \frac{1}{2}a + \frac{1}{2}b\}$ be as in the Example from the Introduction, and consider m_D, where $\{m_D\}_{D\in\mathcal{D}}$ maximizes ξ. First, c is never a unique maximizer $m_D(c) = 0$, a (resp., b) is the unique maximizer if and only if u_1 (resp., u_2) is the realized utility: $m_D(\{a\}) = \xi(u_1) = \frac{1}{2}$ and $m_D(\{b\}) = \xi(u_2) = \frac{1}{4}$. Finally, if c is chosen, then a and b must also be maximizers, so $\{a, b, c\}$ is chosen whenever a is tied with b, or, whenever u_3 is realized: $m_D(\{a, b, c\}) = \xi(u_3) = \frac{1}{4}$. $\{a, b\}$, $\{a, c\}$, $\{b, c\}$ are all chosen with probability 0.

From m_D we can construct ρ^m_D according to (5.1).

$$\begin{aligned}
\rho^m_D(\{a\}) &= m_D\big(\{\{a\}, \{a, c\}, \{a, b\}, \{a, b, c\}\}\big) = \frac{3}{4}\\
\rho^m_D(\{b\}) &= m_D\big(\{\{b\}, \{a, b\}, \{b, c\}, \{a, b, c\}\}\big) = \frac{1}{2}\\
\rho^m_D(\{c\}) &= m_D\big(\{\{c\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}\big) = \frac{1}{4}\\
\rho^m_D(\{a, b\}) &= m_D\big(\{\{a\}, \{b\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}\big) = 1.
\end{aligned}$$

This is exactly ρ_D from the Example in the Introduction. ∎

The relation suggested by the above Example is generalized as follows:

Proposition 5.1. *Let $\{\rho_D\}_{D\in\mathcal{D}}$ maximize ξ, then $\{m_D\}_{D\in\mathcal{D}}$ maximizes ξ if and only if*

$$\rho_D = \rho_D^m,$$

for all $D \in \mathcal{D}$, where ρ^m is defined by (5.1).

So, understanding when a measure over subsets of alternatives arises from a REU is as simple as constructing a choice capacity according to (5.1) and checking the axiomatic characterization below. Moreover, Proposition 5.1 implicitly demonstrates that the set of choice capacities (satisfying our axioms) and the set measures over subsets of choice problems (maximizing some REU) are in bijection—this can be seen by noting that (5.1) is invertible, and appealing to our uniqueness claim, Theorem 2.1.

Status Quo. Often there is an exogenous default implemented in the case of indifference. For example, if the set of acceptable options includes the status quo, then the status quo is implemented. If our primitive observable data is a choice rule defined over a set *and* an observed status quo alternative, then variation in the default can identify a random choice capacity. Assume that these observable data are being generated by a RUM, such that whenever the status quo element is a maximizer, it is definitively chosen (i.e., irrespective of how indifference is broken in other choice problems). In particular: assume for each $x \in \mathbb{R}^n$, and each choice problem D we observe an (additive) random choice rule, ρ_D^x, representing choice from D under status quo x, such that $\rho_D^x(y) \leqslant \rho_D^y(y)$ for all $x, y \in \mathbb{R}^n$, and $D \in \mathcal{D}$. Given a choice problem D, y is chosen more often when it is the status quo than when any other element is.

Say that $\{\rho_D^x\}_{x\in\mathbb{R}^n, D\in\mathcal{D}}$ *maximizes* ξ if

$$\rho_D^x(x) = \xi\big(\{u \mid x \in \arg\max_{z\in D} u(z)\}\big),$$

for all $x \in D$, and $D \in \mathcal{D}$.

Then we can recover a choice capacity as follows:

$$\rho_D^{sq}(\{x\}) = \rho_D^x(\{x\}), \tag{5.2}$$

Although (5.2) defines ρ_D^{sq} only when the choice is a singleton, it is sufficient to identify a unique choice capacity that satisfies our axioms. This result is formally captured by Lemma 5. As always, examples are helpful.

Example. Let ξ and D be as in Example from the Introduction. We will consider some $\{\rho_D^x\}_{x\in\mathbb{R}^n, D\in\mathcal{D}}$ that maximizes ξ. By (5.2) we have

$$\rho_{D'}^{sq}(a) = \rho_{D'}^a(a) = \xi\big(\{u \mid a \in \arg\max_{z\in D} u(z)\}\big) = \xi(\{u_1, u_3\}) = \frac{3}{4}$$
$$\rho_{D'}^{sq}(b) = \rho_{D'}^b(b) = \xi\big(\{u \mid b \in \arg\max_{z\in D} u(z)\}\big) = \xi(\{u_2, u_3\}) = \frac{1}{2}$$
$$\rho_{D'}^{sq}(c) = \rho_{D'}^c(c) = \xi\big(\{u \mid c \in \arg\max_{z\in D} u(z)\}\big) = \xi(\{u_3\}) = \frac{1}{4}.$$

This defines a unique choice capacity, which is, of course, ρ from Example from the Introduction. ∎

Again, we can generalize this observation to a formal result as follows:

Proposition 5.2. *Let $\{\rho_D\}_{D\in\mathcal{D}}$ maximize ξ, then $\{\rho_D^x\}_{x\in\mathbb{R}^n, D\in\mathcal{D}}$ maximizes ξ for each x, if and only if*

$$\rho_D = \rho_D^{sq}$$

where ρ_D^{sq} is as given by (5.2).

Thus, even when the tie breaking rules vary across the population as they depend on some exogenous parameter such as the status quo alternative, identification is still immediate by filtering through our results. There are many other lexicographic costs for which the same argument could be applied in direct analogy. Take as another example choice from a list: when there are multiple acceptable options, then the earliest such option is taken.

References

David S Ahn and Todd Sarver. Preference for flexibility and random choice. *Econometrica*, 81(1):341–361, 2013.

Patrick Billingsley. Probability and measure. 1995. *John Wiley&Sons, New York*, 1995.

Arthur P Dempster. Upper and lower probabilities induced by a multivalued mapping. *The annals of mathematical statistics*, pages 325–339, 1967.

Ding Feng and Hung T Nguyen. Choquet weak convergence of capacity functionals of random sets. *Information Sciences*, 177(16):3239–3250, 2007.

Mira Frick, Ryota Iijima, and Tomasz Strzalecki. Dynamic random utility. 2017.

Bruno Girotto, Silvano Holzer, et al. Weak convergence of bounded, monotone set functions in an abstract setting. *Real Analysis Exchange*, 26(1):157–176, 2000.

Faruk Gul and Wolfgang Pesendorfer. Random expected utility. *Econometrica*, 74 (1):121–146, 2006a.

Faruk Gul and Wolfgang Pesendorfer. Supplement to "random expected utility". *Econometrica*, 74(1):121–146, 2006b.

Faruk Gul and Wolfgang Pesendorfer. Random utility maximization with indifference. Technical report, Working Paper, 2013.

Jun Kawabe. Metrizability of the lévy topology on the space of nonadditive measures on metric spaces. *Fuzzy sets and systems*, 204:93–105, 2012.

Yi-Hsuan Lin. Random expected utility with revealed indifference in choice. 2018. *Working paper.*

Jay Lu. Random choice and private information. *Econometrica*, 84(6):1983–2027, 2016a.

Jay Lu. Supplement to random choice and private information. *Econometrica*, 84(6): 1983–2027, 2016b.

Jay Lu and Kota Saito. Repeated choice: A theory of stochastic intertemporal preferences. 2019. *Working paper*.

Daniel L McFadden. Revealed stochastic preference: a synthesis. *Economic Theory*, 26(2):245–264, 2005.

Hiroki Nishimura and Efe A Ok. Non-existence of continuous choice functions. *Journal of Economic Theory*, 153:376–391, 2014.

R Tyrrell Rockafellar. *Convex Analysis*, volume 28. Princeton University Press, 1970.

Rolf Schneider. *Convex bodies: the Brunn–Minkowski theory*. Number 151. Cambridge university press, 2014.

Larry Alan Wasserman. Prior envelopes based on belief functions. *The Annals of Statistics*, pages 454–464, 1990.

A Proofs

A.1 An Onslaught of Definitions

If A is a convex set and $\text{ext}(A)$ is finite then A is a called at polytope. For a polytope A, let $F \subset A$ be called a face if whenever $\alpha x + (1-\alpha)y \in F$ (for $x, y \in A$) then also $x, y \in F$. Let $\mathbb{F}(A)$ denote the set of all (non-empty) faces of A and $\mathbb{F}^0(A) = \{\text{ri}(F) \mid F \in \mathbb{F}(A)\}$. It is well known that $\mathbb{F}(A)$ is finite and $\mathbb{F}^0(A)$ is a partition of A (Theorems 19.1 and 18.2 of Rockafellar (1970), respectively). A face $F \in \mathbb{F}(A)$ is called exposed if it is the intersection of A with a supporting hyperplane, or, equivalently, if $F = M(A, u)$ for some $u \in \mathbb{R}^n \backslash \mathbf{0}$. Every proper face (i.e., $F \in \mathbb{F}(A)$, $F \neq A$) is an exposed face (Corollary 2.4.2 Schneider (2014)).

If $\lambda K \subseteq K$ for all $\lambda \geqslant 0$ then K is called a *cone*. We say a cone K is generated by A if $K = \{\lambda x \mid x \in A, \lambda \geqslant 0\}$. A cone K is *polyhedral* if it is generated by a polytope; let $\mathcal{K}$ denote all such cones. Let $\mathcal{K}^*$ denote the set of pointed polyhedral cones, those cones with $\mathbf{0} \in \text{ext}(K)$. The face of a polyhedral cone is a polyhedral cone. By proposition 4 of Gul and Pesendorfer (2006a), Ω is the algebra generated by $\mathcal{K}^*$.

Let CB denote an n dimensional cube (the set of unit vectors along each axis). It is true that $\bigcup_{x \in CB} N(CB, x) = \mathbb{R}^n$.

If $A = \{x_1, \ldots x_k\}$ is a set of affinely-independent points then let

$$A^* = \bigcup_{I \subseteq \{1 \ldots k\}} \sum_{i \in I} \frac{x_i}{|I|}$$

The set A^* is a decision problem that has A as the set of extreme points, and contains a point in the relative interior of every face of the decision problem.

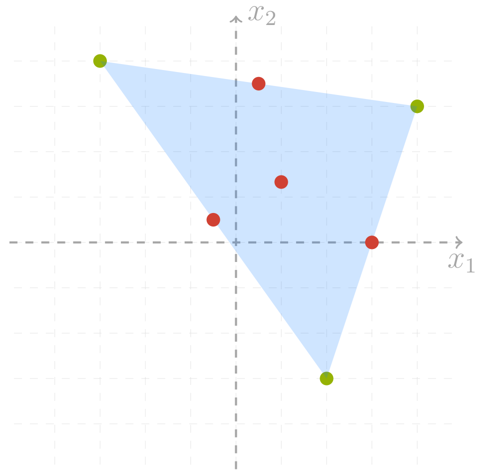

Figure 3: The set A of affinely independent coordinates is given by the green points. The red points are the additional elements of A^*. See, it even looks like a little $*$.

A.2 Lemmas

First we show some properties of the object $\mathrm{pi}(D, A)$. The first lemma shows that $\mathrm{pi}(D, A)$ is the union of the interiors of exposed faces that intersect A. Since exposed faces represent solutions to linear maximization problems, if $x \in \mathrm{pi}(D, A)$ then x being maximal according to u implies that an element of A is also maximal. This property is reflected by ρ, as shown by Lemma 4.

Lemma 1. *Let D be a polytope and $F \in \mathbb{F}(D)$. (i) $u \in N(D, \mathrm{ri}(F))$ implies $F \subseteq M(D, u)$. (ii) $\mathrm{ri}(N(D, \mathrm{ri}(F))) = \{u \in \mathbb{R}^n \mid M(D, u) = F\}$. (iii) $\{\mathrm{ri}(N(D, \mathrm{ri}(F)))\}_{F \in \mathbb{F}(D), F \cap A \neq \varnothing}$ is disjoint and partitions $N(D, A)$.*

Proof. **Part (i)**. Let $u \in N(D, \mathrm{ri}(F))$. Then for some $x \in \mathrm{ri}(F)$, $u(x) \geqslant u(y)$ for all $y \in D$. Let $z \in F$. Since $x \in \mathrm{ri}(F)$, x can be written as $\alpha z + (1 - \alpha)z'$ for some $z' \in F$ and $\alpha \in (0, 1)$. Since x is u maximal, it must be that $u(z) = u(z') = u(x)$, so $z \in M(D, u)$, and hence $F \subseteq M(D, u)$.

Part (ii). Let $u \in \{u \in \mathbb{R}^n \mid M(D, u) = F\}$. Clearly $u \in N(D, \mathrm{ri}(F))$. Moreover, let $v \in N(D, \mathrm{ri}(F))$. By part (i), $F \subseteq M(D, v)$. Hence for $\lambda > 1$ and sufficiently close to 1, $M(D,(1-\lambda)v+\lambda u)=F$. By Theorem 6.4 of Rockafellar (1970), $u \in \mathrm{ri}(N(D,\mathrm{ri}(F)))$.

Let $u \in \mathrm{ri}(N(D,\mathrm{ri}(F))) \subseteq N(D,\mathrm{ri}(F))$. By part (i), $F \subseteq M(D,u)$. By way of contradiction, assume we have some some $y \notin F$ with $y \in M(D,u)$. Since F is a proper face of D, there exists some $v \in \mathbb{R}^n$ such that $F = M(D,v)$. Clearly, $v \in N(D,\mathrm{ri}(F))$. But for all $\lambda > 1$, $((1-\lambda)v+\lambda u)(y) > ((1-\lambda)v+\lambda u)(x)$ for all $x \in \mathrm{ri}(F)$, contradicting the fact that $u \in \mathrm{ri}(N(D,\mathrm{ri}(F)))$. Therefore $M(D,u) \subseteq F$, and hence the two sets are equal.

Part (iii). Disjointness follows from part (ii). That this set partitions $N(D,A)$ follows from the definition of $N(\cdot,\cdot)$ and the fact that $M(D,u) \in \mathbb{F}(D)$ for all u. ∎

Lemma 2. *Let D be a polytope and $A \subseteq D$. Then* $\mathrm{pi}(D,A) = \bigcup_{\{F\in\mathbb{F}(D) | F\cap A\neq\varnothing\}} \mathrm{ri}(F)$.

Proof. Let $x \in \mathrm{pi}(D,A)$. That $A \subseteq \bigcup_{\{F\in\mathbb{F}(D) | F\cap A\neq\varnothing\}} \mathrm{ri}(F)$ follows from the fact that $\mathbb{F}^0(D)$ partitions D. So, take $x = \alpha a + (1-\alpha)y$ with $a \in A$, $y \in D$ and $\alpha \in (0,1)$. Again, since $\mathbb{F}^0(A)$ partitions D, we have that $x \in \mathrm{ri}(F)$ for some $F \in \mathbb{F}(D)$. Moreover, since F is a face, by definition $\{a,y\} \subset F$, so $F \cap A \neq \varnothing$.

Towards the other inclusion, let $x \in \mathrm{ri}(F)$ for some $F \in \mathbb{F}(D)$ with $F \cap A \neq \varnothing$. Let $a \in F \cap A$. Since $\mathrm{ri}(F)$ is convex and relatively open, and $a \in \mathrm{cl}(\mathrm{ri}(F))$, we can write $x = \alpha a + (1-\alpha)y$ for some $y \in \mathrm{ri}(F)$. Hence $x \in \mathrm{pi}(D,A)$. ∎

Lemma 3. *If $N(D,A) \subseteq N(D,B)$, then $A \subseteq \mathrm{pi}(D,B)$.*

Proof. Let $x \in A$. If $x \in \mathrm{ri}(D)$ then the claim holds immediately ($\mathrm{ri}(D) \subseteq \mathrm{pi}(D,x)$ for all x, per Lemma 2). So assume $x \in \mathrm{ri}(F)$ for some $F \in \mathbb{F}(D)$, $F \neq D$. Now every proper face of D is an exposed face, so let u be such that $M(D,u) = F$. Thus, $u \in N(D,A) \subseteq N(D,B) = \bigcup_{y\in B} N(D,y)$; we have $u \in N(D,y)$ for some $y \in B$. This indicates, $y \in M(D,u) = F$ and therefore, by Lemma 2, we have $x \in \mathrm{ri}(F) \subseteq \mathrm{pi}(D,B)$. ∎

Lemma 4. *If ρ satisfies Convex-Modularity, Extremeness, and Monotonicity then for any $A \subseteq D$:*

$$\rho_D(\mathrm{pi}(D, A)) = \rho_D(A).$$

Proof. Let $x \in \mathrm{pi}(D, A)$. We will show that $\rho_D(A \cup x) = \rho_D(A)$, which proves the claim, since for any $y \in \mathrm{pi}(D, A)$, we have also that $y \in \mathrm{pi}(D, A \cup x)$, hence we can proceed iteratively over the finite D. By definition there exists a $z \in D$ and an $\alpha \in (0, 1)$ such that $x \in \alpha A + (1 - \alpha)z$. Let $B = \alpha A + (1 - \alpha)z$ and $D' = D \cup B \cup (\beta A + (1 - \beta)B))$ for some $\beta \in (0, 1)$. We have:

$$\begin{aligned}
\rho_{D'}(A \cup x) &\leqslant \rho_{D'}(A \cup B) \\
&= \rho_{D'}(A) + \rho_{D'}(B) - \rho_{D'}(\beta A + (1 - \beta)(\alpha A + (1 - \alpha)z)) \\
&\leqslant \rho_{D'}(A) + \rho_{D'}(B) - \rho_{D'}((\beta + (1 - \beta)\alpha)A + (1 - \beta)(1 - \alpha)z) \\
&= \rho_{D'}(A)
\end{aligned}$$

The first equality comes from Convex-Modularity and the definition of B; the second inequality from fact that $(\beta + (1 - \beta)\alpha)A + (1 - \beta)(1 - \alpha)z \subseteq \beta A + (1 - \beta)(\alpha A + (1 - \alpha)z)$; the last equality again appeals to Convex-Modularity since both B and $(\beta + (1 - \beta)\alpha)A + (1 - \beta)(1 - \alpha)z$ are mixtures of A and z.

Finally, notice $\mathrm{ext}(D) = \mathrm{ext}(D')$, so by Extremeness: $\rho_D(A \cup x) \leqslant \rho_D(A)$; the other direction is immediate. ∎

We now show that if D is a choice problem then the entirety of ρ_D depends only on the value of singletons in D^*. In other words, if we know the value of $\rho_{D^*}(x)$ for all $x \in D^*$ then we know all choice probabilities out of D.

Lemma 5. *If ρ satisfies Extremeness and Convex-Modularity then ρ_D is uniquely determined by $\{\rho_{D^*}(x) \mid x \in D^*\}$.*

Proof. We will prove $\rho_D(A)$ is identified by induction on the cardinality $A \subseteq D$. Let

$A = \{x\}$. Since D and D^* have the same extreme points, Extremeness states that $\rho_D(x) = \rho_{D^*}(x)$.

Now, assume this was the case for all sets with n or fewer elements, and let $|A| = n+1$. Then $A = B \cup \{x\}$ for some B with $|B| = n$. Extremeness states that $\rho_D(A) = \rho_{D^*}(A)$. Moreover, notice that $(\frac{1}{2}B + \frac{1}{2}x) \subseteq D^*$ by construction. Therefore, appealing to Convex-Modularity delivers,

$$\rho_D(A) = \rho_{D^*}(A) = \rho_{D^*}(B) + \rho_{D^*}(x) - \rho_{D^*}(\frac{1}{2}B + \frac{1}{2}x);$$

each set in question is identified by the inductive hypothesis. ∎

Lemma 6. *If ρ satisfies Monotonicity, Extremeness, and Linearity then $N(D, \{x\}) = N(D', \{x'\})$ implies $\rho_D(x) = \rho_{D'}(x')$.*

Proof. Lemma 1 of Gul and Pesendorfer (2006a). ∎

Lemma 7. *Let $A_1 \dots A_k \subseteq D$. If ρ satisfies Convex-Modularity, then*

$$\rho_{D^*}(\bigcup_{i\leqslant k} A_i) = \sum_{I\subseteq\{1\dots k\}} (-1)^{|I|+1} \rho_{D^*}\left(\sum_{i\in I} \frac{A_i}{|I|}\right).$$

Proof. Follows inductively from successive applications of Convex-Modularity. ∎

Lemma 8. *Let $K_0, K_1 \dots K_k \in \mathcal{K}^*$ be such that* $\mathrm{ri}(K_0) = \bigcup_{i=1\dots k} \mathrm{ri}(K_i)$ *and* $\mathrm{ri}(K_i) \cap \mathrm{ri}(K_j) = \varnothing$ *for $i \neq j \neq 0$. Then (i) $K_0 = \bigcup_{i=1\dots k} K_i$ and (ii)*

$$\sum_{\substack{F\in\mathbb{F}(K_0),\\ F\neq K_0}} \mathbb{1}(\mathrm{ri}(F)) = \sum_{1\leqslant i\leqslant k} \sum_{\substack{F\in\mathbb{F}(K_i),\\ F\neq K_i}} \mathbb{1}(\mathrm{ri}(F)) + \sum_{\substack{I\subseteq\{1\dots k\}\\ |I|\geqslant 2}} (-1)^{|I|+1} \mathbb{1}(\bigcap_I K_i)$$

where $\mathbb{1}$ is the indicator function $\mathbb{R}^n \to \mathbb{R}$ taking a value of 1 on the indicated set and 0 elsewhere.

Proof. Towards (i) Since, $\mathrm{ri}(K_i) \subset \mathrm{ri}(K_0)$ it follows directly that $K_i \subset K_0$ for all i. Thus we need only show that for all $x \in K_0$, $x \in K_i$ for some k. Take $\{x_m\}_{m\in\mathbb{N}} \subset \mathrm{ri}(K_0)$ converging to x. Then there is a subsequence (without relabeling) such that $x_m \in K_i$

for all m (since there are only finitely many K_i). But this subsequence converges to x, so by the fact that K_i is closed, $x \in K_i$ and we are done. Claim (ii) follows directly. ∎

Lemma 9. *For any REU, ξ and choice problem D, define the set of measures*

$$\Gamma(\xi, D) = \left\{ \int_{R^n} t_u \xi(du) : t_u \in \Delta(\mathbb{R}^n),\ \operatorname{supp}(t_u) \subseteq M(D, u) \right\},$$

where $\operatorname{supp}(t)$ *is the support of the measure t. Then, the CC ρ maximizes ξ if and only if $\rho_D(A) = \max_{m \in \Gamma(\xi,D)} m(A)$ for all (D, A).*

Proof. The definition of ρ maximizing ξ indicates that ρ is a plausibility function according to Dempster (1967). The theorem in question follows as a direct corollary of Theorem 2.1 of Wasserman (1990). ∎

A.3 Proofs Omited from the Text

Proof of Proposition 2.1. Let ξ be a REU. Define $\rho_D(A) = \xi(N(D, D \cap A))$. This is clearly a CC and by definition the only such one satisfying $\rho_D(A) = \xi(N(D, D \cap A))$. Let ρ maximize both ξ and ξ'. Then, $\xi(N(D, A)) = \xi'(N(D, A))$ for all (D, A), so by Proposition 4 of GP, for all of Ω^0. Since Ω^0 is a semi-ring there is a unique finitely-additive extension: $\xi = \xi'$. ∎

Proof of Proposition 3.1. The only if direction of the remark follows immediately by setting $\xi(\mathbf{0}) = 1$ and $\tau = \mu$. While trivial, the remark indicated that there is not additional restrictions on an RCR to ensure that it is consistent with an REU and a tie-breaking RCR. The if direction for monotonicity and linearity are obvious. Mixture continuity follows from the fact that $M(\lambda D + \lambda' D, u) = \lambda M(D, u) + \lambda' M(D', u)$, for $\lambda, \lambda' \geqslant 0$. Extremeness follows from the fact that the extreme points of a face of a polytope are extreme points of the polytope itself. ∎

Proof of Theorem 3.2. **Part (i)**. The 'only if' direction follows from Lemma 9. Now, assume that ρ maximizes ξ and dominates μ.

Then for all for all (D, A),

$$\mu_D(A) \leqslant \rho_D(A)$$

and,

$$\mu_D(A) = 1 - \mu_D(A^c) \geqslant 1 - \rho_D(A^c).$$

In the language of Wasserman (1990), μ_D is *comparable* with ρ_D and is therefore contained in $\Gamma(\xi, D)$ (Theorem 2.1 of Wasserman (1990); $\Gamma(\xi, D)$ is defined in Lemma 9).

Therefore, we have that each μ_D can be written as

$$\int_{\mathbb{R}^n} t_u^D \xi(du) \tag{A.1}$$

for some set of measures $\{t_u^D\}_{u \in \mathbb{R}^n}$ such that $\text{supp}(t_u) \subseteq M(D, u)$. For each $D \in \mathcal{D}, D' \subseteq D$ construct the measure

$$\tau_{D'}^D = \begin{cases} \int_{\mathbb{R}^n} t_v^D \xi(du|M(D,u) = D') & \text{if } \xi(u|M(D,u) = D') > 0 \\ \sum_{D'} \frac{\delta_x}{|D'|} & \text{otherwise.} \end{cases}$$

Then $\{\tau_{D'}^D\}_{D \in \mathcal{D}, D' \subseteq D}$ is a tie-breaking rule, and

$$\begin{aligned} \int_{\mathbb{R}^n} \tau_{M(D,u)}^D \xi(du) &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} t_v^D \xi(dv|M(D,v) = M(D,u))\xi(du) \\ &= \int_{\mathbb{R}^n} t_u^D \xi(du) \\ &= \mu_D(A) \end{aligned}$$

so that μ is consistent with (ξ, τ).

Part (ii). Non-Emptiness follows from the fact that $\xi(\mathbf{0}) = 1$ dominates all measures and part (i). We show that the claim holds for a single μ (when I is a singleton); this suffices since the intersection of compact convex sets is compact and convex. For convexity, let ξ^1 and ξ^2 both be consistent with μ. By part (i) this indicates that ρ^1 and ρ^2 both pointwise dominate μ (where ρ^1 and ρ^2 are the unique CCs that maximize ξ^1 and ξ^2). For $\alpha \in (0,1)$, let ρ^α denote the CC maximizing $\alpha\xi^1 + (1-\alpha)\xi^2$. We have for all (D, A),

$$\begin{aligned}
\rho^\alpha_D(A) &= (\alpha\xi^1 + (1-\alpha)\xi^2)(N(D,A)) \\
&= \alpha\xi^1(N(D,A)) + (1-\alpha)\xi^2(N(D,A)) \\
&= \alpha\rho^1_D(A) + (1-\alpha)\rho^2_D(A) \\
&\geqslant \min\{\rho^1_D(A) + \rho^2_D(A)\} \\
&\geqslant \mu_D(A).
\end{aligned}$$

So, ρ^α pointwise dominates μ; by part (i) $\alpha\xi^1 + (1-\alpha)\xi^2$ is consistent with μ.

Towards compactness, we need only to show that the relevant set is closed, since the set of REUs is a compact metric space. Now let $\xi^n \to \xi$ be such that ρ^n (the unique CC maximizing ξ^n) pointwise dominates μ. Since for all (D, A), $N(D, A)$ is closed, Theorem 29.1 in Billingsley (1995) provides

$$\begin{aligned}
\rho_D(A) &= \xi(N(D,A)) \\
&\geqslant \limsup \xi^n(N(D,A)) \\
&\geqslant \limsup \rho^n_D(A) \\
&= \mu_D(A),
\end{aligned}$$

where ρ is the unique CC that maximizes ξ. Therefore, by part (i) μ is consistent with ξ. The set of consistent REUs is therefore closed, and hence compact. ∎

Proof of Proposition 3.3. Let $X(\mu)$ denote the set of REUs consistent with μ. Fix some $\zeta \in X(\mu)$ and (A, D). Then there exists a μ-minimal η such that $\rho^\eta_D(A) \leqslant \rho^\zeta_D(A)$. Indeed, consider $\inf_{\eta \in X(\mu)} \eta_D(A) \leqslant \rho^\zeta(A)$; by the compactness of $X(\mu)$, from Theorem 3.2(ii), this infimum is attained, say by η. It is immediate that η is μ-minimal, since ρ^η cannot strictly pointwise dominate any other element of $X(\mu)$. This proves both claims: the first since $X(\mu)$ is non-empty by Theorem 3.2(ii), and the second by contradiction—if ξ is uniquely μ-minimal but $\zeta \in X(\mu)$ with $\rho^\zeta_D(A) < \rho^\xi_D(A)$ for some (A, D), then by the claim, there is a μ-minimal $\eta \neq \xi$. ∎

Proof of Theorem 3.4. Let μ be consistent with (ξ, τ) where τ displays a lexicographic preference for hedging. Construct ρ via $(\star)$. We have

$$\begin{aligned}
\rho_D(A) &= \sup_{D' \in \mathrm{cv}(D)} \mu_{D'}(\mathrm{pi}(D, A)) && \text{(A.2)}\\
&= \sup_{D' \in \mathrm{cv}(D)} \int_{R^n} \tau^{D'}_{M(D',u)} \xi(du) && \text{(A.3)}\\
&= \sup_{D' \in \mathrm{cv}(D)} \sum_{F \in \mathbb{F}(D)} \xi(\mathrm{ri}(N(D, \mathrm{ri}(F)))) \tau^{D'}_F(\mathrm{pi}(D, A)) && \text{(A.4)}\\
&= \sup_{D' \in \mathrm{cv}(D)} \sum_{F \in \mathbb{F}(D), F \cap A \neq \varnothing} \xi(\mathrm{ri}(N(D, \mathrm{ri}(F)))) \tau^{D'}_F(\mathrm{pi}(D, A)) && \text{(A.5)}\\
&= \sum_{F \in \mathbb{F}(D), F \cap A \neq \varnothing} \xi(\mathrm{ri}(N(D, \mathrm{ri}(F)))) && \text{(A.6)}\\
&= \xi(N(D, A)), && \text{(A.7)}
\end{aligned}$$

where the equality between (A.3) and (A.4) comes from Lemma 1 part (ii), the equality between (A.4) and (A.5) comes from the fact that if $A \cap F = \varnothing$ then $\mathrm{pi}(D, A) \cap F = \varnothing$ (Lemma 2) so $\tau^{D'}_F(\mathrm{pi}(D, A)) = 0$, the penultimate equality comes from the lexicographic preference for hedging of τ, and the final equality comes from Lemma 1 part (iii). So ρ maximizes ξ.

Now assume that ρ constructed via $(\star)$ maximizes ξ. By construction, ρ dominates μ so by Theorem 3.2 there exists a tie breaking rule, τ, such that

$$\mu_D = \int_{R^n} \tau^D_{M(D,u)} \xi(du) \tag{A.8}$$

We have (using the same reasoning as above, but exchanging the order of equalities),

$$\sup_{D' \in \mathrm{cv}(D)} \sum_{\substack{F \in \mathbb{F}(D), \\ F \cap A \neq \varnothing}} \xi(\mathrm{ri}(N(D, \mathrm{ri}(F)))) \tau_F^{D'}(\mathrm{pi}(D, A)) \tag{A.9}$$

$$= \sup_{D' \in \mathrm{cv}(D)} \sum_{F \in \mathbb{F}(D)} \xi(\mathrm{ri}(N(D, \mathrm{ri}(F)))) \tau_F^{D'}(\mathrm{pi}(D, A)) \tag{A.10}$$

$$= \sup_{D' \in \mathrm{cv}(D)} \int_{R^n} \tau_{M(D,u)}^{D} \xi(du) \tag{A.11}$$

$$= \sup_{D' \in \mathrm{cv}(D)} \mu_{D'}(\mathrm{pi}(D, A)) \tag{A.12}$$

$$= \rho_D(A) \tag{A.13}$$

$$= \xi(N(D, A)). \tag{A.14}$$

Moreover, by Lemma 1 part (iii) and the finite additivity of ξ, $\sum_{\substack{F \in \mathbb{F}(D), \\ F \cap A \neq \varnothing}} \xi(N(D, \mathrm{ri}(F))) = \xi(N(D, A))$. Therefore it must be that $\tau_F^{D'}(\mathrm{pi}(D, A)) = 1$ for all $F \in \mathbb{F}(D)$, $F \cap A \neq \varnothing$ such that $\xi(\mathrm{ri}(N(D, \mathrm{ri}(F)))) > 0$. Of course, for any F such that $\xi(\mathrm{ri}(N(D, \mathrm{ri}(F)))) = 0$, we can set $\tau_F^{D'}$ arbitrarily without affecting (A.8), so without loss of generality, we can assume $\tau_F^{D'}(\mathrm{pi}(D, A)) = 1$. Since for all u, $M(D, u) \in \mathbb{F}(D)$, we can also set $\tau_B^{D'}$ for $B \notin \mathbb{F}(D)$ arbitrarily without affecting (A.8). And so, we have found our lexicographic tie-breaking rule. ∎

Proof of Proposition 5.1. Let $\{\rho_D\}_{D \in \mathcal{D}}$ maximize ξ and $\rho_D = \rho_D^m$ (as defined by (5.1)) for all $D \in \mathcal{D}$. We will show that m_D maximized ξ. This proves the claim: the if part directly, the only if part in light of the straightforward fact that if m_D and m'_D both maximize ξ then $m_D = m'_D$, that (5.1) is invertible, and the uniqueness of ρ_D as a maximizer of ξ.

We can rewrite (5.1) as

$$\rho_D(A) = 1 - \sum_{\substack{B \in 2^D, \\ B \subseteq A^c}} m_D(B).$$

The proof in by structural induction in the number of elements in A. Let $A = \{x\}$. Then

$$m_D(\{x\}) = 1 - \rho_D^m(A^c) = 1 - \xi\big(\{u \mid \arg\max_{z\in D} u(z) \cap A^c \neq \varnothing\}\big),$$

which, of course, is exactly the probability of drawing a u that is maximized uniquely by x. Now, assume this holds for all A with n or fewer elements. Then,

$$m_D(A) = 1 - \rho_D^m(A^c) - \sum_{\substack{B\in 2^D,\\ B\subsetneq A}} m_D(B),$$

which by our inductive hypothesis and the definition of ξ is equal to

$$1 - \xi\big(\{u \mid \arg\max_{z\in D} u(z) \cap A^c \neq \varnothing\}\big) - \sum_{\substack{B\in 2^D,\\ B\subsetneq A}} \xi\big(\{u \mid \arg\max_{z\in D} = B\}\big)$$

And, as desired, this last line is the 1 minus the probability that the set of maximizers is larger than A, minus the probability the set is smaller than A. ∎

Proof of Proposition 5.2. First, assume that $\{\rho_D\}_{x\in\mathbb{R}^n, D\in\mathcal{D}}$ maximizes ξ. Then,

$$\rho_D(x) = \xi\big(\{u \mid \arg\max_{z\in D} u(z) \cap \{x\} \neq \varnothing\}\big) = \xi\big(\{u \mid x \in \arg\max_{z\in D} u(z)\}\big) = \rho_D^x(x),$$

and (5.2) holds.

Next assume that (5.2) holds. Therefore, for each $D \in \mathcal{D}$ and $x \in D$ we have

$$\rho_D^x(x) = \rho_D(x) = \xi\big(\{u \mid \arg\max_{z\in D} u(z) \cap \{x\} \neq \varnothing\}\big) = \xi\big(\{u \mid x \in \arg\max_{z\in D} u(z)\}\big),$$

so $\{\rho_D\}_{x\in\mathbb{R}^n, D\in\mathcal{D}}$ maximizes ξ. ∎

A.4 Proof of Theorem 4.1

Necessity

The necessity of the GP axioms are essentially the same as in GP. Recall that $N(D, A\cup B) = N(D, A) \cup N(D, B)$ and $N(D, A) \cap N(D, B) = N(D, \alpha A + (1-\alpha)B)$

for $\alpha \in (0,1)$.

Fix D and $A, B \subseteq D$ and $\alpha \in (0,1)$. Since D is fixed, we will write $N(A)$ rather than $N(D, A)$ for the remainder of the claim.

These above properties imply that

$$\xi(N(A \cup B)) + \xi(N(\alpha A + (1-\alpha)B)) = \xi(N(A) \cup N(B)) + \xi(N(A) \cap N(B)).$$

Now, $N(A) \cup N(B)$ is equal to the (piecewise disjoint) expression

$$\Big(N(A)\backslash(N(A) \cap N(B))\Big) \cup \Big(N(B)\backslash(N(A) \cap N(B))\Big) \cup N(A) \cap N(B),$$

so by the additivity of ξ we have that

$$\begin{aligned}
&\xi(N(A \cup B)) + \xi(N(A) \cap N(B)) = \\
&\xi\Bigg(\Big(N(A)\backslash(N(A) \cap N(B)\Big) \cup \Big(N(A) \cap N(B)\Big)\Bigg) + \\
+\, &\xi\Bigg(\Big(N(B)\backslash(N(A) \cap N(B))\Big) \cup \Big(N(A) \cap N(B)\Big)\Bigg) = \\
&\xi(N(A)) + \xi(N(B))
\end{aligned}$$

indicating that Convex-Modularity must hold.

Sufficiency

First, define $\Omega^0 = \{\mathrm{ri}(K) | K \in \mathcal{K}\}$. GP show that Ω^0 is a semi-ring and that Ω can be reclaimed by taking finite disjoint unions over Ω^0. Thus it suffices to define a finitely additive ξ over Ω^0, as it will extend uniquely to Ω.

We will construct ξ inductively on the dimension of K. Let $\xi(\varnothing) = 0$. Let K be 0 dimensional so that $K = \mathrm{ri}(K) = \{\mathbf{0}\}$. We have that $\mathbf{0} = N(CB^*, \mathbf{0})$. Set

$$\xi(\mathbf{0}) = \rho_{CB^*}(\mathbf{0}).$$

Lemma 6 ensures this is well defined. Now assume that this process has been completed for all K with dimension k or less.

Consider a K of dimension $k+1$. By Proposition 4 of GP, $K = N(D, x)$ for some (D, x), with Lemma 6 ensuring it does not matter which such (D, x) we choose. Set

$$\xi(\text{ri}(K)) = \rho_D(x) - \sum_{F \in \mathbb{F}(K), F \neq K} \xi(\text{ri}(F)), \tag{A.15}$$

where the latter is previously set by the inductive hypothesis and the fact that for all $F \in \mathbb{F}(K)$, $F \neq K$, F is a polyhedral cone such that $dim(F) < k+1$ (corollary 18.1.3 of Rockafellar (1970)).

Lemma 10. *ξ is finitely additive.*

Proof. We will prove the claim by induction on the dimension of the sets in question. When $dim(K) = 0$ there is a single convex cone, to wit, $\mathbf{0}$, so the claim holds trivially. Assume that ξ is finitely additive over any sets of whose union is of dimension m or less. Let $K_0, K_1 \ldots K_k \in \mathcal{K}^*$ be such that $\text{ri}(K_0) = \bigcup_{i=1\ldots k} \text{ri}(K_i)$ and $\text{ri}(K_i) \cap \text{ri}(K_j) = \varnothing$ for $i \neq j \neq 0$, with K_0 of dimension $m+1$. From Lemma 8, $K_0 = \bigcup_{i=1\ldots k} K_i$.

For the first half of the claim, we will follow the general logic of GP's lemma 4. By Proposition 4 of GP, we can find $D_i \in \mathcal{D}$ and $x_i \in D_i$ such that $K_i = N(D_i, x_i)$ for $i = 0 \ldots m$. Let $D = D_0 + D_1 + \ldots + D_k$. For $y \in \bigcup_{j=0}^{k} D_j$, construct the sets:

$$Z(y) = \left\{ z = (z^0 \ldots z^k) \in \prod_{j=0}^{k} D_j \mid z^j = y, \text{ for some } j \right\}$$

and

$$G(y) = \left\{ y' \in D \mid y' = \sum_{j=0}^{k} z^j, z \in Z(y) \right\}.$$

Using Mixture-Continuity, GP show that $\rho_D(G(y)) = \rho_{D_i}(y)$.

Now, by construction $N(D, G(x_0)) = N(D, \bigcup_{i \leqslant k} G(x_i))$. By Lemma 3 this implies $G(x_0) \subseteq \text{pi}(D, \bigcup_{i \leqslant k} G(x_i))$ and $\bigcup_{i \leqslant k} G(x_i) \subseteq \text{pi}(D, G(x_0))$. Therefore, by Lemma 4,

$\rho_D(G(x_0)) = \rho_D(\bigcup_{i \leqslant k} G(x_i))$. Now this implies, by the construction of ξ, via (A.15),

$$\begin{aligned}
\xi(\mathrm{ri}(K_0)) &= \rho_{D_0}(x_0) - \sum_{\substack{F \in \mathbb{F}(K_0), \\ F \neq K_0}} \xi(\mathrm{ri}(F)) \\
&= \rho_D(G(x_0)) - \sum_{\substack{F \in \mathbb{F}(K_0), \\ F \neq K_0}} \xi(\mathrm{ri}(F)) \\
&= \rho_D(\bigcup_{i \leqslant k} G(x_i)) - \sum_{\substack{F \in \mathbb{F}(K_0), \\ F \neq K_0}} \xi(\mathrm{ri}(F)). \qquad \text{(A.16)}
\end{aligned}$$

Appealing to Lemma 7, we can rewrite (A.16):

$$\begin{aligned}
\xi(\mathrm{ri}(K_0)) &= \sum_{I \subseteq \{1 \ldots k\}} (-1)^{|I|+1} \rho_{D*}\left(\sum_{i \in I} \frac{G(x_i)}{|I|}\right) - \sum_{\substack{F \in \mathbb{F}(K_0), \\ F \neq K_0}} \xi(\mathrm{ri}(F)) \\
&= \sum_k \rho_D(G(x_i)) + \sum_{\substack{I \subseteq \{1 \ldots k\} \\ |I| \geqslant 2}} (-1)^{|I|+1} \rho_{D*}\left(\sum_{i \in I} \frac{G(x_i)}{|I|}\right) - \sum_{\substack{F \in \mathbb{F}(K_0), \\ F \neq K_0}} \xi(\mathrm{ri}(F)) \qquad \text{(A.17)} \\
&= \sum_k \xi(\mathrm{ri}(K_i)) + \sum_k \sum_{\substack{F \in \mathbb{F}(K_i), \\ F \neq K_i}} \xi(\mathrm{ri}(F)) + \\
&\quad + \sum_{\substack{I \subseteq \{1 \ldots k\} \\ |I| \geqslant 2}} (-1)^{|I|+1} \rho_{D*}\left(\sum_{i \in I} \frac{G(x_i)}{|I|}\right) - \sum_{\substack{F \in \mathbb{F}(K_0), \\ F \neq K_0}} \xi(\mathrm{ri}(F)) \qquad \text{(A.18)} \\
&= \sum_k \xi(\mathrm{ri}(K_i)) + \sum_k \sum_{\substack{F \in \mathbb{F}(K_i), \\ F \neq K_i}} \xi(\mathrm{ri}(F)) + \\
&\quad + \sum_{\substack{I \subseteq \{1 \ldots k\} \\ |I| \geqslant 2}} (-1)^{|I|+1} \xi\left(\bigcap_I K_i\right) - \sum_{\substack{F \in \mathbb{F}(K_0), \\ F \neq K_0}} \xi(\mathrm{ri}(F)). \qquad \text{(A.19)}
\end{aligned}$$

Notice the equality between (A.17) and (A.18) appeals again to the construction of ξ, via (A.15), and between (A.18) and (A.19) appeals to the well known fact that $N(D, \alpha x + (1 - \alpha y)) = N(D, x) \cap N(D, y)$ for $\alpha \in (0, 1)$.

Finally, notice that Lemma 8 indicates that

$$\sum_{1\leqslant i\leqslant k}\sum_{\substack{F\in\mathbb{F}(K_i),\\ F\neq K_i}}\mathbb{1}(\mathrm{ri}(F))+\sum_{\substack{I\subset\{1\ldots k\}\\ |I|\geqslant 2}}(-1)^{|I|+1}\mathbb{1}\left(\bigcap_I K_i\right)-\sum_{\substack{F\in\mathbb{F}(K_0),\\ F\neq K_0}}\mathbb{1}(\mathrm{ri}(F))=0$$

so, by the inductive hypothesis that ξ is additive over such a domain,[7] we can conclude that the sum of all but the first term of (A.19) equals 0, so that $\sum_k \xi(\mathrm{ri}(K_i)) = \xi(\mathrm{ri}(K_0))$. ∎

Extend ξ from Ω^0 to Ω in the usual way. Since $\bigcup_{y\in CB} N(CB,y) = \mathbb{R}^n$ and $\rho_{CB}(CB) = 1$ we have that ξ is a finitely additive measure.

Lemma 11. *ρ maximizes ξ.*

Proof. Consider (D,x) with $dim(D) = n$ so that $N(D,x) = K \in \mathcal{K}$. Recall, that (i) $\mathbb{F}^0(K)$ partitions K and (ii) $K \in \mathbb{F}(K)$. Therefore, we have

$$\rho_D(x) = \xi(\mathrm{ri}(K)) + \sum_{F\in\mathbb{F}(K),F\neq K}\xi(\mathrm{ri}(F)) = \sum_{A\in\mathbb{F}^0(K)}\xi(A) = \xi(N(D,x)).$$

By Lemma 5, the entirety of ρ_D is determined by ρ's value on singletons, hence ρ maximizes ξ on all n dimensional problems.

Let D of dimension less than n, and consider $D+\alpha CB$, the dimension of the later object is n. We have

$$\begin{aligned}\rho_{D+\alpha CB}(x+\alpha CB) &= \xi(N(D+\alpha CB, x+\alpha CB))\\ &= \xi(\bigcup_{y\in CB} N(D+\alpha CB, x+\alpha y))\\ &= \xi(\bigcup_{y\in CB}(N(D,x)\cap N(CB,y))\\ &= \xi(N(D,x)).\end{aligned}$$

[7]Notice that all sets in question are subsets of the boundary of cones themselves of dimension at most $m+1$. Further, while the value of ξ was not explicitly defined on the cones of dimension m or less, the fact that such objects are partitioned into relative interiors of faces, and the inductive hypothesis of additivity, indicates that ξ is implicitly defined over such objects.

The first equality follows the fact that ρ maximizes ξ for n dimensional problems, the second from the definition of $N(D, A)$, the third from properties of normal cones, and the final equality from the fact that $\bigcup_{y \in CB} N(CB, y) = \mathbb{R}^n$. Appealing to Mixture-Continuity—letting α tend to 0—we conclude that $\rho_D(x) = \xi(N(D, x))$, as desired. ∎