WEIGHTLESS LEARNING AND DECISION MAKERS

MAXWELL B. STINCHCOMBE

Abstract. Sequences of observations drawn from a purely finitely additive, i.e. weightless, probability are always consistent uncountably many probabilities having mutually disjoint supports. Such observations generally have no value for expected utility maximizers. Using the probabilities consistent with observations as a set of priors in ambiguous choice: we can model only a negligible set of problems when there are three or more outcomes; we may partially, or completely, convexify the set of problems that can be modeled using the core of the complete ignorance capacity; and we may model problems that cannot be modeled using the core of any convex capacity.

1. Introduction

A commonly used model for decisions made in the face of uncertainty consists of a set of available actions or strategies, A, a measure space of of states, $(X, \mathcal{X})$, a (normalized) expected utility function, $u : A \times X \to [0, 1]$, and information about X contained in a probability, μ, on $\mathcal{X}$. The decision maker takes an action in the set $a^*(\mu) := \text{argmax}_{a \in A} \int_X u(a, x)\, d\mu(x)$. Often the set A is a class of **acts**, that is, a set of functions on X taking values in a space of consequences, $\mathbb{W}$ (often thought of as wealth), and $u(a(\cdot), x)$ is equal to $v(a(x))$ where $v : \mathbb{W} \to [0, 1]$ is a Bernoulli utility function.

Variants of this model of decisions with μ finitely additive have an extensive history in statistical decision theory. For Bayesian statisticians and decision makers, the μ is the posterior distribution after information has been gathered. In this vein, Heath and Sudderth [23] show that if A is a set of statistical decision rules, then a^* is extended admissible if and only if it is of the form $a^*(\mu)$ for some finitely additive μ. Interest in the dependence of a^* on μ, both from a frequentist and a Bayesian point of view, leads to interest in models of learning about μ.

Learning weightless probabilities. Δ^{fa} denotes the set of finitely additive probabilities, those satisfying $\mu(A \cup B) = \mu(A) + \mu(B)$ for disjoint $A, B \in \mathcal{X}$. Within Δ^{fa} are the purely finitely additive probabilities, felicitously called **weightless** by Maharam [28] because there exists a sequence of disjoint events, F_m, with $\cup_m F_m = X$ even though $\mu(F_m) \equiv 0$. Letting $H_n = \cup_{m \leq n} F_m$, $H_n \uparrow X$ even though $\mu(H_n) \equiv 0$, and, taking complements, if $E_n = \cup_{m \geq n} F_m$, $E_n \downarrow \varnothing$ even though $\mu(E_n) \equiv 1$.

To circumvent severe interpretational difficulties, we study the process of learning a weightless probability by studying the extent to which its values on $\mathcal{X}$ are determined by its values on a subclass of sets, $\mathcal{C}$. Al-Najjar [1] gives a partial indeterminacy result about the values of $\mu(E)$ for $E \notin \mathcal{C}$ when X is a discrete

Date: August 21, 2013
Key words and phrases. Finitely additive probabilities, learning models, Vapnik-Červonenkis classes, expected utility, ambiguous choice theory, cores of convex capacities.

metric space, μ belongs to a particular class of weightless probabilities, and $\mathcal{C}$ is a Vapnik-Červonenkis (VC) class.

He interprets the indeterminacy in two ways. First, he argues that it implies that large finite learning problems are more difficult than learning countably additive probabilities on complete separable metric (Polish) spaces. Second, he suggests that the set of probabilities agreeing with μ on $\mathcal{C}$ provides a learning-based underpinning for the various multiple prior models of preferences over acts in the presence of ambiguity.

An impossibility result. A four-fold extension of the indeterminacy result in [1] shows that learning weightless probabilities is impossible and calls into question the offered interpretations. Al-Najjar showed that (i) when $(X, \mathcal{X})$ is a discrete metric space and its Borel σ-field, and (ii), $\mathcal{C}$ is a VC class, (iii) there exists a weightless density measure μ that has uncountably many weightless perturbations, ν, with (iv) a half-way indeterminacy, $|\mu(E_r) - \nu(E_r)| = r$ for uncountably many E_r provided that $r \in (0, \frac{1}{2}]$. By contrast, we will see that (i′) when $(X, \mathcal{X})$ is any infinite measure space, and (ii′), $\mathcal{C}$ belongs to a class of events much more general than the VC classes, (iii′) for any weightless μ, there are uncountably many ν agreeing with μ on $\mathcal{C}$ with (iv′) complete indeterminacy, $|\mu(E_r) - \nu(E_r)| = r$ for uncountably many E_r for any $r \in (0, 1]$.

Taking these in turn: the change of (i) to (i′) shows that the indeterminacy has nothing to do with the structure of the underlying measure space; the change of (ii) to (ii′) shows that the indeterminacy is unrelated to VC-based learning theory; the change of (iii) to (iii′) shows that all weightless probabilities demonstrate indeterminacy, not just a special class of density measures; and the change of (iv) to (iv′) shows the indeterminacy is total, taking $r = 1$, one cannot even identify the support set of a weightless probability from observations of its values on $\mathcal{C}$.

For expected utility maximizers, not knowing the support of a probability means that one has, in general, no useful information. By contrast, a sequence of i.i.d. observations from a countably additive probability yields sequential/frequentist learning delivering perfect expected utility maximization. Further, the rate of convergence to the optimum provides a measure of the difficulty of the decision problem.

A negligibility result. Al-Najjar [1, §4] suggests, but does not examine, the idea that the set of probabilities consistent with μ on a VC class might be useful as a set of priors in multiple prior models of choice, potentially providing an attractive learning interpretation. The indeterminacy implies that the set of ambiguous choice problems that can be modeled using such sets is, in general, negligibly small. For example, if $\mathcal{C}$ is one of several of the well-known VC classes and $(X, \mathcal{X})$ belongs to the class of discrete metric spaces in which Al-Najjar examines weightless probabilities, then no choice problems involving non-degenerate risk can be modeled using the suggested set of priors, nor can problems involving any non-trivial comparisons of likelihoods. However, for any infinite measure space, there are classes $\mathcal{C}$, probabilities μ, and classes of outcome spaces for which the set of problems that can be modeled this way compares favorably with the set of problems that can be modeled using the core of a convex capacity as the set of priors.

1.1. **Uniformly Learnable Classes of Events.** If P is a countably additive probability on a measure space $(X, \mathcal{X})$ and P_n is the random empirical distribution of n independent draws according to P, then the strong law of large numbers (SLLN) implies that for any measurable set E, $P_n(E) \to P(E)$ (equivalently,

$\int f\, dP_n \to \int f\, dP$ for any bounded measurable f), and the central limit theorem (CLT) shows that the convergence is at a $1/\sqrt{n}$ rate. The question that Vapnik and Červonenkis [40] answered is "How large can a class of measurable sets $\mathcal{C}$ be and still have the convergence of $P_n(C)$ to $P(C)$ be uniform over $\mathcal{C}$?"[1]

The canonical example of a VC class has $X = [0,1]$, $\mathcal{X}$ the Borel σ-field, and has $\mathcal{C}$ being the Glivenko-Cantelli class of initial intervals, $\mathcal{C}_{GC} = \{[0,r] : r \in [0,1]\}$. With P being any element of Δ^{ca}, the countably additive probabilities, and $P_n(\cdot)$ denoting the random empirical distribution of an i.i.d. sequence of draws having distribution P, the SLLN is the Glivenko-Cantelli theorem, $\lim_n \sup_{[0,r]\in\mathcal{C}} |P_n([0,r]) - P([0,r])| = 0$ almost everywhere (a.e.). The associated CLT is the covergence of $\frac{1}{\sqrt{n}}[P_n([0,r]) - P([0,r])]$ to a Gaussian process indexed by r (a Brownian bridge if P is non-atomic).

Dudley, Giné, and Zinn [16, Proposition 11] showed that, subject to a measurability condition, $\mathcal{C}$ is a VC class if and only if

$$\sup_{P\in\Delta^{ca},\ C\in\mathcal{C}} |P_n(C) - P(C)| \overset{a.e.}{\to} 0. \tag{1}$$

Further, the associated CLT is uniform. To summarize, if we ask for learnability of a class of events, then only VC classes can be learned uniformly, the uniformity is over the class of sets and over the class of countably additive probabilities, and we have a $1/\sqrt{n}$-rate of convergence from the associated uniform CLT.

1.2. **Deficiency and Indeterminacy.** A more complete investigation will show that the indeterminacy depends on the degree to which the probability fails countable additivity. The degree of failure of $\mu \in \Delta^{fa}$ is given by Kingman's [27] **deficiency**, $\delta(\mu) := \sup\{\epsilon \geq 0 : \exists E_n \downarrow \varnothing,\ \mu(E_n) \geq \epsilon\}$, the supremum being taken over all decreasing sequences of sets. Weightless probabilities have deficiency 1, while countably additive probabilities have deficiency 0 because continuity from above at the empty set is equivalent to countable additivity. The extension of the indeterminacy result shows that, for a class of $\mathcal{C}$'s much larger than set of VC classes, if the deficiency of μ is δ, then for any $r \in (0,\delta]$, then there are uncountably many $\nu \in \Delta^{fa}$, and for each of these ν, uncountably many different sets E_r with $|\mu(E_r) - \nu(E_r)| = r$ even though $|\mu(C) - \nu(C)| = 0$ for all $C \in \mathcal{C}$.

It is worth re-emphasizing the extent to which the indeterminacy is unrelated to VC learning theory or the underlying space. First, the size of the indeterminacy is given by the deficiency of the probability, this without any reference the space in which the analysis is carried out. Second, in §3.7, we will see that it is the μ-separability of VC classes that is behind the indeterminacy result, not their uniform learnability, nor the class of spaces in which they are set.

1.3. **The Decisions of Expected Utility Maximizers.** The modeling choice to use weightless probabilities has serious negative consequences for understanding the effects of learning on optimal behavior. The decision problem $\max_{a\in A} \int u(a,x)\, d\mu(x)$ is **well-behaved** if A is a compact metric space, $(X,\mathcal{X})$ is a measure space,

---

[1] In econometrics, VC classes/stochastic equicontinuity began to appear in the mid 1980's, see Pollard [31] and Andrews [3]. Their appearance in the *Handbook of Econometrics* [4] in the mid-1990's cemented their place in econometricians' tool-kit. An early exposition is Pollard's textbook, [30], van der Vaart and Wellner [39] is a more recent textbook, and Dudley's *Uniform Central Limit Theorems* [14] is a very thorough monograph from one of the pioneers of the field. VC classes are often defined in terms of an equivalent combinatorial property.

$0 \le u(a,x) \le 1$ for all $(a,x) \in A \times X$, each slice $u(a,\cdot)$ is measurable, and each slice $u(\cdot,\omega)$ is continuous. These assumptions imply that $u(\cdot,\cdot)$ is jointly measurable, and that, for each $a \in A$, the mapping $\mu \mapsto \int u(a,x)\,d\mu(x)$ from Δ^{fa} to $\mathbb{R}$ is continuous when Δ^{fa} is given the weak* topology.[2]

Fix a weightless μ, a separable class $\mathcal{C}$, and let $\mathcal{C}^\circ$ denote the smallest field containing $\mathcal{C}$. We will see that the complete indeterminacy result for weightless probabilities implies that for any weightless μ, there always exists a well-behaved decision problem and uncountably many disjoint support ν agreeing with μ on $\mathcal{C}^\circ$ such that

$$\int u(a^*(\mu),x)\,d\mu(x) = 1, \ \int u(a^*(\mu),x)\,d\nu(x) = 0, \text{ and}$$
$$\max_{a\in A} \int u(a,x)\,d\nu(x) = 1. \tag{2}$$

If μ is weightless, but a $\mathcal{C}^\circ$-indistinguishable ν is the true distribution, then the utility cost of the decision error is as large as possible.

1.4. **Countably Additive Learning and Difficult Decisions.** The contrast with the countably additive case is quite stark. In the weightless case, it is impossible to make an informed decision after observing an entire infinite sequence of i.i.d. realizations. In the countably additive case, the following uniform convergence result implies that maximizing against the n'th empirical distribution gives a sequence of solutions that converges to the true solution set.[3]

Lemma 1. *If P and each P_n, $n \in \mathbb{N}$, is countably additive and $\int f\,dP_n \to \int f\,dP$ for every bounded, measurable f, then for any well-behaved $u : A \times X \to [0,1]$, $|\max_{a\in A} \int u(a,x)\,dP_n(x) - \max_{b\in A} \int u(b,x)\,dP(x)| \to 0$.*

By the SLLN, the condition that $\int f\,dP_n \to \int f\,dP$ for any bounded measurable f is satisfied with probability 1 if P_n is the random empirical distribution of an i.i.d. sample drawn from a countably additive P. Therefore, a frequentist approach to learning for decision problems, maximizing against the n'th empirical distribution, delivers, with probability 1, a sequence of decisions, a_n, having a value converging to the true value. Further, it can be shown that the distance between a_n and the solution set at the true P converges to 0.

At this level of generality, learning a countably additive probability for decisions is simple, maximize against the empirical distributions. This conclusion does not depend on the measure space, $(X, \mathcal{X})$, nor the problem, $u(\cdot,\cdot)$. However, such generality may conceal more than it reveals.

If $\mathcal{U}_X := \{u(\cdot,x) : x \in X\}$ has compact closure in $C(A)$, the continuous functions on A with the supnorm metric, then $\mathcal{U}_A := \{u(a,\cdot) : a \in A\}$ is a VC class of functions (by the bracketing arguments covered in [39, Ch. 2]). The uniform CLT yields convergence at rate $1/\sqrt{n}$ with probability 1 in Lemma 1, uniformly in P. On the other hand, if e.g. the closure of $\mathcal{U}_X$ has non-empty interior in $C(A)$, then $\mathcal{U}_A$ is not a VC class, and the worst case convergence will be slower.

[2]The weak* topology on Δ^{fa} is the weakest topology making $\mu \mapsto \int f\,d\mu$ continuous for each bounded measurable $f : X \to \mathbb{R}$. The SLLN implies that the empirical distributions of an i.i.d. sequence of draws made according to a countably additive distribution weak* converge.

[3]The result cannot be novel, but I have been unable to find in this exact form, perhaps because interest in statistical decision theory focuses on unbounded loss functions. A proof is in the appendix.

For decision problems, the difficulty of learning comes from an interaction of the complexity of the set of utility functions and the probability distribution. Slow convergence of the optimal values is an indication of a more difficult learning problem. Connections between difficult learning problems and difficult estimation problems are well-exposited in Cucker and Smale [11].

1.5. **Implications for Multiple Prior Models.** Let $\Pi(\mathcal{C}, \mu)$ denote the set of probabilities in Δ^{fa} that agree with μ on $\mathcal{C} \subset \mathcal{X}$. As Al-Najjar [1, §4] noted, if $\Pi(\mathcal{C}, \mu)$ has properties that make it useful for modeling choice in the presence of ambiguity, then weightless learning provides a fascinating interpretation of multiple prior models. We will see that there is a limited, but not trivial, set of problems that a set such as $\Pi(\mathcal{C}, \mu)$ can model when μ is weightless.

1.5.1. *A Change of Variable Analysis.* For multiple prior models, one replaces the single probability, μ, with a set of priors, e.g. $\Pi = \Pi(\mathcal{C}, \mu)$. Moderately general multiple prior preferences over acts can be represented by the α-Minmax Expected Utility (α-MEU) function

$$U(a) := \alpha \cdot \min_{\mu \in \Pi} \int_X v(a(x))\, d\mu(x) + (1-\alpha) \cdot \max_{\nu \in \Pi} \int_X v(a(x))\, d\nu(x) \tag{3}$$

where $0 \leq \alpha \leq 1$ and $v : \mathbb{W} \to [0,1]$ is a Bernoulli utility function. This allows the modeler to account for some forms of a decision maker's reactions the probabilities of some events in the state space being only partially known.

A change of variables is informative. Let $a(\Pi)$ be the set of probability on $\mathbb{W}$ induced by the function $a(\cdot)$, that is, $a(\Pi) = \{p \in \Delta(\mathbb{W}) : (\exists \mu \in \Pi)[p(\cdot) = \mu(a^{-1}(\cdot))]\}$. Letting $A = a(\Pi)$, the utility function in (3) can be re-written as

$$U(A) = \alpha \cdot \min_{p \in A} \int_{\mathbb{W}} v(c)\, dp(c) + (1-\alpha) \cdot \max_{q \in A} \int_{\mathbb{W}} v(c)\, dq(c). \tag{4}$$

All acts $a(\cdot)$ with $a(\Pi) = A$ are indistinguishable for α-MEU preferences, and for all of the other multiple prior preferences known to the author.

The **descriptive range** of Π is the class of sets of probabilities of the form $a(\Pi)$. For this class of optimization problems to be broadly applicable, the descriptive range of Π should be a large class of sets.

1.5.2. *Descriptive Range.* A first, mistaken, intuition is that larger sets of priors allow for more sets to be modeled in decision problems, and this would make the strong indeterminacy result "good news" for the weightless learning interpretation of choice under ambiguity. This is not correct, and the descriptive range of $\Pi(\mathcal{C}, \mu)$ is, in general, very small. However, for some classes $\mathcal{C}$ and probabilities μ, the descriptive range of $\Pi(\mathcal{C}, \mu)$ compares favorably to the cores of convex capacities in potentially important fashions.

The next section shows that if $\mathcal{C}$ is one of several of the well-known VC classes of subsets of the integers, then i.i.d. observations from any weightless probability are consistent with *all* weightless probabilities. In this case, it is easy to show that if $\mathbb{W}$ is finite, then $a(\Pi)$ can only be a face of the simplex $\Delta(\mathbb{W})$, i.e. $a(\Pi) = \Delta(K)$ for some non-empty $K \subset \mathbb{W}$. This has two strikingly negative implications for the class of problems that can be modeled: first, no non-degenerate risky problems can be modeled because $\{p\} \neq \Delta(K)$ unless $K = \{w\}$ and p is point mass on w; and second, no non-trivial comparisons of likelihoods can be modeled, e.g. the set of

$p \in \Delta(\mathbb{W})$ for which outcome $w \in \mathbb{W}$ is at least as likely as outcome w' is not of the form $\Delta(K)$ for any K.

Though small in many senses,[4] the class of problems need not be quite so limited as this suggests. We will see that the descriptive range of Π is a subset of the convex combinations of the faces in $\Delta(\mathbb{W})$ where the set of convex combinations is given by the closure of the set of values $\{\mu(C) : C \in \mathcal{C}^\circ\}$ where $\mathcal{C}^\circ$ is the field generated by $\mathcal{C}$. For many combinations of $\mathcal{C}$ and μ, *some* non-degenerate risky problems can be modeled, and if the closure of $\{\mu(C) : C \in \mathcal{C}^\circ\}$ is all of $[0, 1]$, then *all* risky problems can be modeled, as can all convex combinations of problems in which the support of the probability is all that is known. By contrast, the core of a convex capacity does not contain these problems in its descriptive range.

1.6. **Cardinality and Deficiency Intuitions.** There is a canonical compact Hausdorff space $\widehat{X}$ containing X as a dense imbedded subset and defined by the property that every bounded measurable function, f, on X has a unique continuous extension, $\widehat{f}$, to $\widehat{X}$. The representation theorems for integrals tell us that for every μ in Δ^{fa}, there exists a unique countably additive $\widehat{\mu}$ on $\widehat{X}$ determined by $\int_{\widehat{X}} \widehat{f}\, d\widehat{\mu} = \int_X f\, d\mu$, and that every countably additive $\widehat{\mu}$ on $\widehat{X}$ arises from a $\mu \in \Delta^{fa}$ in this fashion. There are two strong intuitions about the indeterminacy, one related to deficiency, and one related to the cardinality of $\widehat{X}$.

Deficiency. If δ is the deficiency of μ, then $\widehat{\mu}(\widehat{X} \setminus X) = \delta$, that is, δ of the mass is carried on $\widehat{X} \setminus X$. Since points in $\widehat{X} \setminus X$ are not observable, nothing about δ of the mass in μ can be learned.

Cardinality. The space $\widehat{X}$, and therefore the space Δ^{fa}, has cardinality at least $2^{\mathfrak{c}}$ where $\mathfrak{c}$ is the cardinality of the continuum. VC classes are, up to closure, compact pseudo-metric spaces, hence separable. One can not distinguish $2^{\mathfrak{c}}$ points on the basis of countably many pieces of information because $\{0, 1\}^{\mathbb{N}}$ has cardinality $\mathfrak{c}$, which is strictly smaller.

1.7. **Outline.** Section 2 begins with a discussion of the difficulties interpreting "i.i.d. observations" drawn from a weightless probability. It then turns to three special cases of the general indeterminacy result and the implications for decision theory, including extensive comparisons of descriptive ranges with those of the cores of convex capacities. With the exception of a brief appeal to a later result, the proofs in this section use little more than introductory real analysis techniques.

Section 3 gives, in Theorem 1, the general form of the complete indeterminacy result, and its Corollary containing the implications for decision theory. The non-elementary proofs of this section make extensive use of the properties of the space $\widehat{X}$, and are relegated to the appendix. The final summarizes, evaluates Al-Najjar's [1] conclusions in the light of these results, and gives contact points with the literature on weightless probabilities in decision theory, game theory, statistics, and stochastic processes.

---

[4]When the space of outcomes has three or more elements, the class of problems that can be modeled is closed and Gauss/Aronszajn null, hence also Haar null and porous. See Benyamini and Lindenstrauss [5, Ch. 6] for a unified treatment of these concepts of the "smallness," see Anderson and Zame [2] for the unexpectedly subtle extension of Haar nullity to infinite dimensional convex sets, and see Stinchcombe [35] for the failure of Haar nullity to have even an approximate Bayesian interpretation.

2. Three Examples of Weightless Learning

The first two examples in this section demonstrate how complete indeterminacy arises in the class of discrete metric spaces used by Al-Najjar [1], the third example covers complete indeterminacy in the unit interval. The descriptive range in the second example is a partial convexification of the descriptive range in the first, in the third example, it is the convexification (and closure).

For this section, we restrict attention to the case that the acts take their values in finite spaces, but, as we will see in the next section, the results extend to much more general spaces of consequences. Before turning to the examples, we discuss an interpretational difficulty with the notion of "observing an i.i.d. sequence ξ_t with distribution μ" when μ is weightless.

2.1. A Difficult Job of Interpretation.

The problems come in two flavors, "What is meant by 'observing'?" and "What is meant by 'i.i.d.'?"

2.1.1. *Observations.*

We see the difficulty of interpreting "observing … ξ_t" most clearly when μ is a weightless probability on the integers and $\mathcal{C}$ is the class of initial intervals. Because μ is weightless, at every t, one learns that the random "integer," ξ_t, does not belong to any of the sets $\{1,\ldots,n\}$ because $\mu(\{1,\ldots,n\}) \equiv 0$. To put it another way, the empirical cumulative distribution function (cdf) of the observations is identically equal to 0, that is, none of the ξ_t are ever observed.

The case that $X = [0,1]$ has both a related and a very different set of interpretational difficulties. If μ's cdf has, for any r, $F_\mu(r-) := \lim_{s\uparrow r} F_\mu(s) < F_\mu(r)$ or $F_\mu(r) < F_\mu(r+) := \lim_{s\downarrow r} F_\mu(s)$, then with probability given by the size of the discrepancies, with observe that the random "real number," ξ_t is not equal to r but still satisfies $|r - \xi_t| < \epsilon$ for every $\epsilon > 0$. An entirely different class of problems arise if $F_\mu(\cdot)$ is continuous and there is no evidence that μ is deficient.

If μ is *not* deficient, i.e. it is countably additive, and $\mathcal{C}$ is the Glivenko-Cantelli class of sets, then we learn its cdf with a sup norm error proportional to $1/\sqrt{n}$ where n is the number of observations. However, if μ is weightless, then there is complete indeterminacy and it is impossible to infer even the support set of μ from the ξ_t. Here the degree of learnability of μ depends entirely on the assumed degree of deficiency, and observations can provide no guidance on the validity of the assumption.

A solution to the observational problems is to replace the notion of directly observing ξ_t at time t with the notion of observing the value of $1_C(\xi_t)$ for each $C \in \mathcal{C}$ and each t. The **observation space**, $\{0,1\}^{\mathcal{C}}$, is compact, and even metrizable if $\mathcal{C}$ is countable. Whether or not the class of sets $\mathcal{C}$ is countable, the following standard result tells us that using this observation space means that the distribution of the observations is countably additive (see e.g. Billingsley [9, §2, Thm. 2.3] for the case of a countable $\mathcal{C}$).

Lemma 2. *Every probability on $\{0,1\}^{\mathcal{C}}$ is countably additive. In particular, if $\mu \in \Delta^{fa}$ and $\mathcal{C} \subset \mathcal{X}$, then the image of μ under the mapping $x \mapsto (1_C(x))_{C\in\mathcal{C}}$ is countably additive probability on $\{0,1\}^{\mathcal{C}}$.*

The the useful classes $\mathcal{C}$, the mapping $x \mapsto (1_C(x))_{C\in\mathcal{C}}$ is one-to-one. We can understand both the observational problem for weightless probabilities and the indeterminacy result as arising from the fact that, even if $x \mapsto (1_C(x))_{C\in\mathcal{C}}$ is one-to-one, the mapping $\mu \mapsto (1_C(\mu))_{C\in\mathcal{C}}$ is uncountable-to-one for weightless probabilities. This happens because a weightless μ is carried on $\widehat{X} \setminus X$, and a function being one-to-one on X has no such implication for its extension to $\widehat{X}$.

2.1.2. *Independence.* The countable additivity guaranteed by Lemma 2 and Kolmogorov's extension theorem (e.g. Dudley [15, §12.1]), implies that there is a unique probability on the countable product of the observation space corresponding to the distribution of an i.i.d. sequence of observations. By contrast, if μ is a weightless probability on, say the integers, $\mathbb{N}$, then, for any $r \in [0, 1]$, there are extensions, μ_r^∞, of $\mu \times \mu \times \cdots$ from the product field of subsets of $\mathbb{N} \times \mathbb{N} \times \cdots$ to the product σ-field for which $\mu_r^\infty(\{(n_1, n_2, \ldots) : n_1 < n_2 < \cdots\}) = r$.

Even though μ_1^∞ seems to give a model of an "i.i.d." sequence of integers that is, with probability 1, strictly increasing, one should be careful of this interpretation. Purves and Sudderth [32] show that it is possible to pick an extension that accords with limit intuitions developed in the countably additive case. Further, the root of this odd situation is the observability problem — if none of the values of the ξ_t are observable, then "observing" that $\xi_1 < \xi_2 < \xi_3 < \cdots$ is problematic.

While using the space $\{0, 1\}^{\mathcal{C}}$ allows us to proceed with the analysis, it does obscure some of the interpretational difficulties, and we will return to them in §3.5.

2.2. **Integer Indeterminacy for Some Well-Known VC Classes.** The first two examples work with $(X, \mathcal{X})$ being a discrete metric space with its Borel σ-field.[5] This means that X is a Polish space, that the σ-field, $\mathcal{X}$, is simultaneously the topology, the Borel σ-field generated by the topology, and the class of all subsets of X, and when X is countable (as in the proofs in [1]), there is no loss in assuming that X is $\mathbb{N}$, the set of integers.

The following three classes of sets are VC classes when $X = \mathbb{N}$: the Glivenko-Cantelli class of initial intervals, $\mathcal{C}_{GC} = \mathcal{C}_{\leq} = \{\{1, \ldots, n\} : n \in \mathbb{N}\}$; the class of tail intervals, $\mathcal{C}_{\geq} = \{\{n, n + 1, \ldots\} : n \in \mathbb{N}\}$; and for each integer M, the class of sets having M or fewer elements, $\mathcal{C}_M = \{E \subset \mathbb{N} : \#E \leq M\}$. The smallest field containing any one of these VC classes is $\mathcal{C}_{cof}$, the unlearnable field of sets that are either finite or cofinite (have finite complement). For any countably additive P and $\mathcal{C}$ being any one of these classes of sets, $\Pi(\mathcal{C}, P) = \{P\}$, that is, $\mathcal{C}$ **determines countably additive probabilities**. If μ is weightless, Proposition A (below) tells us that $\Pi(\mathcal{C}, \mu)$ is the set of all weightless probabilities, essentially the opposite result.

A probability ν is **full valued** if for every $E \in \mathcal{X}$ with $\nu(E) > 0$ and any $r \in (0, \nu(E))$, there exists an $E' \subset E$ with $\nu(E') = r$. For countably additive probabilities, being full valued and being non-atomic are equivalent, but this is not true for finitely additive probabilities (e.g. Maharam [28, §3], or Rao and Rao [8, Ch. 5] for more detailed information on the ranges of finitely additive measures). For $\nu \in \Delta^{fa}$, two sets, E, E', are ν**-distinct** if $\nu(E \setminus E') + \nu(E' \setminus E) > 0$.

---

[5] There are many versions of this metric. The easiest is $\rho(n, m) = 1$ if $n \neq m$ and $\rho(n, n) = 0$. Two with a simple geometric formulation are $d(n, m) = |\frac{1}{n} - \frac{1}{m}|$ and $r(n, m) = |e^{-n} - e^{-m}|$. More general geometric formulations can be had as follows: if φ is a one-to-one imbedding of X in a metric space (M, d_M) and no $\varphi(n)$ is a cluster point of $\varphi(X)$, then X and $\varphi(X)$ are homeomorphic so that $d_M(\varphi(n), \varphi(m))$ is a metric for this topology.

2.2.1. *Indeterminacy.* The last part of the following tells us that one cannot even learn the support set of a weightless μ from its values on $\mathcal{C}_{cof}$.

Proposition A. *If μ is a weightless probability on the integers, then $\Pi(\mathcal{C}_{cof}, \mu)$ is the set of all weightless probabilities. This implies that there are uncountably many $\nu \in \Pi(\mathcal{C}_{cof}, \mu)$ with disjoint support, and that for each of them, for any $r \in (0, 1]$, there are uncountably many ν-distinct E_r with $|\mu(E_r) - \nu(E_r)| = r$.*

Proof. For the first part, it is sufficient to show that for every weightless μ and every finite $C \in \mathcal{C}_{cof}$, $\mu(C) = 0$. To this end, pick an arbitrary weightless μ. There exists a sequence of sets, $H_m \uparrow \mathbb{N}$ with $\mu(H_m) \equiv 0$. Every finite C has $\mu(C) = 0$ because C is a subset of H_m for all sufficiently large m.

For the second part, start with a subclass $\mathcal{E} \subset \mathcal{X}$ that consists of an uncountable collection of infinite subsets of X with the property that any pair, $E \neq E' \in \mathcal{E}$ have $E \cap E'$ finite.[6] For any weightless μ, $\mu(E \cap E') = 0$ so that $\mu(E) = 0$ for uncountably many $E \in \mathcal{E}$. Let E' be one of them. There are uncountably many full valued weightless ν's with $\nu(E') = 1$. For any such ν and any $r \in (0, 1)$, there are uncountably many ν-distinct $E_r \subset E$ with $\nu(E_r) = r$. For any such E_r, $|\mu(E_r) - \nu(E_r)| = |0 - r| = r$. For $r = 1$, let E_1 be the union of E' and any of the uncountably many other $E'' \in \mathcal{E}$ with $\mu(E'') = 0$. □

2.2.2. *Weightless Learning for Expected Utility Maximizers.* Because the values of μ on the field $\mathcal{C}_{cof}$ do not identify the support of μ, learning them may be entirely useless for the purposes of making decisions.

Corollary A.1. *If μ is a weightless probability on the integers, there there exists a well-behaved $u : \{0, 1\} \times X \to [0, 1]$ and uncountably many $\nu \in \Pi(\mathcal{C}_{cof}, \mu)$ with mutually disjoint support sets such that $\int u(1, x)\, d\mu(x) = 1$, $\int u(1, x)\, d\nu(x) = 0$, and $\int u(0, x)\, d\nu(x) = 1$.*

Proof. As in the proof of Proposition A, let $\mathcal{E}$ be an uncountable collection of infinite subsets of X with $E \cap E'$ finite for any distinct pair $E, E' \in \mathcal{E}$. For at least one $E \in \mathcal{E}$, $\mu(E) = 0$. Set $u(1, x) = 1_{E^c}(x)$ and $u(0, x) = 1_E(x)$. Since E is homeomorphic to X, Proposition A completes the argument. □

2.2.3. *Indeterminacy and Multiple Prior Models.* Fix a finite space of utility relevant consequences, $\mathbb{W}$. For any $S \subset \Delta^{fa}$, $\mathcal{R}_{\mathbb{W}}(S)$ denotes the **descriptive range** of S, that is, the class of all sets that are of the form $\{f(\mu) : \mu \in S\}$ for some function f. For any non-empty $K \subset \mathbb{W}$, $\Delta(K)$ denotes the set of probabilities on $\mathbb{W}$ with $p(K) = 1$.

Corollary A.2. *If $\mathbb{W}$ is a finite space and Π is the set of weightless probabilities or the set of full valued weightless probabilities, then $\mathcal{R}_{\mathbb{W}}(\Pi)$ is the class of sets of the form $\Delta(K)$ where K is a non-empty subset of $\mathbb{W}$.*

Proof. Let $f : X \to \mathbb{W}$ and define $K_f = \{w \in \mathbb{W} : f^{-1}(w) \text{ is infinite}\}$. For each $x \in K_f$, all of the uncountably many weightless μ, full valued or not, with $\mu(f^{-1}(w)) = 1$ have the property that $f(\mu)$ is point mass on w. For each $w' \notin K_f$

[6] To see that such classes exist, let φ be a bijection between X and the rationals, $\mathbb{Q} \subset \mathbb{R}$. For each of the uncountably many irrationals, $x \in \mathbb{R}$, let $q_k^x \uparrow x$ be a sequence in $\mathbb{Q}$ converging up to x and define E_x as $\varphi^{-1}(\{q_k^x : k \in \mathbb{N}\})$. The collection $\mathcal{E} := \{E_x : x \text{ irrational}\}$ has the requisite properties.

and each weightless μ, full valued or not, $\mu(f^{-1}(w')) = 0$. Since the set of (full-valued) weightless probabilities is convex, $f(\Pi) = \Delta(K_f)$. $\square$

As an example, using $\Pi = \Pi(\mathcal{C}_{cof}, \mu)$ as a set of priors when $\mathbb{W} = \{w, w'\}$, we can model only the following three situations: a decision maker who is sure that w will happen; a decision maker who is sure that w' will happen; and a decision maker who believes that w' will happen with some probability. In general, using Π (or the full valued elements of Π) as the set of priors, no non-degenerate risky problems can be modeled because $\{p\}$ is not of the form $\Delta(K)$ unless K is a singleton set and p is a point mass on that singleton. Further, there are no non-degenerate comparisons of likelihood that can be modeled, e.g. the set of probabilities satisfying "w' is at least as likely as w" is not of the form $\Delta(K)$, hence is not in the descriptive range of Π.

2.2.4. *Indeterminacy and the Ignorance Core.* A **capacity** on $(X, \mathcal{X})$ is a set function $c : \mathcal{X} \to [0,1]$ with $c(\varnothing) = 0$, $c(X) = 1$, and $c(E) \leq c(F)$ for all $E \subset F$, $E, F \in \mathcal{X}$. A capacity is **convex** (or super-additive) if for all $E, F \in \mathcal{X}$, $c(E \cup F) + c(E \cap F) \geq c(E) + c(F)$. The **core** of a convex capacity is $\mathcal{P}_c := \{p \in \Delta^{fa} : (\forall E \in \mathcal{X})[p(E) \geq c(E)]\}$. If c is itself a probability, then $\mathcal{P}_c = \{c\}$, if c is convex and not a probability, then $\mathcal{P}_c$ is a non-degenerate convex set of probabilities, and $c(\cdot)$ is the **lower envelope of** $\mathcal{P}_c$, that is, $c(E) = \inf\{p(E) : p \in \mathcal{P}_c\}$ for all measurable E.

For the case that Π is the set of weightless probabilities and $\mathbb{W}$ is finite, we will compare $\mathcal{R}_{\mathbb{W}}(\Pi)$ to $\mathcal{R}_{\mathbb{W}}(\mathcal{P}_c)$ for the core of the "complete ignorance" capacity. This is defined by $c_{ig}(E) = 0$ if $E^c \neq \varnothing$ and $c_{ig}(X) = 1$. It has its name from the observation that its core is the set of all probabilities on X.

Lemma 3. *If $\mathbb{W}$ is finite, then $\mathcal{R}_{\mathbb{W}}(\Pi) = \mathcal{R}_{\mathbb{W}}(\mathcal{P}_{c_{ig}})$.*

Proof. By Corollary A.2, we must show that any set in $\mathcal{R}_{\mathbb{W}}(\mathcal{P}_{c_{ig}})$ is of the form $\Delta(K)$ for some non-empty $K \subset \mathbb{W}$. Let $f : X \to \mathbb{W}$ and note that $f(\Delta^{fa})$ is $\Delta(K_f)$ where $K_f := \{w \in \mathbb{W} : f^{-1}(w) \neq \varnothing\}$ is, here, the range of the function f. $\square$

For the purposes of multiple prior modeling ambiguous choice problems, there is no difference between using $\Pi(\mathcal{C}_{cof}, \mu)$ for a weightless μ or the core of the complete ignorance capacity as the set of priors. This makes intuitive sense because there are no significant restrictions on the sets of priors in either case: $\Pi(\mathcal{C}_{cof}, \mu)$ is equivalent to the set of countably additive probabilities on $\widehat{X} \setminus X$; and $\mathcal{P}_{c_{ig}}$ is the set of all probabilities. The image, under a function f, of the set of all probabilities on a space Ω is the set of probabilities on $f(\Omega)$. For this reason, it is impossible to model restrictions on the sets of probabilities on consequences in either of these cases.

We now turn to the simplest additional restriction that $\Pi(\mathcal{C}, \mu)$ might satisfy, one that allows us to model some restrictions.

2.3. **Integer Indeterminacy with an Intermediate Mass Set.** The weightless probabilities on a countable set assign either mass 0 or mass 1 to all sets in $\mathcal{C}_{cof}$, that is, $\{\mu(C) : C \in \mathcal{C}_{cof}\} = \{0, 1\}$. Corollary 1.2 gives the general result that if $\mathcal{C}^{\circ}$ is the field generated by $\mathcal{C}$, then the set of weights, $\{\mu(C) : C \in \mathcal{C}^{\circ}\} \subset [0,1]$, determines the extent to which one can convexify the descriptive range. To see how this works, we add one set to $\mathcal{C}_{cof}$.

Let $\mathcal{C}_{+E}$ denote the smallest field containing $\mathcal{C}_{cof}$ and one infinite set E having an infinite complement, and let μ be a weightless probability with $\mu(E) = \beta \in (0,1)$.

This partially convexifies $\{\mu(C) : C \in \mathcal{C}_{cof}\} = \{0, 1\}$, expanding it to the set $\{\mu(C) : C \in \mathcal{C}_{+E}\} = \{0, \beta, (1-\beta), 1\}$.

2.3.1. *Indeterminacy.* Let $\Delta^{pfa}(E)$ denote the weightless probabilities on E and $\Delta^{pfa}(E^c)$ the weightless probabilities on E^c.

Proposition B. *If μ is a weightless probability on the integers and assigns mass $\beta \in (0,1)$ to an infinite set E having infinite complement, then $\Pi(\mathcal{C}_{+E}, \mu) = \beta\Delta^{pfa}(E) + (1-\beta)\Delta^{pfa}(E^c)$. This implies that there are uncountably many $\nu \in \Pi(\mathcal{C}_{+E}, \mu)$ with disjoint support, and for each of them, for any $r \in (0, 1]$, there are uncountably many ν-distinct sets E_r with $|\mu(E_r) - \nu(E_r)| = r$.*

Proof. Since $\mu(\cdot) = \beta\mu(\cdot|E) + (1-\beta)\mu(\cdot|E^c)$ and both E and E^c are homeomorphic images of X, this follows from Proposition A. $\square$

2.3.2. *Indeterminacy and Learning for Decision.* Unlike the cofinite case, it is now possible to specify decision problems in which knowledge of $\mu(C)$, $C \in \mathcal{C}_{+E}$, is of use. For example, when $u(0, x) = 1_E(x)$ and $u(1, x) = 1_{E^c}(x)$, then every ν that agrees with μ on $\mathcal{C}_{+E}$ has the same optimal choice set. However, applying Corollary A.1 to $\mu(\cdot|E)$ and $\mu(\cdot|E^c)$ independently yields the following "weightless learning may be useless" result.

Corollary B.1. *If μ is weightless and assigns mass $\beta \in (0,1)$ to an infinite set E having infinite complement, then there is a jointly continuous $u : \{0,1\} \times X \to [0,1]$ and uncountably many $\nu \in \Pi(\mathcal{C}_{+E}, \mu)$ with disjoint support sets such that $\int u(1,x)\, d\mu(x) = 1$, $\int u(1,x)\, d\nu(x) = 0$, and $\int u(0,x)\, d\nu(x) = 1$.*

2.3.3. *Indeterminacy and Multiple Prior Models.* Of present interest is the observation that the descriptive range has expanded significantly. The analysis of Corollary A.2 applies to $\mu(\cdot|E)$ and to $\mu(\cdot|E^c)$, which delivers the following.

Corollary B.2. *If μ is a weightless probability on the integers and assigns mass $\beta \in (0,1)$ to an infinite set E having infinite complement and $\mathbb{W}$ is a finite space, then $\mathcal{R}_{\mathbb{W}}(\Pi(\mathcal{C}_{+E}, \mu)$ is the class of all sets of the form $\beta\Delta(K) + (1-\beta)\Delta(K')$ where K and K' are non-empty subsets of $\mathbb{W}$.*

By contrast with the cofinite case, *some* non-degenerate risky problems can be modeled. Taking $K = \{w\}$ and $K' = \{w'\}$ to be singleton sets, $\beta\Delta(K) + (1-\beta)\Delta(K')$ models the risky outcome that puts mass β on w and mass $(1-\beta)$ on w'. However, it is still the case that no non-trivial comparisons of likelihood can be modeled, "w is at least as likely as w'" is not descriptive of any set of the form $\beta\Delta(K) + (1-\beta)\Delta(K')$.

2.3.4. *Indeterminacy and Convex Distortion Cores.* Let τ be a full-valued probability. If $\varphi : [0,1] \to [0,1]$ is a strictly increasing, onto, convex function with $\varphi(t) < t$ for some t, then we define the **φ-distortion capacity** $c_\varphi : \mathcal{X} \to [0,1]$ by $c_\varphi(E) = \varphi(\tau(E))$ for $E \in \mathcal{X}$. The more convex is φ, the larger is the core of c_φ.

It is easiest to compare the descriptive ranges in the two outcome case, $\mathbb{W} = \{w, w'\}$. We can represent closed convex subsets of $\Delta(\{w, w'\})$ as intervals, $[r, s]$, $0 \le r \le s \le 1$, equivalently, as points in the triangle $T = \{(r, s) : 0 \le r \le s \le 1\}$ where, by convention, r and s represent probabilities of the outcome w'.

The set $\mathbb{W} = \{w, w'\}$ has three non-empty subsets, $\{w\}$, $\{w'\}$, and $\{w, w'\}$. From Corollary B.2, we know that $\mathcal{R}(\Pi)$ is the set of β-weighted combinations of pairs of these. Therefore, nine distinct problems can be modeled if $\beta = \mu(E) \neq \frac{1}{2}$. By comparison, a one-dimensional (Lipschitz) manifold of problems can be modeled using $\mathcal{P}_{c_\varphi}$. The manifold contains no non-degenerate risky problems nor complete lack of knowledge about the probabilities.

Lemma 4. *If $\mathcal{P}_c$ is the core of a convex distortion capacity c_φ, then $\mathcal{R}(\mathcal{P}_c)$ is the one-dimensional Lipschitz manifold $\{[\varphi(t), 1-\varphi(1-t)] : t \in [0,1]\}$.*

Proof. Pick an arbitrary $f : X \to \{w, w'\}$ and let $E = f^{-1}(w)$. The minimal probability that can be assigned to E by any p in the core of c_φ is $\varphi(\tau(E))$. One finds the maximal probability that can be assigned to E by any q in the core by noting that the minimal probability that can be assigned to E^c is given by $q(E^c) = 1 - q(E) \geq \varphi(1 - \tau(E))$, that is, $q(E) \leq 1 - \varphi(1 - \tau(E))$. Since τ is full valued, $t := \tau(E)$ can be any value in $[0, 1]$. Therefore, the class of problems that can be modeled using $\mathcal{P}_{c_\varphi}$ is given by $\{[\varphi(t), 1-\varphi(1-t)] : t \in [0,1]\}$. □

In this, the simplest, two-outcome case, the sets $\mathcal{R}(\mathcal{P}_c)$ and $\mathcal{R}(\Pi(\mathcal{C}_{+E}, \mu))$ are both negligible, but negligible sets are not all created equal.

Example 1. *The nine $\circ$'s in Figure 1 represent the elements of $\mathcal{R}(\Pi)$ in T for $\beta = \mu(E) = 0.3$, and the curved line joining $(0,0)$ to $(1,1)$ represents $\mathcal{R}(\mathcal{P}_{c_\varphi})$. All Choquet integral preferences on intervals $[r,s] \subset \Delta(\{w,w'\})$ are given by $\alpha r + (1-\alpha)s$ for some $\alpha \in [0,1]$. Restricted to the problems modeled by the convex distortion capacity, all α give the same preference ordering. By contrast, the risky outcome $[0.3, 0.3]$ is strict prefered to the ambiguous outcome $[0, 0.7]$ if and only if $\alpha > \frac{4}{7}$.*

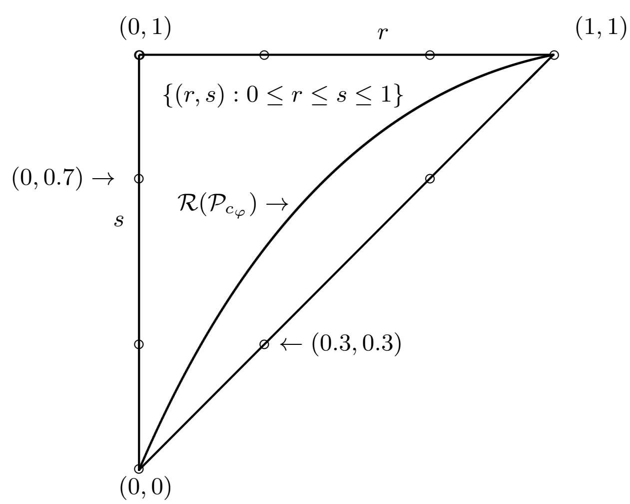

Figure 1

2.4. **Weightless Learning in the Unit Interval.** The indeterminacy result for the discrete case carries over to the case that X is the unit interval, $\mathcal{X} = \mathcal{B}$ is the Borel σ-field, and $\mathcal{C}^\circ_{GC}$ is the smallest field containing the Glivenko-Cantelli class $\mathcal{C}_{GC} = \{[0,r] : r \in [0,1]\}$. The implication for the value of learning a weightless probability is identical, it may be 0. We will see that if μ's values on the field generated by $\mathcal{C}^\circ_{GC}$ are a rich set, then $\Pi(\mathcal{C}^\circ_{GC}, \mu)$ will be useful as a set of priors to model ambiguous problem with two utility relevant outcomes, and may compare favorably to sets of priors that are the core of a convex capacity. These possibilities can be most easily seen in the case of a weightless μ with continuous cdf $r \mapsto F_\mu(r) := \mu([0, r])$ and $F_\mu(0) = 0$.

2.4.1. *Indeterminacy.* There are uncountably many countable dense $I \subset [0, 1]$. Any one of these sets can be the support set if μ is weightless, all of them have mass 0 is μ is countably additive.

Proposition C. *If μ is a weightless probability on $([0, 1], \mathcal{B})$ with a continuous cdf satisfying $F_\mu(0) = 0$, then there are uncountably many $\nu \in \Pi(\mathcal{C}^\circ_{GC}, \mu)$ such that for any $r \in (0, 1]$, there are uncountably many ν-distinct sets E_r with $|\mu(E_r) - \nu(E_r)| = r$.*

Proof. For any countable dense $I \subset [0, 1]$, define a weightless ν_I on the field of subsets of I generated by $\{(a, b] \cap I : a < b\}$ by setting $\nu_I((a, b] \cap I) = F_\mu(b) - F_\mu(a)$ and using finite additivity. By the Hahn-Banach theorem, ν_I has an extension, also denoted ν_I, to the class of all subsets of I. Because I is countable, the extension must also be weightless. The space $[0, 1]$ can be partitioned into uncountably many disjoint countable dense subsets, and $\mu(I) = 0$ for uncountably many of them. Each such I has uncountably many ν_I's, and each full valued ν_I uncountably many ν_I-distinct E_r's with $\nu_I(E_r) = r$. □

The condition that the cdf F_μ is continuous and satisfies $F_\mu(0) = 0$ guarantees that μ is full-valued. At the cost of a non-elementary proof, we can change the conditions on the cdf to μ being full-valued, both in Proposition C and its Corollaries.

2.4.2. *Indeterminacy and Learning for Decisions.* Again, not being able to identify the support means that learning a weightless probability may be useless.

Corollary C.1. *If μ is a weightless probability on $([0, 1], \mathcal{B})$ and has a continuous cdf with $F_\mu(0) = 0$, then there is a well-behaved $u : \{0, 1\} \times X \to [0, 1]$ and uncountably many $\nu \in \Pi(\mathcal{C}_{+E}, \mu)$ with disjoint support sets such that $\int u(1, x)\, d\mu(x) = 1$, $\int u(1, x)\, d\nu(x) = 0$, but $\int u(0, x)\, d\nu(x) = 1$.*

Recall that $x \in [0, 1]$ is a **condensation point** $I \subset [0, 1]$ if for all $\epsilon > 0$, the intersection, $(x - \epsilon, x + \epsilon) \cap I$, is uncountable. The following Lemma (proved in the appendix) will be used.

Lemma 5. *There exists a partion, $\mathcal{I} = \{I_\gamma : \gamma \in [0, 1]\}$ of $[0, 1]$ into uncountable, measurable sets with the property that every $x \in [0, 1]$ is a condensation point of each I_γ, and each I_γ contains uncountably many countable, disjoint, dense subsets of $[0, 1]$.*

Proof of Corollary C.1. Pick an arbitrary weightless μ. Let $\mathcal{I}$ be the partition from Lemma 5. Since μ can assign mass greater than $1/n$ to at most n of the I_γ, there is an uncountable collection of elements of $\mathcal{I}$ with $\mu(I_\gamma) = 0$. Any I_γ with $\mu(I_\gamma) = 0$ contains an uncountable collection of countable, disjoint, dense sets. As in the proof of Proposition C, for any such set, J, define a weightless ν_J on J by $\nu_J((a, b] \cap J) = F_\mu(b) - F_\mu(a)$, extend this to the field generated by the half-intervals by additivity, and use the Hahn-Banach theorem to extend this to a weightless ν_I on the collection of all subsets of J. Let $H = I^c_\gamma$, define $u(1, x) = 1_H(x)$ so that

$\int u(1,x)\,d\mu(x) = 1$ and $\int u(1,x)\,d\nu_J(x) = 0$ for each J, and define $u(0,x) = 1_{I_\gamma}(x)$ so that $\int u(0,x)\,d\mu(x) = 0$ and $\int u(1,x)\,d\nu_J(x) = 1$ for each J. $\square$

2.4.3. *Indeterminacy and Multiple Prior Models.* The descriptive range of $\Pi(\mathcal{C}^{\circ}_{GC}, \mu)$ is much larger than the previous ranges. For example, an implication of the following is that, in the two outcome case, $\mathbb{W} = \{w, w'\}$, the descriptive range is *all* of the ambiguous outcomes, $[r,s]$, $0 \leq r < s \leq 1$, a set that includes all the risky outcomes, $[r,r]$, $0 \leq r \leq 1$, as well as complete ignorance of the probability distribution, $[0,1] \subset \Delta(\mathbb{W})$.

Corollary C.2. *If μ is a weightless probability on $([0,1], \mathcal{B})$, has a continuous cdf with $F_\mu(0) = 0$, and the outcome space $\mathbb{W}$ is finite, then $\mathcal{R}(\Pi(\mathcal{C}^{\circ}_{GC}, \mu))$ is the set of all convex combinations of faces of $\Delta(\mathbb{W})$.*

Proof. We must first show that the descriptive range contains all sets of the form $\sum_{K \subset \mathbb{W}} \alpha_K \Delta(K)$, where the sum is over non-empty K, the α_K are non-negative and sum to 1. Pick an $f : X \to \mathbb{W}$ such that for all $K \subset \mathbb{W}$, $\cap_{w \in K} \mathrm{cl}(f^{-1}(w)))$ is a (possibly empty) finite union of intervals $[a,b] \subset [0,1]$ where $\mathrm{cl}(E)$ denotes the (usual) closure of $E \subset [0,1]$. Define $\alpha_K = \mu(\cap_{w \in K} \mathrm{cl}(f^{-1}(w)))$ so that $\alpha_K \geq 0$ and $\sum_{K \subset \mathbb{W}} \alpha_K = 1$. Using ν's supported on disjoint dense subsets of $\cap_{w \in K} \mathrm{cl}(f^{-1}(w)))$, we see that $f(\Pi) = \sum \alpha_K \Delta(K)$. Further, all collections of convex weights can be had by some f with the given properties. Corollary 1.2 (below) shows that this exhausts the descriptive range. $\square$

As we have seen, if the space $\mathbb{W} = \{w, w'\}$, then the class of all closed convex subsets of $\Delta(\mathbb{W})$ is two dimensional with extreme sets $[0,0] = \{\delta_w\}$, $[0,1] = \Delta(\mathbb{W})$, and $[1,1] = \{\delta_{w'}\}$. In this case, $\mathcal{R}_{\mathbb{W}}(\Pi(\mathcal{C}_{GC}, \mu))$ is the set of all closed convex subsets. More generally, if a finite space of consequences has 3 or more elements, then the class of closed convex sets is infinite dimensional while $\mathcal{R}_{\mathbb{W}}(\Pi)$ is a closed, finite dimensional, convex subset of this class.

2.4.4. *Indeterminacy and the Cores of Convex Capacities.* We will see that the conclusions of Corollary C.2 extend to the case that the space of utility relevant consequences is a compact metric space. Taking $\mathbb{W}$ to be the compact metric space $[0,1]$, this means that $\Delta([0,1])$ belongs to $\mathcal{R}(\mathcal{C}_{GC}, \mu)$ as do all of the risky outcomes, $\{p\}$, $p \in \Delta([0,1])$. Example 2 will show that there are non-convex capacities containing both $\Delta([0,1])$ and each $\{p\}$ in their descriptive range, Lemma 6 shows that no convex capacity can have this property.

Because we are presently interested in using Fubini's theorem, the uniqueness of product measures, and the universal measurability of Borel images, we follow Huber and Strassen [24] and restrict attention to countably additive probabilities on Polish spaces for this part of the investigation. The argument for the non-convexity of the lower envelope $C_{ig\times\lambda}(E)$ in the following contains most of the proof of Lemma 6.

Example 2. *Let c_{ig} be the ignorance capacity on $([0,1], \mathcal{B})$. For each countably additive $p \in \mathcal{P}_{c_{ig}}$, let $p \times \lambda$ denote the product of p and Lebesgue measure, λ, on $[0,1] \times [0,1]$ with the usual product σ-field, $\mathcal{B} \otimes \mathcal{B}$. Define $\mathcal{P}_{ig\times\lambda} = \{p \times \lambda : p \in \mathcal{P}_{c_{ig}}\}$ so that its lower envelope is $C_{ig\times\lambda}(E) = \inf\{q(E) : q \in \mathcal{P}_{ig\times\lambda}\}$.*

1. *If $f(y,u) = y$, then $f(\mathcal{P}_{ig\times\lambda}) = \Delta([0,1])$, and if $g(y,u) = h_p(u)$ where $h_p(\lambda) = p$, then $g(\mathcal{P}_{ig\times\lambda}) = \{p\}$.*

2. *$C_{ig\times\lambda}(\cdot)$ is not convex, for the closed sets E and F in Figure 2, $C_{ig\times\lambda}(E) = C_{ig\times\lambda}(F) = \frac{1}{2}$, while $C_{ig\times\lambda}(E\cap F) = 0$ and $C_{ig\times\lambda}(E\cup F) = \frac{1}{2}$ so that $C_{ig\times\lambda}(E) + C_{ig\times\lambda}(F) = 1 > C_{ig\times\lambda}(E\cap F) + C_{ig\times\lambda}(E\cup F) = \frac{1}{2}$*

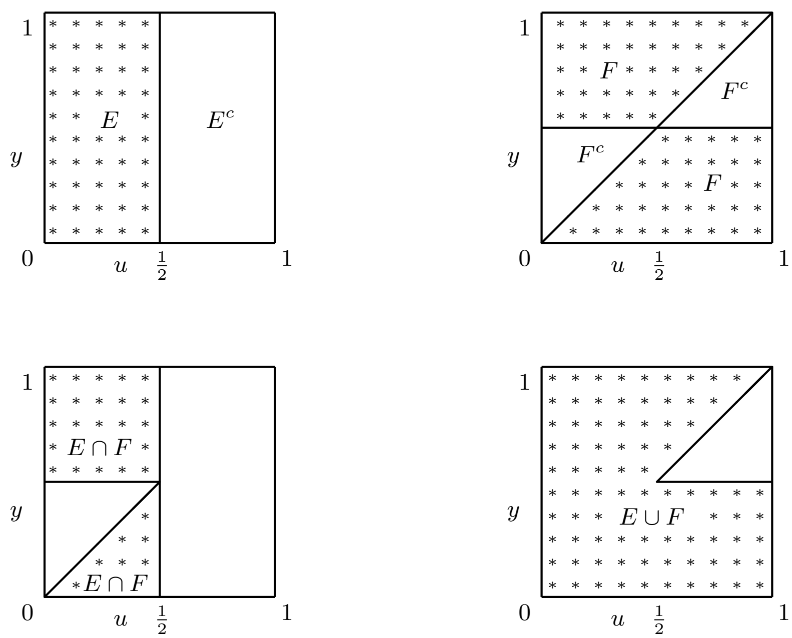

Figure 2

Lemma 6. *If $(X, \mathcal{X})$ is a Blackwell space,*[7] *$\mathcal{P}$ is a set of countably additive probabilities on $\mathcal{X}$ with descriptive range containing $\Delta([0,1])$ and Lebesgue measure, $\{\lambda\}$, then its lower envelope is not a convex capacity.*

Proof. Let $f_\Delta : X \to [0,1]$ satisfy $f_\Delta(\mathcal{P}) = \Delta([0,1])$ and $f_\lambda : X \to [0,1]$ satisfy $f_\lambda(\mathcal{P}) = \lambda$. For any $y \in [0,1]$, the point mass δ_y belongs to $\Delta([0,1])$, so there exists a $p_y \in \mathcal{P}$ with $p_y(f_\Delta^{-1}(y)) = 1$. Define $g : X \to [0,1]\times[0,1]$ by $g(x) = (f_\Delta(x), f_\lambda(x))$ and note that $G := g(X)$ is a universally measurable subset of $[0,1]\times[0,1]$. Further, since $f_\lambda(p_y) = \lambda$, the cross-section $G_y = \{u : (u,y) \in G\}$ is universally measurable and has full Lebesgue measure. For the sets E and F given in Example 2, define $E' := g^{-1}(E)$ and $F' = g^{-1}(F)$, and let $c_\mathcal{P}$ denote the lower envelope of $\mathcal{P}$. We then have $c_\mathcal{P}(E') + c_\mathcal{P}(F') = 1 > c_\mathcal{P}(E'\cap F') + c_\mathcal{P}(E'\cup F') = \frac{1}{2}$, so that $c_\mathcal{P}$ is not convex. □

2.5. **The Role of Deficiency.** Yosida and Hewitt [41, Theorem 1.22] implies that for any $\mu \in \Delta^{fa}$ with deficiency $\delta < 1$, there is a unique decomposition of μ into a

[7] A Blackwell space is a measure space that is measurably isomorphic to an analytic subset of a Polish measure space. See Dellacherie and Meyer [13, Ch. 2] or Dudley [15, Ch. 13] for developments of the theory of analytic sets, see Stinchcombe and White [37] for a summary of, and applications to, economic theory and econometrics of many of the useful measure theoretic results involving analytic sets.

countably additive η_μ and a weightless γ_μ, $\mu = (1-\delta)\eta_\mu + \delta\gamma_\mu$. The arguments for the following closely parallel the arguments given above.

Corollary C.3. *If μ is a probability on any of the spaces discussed in this section, has deficiency δ, and $\mathcal{C}$ is a countable collection of measurable sets determining countably additive probabilities, then:*

1. *$\nu \in \Pi(\mathcal{C}, \mu)$ iff $\nu = (1-\delta)\eta_\mu + \delta\gamma$ for some weightless $\gamma \in \Pi(\mathcal{C}, \gamma_\mu)$;*
2. *there exists a well-behaved $u : \{0,1\} \times X \to [0,1]$ and uncountably many $\nu \in \Pi(\mathcal{C}, \mu)$ with disjoint support sets such that $\int u(1,x)\, d\mu(x) = 1$, $\int u(1,x)\, d\nu(x) = (1-\delta)$, but $\int u(0,x)\, d\nu(x) = 1$; and*
3. *if $\mathbb{W}$ is finite, then $\mathcal{R}_{\mathbb{W}}(\Pi(\mathcal{C}, \mu))$ contains only sets of the form $(1-\delta)p + \delta\Delta(K)$ where K is a non-empty subset of $\mathbb{W}$ and p is the image of η_μ under some measurable $f : X \to \mathbb{W}$.*

Comments: Corollary C.3.1 tells us that the indeterminacy in what can be learned about μ is entirely due to its deficiency; Corollary C.3.2 tells us that the less deficient μ is, the larger is the lower bound for the effective of learning about μ; Corollary C.3.3 tells us that the class of sets that can be modeled using what can be learned as a set of priors turns to a deficiency weighted convex combination of a probability and complete uncertainty about what distribution on a non-empty set obtains. Taking K to be a singleton set, this last point tells us that some risky problems can always be modeled using $\Pi(\mathcal{C}, \mu)$ if μ is not completely deficient, although it still the case that no non-degenerate likelihood comparisons can be modeled.

3. Weightless Learning Theory for General Spaces

This first part of this section gives a definition of, and basic results for, $(\widehat{X}, \widehat{\mathcal{X}})$. The following three parts give the general indeterminacy result, its implications for learning, and its implications for multiple prior models. The last part of this section shows that VC classes are a special case of the general results. The proofs for this section are less elementary, and are gathered in the appendix.

3.1. **Weightless Probabilities on General Spaces.**

Throughout, X is a non-empty, infinite set, and $\mathcal{X}$ is a σ-field on subsets of X with $\{x\} \in \mathcal{X}$ for all $x \in X$. $M_b = M_b(X, \mathcal{X})$ denotes the set of $\mathcal{X}$-measurable, bounded, $\mathbb{R}$-valued functions on X with the supnorm, $\|f - g\| = \sup_{x \in X} |f(x) - g(x)|$.

3.1.1. *The Probabilities.*

From the basic integral representation theorems, there are two fully equivalent ways to charaterize a probability $\mu \in \Delta^{fa}$:

1. as a function $\mu : \mathcal{X} \to [0,1]$ satisfying $\mu(\varnothing) = 0$, $\mu(X) = 1$, and $\mu(A \cup B) = \mu(A) + \mu(B)$ if $A \cap B = \varnothing$; or
2. as a supnorm continuous linear functional $L_\mu : M_b \to \mathbb{R}$ satisfying $L_\mu(f) \geq 0$ for every $f \geq 0$ and $L_\mu(1_X) = 1$.

The bijective relation between the representations is given by $L_\mu(f) = \int f\, d\mu$.

Following the usual dual space conventions, the weak* topology on Δ^{fa} is defined as the weakest topology making the mappings $\mu \mapsto \int f\, d\mu$ continuous for all $f \in M_b$. The weak* topology is Hausdorff since M_b contains the indicator functions, and Δ^{fa} is compact in the weak* topology (e.g. [20, Thm. V.4.1.2]).

3.1.2. *The Space* $\widehat{X}$. A **point mass** is a probability $\mu \in \Delta^{fa}$ with $\mu(E) = 0$ or $\mu(E) = 1$ for all $E \in \mathcal{X}$. Yosida and Hewitt [41] define $\widehat{X}$ as the point masses in the closure of Δ^{fa} in the compact Hausdorff product space $[0,1]^{\mathcal{X}}$. This is the easiest representation of $\widehat{X}$ to define, but we use the textbook development of Dunford and Schwarz [20, Ch. III] because it makes several parts of the later proofs more transparent.

For each $f \in M_b$, let I_f denote the interval $[\inf_{x\in X} f(x), \sup_{x\in X} f(x)]$. Imbed each $x \in X$ in the compact product space $Z := \times_{f\in M_b} I_f$ by identifying x with the unique infinite length vector $\psi(x) := (f(x))_{f\in M_b}$. Define $\widehat{X}$ as the closure of $\psi(X)$ in Z. For any $E \in \mathcal{X}$, $1_E \in M_b$ and one can define $\widehat{E}$ in two equivalent fashions: $\widehat{E}$ is the closure of E in $\widehat{X}$; $\widehat{E} = \widehat{f}^{-1}(1)$ where $\widehat{f}$ is the unique continuous extension of 1_E to $\widehat{X}$.

3.1.3. *From* μ *to* $\widehat{\mu}$. The product topology on Z is defined as the weakest topology making each projection mapping $z \mapsto \mathrm{proj}_f(z)$ continuous. Thus, each $f \in M_b$ corresponds to the continuous function $\widehat{f} : \widehat{X} \to \mathbb{R}$ defined by $\widehat{f}(\widehat{x}) = \mathrm{proj}_f(\widehat{x})$. Since X is dense in its closure, $\widehat{X}$, the continuous extension is unique, and $\max_{\widehat{x}\in\widehat{X}} |\widehat{f}(\widehat{x}) - \widehat{g}(\widehat{x})| = \sup_{x\in X} |f(x) - g(x)|$ for all $f, g \in M_b$.

The identification of f with $\widehat{f}$ makes sup norm continuous linear functionals on M_b into sup norm continuous linear functionals on $C(\widehat{X})$, the continuous functions on $\widehat{X}$. The Riesz representation theorem tells us that the continuous linear functionals on $C(\widehat{X})$ correspond to integration against countably additive signed Baire measures on the compact space $\widehat{X}$. Because $\widehat{X}$ is a compact Hausdorff space, every countably additive Baire measure has a unique extension to the Borel σ-field, here denoted $\widehat{\mathcal{X}}$. Thus, each $\mu \in \Delta^{fa}$ corresponds uniquely to a a countably additive $\widehat{\mu}$ on $(\widehat{X}, \widehat{\mathcal{X}})$.

3.1.4. *Some Properties of* $(\widehat{X}, \widehat{\mathcal{X}})$. We first give the foundations for the cardinality and deficiency intuitions discussed in §1.6.

Cardinality. If $E \in \mathcal{X}$ is a countably infinite subset of X, then every subset of E belongs to $\mathcal{X}$ so that $\widehat{E}$ is homeomorphic to the Stone-Čech compactification of the integers, $\beta\mathbb{N}$. Since the cardinality of $\beta\mathbb{N}$ is $2^{\mathfrak{c}}$, there are $2^{\mathfrak{c}}$ weightless ν with $\nu(E) = 1$. Knowing $\{\int f_n\, d\mu : n \in \mathbb{N}\}$ for a countable subset of M_b cannot identify a weightless $\mu \in \Delta^{fa}$ because $\times_{n\in\mathbb{N}} I_{f_n}$ has cardinality $\mathfrak{c}$, while the set of weightless point masses has cardinality at least $2^{\mathfrak{c}}$.

Deficiency. Recall that the deficiency of a $\mu \in \Delta^{fa}$ is defined by $\delta(\mu) = \sup\{\epsilon \geq 0 : \exists E_n \downarrow \varnothing,\ \mu(E_n) \geq \epsilon\}$, the supremum being taken over all sequences E_n decreasing to the empty set. From Yosida and Hewitt [41, Theorem 1.22], the supremum is achieved, say for a sequence E_n. The corresponding sequence of compact sets $\widehat{E}_n$ has the finite intersection property so that $\widehat{E} := \cap_n \widehat{E}_n$ is a non-empty subset of $\widehat{X} \setminus X$. Further, since $\widehat{\mu}$ is countably additive, $\widehat{\mu}(\widehat{E}) = \delta(\mu)$. Thus, the deficient part of μ is carried on $\widehat{X} \setminus X$, and weightless μ's are completely carried on $\widehat{X} \setminus X$.

We will make use of the following extension result.[8]

[8] This is a, basically, a well-known result about compactifications, e.g. Dugundji [17, Theorem 8.2(1)], but for completeness, a proof is provided in the appendix.

Lemma 7 (Extension)**.** *If Y is a compact Hausdorff space and $f : X \to Y$ is measurable, then there is a unique continuous $F : \widehat{X} \to Y$ satisfying $F(x) = f(x)$ for each $x \in X$.*

3.1.5. *Examples of Probabilities on $\widehat{X}$.* Following Maharam [28, §2], there are three easily analyzed classes of weightless probabilities: the convex combinations of weightless point masses; the density measures; and the mapping measures.

Point masses and their convex combinations. The class of sets $\widehat{E}$, $E \in \mathcal{X}$, generates the Borel σ-field on $\widehat{X}$, and so determines points in $\widehat{X}$. Therefore, a point mass on $\widehat{X}$ corresponds to a probability with $\mu \in \Delta^{fa}$ with $\mu(E)$ always being either 0 or 1. For such a μ, the class of $\mathfrak{E} = \{E \in \mathcal{X} : \mu(E) = 1\}$ is an ultrafilter on $\mathcal{X}$: it is closed under finite intersection; if $E' \subset E$ and $E \in \mathfrak{E}$, then $E' \in \mathfrak{E}$; and for all $E \in \mathcal{X}$, either $E \in \mathfrak{E}$ or $E^c \in \mathfrak{E}$. Being a point mass on $\widehat{X} \setminus X$ corresponds to the point mass being weightless, which in turn corresponds to the ultrafilter being free, i.e. $\bigcap\{E : E \in \mathfrak{E}\} = \varnothing$. Conversely, every free ultrafilter in $\mathcal{X}$ corresponds to a point mass on $\widehat{X} \setminus X$. If $\{\delta_k : k \in \mathbb{N}\}$ is a collection of weightless point masses and $\{\alpha_k : k \in \mathbb{N}\}$ a collection of non-negative numbers summing to 1, then $\mu := \sum_k \alpha_k \delta_k$ is a weightless convex combination.

Density measures. Let $E = \{e_1, e_2, \ldots\}$ be a countable subset of X. For each $n \in \mathbb{N}$, let m_n be a probability distribution on $\{e_1, \ldots, e_n\}$, and suppose that for any finite F, $\lim_n m_n(F) = 0$. Because the set of countably additive probabilities on $(\widehat{X}, \widehat{\mathcal{X}})$ is weak* compact, the infinite set $\{m_n : n \in \mathbb{N}\}$ has an accumulation point. Any weak* accumulation point is weightless, and if it is also full valued, then it is called a density measure. Taking m_n to be the uniform distribution on $\{e_1, \ldots, e_n\}$ is a frequently made choice yielding a density measure. One can as easily start with e_n belonging to $\widehat{X} \setminus X$ in the construction, and we will exploit this below.

Mapping measures. Give $Y = \{0, 1\}^{\mathbb{R}}$ the product topology so that it is a compact Hausdorff space with a countable dense subset, D. Let $\mathcal{Y}^\circ$ be the field of subsets of Y that are both open and closed. For each $\theta \in (0, 1)$, let p_θ denote the unique non-atomic countably additive probability on Y determined by the property that the finite dimensional distributions are i.i.d. products of Bernoulli(θ) distributions. For any measurable onto $f : X \to D$, let $F : \widehat{X} \to Y$ be its unique continuous extension, and define $\widehat{\mathcal{X}}_F = F^{-1}(\mathcal{Y}^\circ)$ as the field of closed and open subsets of $\widehat{X}$ of the form $F^{-1}(K)$ for some $K \in \mathcal{Y}^\circ$. Define $\widehat{\mu}_\theta$ on $\widehat{\mathcal{X}}_F$ by $\widehat{\mu}_\theta(F^{-1}(K)) = p_\theta(K)$ for $K \in \mathcal{Y}^\circ$. Any Hahn-Banach extension of $\widehat{\mu}_\theta$ to $\widehat{\mathcal{X}}$ is called a **mapping measure**, and the corresponding μ_θ is weightless and full valued.

3.2. **Indeterminacy.** The following contains the previous indeterminacy results as special cases, but the proof here is quite different from the ones that preceded it. The arguments for Propositions A and B were essentially based on the deficiency intuition and used the simple structure of weightless probabilities restricted to the cofinite sets. The arguments for Proposition C used a special decomposition of X into uncountably many sets that are both dense and countable, and that interact with $\mathcal{C}_{GC}$ in a special fashion, and we do not presently have that luxury. Recall that the field generated by a countable collection of sets is countable.

Theorem 1. *If $\mathcal{C}$ is a countable field of subsets of X, then for every weightless μ, there are uncountably many $\nu \in \Pi(\mathcal{C}, \mu)$ having disjoint support and the property* *that for any $r \in (0,1]$, there are uncountably many ν-distinct sets E_r with $|\mu(E_r) - \nu(E_r)| = r$.*

3.3. **Indeterminacy and Learning for Decisions.** The earlier disjoint support arguments apply here too.

Corollary 1.1. *For every weightless μ and countable $\mathcal{C} \subset \mathcal{X}$, there is a well-behaved $u : \{0,1\} \times X \to [0,1]$ and uncountably many $\nu \in \Pi(\mathcal{C}, \mu)$ with disjoint supports such that $\int u(1,n)\, d\mu(n) = 1$, $\int u(1,n)\, d\nu(n) = 0$, but $\int u(0,n)\, d\nu(n) = 1$.*

3.4. **Indeterminacy and Multiple Prior Models.** Assume now that the space of utility relevant consequences, $\mathbb{W}$, is a compact metric space. We are again interested in the structure of $\mathcal{R}_{\mathbb{W}}(\Pi)$ where $\Pi = \Pi(\mathcal{C}, \mu)$ is the set of probabilities consistent with a weightless μ on a countable field $\mathcal{C}$.

Let $\mathcal{A} = \mathcal{A}(\mathcal{C})$ denote the supnorm separable closure of the simple $\mathcal{C}$-measurable functions and let $\widehat{X}_{\mathcal{A}}$ denote the compactification of X determined by the property that $\mathcal{A}$ is the set of bounded, measurable functions on X that have unique continuous extensions from X to $\widehat{X}_{\mathcal{A}}$. This is the image of $\widehat{X}$ under the mapping $\text{proj}_{\mathcal{A}} : \times_{f \in M_b} I_f \to \times_{f \in \mathcal{A}} I_f$.

Restricted to $\mathcal{A}$, the linear functional $f \mapsto \int f\, d\mu$ determines a unique countably additive Borel probability, $\widehat{\mu}_{\mathcal{A}}$, on the compact metric space $\widehat{X}_{\mathcal{A}}$. The probability $\widehat{\mu}_{\mathcal{A}}$ is a representation of everything about μ that can be learned from $\mu_{|\mathcal{C}}$ and is central in determining the range of problems that can be modeled using $\Pi(\mathcal{C}, \mu)$ as a class of priors for ambiguous choice problems.[9]

Corollary 1.2. *If $\mathbb{W}$ is a compact metric space, μ is weightless, $\mathcal{C}$ is a countable subset of $\mathcal{X}$, and $\Pi = \Pi(\mathcal{C}, \mu)$, then every set in $\mathcal{R}_{\mathbb{W}}(\Pi)$ is of the form $\int_{\widehat{X}_{\mathcal{A}}} \Delta(K(x_{\mathcal{A}}))\, d\widehat{\mu}_{\mathcal{A}}(x_{\mathcal{A}})$ where $x_{\mathcal{A}} \mapsto K(x_{\mathcal{A}})$ is a measurable mapping from $\widehat{X}_{\mathcal{A}}$ to the non-empty closed subsets of $\mathbb{W}$.*

3.5. **Special Cases.** Analysis of special cases is informative, not only for the workings of Corollary 1.2, but for a deeper understanding of the range and depth of the interpretational difficulties discussed in §2.1. As above, ξ_t, $t = 1, 2, \ldots$ denotes the i.i.d. sequence of random variables with distribution μ.

3.5.1. *Integer Indeterminacy Redux.* If $X = \mathbb{N}$ and $\mathcal{C}_{cof}$ is the class of subsets of X that are finite or have finite complement as in §2.2, then $\mathcal{A}(\mathcal{C}_{cof})$ is the set of $f : \mathbb{N} \to \mathbb{R}$ such that $\lim_{n \uparrow \infty} f(n)$ exists. In this case, $\widehat{X}_{\mathcal{A}}$ is the (Alexandroff) one point compactification of X, $\widehat{X}_{\mathcal{A}} = \mathbb{N} \cup \{\infty\}$, and $\widehat{\mu}_{\mathcal{A}}$ is point mass on ∞. For this $\widehat{\mu}_{\mathcal{A}}$, any measurable $x_{\mathcal{A}} \mapsto K(x_{\mathcal{A}})$ is almost everywhere constant, recovering the descriptive range conclusions of Corollary A.2.

Suppose now that μ is a weightless probability on $X = \mathbb{R}$ with cdf satisfying either $F_\mu(r) = \mu((-\infty, r]) \equiv 0$ or $F_\mu(r) \equiv 1$. In the first case, we observe that the i.i.d. sequence of "real numbers" ξ_t belongs to none of the intervals $(-\infty, r]$, in the second case that it belongs to all of them. Let $\mathcal{C}^\circ$ be the field generated by $\mathcal{C}_{\mathbb{Q}} := \{(-\infty, q] : q \in \mathbb{Q}\}$, $\mathbb{Q}$ being the rational numbers in $\mathbb{R}$. For any $f \in \mathcal{A}(\mathcal{C}_{\mathbb{Q}})$, $\lim_{x \uparrow \infty} f(x)$ exists as does $\lim_{x \downarrow \infty} f(x)$. This means that $\widehat{X}_{\mathcal{A}}$ contains a new point,

[9] The metric on $\widehat{X}_{\mathcal{A}}$ can be given by $d(x_{\mathcal{A}}, x'_{\mathcal{A}}) = \sum_n 2^{-n} \min\{1, |\widehat{f}_n(x_{\mathcal{A}}) - \widehat{f}_n(x'_{\mathcal{A}})|$ where $\{f_n : n \in \mathbb{N}\}$ is a sup norm dense subset of $\mathcal{A}$. If $\mathcal{C}$ is a field of sets, then the span of the observation space, $\{0,1\}^{\mathcal{C}}$, is dense in $\mathcal{A}(\mathcal{C})$, which shows that $\widehat{\mu}_{\mathcal{A}}$ contains all that can be learned about μ from its values on $\mathcal{C}$.

denoted $+\infty$, "just to the right of" $\mathbb{R}$ and a new point, denoted $-\infty$, "just to the left of" $\mathbb{R}$. If $F_\mu(r) \equiv 0$, then $\widehat{\mu}_{\mathcal{A}}$ is point mass on $+\infty$, if $F_\mu(r) \equiv 1$, then $\widehat{\mu}_{\mathcal{A}}$ is point mass on $-\infty$. For either of these $\widehat{\mu}_{\mathcal{A}}$'s, any measurable $x_{\mathcal{A}} \mapsto K(x_{\mathcal{A}})$ is almost everywhere constant, recovering the descriptive range of Corollary A.2.

3.5.2. *Intermediate Set Indeterminacy Redux.* Suppose now that $X = \mathbb{N}$ and that $\mathcal{C}_{+E}$ is the smallest field of subsets of X containing $\mathcal{C}_{cof}$ and a set E that is infinite and has infinite complement, as in §2.3. In this case, $\mathcal{A}(\mathcal{C}_{+E})$ is the set of $f : \mathbb{N} \to \mathbb{R}$ such that $\lim_{n\uparrow\infty, n\in E} f(n)$ and $\lim_{n\uparrow\infty, n\in E^c} f(n)$ both exist, but may be different. In this case, $\widehat{X}_{\mathcal{A}}$ is a two point compactification of X, $\mathbb{N} \cup \{\infty_E\} \cup \{\infty_{E^c}\}$, and $\widehat{\mu}_{\mathcal{A}}$ puts mass β on ∞_E and mass $(1-\beta)$ on ∞_{E^c}. For this $\widehat{\mu}_{\mathcal{A}}$, any measurable $x_{\mathcal{A}} \mapsto K(x_{\mathcal{A}})$ takes on at most two values with positive probability, recovering the descriptive range of Corollary B.1.

Suppose now that μ is a weightless probability on $X = \mathbb{R}$ with $F_\mu(r) \equiv \beta$ for some $\beta \in (0,1)$. Here β of the "real numbers" ξ_t belong to each of the intervals, $(-\infty, r]$, while the rest of them belong to none of the intervals. Following the analysis above with $\mathcal{A}(\mathcal{C}_{\mathbb{Q}})$, $\widehat{\mu}_{\mathcal{A}}$ puts mass β on $-\infty$ and mass $(1-\beta)$ on $+\infty$. For this $\widehat{\mu}_{\mathcal{A}}$ as well, any measurable $x_{\mathcal{A}} \mapsto K(x_{\mathcal{A}})$ takes on at most two values with positive probability, recovering the descriptive range in Corollary B.2.

Now let $X = \mathbb{N}$, and let $\varphi : \mathbb{N} \leftrightarrow \mathbb{Q}$ be a bijection between the integers and the rational numbers in $\mathbb{R}$. For each interval, $(-\infty, x] \subset \mathbb{R}$, define $E_x \subset \mathbb{N}$ as $\varphi^{-1}((-\infty, x])$ so that $\mathcal{C}_{\mathbb{N}} := \{E_x : x \in \mathbb{R}\}$ generates the discrete metric σ-field on $\mathbb{N}$. Being linearly ordered by set inclusion, this is a class VC class (with VC dimension 1). Define μ to satisfy $\mu(E_x) \equiv \beta$. Here, β of the "integers" ξ_t belong to each E_x, the rest of them belong to none of the E_x. If $f \in \mathcal{A}(\mathcal{C}_{\mathbb{N}})$, then $\lim_{x\uparrow\infty} f(E_x)$ and $\lim_{x\downarrow\infty} f(E_x)$ exist, and $\widehat{X}_{\mathcal{A}}$ contains two limit points to represent $\widehat{f}$'s value at these limits. For this $\widehat{\mu}_{\mathcal{A}}$ as well, any measurable $x_{\mathcal{A}} \mapsto K(x_{\mathcal{A}})$ takes on only two values with positive probability, recovering the descriptive range in Corollary B.2.

3.5.3. *Weightlessness in the Unit Interval Redux.* Suppose now that the weightless μ is a probability on $\mathbb{R}$ or $[0,1]$, and has a continuous cdf consistent with μ being non-atomic countably additive probability (as in §2.4). In this case, the countably additive $\widehat{\mu}_{\mathcal{A}}$ is also non-atomic. Taking each $K(x_{\mathcal{A}})$ to be a singleton set, $\mathcal{R}_{\mathbb{W}}(\Pi)$ contains all of the risky outcomes because every probability in $\Delta(\mathbb{W})$ can be expressed as an integral of point masses with respect to a non-atomic distribution.

If the space of consequences has two elements, then the class of closed convex subsets of $\Delta(\mathbb{W})$ is two-dimensional set with extreme sets $[0,0]$, $[0,1]$, and $[1,1]$. In this case, $\mathcal{R}_{\mathbb{W}}(\Pi)$ is the set of all closed convex subsets. This provides the circumstance in which one will have the most successful use of a set $\Pi(\mathcal{C}, \mu)$ as a set of priors for ambiguous problems: two utility relevant consequences; and μ has within it the behavior of a non-atomic countably additive probability on a compact metric space.

The need for there to be only two outcomes is acute: if $\mathbb{W}$ is a finite space with 3 or more elements, then the class of closed convex sets is infinite dimensional while $\mathcal{R}_{\mathbb{W}}(\Pi)$ is a closed, finite dimensional, empty interior, convex subset of this class; if the consequence space is infinite, then $\mathcal{R}_{\mathbb{W}}(\Pi)$ is non-generic in every sense known to the author.

3.6. **The Role of Deficiency.** Applying Yosida and Hewitt's [41] unique decomposition of a finitely additive measure into a countably additive part and a purely finitely additive part to probabilities, for any μ with deficiency less than 1, there is a decomposition of μ into a countably additive η_μ and a weightless γ_μ, $\mu = (1-\delta)\eta_\mu + \delta\gamma_\mu$. The arguments for the following closely parallel the arguments given above.

Corollary 1.3. *If μ has deficiency δ, and $\mathcal{C}$ is a countable collection of measurable sets that determines countably additive probabilities, then we have the following:*

1. *if $\nu \in \Pi(\mathcal{C}, \mu)$, then $\nu = (1-\delta)\eta_\mu + \delta\gamma$ for some weightless $\gamma \in \Pi(\mathcal{C}, \gamma_\mu)$;*
2. *there exists a well-behaved $u : \{0,1\} \times X \to [0,1]$ and uncountably many $\nu \in \Pi(\mathcal{C}, \mu)$ with disjoint support sets such that $\int u(1,n)\, d\mu(n) = 1$, $\int u(1,n)\, d\nu(n) = (1-\delta)$, but $\int u(0,n)\, d\nu(n) = 1$; and*
3. *if $\mathbb{W}$ is a compact metric space, then $\mathcal{R}_{\mathbb{W}}(\Pi(\mathcal{C}, \mu))$ is the class of sets of the form $(1-\delta)p + \delta\Delta(K)$ where K is a non-empty, closed subset of $\mathbb{W}$ and p is the image of η_μ under some measurable $f : X \to \mathbb{W}$.*

3.7. **VC Class Indeterminacy.** The results above show that countable classes of sets leave purely finitely additive probabilities indeterminate in the strong sense that not even their support sets are identified. We now show that replacing a VC class with a countable subclass can always be done. Specifically, we show that for any VC class $\mathcal{C}$ and any $\mu \in \Delta^{fa}$, there exists a countable subclass, $\mathcal{C}^\circ$, with $\Pi(\mathcal{C}^\circ, \mu) = \Pi(\mathcal{C}, \mu)$.

For a pseudo-metric space (S, d) and $\epsilon > 0$, define $D(\epsilon, S, d)$ as the maximum number of points in S that are all more than ϵ apart. For a probability Q and measurable sets E, F, we have the pseudo-metric $d_{2,Q}(E, F) = Q(E\Delta F)^{1/2} = \left(\int (1_{E_1} - 1_{E_2})^2\, dQ\right)^{1/2}$ where $E\Delta F$ is the symmetric difference of E and F. For a class of sets $\mathcal{C}$, $D^{(2)}(\epsilon, \mathcal{C})$ is defined as the supremum of $D(\epsilon, \mathcal{C}, d_{2,Q})$ where the supremum is taken over finitely supported Q. From Dudley [14, Theorem 10.1.7], if $\mathcal{C}$ is a VC class, we have the following metric entropy result,

$$\log D^{(2)}(\epsilon, \mathcal{C}) = O(\epsilon^{-2}) \text{ as } \epsilon \downarrow 0. \tag{5}$$

We now show the following.

1. $\mathcal{C}$ satisfies (5) iff $\widehat{\mathcal{C}}$ satisfies (5) where $\widehat{\mathcal{C}} = \{\widehat{C} : C \in \mathcal{C}\}$.
2. The supremum in $D^{(2)}(\epsilon, \widehat{\mathcal{C}})$ can be taken over all countably additive probabilities on $\widehat{\mathcal{X}}$.
3. The previous implies that for any purely finitely additive μ, $\mathcal{C}$ is totally bounded in the $d_\mu(C_1, C_2) := \mu(C_1 \Delta C_2)^{1/2}$ pseudo-metric; and
4. if $\mathcal{C}_0$ is a countable d_μ-dense subset of $\mathcal{C}$, then $\Pi(\mathcal{C}_0, \mu) = \Pi(\mathcal{C}, \mu)$.

Lemma 8. *$\mathcal{C}$ satisfies* (5) *iff $\widehat{\mathcal{C}}$ satisfies* (5)*.*

From the usual weak* approximation of all countably additive probabilities on a compact Hausdorff space by finitely supported probabilities on dense subsets, this means that the supremum $\sup_Q D(\epsilon, \widehat{\mathcal{C}}, d_{2,Q})$ could just as well be taken over all $Q \in \Delta^{ca}(\widehat{\mathcal{X}})$. Therefore, for any weightless μ, we know that

$$\log D(\epsilon, \widehat{\mathcal{C}}, d_{2,\widehat{\mu}}) = \log D(\epsilon, \mathcal{C}, d_{2,\mu}) \leq O(\epsilon^{-2}) \text{ as } \epsilon \downarrow 0. \tag{6}$$

Since $\widehat{\mu}(\widehat{C}) = \mu(C)$ for all $C \in \mathcal{C}$, this means that $\mathcal{C}$ is totally bounded in the pseudo-metric $d_{2,\mu}$. Let $\mathcal{C}_0$ be a countable d_μ-dense subset of $\mathcal{C}$. It is immediate that for any countable, d_μ-dense $\mathcal{C}_0 \subset \mathcal{C}$, $\Pi(\widehat{\mathcal{C}}_0, \widehat{\mu}) = \Pi(\widehat{\mathcal{C}}, \widehat{\mu})$.

4. Summary and Complements

We begin with a summary of the results in this paper, and then turn to a comparison with, and evaluation of, Al-Najjar [1], who initiated the study of weightless learning models. We follow this with a more general discussion of the virtues and vices of weightless probabilities in economic models.

4.1. **Summary.**

The main result here is Theorem 1, which tells us that it is impossible to recover even the support set of a weightless probability from its values on a separable class of sets. We previewed this result in three special cases: Proposition A covered weightless probabilities on the integers with classical VC classes of sets; Proposition B complicated the previous by adding one set that is infinite and has infinite complement; and Proposition C, which covered the case of the unit interval when the weightless probability has a cdf consistent with a non-atomic countably additive probability.

4.1.1. *Intuitions.*

There are two strong intuitions for the impossibility result, one that works through deficiency, and one that works through cardinality. In his review of the B. Rao and B. Rao's [8] monograph on finitely additive measures, Uhl [38] wrote the following.

> There are essentially three ways to prove theorems about finitely additive measures: The easiest is usually proof via the Stone representation theorem which allows a direct transfer of the finitely additive case to the countably additive case. ... The second is Drewnowski's principle ... both show that a finitely additive measure is just a countably additive measure that was unfortunate enough to have been cheated on its domain. The third approach (followed by Rao and Rao) is to prove everything directly with absolutely no reference to the countably additive case.

The construction of $\widehat{X}$ given above in §3.1 is the "Stone representation" that Uhl refers to. The "cheated on its domain" part is the observation that if the deficiency of μ is δ, then $\widehat{\mu}(\widehat{X} \setminus X) = \delta$. If μ is weightless, then it has been totally cheated on its domain, and one cannot observe random draws according to μ using the space X because the draws take values in $\widehat{X} \setminus X$ with probability 1. The interpretational difficulties highlighted in §2.1 and §3.5 give a sense of what is involved.

To circumvent these difficulties, we convert ξ_t into the infinite length vector, $\{1_C(\xi_t) : C \in \mathcal{C}\} \in \{0,1\}^{\mathcal{C}}$. This has the effect of making the distributions of the observations countably additive. Since the cardinality of countable products of $\{0,1\}^{\mathcal{C}}$ is $\mathfrak{c}$ when $\mathcal{C}$ is countable, this clarifies why one cannot use sequences of observations to distinguish between the $2^{\mathfrak{c}}$ different weightless probabilities.

4.1.2. *Corollaries About Decision Makers.*

We studied two classes of decision makers, expected utility maximizers facing a decision problem knowing what they had learned from the sequence of observations, and multiple prior decision makers in the same situation facing a choice between acts.

Expected utility maximizers. If what one has learned one does not even identify the support of the probability μ that one is facing when trying to solve the problem

$$\max_{a \in A} \int u(a, x)\, d\mu(x), \tag{7}$$

then it should come as no surprise that what one has learned is, in general, useless for an expected utility maximizer. This is the content of Corollary 1.1, and its special case predecessors, Corollaries A.1, B.1, and C.1.

In particular, if μ is weightless, this means that one could not, in any generality, predict an expected utility maximizer's choice, $a^*(\mu)$, in (7) if one knew both the utility function and the sequence of observations that they had made. This is in stark contrast to the countably additive case — Lemma 1 showed that for any measure space $(X, \mathcal{X})$ and any countably additive P on $\mathcal{X}$, if P_n is a random sequence of empirical distributions from the observations ξ_t, $t = 1, \ldots, n$, then optimizing against the P_n converges to the solution set for the problem in (7).

Multiple prior decision makers. Many models of choice under ambiguity substitute a set S of priors for the single prior and use this set to rank choices between functions f mapping states to utility relevant consequences. For such models, the choice between the pair of functions f and g models a situation in which the decision maker believes that they are choosing between two sets of distributions, $A := f(S)$ and $B := g(S)$. For such a model to be useful, the class of sets of the form $f(S)$ should contain the decision problems of interest.

We have seen the circumstances in which one will most successfully be able to use of a set $\Pi(\mathcal{C}, \mu)$ as a set of priors for ambiguous problems: it involves there being two utility relevant consequences; it also involves μ behaving like a non-atomic countably additive probability with respect to a supnorm separable sub-algebra of bounded measurable functions. However, if there are three or more utility relevant consequences, the set of decision problems that can be modeled is extremely small in many formal senses.

There do exist sets of priors with which one can model all relevant problems. A set of priors, S, is **descriptively complete** if every closed and convex subset of probabilities on any compact metric space of utility relevant consequences is of the form $f(S)$ for some measurable function f. Dumav and Stinchcombe [18] give sufficient, and nearly necessary, conditions for a set of priors to be descriptively complete. Some of the choice theoretic implications of descriptive completeness are covered in Dumav and Stinchcombe [19].

If a set of priors is descriptively complete, then the details of S and the functions f and g do not matter — there always exist many functions f and g giving rise to the closed convex sets of probabilities A and B. Examples in [19] show that many well-known sets of priors can only model negligible sets of problems, while [19, Theorem 3] shows that a set of priors failing descriptive completeness *always* fails to model a dense set of problems when the outcome space, $\mathbb{W}$, is e.g. an interval subset of $\mathbb{R}$. When we move away from the two-outcome case, we have seen that we are only able to model a null set of problems using $\Pi(\mathcal{C}, \mu)$ as a set of priors.

4.1.3. *Comparisons with Convex Capacities.* There are three comparisons in §2 between the ambiguous choice problems that can be modeled using $\Pi(\mathcal{C}, \mu)$ and using the core of a convex capacity, $\mathcal{P}_c$.

Lemma 3 shows that if one what one has learned about a weightless probability μ (on the integers) is that its cdf is identically equal to 0, then the set of problems that can be modeled is exactly the same as the set of problems that can be modeled using the core of the complete ignorance capacity. Here the intuition is clear, in neither case does one have any information about the set of probabilities, so that what can be modeled is quite limited, only sets of probabilities of the form $\Delta(K)$, K a closed set.

Lemma 4 and Example 1 show that, in the two outcome case, what can be modeled using a convex distortion capacity compares unfavorably to what can be modeled using $\Pi(\mathcal{C}_{+E}, \mu)$.

Lemma 6 shows that no convex capacity has a core that can model both a risky problem with a uniform distribution on $[0, 1]$ and an ambiguous problem in which ignorance about what distribution on $[0, 1]$ applies. By contrast, Corollary 1.2 shows that if μ contains within it a part that is compatible with a non-atomic countably additive probability on $\mathcal{C}$, then $\mathcal{R}(\Pi(\mathcal{C}, \mu))$ contains all (generalized) convex combinations of sets of the form $\Delta(K)$. In particular, it contains $\Delta([0, 1])$ and the uniform distribution.

To summarize, what can be represented using $\Pi(\mathcal{C}, \mu)$ always includes ignorance about what distribution with support set $K \subset \mathbb{W}$ obtains, and may contain some, or many, convex combinations of these sets. In the case that the descriptive range of $\Pi(\mathcal{C}, \mu)$ is large as it can be, no convex capacity has a core with descriptive range containing these convex combinations.[10]

4.2. **Comparison and Evaluation.** Motivated as an attempt to understand the comparative difficulty of learning problems, [1] partially investigated the structure of $\Pi(\mathcal{C}, \mu)$, noting that his indeterminacy result produces "an explicit model of learning to derive a set of probability measures ... consistent with empirical evidence." We have now seen that his model of weightless learning sometimes covers interesting classes of choice problems in the face of ambiguity. We turn to his analysis of the comparative difficulties of different types of learning problems.

In his notation: discrete spaces are $(X_d, \mathcal{X}_d)$, X_d being an infinite, non-empty set, and $\mathcal{X}_d$ being the class of all subsets of X_d, and he considers only weightless probabilities on such spaces; the "standard outcome spaces" are $(X_c, \mathcal{X}_c)$, X_c being a complete separable metric (Polish) space and $\mathcal{X}_c$ its Borel σ-field, and he claims to consider only countably additive probabilities on such spaces. He presents two main results, and several interpretations.

Al-Najjar's Theorem 3 notes that for every Polish space, X_c, there is a measurable $\varphi : X_c \to [0, 1]$ with a measurable range. Since the Glivenko-Cantelli class of sets is a VC class for $[0, 1]$, the class of sets $\mathcal{C} := \varphi^{-1}([0, r])$, $r \in [0, 1]$ is a VC class for countably additive probabilities on X_c. He offers the following interpretation.

> Theorem 3 shows that the known theory of uniform learning has no bite in the limit. Specifically, in standard outcome spaces, which I take to be complete separable metric spaces with countably additive probabilities, all statistical ambiguity disappears in the limit. In these spaces, the tension between the availability of data and statistical complexity disappears. This is at odds with the central role this tension plays in finite settings in distinguishing between simple and hard learning problems. (Al-Najjar, p. 1373)

Al-Najjar's Theorem 4 and its Corollary show that for every discrete space, X_d and every VC class $\mathcal{C}$, there exists a weightless density measure, μ, that has uncountably

[10] The characterization of the extreme points of the cores of convex capacities in El Kaabouchi [21] should yield a characterization of these descriptive ranges.

many perturbations, ν, that agree on $\mathcal{C}$ and have $|\mu(E_r) - \nu(E_r)| = r$ for uncountably many E_r and any $r \in (0, \frac{1}{2}]$. The offered interpretation is "that statistical ambiguity persists in the form of a set of probability measures representing beliefs that are not contradicted by data."

The essential conclusion that he draws is "that the asymptotic elimination of statistical ambiguity in standard outcome spaces (the X_c) is a consequence of implicit structural restrictions these spaces impose. For example, the Borel events on $[0, 1]$ are defined in terms of a topology that embeds a notion of similarity between outcomes." Put in another fashion, he argues that the learnability of a probability arises from properties of the space on which it is defined, from the "inductive biases involving notions of distance, ordering, or similarity."

Aside from the doubts raised by any argument leading from human "inductive biases" to the truth or falsity of mathematical statements, there are several deeper reasons to doubt these conclusions and interpretations.

First, Al-Najjar's halfway indeterminacy result is in fact a complete indeterminacy result that holds for all weightless probabilities on all measure spaces. The "implicit structural restrictions" play no role.

Second, for decision theory and estimation, "the tension between the availability of data and statistical complexity" does not "disappear." Rather, the tension takes the form of different rates of convergence. If one takes limits, then by definition, one goes infinitely far beyond any finite model. That much data solves all finite problems, as well as all of the infinite problems using countably additive probabilities.

Third, the discrete spaces that Al-Najjar uses are complete metric spaces, the only ones appearing in the proof of his Theorem 4 are also separable. Comparisons of weightless probabilities on a special class of Polish spaces to countably additive probabilities on Polish spaces cannot possibly throw light on the role of failures of metrizability.

Fourth, the argument that "Borel events on $[0, 1]$ are defined in terms of a topology that embeds a notion of similarity between outcomes" is contradicted by the Borel isomorphism theorem, a weak version of which is the crucial ingredient in the proof of Al-Najjar's Theorem 3. While it is certainly true that it is *conventional* to define the Borel σ-field on $[0, 1]$ in terms of the usual metric topology, the Borel isomorphism theorem tells us that all uncountable Borel measurable subsets of Polish spaces are in fact the same measure space in different disguises: if X and Y are measurable subsets of Polish spaces and have the same cardinality, then there exists a measurable isomorphism between them, that is, there exists a one-to-one, onto measurable $\varphi : X \leftrightarrow Y$ with a measurable inverse.

To get a sense of what is involved, suppose that $\varphi : X_c \leftrightarrow [0, 1]$ is the measurable isomorphism between an uncountable Polish space and $[0, 1]$. The Borel σ-field on $[0, 1]$ is also the Borel σ-field generated by the metric topology for the metric $d_\varphi(r, r') = \rho_{X_c}(\varphi^{-1}(r), \varphi^{-1}(r'))$ where ρ_{X_c} is the metric on X_c. The following shows that none of the usual notions of "similarity between outcomes" are at work in the Borel σ-field on $[0, 1]$.

Lemma 9. *If X_c is a Banach space of dimension 2 or higher and $\varphi : X_c \leftrightarrow [0, 1]$ is a measurable isomorphism, then the set of discontinuities of φ is dense in X_c,* *and for any $\epsilon > 0$, there is an interval subset $(a, b) \subset [0, 1]$ having diameter less than ϵ with the diameter of $\varphi^{-1}((a, b))$ infinite.*

4.3. **On Failures of Countable Additivity.** Probabilities that fail countable additivity have some virtues and some compensating vices. We discuss some of these in turn, and with these perspectives in hand, turn to some of the heated rhetoric that surrounds these issues.

4.3.1. *Virtues.* Weightless probabilities have at least four advantages: the Bayes optimality of certain 'natural' statistical procedures; uncovering the basic structures of probabilistic limit arguments in the law of large numbers, the central limit theorem, and the 0-1 laws; being able to dispense with measurability considerations; and being able to represent equilibria in some classes of discontinuous normal form games.

Natural statistical procedures. There are many settings in which natural, non-Bayesian statistical procedures are not admissible unless one allows for uniform priors on 'arbitrarily large intervals in $\mathbb{R}$,' e.g. an accumulation point of the sequence of uniform distributions on $[-n, +n] \subset \mathbb{R}$ (see Jaynes [25, Ch. 12, 13] for an extensive development of these ideas, or the postscript to Blackwell and Diaconis [10] for a brief discussion and some references). In these contexts, the natural procedures are the limits, as $n \to \infty$, of what a Bayesian would do if one's true prior distribution were uniform on $[-n, +n]$. Because the optimal actions at n converge to the optimal actions in the purely finitely additive limit, the deficiency of the limit has little effect beyond making some interpretations rather a stretch.[11]

Structures of limit theorems. The usual proofs of the basic limit theorems of probability, the strong law of large numbers, the central limit theorem, the various 0-1 laws, depend in a crucial fashion on countable additivity. By replacing countable additivity with succesively stronger restrictions on finitely additive models of probability, one can recover the limit theorems (see Berti et. al. [6] and Berti and Rigo [7] for an overview and recent progress in this area). This has the potential to make clearer the structure of the fundamental arguments behind the limit results.

Measurability. By the Hahn-Banach theorem, it is always possible to extend a probability from a field or σ-field to the collection of all subsets of a space provided that one is willing to forego both the uniqueness and the countable additivity of the extension. This means that one can, if one wishes, work with probabilities in a setting where all sets are measurable if one is willing to work with weightless probabilities. The work here has provided some extra insight into the extent of the failure of uniqueness.

Discontinuous games. Consider games with a finite set of players, $\{1, \ldots, I\}$, infinite action spaces A_i, $i \in I$, and expected utility functions $u_i : \times_{j \in I} A_j \to [0, 1]$. For some utility functions, it is possible to integrate a product of weightless probabilities. For games with such utility functions, Marinacci [29] showed that equilibria exists in $\times_{j \in I} \Delta^{fa}(A_i)$, and from this deduced the existence of ϵ-Nash equilibria in countably additive strategies.

Harris et al. [22] give measure theoretic, algebraic, functional analytic, and finite approximability characterizations of the set of utility functions for which the

[11] Having one's prior probability put mass 1 on the absolute value of an estimated mean yearly income per capita in the US being more than $\$10^{10^{10}}$ might be interpreted as certainty that all of the Census Bureau figures are off by many orders of magnitude.

integration of products of weightless probabilities is well-defined. Using arguments similar to those in Corollary 1.2, they also showed that in such games, it is always possible to imbed the A_i as dense subsets of compact metric spaces in such a fashion that the u_i have jointly continuous extensions. From this, the usual upper hemi-continuity results about equilibrium sets are available, especially those about the upper and lower hemi-continuity of the approximate equilibrium correspondence when finite approximations to the A_i are used. Unfortunately for the general theory of infinite games, Stinchcombe [36, §1.2] shows that the class of normal form games covered by this approach does not include the normal forms of most extensive form games with infinite choice sets.

4.3.2. *Vices.* Modeling with weightless probabilities on a measure space $(X, \mathcal{X})$ corresponds to modeling with probabilities defined on a space, $\widehat{X} \setminus X$, that is not part of the model. This causes the problems interpreting random draws discussed above, as well as a variety of failures of continuity and extension central to modeling random phenomena. The failure of continuity affects *inter alia* the meaning of jump processes in stochastic process theory, the convergence of sample averages to true means, the central limit theorem, 0-1 laws, the martingale convergence theorem. The failure of extension impacts Nash equilibrium and independence at the level of basic definitions, and putting weight on unmodeled parts of a state space affects the basic logic of choice in the presence of risk.

Measurability and limit theorems redux. One cost of being able to avoid measurability arguments by using weightless probabilities is the loss of Lebesgue's dominated convergence theorem, the essential continuity result for much the of probabilistic reasoning that economists do. This result says that if P is countably additive, f_n is a dominated sequence of integrable random variables, and $P(\{x \in X : f_n(x) \to f(x)\}) = 1$, then averages converge, $\int f_n \, dP \to \int f \, dP$. If μ is weightless, not even the stronger assumption that $f_n(x) \to f(x)$ for all $x \in X$ is sufficient — take $f_n(x) = 1_{E_n}(x)$ where $E_n \downarrow \varnothing$ and $\mu(E_n) \equiv 1$.

The problem is that assumptions about the behavior of the f_n on points in X may have no bearing on the behavior of f on the points on which μ is carried, $\widehat{X} \setminus X$. One can, at a cost, recover the convergence of averages: either strengthen the convergence assumption e.g. to $\sup_{x \in X} |f_n(x) - f(x)| \to 0$, an assumption that fails if $f_n(x)$ is the average of n non-degenerate i.i.d. random variables; or assume, in some form or another, that μ behaves as if it were countably additive relative to the sequence f_n.

Stochastic processes. Kingman [27] shows that when X is the set of polynomial times paths on $[0, \infty)$, there are weightless probabilities on X with the finite dimensional distributions of a point process. In other words, there are weightless probabilities putting mass 1 on the set of polynomial paths, but their finite dimensional distributions, i.e. the distribution of what can be observed at points in time $t_1, t_2, \ldots, t_N$, are only consistent with jumps. In a similar fashion, there are weightless models of Brownian motions having only polynomial paths.

Product measures. Fundamental to the definition of a Nash equilibrium and to the definition of an i.i.d. sequence of random variables is the extension of a product of distributions on A_i, $i = 1, \ldots I$ or $i = 1, 2, \ldots$, from the product field to the product

σ-field. When the probabilities are countably additive we can use Kolmogorov's extension theorem, when the probabilities are weightless, information about behavior on $\times_i A_i$ often carries little information about about their behavior on $\times_i \widehat{A}_i$.

For example, if μ is a weightless full-valued probability on $\mathbb{N}$, the set of extensions of $\mu \times \mu$ to the product σ-field on $\mathbb{N} \times \mathbb{N}$ contains probabilities assigning mass p to the event that the first draw is larger than the second, this for any $p \in [0, 1]$. Only those extensions that have $p = \frac{1}{2}$ accord with notions of independence. Purves and Sudderth [32] show that one can pick between the different extensions in a fashion that allows the study of stochastic processes to go forward, Stinchcombe [36] shows that one can understand and evaluate the plausibility, for equilibrium analysis, of different extensions.

The logic of bets. As Ambrose Bierce observed, there are good bets that turn out badly, but this does not affect our estimation of the rationality of having taken the bet.[12] In surprisingly simple betting situations, this dynamic consistency logic is lost if one uses a weightless full valued probability as a prior (see [34] for citations to sources of this class of examples). All that is required is a model of decision making rich enough to include a single infinite support distribution, e.g. a Gaussian or a Poisson distribution.

Consider someone with low enough risk aversion that they would be willing to pay \$5 to take a bet on a fair coin that pays them \$100 if Heads comes up and costs them \$10 if Tails comes up. If they take the bet and Tails occurs, they would be willing to pay any amount up to \$10 to get out of having taking the bet, but if Heads occurs, they would need to be paid at least \$100 to give up their claim to the winnings. Let us further suppose that if the odds against Heads were $1 : 10,000$, they would pay at least \$9 to be rid of the bet.

Let $E \subset \mathbb{N}$ be the evens, $E = \{2k : k \in \mathbb{N}\}$ and let O be its complement, $O = \{2k-1 : k \in \mathbb{N}\}$. If μ is a weightless full valued prior on a state space $(X, \mathcal{X})$, then there exists a function $f : X \to \mathbb{N}$ with the following properties: (i) $\mu(f^{-1}(E)) = \mu(f^{-1}(O)) = \frac{1}{2}$; (ii) for every $k \in \mathbb{N}$, $\mu(f^{-1}(\{2k-1, 2k\}) > 0$; and (iii), for every $k \in \mathbb{N}$,

$$\mu(f \in E \big| f \in \{2k-1, 2k\}) / \mu(f \in O \big| f \in \{2k-1, 2k\}) = 1/10,000. \tag{8}$$

If a bet pays \$100 if $f(x) \in E$ and costs \$10 if $f(x) \in O$ occurs, the decision maker would be willing to buy the bet for a price of \$5. However, after having being told that $f(x)$ belongs to an interval $\{2k-1, 2k\}$ for *any* $k \in \mathbb{N}$, they would be willing to pay at least \$9 to get out of the obligation.

Apparently, a pair of willing decisions leads to a certain loss of at least \$14, and we can replace "\$14" with any amount at which the utility function is still strictly increasing. The problem is that the weightlessness of μ means that the states in which the decision maker is happy to have taken the bet have been mis-laid, they have been moved outside of the model. In the space $\widehat{X} \setminus X$, there is a set having mass $\frac{1}{2} \cdot \frac{9,999}{10,000}$ on which the decision maker would need to be paid at least \$100 to

[12] Bierce's short poem, "A Lacking Factor," is cited in Jaynes [25, Ch. 13].

> "You acted unwisely," I cried, "as you see
> By the outcome." He calmly eyed me:
> "When choosing the course of my action," said he,
> "I had not the outcome to guide me."

give up their claim, and on which the unique extension of f takes values in $\widehat{E} \setminus E$, points not to be found in $\mathbb{N}$.

The first three vices can be overcome at what seems, to the author, an extravagant cost in terms of mathematical complexity, but the fourth vice seems to strike more directly at the foundations of economic modeling. If one models decision makers as having weightless priors and uses only the state space and the range space in the model, then the decision makers are so dynamically inconsistent as to be happy to give away essentially all of their money. This is a bad ingredient for models of markets or trade.[13]

4.3.3. *Rhetorical Heat.* Given the difficulties of using and interpreting weightless probabilities, it seems odd to the author that they should still be found in economic models. This puzzling persistence of weightlessness in decision theory is perhaps due to Savage's [33] and de Finetti's [12] ardent advocacy: one must have "gone mad" to advocate for countable additivity; insisting on countable additivity is not "scientific." The arguments have a surface plausibility — since Δ^{ca} is a strict subset of Δ^{fa}, it might seem that using Δ^{fa} is more mathematically cautious, more "scientific" as de Finetti would have it.

The deficiency arguments and the money pump examples tell us that one only arrives at Δ^{fa} from Δ^{ca} by moving mass in the model to someplace outside of the model. The indeterminacy results in this paper provide one more example of a consequence of arriving at Δ^{fa} by this kind of movement. Here however, there is a silver lining: we can now model some convex combinations of ignorance about consequences using, as sets of priors, the result of a model of learning.

References

[1] Nabil Al-Najjar. Decision makers as statisticians: diversity, ambiguity, and learning. *Econometrica*, 77(5):1371–1401, 2009.

[2] Robert M. Anderson and William R. Zame. Genericity with infinitely many parameters. *Adv. Theor. Econ.*, 1:Art. 1, 64 pp. (electronic), 2001.

[3] Donald W. K. Andrews. Chi-square diagnostic tests for econometric models: theory. *Econometrica*, 56(6):1419–1453, 1988.

[4] Donald W. K. Andrews. Empirical process methods in econometrics. In *Handbook of econometrics, Vol. IV*, volume 2 of *Handbooks in Economics.*, pages 2247–2294. North-Holland, Amsterdam, 1994.

[5] Yoav Benyamini and Joram Lindenstrauss. *Geometric nonlinear functional analysis. Vol. 1*, volume 48 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2000.

[6] Patrizia Berti, Eugenio Regazzini, and Pietro Rigo. Modes of convergence in the coherent framework. *Sankhyā*, 69(2):314–329, 2007.

[7] Patrizia Berti and Pietro Rigo. Finitely additive uniform limit theorems. *Sankhyā*, 68(1):24–44, 2006.

[8] K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges: A study of finitely additive measures*, volume 109 of *Pure and Applied Mathematics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983.

[9] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[13] In some contexts, neither the countably additive probabilities on $([0,1],\mathcal{B})$ nor the weightless probabilities on $([0,1],2^{[0,1]})$ are sufficient for modeling purposes, see e.g. Khan and Sun [26], who discuss this issue for arbitrage theory in models of economies with a continuum of agents.

[10] David Blackwell and Persi Diaconis. A non-measurable tail set. In *Statistics, probability and game theory*, volume 30 of *IMS Lecture Notes Monogr. Ser.*, pages 1–5. Inst. Math. Statist., Hayward, CA, 1996.
[11] Felipe Cucker and Steve Smale. On the mathematical foundations of learning. *Bull. Amer. Math. Soc. (N.S.)*, 39(1):1–49 (electronic), 2002.
[12] Bruno de Finetti. *Probability, induction and statistics. The art of guessing.* John Wiley & Sons, London-New York-Sydney, 1972. Wiley Series in Probability and Mathematical Statistics.
[13] Claude Dellacherie and Paul-André Meyer. *Probabilities and potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1978.
[14] Richard M. Dudley. *Uniform central limit theorems*, volume 63 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999.
[15] Richard M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.
[16] Richard M. Dudley, Evarist Giné, and Joel Zinn. Uniform and universal Glivenko-Cantelli classes. *J. Theoret. Probab.*, 4(3):485–510, 1991.
[17] James Dugundji. *Topology*. Allyn and Bacon Inc., Boston, Mass., 1966.
[18] Martin Dumav and Maxwell B. Stinchcombe. Skorokhod's representation theorem for sets of probabilities. Working Paper, Department of Economics, U.T. Austin, 2013.
[19] Martin Dumav and Maxwell B. Stinchcombe. The von Neumann/Morgenstern approach to ambiguity. Working Paper, Department of Economics, U.T. Austin, 2013.
[20] Nelson Dunford and Jacob T. Schwartz. *Linear operators. Part I: General theory*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1988.
[21] Aziz El Kaabouchi. Mesures dominées par une capacité alternée d'ordre 2. *Proc. Amer. Math. Soc.*, 121(3):823–832, 1994.
[22] Christopher J. Harris, Maxwell B. Stinchcombe, and William R. Zame. Nearly compact and continuous normal form games: characterizations and equilibrium existence. *Games Econom. Behav.*, 50(2):208–224, 2005.
[23] David Heath and William Sudderth. On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.*, 6(2):333–345, 1978.
[24] Peter J. Huber and Volker Strassen. Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Statist.*, 1:251–263, 1973.
[25] Edwin T. Jaynes. *Probability theory: The logic of science*. Cambridge University Press, Cambridge, 2003.
[26] M. Ali Khan and Yeneng Sun. Asymptotic arbitrage and the APT with or without measure-theoretic structures. *Journal of Economic Theory*, 101(1):222–251, 2001.
[27] J. F. C. Kingman. Additive set functions and the theory of probability. *Proc. Cambridge Philos. Soc.*, 63:767–775, 1967.
[28] Dorothy Maharam. Finitely additive measures on the integers. *Sankhyā Ser. A*, 38(1):44–59, 1976.
[29] Massimo Marinacci. Finitely additive and epsilon Nash equilibria. *Internat. J. Game Theory*, 26(3):315–333, 1997.
[30] David Pollard. *Convergence of stochastic processes*. Springer Series in Statistics. Springer-Verlag, New York, 1984.
[31] David Pollard. New ways to prove central limit theorems. *Econometric Theory*, 1(3):295–313, 1985.
[32] Roger A. Purves and William D. Sudderth. Some finitely additive probability. *Ann. Probability*, 4(2):259–276, 1976.
[33] Leonard J. Savage. *The foundations of statistics*. Dover Publications Inc., New York, revised edition, 1972.
[34] Maxwell B. Stinchcombe. Countably additive subjective probabilities. *Rev. Econom. Stud.*, 64(1):125–146, 1997.
[35] Maxwell B. Stinchcombe. The gap between probability and prevalence: loneliness in vector spaces. *Proc. Amer. Math. Soc.*, 129(2):451–457, 2001.
[36] Maxwell B. Stinchcombe. Nash equilibrium and generalized integration for infinite normal form games. *Games Econom. Behav.*, 50(2):332–365, 2005.

[37] Maxwell B. Stinchcombe and Halbert White. Some measurability results for extrema of random functions over random sets. *Rev. Econom. Stud.*, 59(3):495–514, 1992.
[38] J. Jerry Uhl. Review of Rao and Rao (1983). *Bull. Lond. Math. Soc.*, 16(4):431–432, 1984.
[39] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes: With applications to statistics.* Springer Series in Statistics. Springer-Verlag, New York, 1996.
[40] Vladimir N. Vapnik and A. Ja. Červonenkis. On the uniform convergence of frequencies of the appearance of events to their probabilities. *Theor. Probability Appl.*, 16:264–280, 1971.
[41] Kôsaku Yosida and Edwin Hewitt. Finitely additive measures. *Trans. Amer. Math. Soc.*, 72:46–66, 1952.

Appendix A. Proofs

Proof of Lemma 1. Because $\left|\max_{a\in A}\int u(a,x)\,dP_n(x) - \max_{b\in A}\int u(b,x)\,dP(x)\right|$ is less than or equal to $v_n := \max_{a\in A}\left|\int u(a,x)\,dP_n(x) - \int u(a,x)\,dP(x)\right|$, it is sufficient to show $v_n \to 0$. For this, it is in turn sufficient to show that every subsequence, $v_{n'}$ has a further subsequence, $v_{n''}$, with $v_{n''} \to 0$. Along the subsequence n', let $a_{n'}$ maximize $v_{n'}$. Because A is compact, there exists an $a^* \in A$ and a further subsequence, $a_{n''} \to a^*$. Relabeling the subsubsequence $a_{n''}$ as a_m, we have $a_m \to a^*$, and we must show that $v_m = |\int u(a_m,x)\,dP_m(x) - \int u(a_m,x)\,dP(x)| \to 0$. Pick $\epsilon > 0$. We will show that there exists an M such that for all $m \geq M$, $v_m < 7\cdot\epsilon$.

Adding and subtracting $\int u(a^*,x)\,dP_m(x)$ and $\int u(a^*,x)\,dP(x)$ to v_m, we have

$$v_m = \left|\int u(a_m,x)\,dP_m(x) - \int u(a_m,x)\,dP(x)\right| \leq \tag{9}$$

$$\left|\int u(a_m,x)\,dP_m(x) - \int u(a^*,x)\,dP_m(x)\right| + \tag{10}$$

$$\left|\int u(a^*,x)\,dP_m(x) - \int u(a^*,x)\,dP(x)\right| + \tag{11}$$

$$\left|\int u(a^*,x)\,dP(x) - \int u(a_m,x)\,dP(x)\right|. \tag{12}$$

For the term in (10): The pointwise, hence almost everywhere, convergence of $u(a_m,x)$ to $u(a^*,x)$ implies that there exists an $E \in \mathcal{X}$ with $P(E) > 1-\epsilon$, and an M_1 such that for all $m \geq M_1$ and all $x \in E$, $|u(a_m,x) - u(a^*,x)| < \epsilon$. Integrating separately over E and E^c, this term is less than or equal to

$$\left|\int_E u(a_m,x)\,dP_m(x) - \int_E u(a^*,x)\,dP_m(x)\right| + \tag{13}$$

$$\left|\int_{E^c} u(a_m,x)\,dP_m(x) - \int_{E^c} u(a^*,x)\,dP_m(x)\right|. \tag{14}$$

For $m \geq M_1$, the first of these, in (13), is less than ϵ. By assumption, $\int f\,d \to \int f\,dP$ for all measurable bounded f. Taking $f = 1_{E^c}$, we can pick M_2 such that for all $m \geq M_2$, $|\int 1_{E^c}\,dP_m - \int 1_{E^c}\,dP| < \epsilon$. Because $0 \leq u(a,x) \leq 1$ for all a and x and

$P(E^c) < \epsilon$, we have the following bounds for (14),

$$\left|\int_{E^c} u(a_m, x)\, dP_m(x) - \int_{E^c} u(a^*, x)\, dP_m(x)\right| \leq \tag{15}$$

$$\left|\int_{E^c} u(a_m, x)\, dP_m(x)\right| + \left|\int_{E^c} u(a^*, x)\, dP_m(x)\right| \leq \tag{16}$$

$$\left|\int_{E^c} 1\, dP_m(x)\right| + \left|\int_{E^c} 1\, dP_m(x)\right| < 2 \cdot \epsilon + 2 \cdot \epsilon. \tag{17}$$

Combining, for all $m \geq \max\{M_1, M_2\}$, the term in (10) is less than $5 \cdot \epsilon$.

For the term in (11): Note that $x \mapsto u(a^*, x)$ is measurable and bounded. By assumption, there exists an M_3 such that for all $m \geq M_3$, this is less than ϵ.

For the term in (12): By Lebesgue's dominated convergence theorem, there exists an M_4 such that for all $m \geq M_4$, this term is less than ϵ.

Combining, for all $m \geq \max\{M_1, M_2, M_3, M_4\}$, we have $v_m < 7 \cdot \epsilon$. □

Proof of Lemma 5. We first give the requisite partition. Every irrational $x \in [0, 1]$ has a unique dyadic expansion, $x = \sum_n r_n(x)/2^n$ where $\{r_n(x) : n \in \mathbb{N}\}$ is a sequence in $\{0, 1\}$. For irrational x, define $f(x) = \limsup_N (\sum_{n=1}^N r_n(x))/N$, and for rational x, define $f(x) = \frac{1}{2}$. The function f is measurable, the requisite partition can be defined by setting $I_\gamma = f^{-1}(\gamma)$.

Suppose now that every $x \in [0, 1]$ is a condensation point of $I \subset [0, 1]$. For each $n \in \mathbb{N}$ and $k \in \{0, 1, \ldots, 2^n - 1\}$, $I \cap (k/2^n, (k+1)/2^n]$ is uncountable, hence there exists a bijection, $\varphi_{n,k} : I \cap (k/2^n, (k+1)/2^n] \leftrightarrow [0, 1]$. For each $r \in [0, 1]$, let $J_r = \{\varphi_{n,k}^{-1}(r) : n \in \mathbb{N}, k \in \{0, \ldots, 2^n - 1\}\}$. Since J_r intersects every dyadic interval subset of $(0, 1]$, it is dense in $[0, 1]$, and by construction, for $r \neq r'$, $J_r \cap J_{r'} = \emptyset$. □

Proof of Lemma 7. Since Y is a compact Hausdorff space, $C(Y)$ is a sup norm closed algebra of bounded functions. Therefore $\mathcal{A} := \{h \circ f : h \in C(Y)\} \subset M_b$ is a sup norm closed algebra of bounded functions on X. For each $\widehat{x} \in \widehat{X}$ and each $h \in C(Y)$, $h(F(\widehat{x}))$ can be defined as $\text{proj}_{h \circ f}(\widehat{x})$. Since a point $y \in Y$ is uniquely determined by the values of $h(y)$ for $h \in C(Y)$, this uniquely determines $F(\widehat{x})$. The continuity of F follows from the observation that $F(\widehat{x}) = \text{proj}_{\mathcal{A}}(\widehat{x})$. The fact that F is an extension of f follows by construction. □

Proof of Theorem 1. Let $\mathcal{C}$ be a countable field of measurable sets, and let μ be a weightless probability on $(X, \mathcal{X})$. Enumerate $\mathcal{C}$ as $\{C_m : m \in \mathbb{N}\}$, and for each $M \in \mathbb{N}$, let $\mathcal{C}_M$ denote the partition of X generated by $\{C_m : m \leq M\}$.

Since μ is weightless, for each infinite $C \in \mathcal{C}_M$, we can pick a countably infinite $D_C \subset C$ with $\mu(D_C) = 0$ (e.g. by Proposition A). Let $\widehat{D}_M = \bigcup\{\widehat{D}_C : C \in \mathcal{C}_M, C \text{ is infinite}\}$ so that $\widehat{\mu}(\widehat{D}_M) = 0$. Define $\widehat{D}_\infty = \cup_M \widehat{D}_M$. Because $\widehat{\mu}$ is countably additive, $\widehat{\mu}(\widehat{D}_\infty) = 0$. The proof will be complete once we show that there are uncountably many non-atomic $\widehat{\nu}$ carried by $\widehat{D}_\infty$ and agreeing with $\widehat{\mu}$ on $\widehat{\mathcal{C}} := \{\widehat{C} : C \in \mathcal{C}\}$.

Returning to the definition of the D_C, let $\widehat{E}_C = \{\widehat{e}_1, \ldots, \widehat{e}_M\}$ be a subset of $\widehat{D}_C \setminus X$ with M elements. Let $\widehat{\nu}_C$ be the uniform distribution on $\widehat{E}_C$. Define $\widehat{\nu}_M = \sum_{\{C \in \mathcal{C}_M : C \text{ is infinite}\}} \mu(C) \cdot \widehat{\nu}_C$ so that for each $C \in \mathcal{C}_M$, $\widehat{\nu}_M(\widehat{C}) = \widehat{\mu}(\widehat{C})$. Any accumulation point of $\{\widehat{\nu}_M : M \in \mathbb{N}\}$ is non-atomic, assigns mass 1 to $\widehat{D}_\infty$, and agrees with $\widehat{\mu}$ on $\widehat{\mathcal{C}}$. Since the cardinality of each set $\widehat{D}_C \setminus X$ is at least $2^{\mathfrak{c}}$, there are uncountably many such accumulation points. The rest of the proof follows that of Proposition A. □

Proof of Corollary 1.2. Enumerate $\mathcal{C}$ as $\{C_m : m \in \mathbb{N}\}$, for each $M \in \mathbb{N}$, let $\mathcal{C}_M$ denote the partition of X generated by $\{C_m : m \leq M\}$, and let $\Pi_M = \Pi(\mathcal{C}_M, \mu)$. Each Π_M is a compact convex set of probabilities, and as $\Pi_M \downarrow \Pi$, $\mathcal{R}_{\mathbb{W}}(\Pi_M) \uparrow \mathcal{R}_{\mathbb{W}}(\Pi)$. Let $f : X \to \mathbb{W}$ be an $\mathcal{X}$-measurable function, and let F be its unique continuous extension from X to $\widehat{X}$. For $C \in \mathcal{C}_M$, if $\mu(C) > 0$, then C is infinite. From Corollary A.2, $F(\Pi_M) = \sum_{\{C \in \mathcal{C}_M : \mu(C) > 0\}} \mu(C) \cdot \Delta(K_C)$ where each K_C is a non-empty closed subset of $\mathbb{W}$. Any limit of such $F(\Pi_M)$'s has the given integral representation. □

Proof of Lemma 8. It is sufficient to show that $\sup_p D(\epsilon, \mathcal{C}, d_{2,p}) = \sup_q D(\epsilon, \widehat{\mathcal{C}}, d_{2,q})$ where the first supremum is taken over finitely supported probabilities in X and the second is taken over finitely supported probabilities in $\widehat{X}$. This follows from the openness of each $\widehat{C}$ and denseness of X in $\widehat{X}$. □

Proof of Lemma 9. Let D be the set of discontinuities of φ, suppose, for the purposes of contradiction, that D^c contains on open set. Let $G \subset D^c$ be an open ball. Because $\varphi : G \to [0, 1]$ is continuous and G is connected, so is $V := \varphi(G)$. Connected subsets of $[0, 1]$ are intervals, and because φ is one-to-one, V is a non-degenerate interval. Remove from V an interior point, x, and denote the resulting set V'. $\varphi^{-1}(V')$ is the connected set $G' := G \setminus \{\varphi^{-1}(x)\}$. But this implies that $V' = \varphi(G')$ must be connected, a contradiction. For the second part, if the diameter of each $\varphi^{-1}((a, a+\epsilon))$ is finite, then we can cover a Banach space by a finite number of balls of finite diameter. □

Dept. of Economics, University of Texas, Austin, e-mail: max.stinchcombe@gmail.com