STOCHASTIC CHOICE AND REVEALED PERTURBED UTILITY

By Drew Fudenberg, Ryota Iijima, and Tomasz Strzalecki[1]

Perturbed utility functions—the sum of expected utility and a nonlinear perturbation function—provide a simple and tractable way to model various sorts of stochastic choice. We provide two easily understood conditions each of which characterizes this representation: One condition generalizes the acyclicity condition used in revealed preference theory, and the other generalizes Luce's IIA condition. We relate the discrimination or selectivity of choice rules to properties of their associated perturbations, both across different agents and across decision problems. We also show that these representations correspond to a form of ambiguity-averse preferences for an agent who is uncertain about her true utility.

Keywords: Control cost, preference for randomization, ambiguity aversion.

1. INTRODUCTION

Deterministic theories of choice cannot accommodate the fact that observed choices in many settings seem to be stochastic. This raises the question of the extent to which stochastic choice follows a consistent principle that can be given a simple theoretical foundation. Here we provide conditions under which stochastic choice corresponds to the maximization of the sum of expected utility and a perturbation function

$$(1) \qquad P(A) = \arg\max_{p \in \Delta(A)} \sum_{z \in A} u(z) p(z) - c(p(z)),$$

where $P(A)$ is the probability distribution of choices from the set A, u is the utility function of the agent, and c is a convex perturbation function that may reward the agent for randomizing; we call this an *Additive Perturbed Utility* (APU) representation. Such perturbed utility functions and their variants have been previously used by, for example, Harsanyi (1973b), Machina (1985), Rosenthal (1989), Clark (1990), Mattsson and Weibull (2002), and Swait and Marley (2013).[2] Because we want to apply the perturbed-utility representation

[2]Perturbed utility has also been used in the theory of learning in games. Fudenberg and Levine (1995) showed how this leads to "stochastic fictitious play," and generates Hannan-consistent choice, meaning that its long-run average payoff is at least as good as the best response to the time average of the moves of Nature and/or other players (Hannan (1957)). Hofbauer and Sandholm (2002), Benaim, Hofbauer, and Hopkins (2009), and Fudenberg and Takahashi (2011) used perturbed utility to construct Lyapunov functions for stochastic fictitious play, van Damme and Weibull (2002) and Iijima (2014) studied perturbed utility in evolutionary models, and Noguchi (2015) studied convergence to Nash equilibrium in repeated perturbed games.

to choice sets A of varying size, we adopt an additive form for the perturbation function, as opposed to allowing general functions on $\Delta(A)$. The additive specification has content because we require the cost function c to depend only on p. At the other extreme, additivity is vacuous if we allow c to depend on A and z as well as $p(z)$; we mention some intermediate cases in Section 4.

In contrast to past work on nonlinear perturbed utility, we take a revealed preference approach: we suppose that the analyst observes the agent's choice probabilities from some (but not necessarily all) menus, and show that various restrictions on the probabilities correspond to particular forms of the perturbation function. In particular, we relate restrictions on the perturbation function to whether the agent's choices satisfy various sorts of internal consistency conditions. We argue that the perturbation-function approach provides a simple and tractable way to model stochastic choice, and that it helps us organize the empirical evidence and evaluate how much it pushes the boundaries of "rational" behavior.

We develop two alternative conditions that characterize the APU representation. The first condition, *Acyclicity*, extends the Strong Axiom of Revealed Preference to stochastic choice. Acyclicity implies that $P(x|A) \geq P(y|A)$ if and only if $P(x|B) \geq P(y|B)$, so that the observed choice probabilities P induce an ordinal ranking of the items. It also implies that the choice probabilities induce an ordinal ranking of the menus: menu A is weaker than menu B if, for any $x \in A \cap B$, $P(x|A) \geq P(x|B)$. Acyclicity has more bite than these two implications: It also ensures that the rankings on items and on menus "agree" with each other. Our second characterization of APU, *Ordinal IIA*, requires that the observed choice probabilities can be rescaled to satisfy Luce's (1959) IIA condition. Either of these conditions implies that the observed choice data are consistent with APU. To pin down the sense in which the representation is unique, we assume that the choice data are "rich" and satisfy a continuity property.

The most commonly used cost function in the literature is the entropy function $c(q) = \eta q \log q$. This cost function generates logistic choice, and so implies that the choice probabilities satisfy Luce's IIA, which requires that the ratio of the choice probabilities of x and y is the same in any menu that contains both of them. The more general cost functions allowed in APU let the model describe a broader range of behavior, and permit tractable conditions that can be used to organize alternative classes of choice rules. In addition, the uniqueness of the representation lets us relate the discrimination or selectivity of choice rules to properties of their associated cost functions, both across different agents and across decision problems. One focus here is on how the selectivity of a rule—that is, the weight it places on more frequently chosen items—varies with the size and attractiveness of the menu.

One interpretation of representation (1) is that agents facing a decision problem randomize to maximize their non-EU preferences on lotteries, as

in Machina (1985);[3] recent experimental evidence (Agranov and Ortoleva (2015), Dwenger, Kubler, and Weizsacker (2014)) indicates that stochastic choice sometimes reflects deliberate randomization by subjects, rather than random variation in their expected utility functions. In Section 5.1, we show that such preference for randomization may arise due to uncertainty about the true utility function. Specifically, we show that the perturbed-utility objective function corresponds to a game in which the agent has a form of variational preferences (Maccheroni, Marinacci, and Rustichini (2006)) and so randomizes to guard against moves by a malevolent Nature. Another interpretation is that stochastic choice arises due to inattention or implementation costs: It may be costly to take care to implement the desired choice, so that the agent trades off the probability of errors against the cost of avoiding them, as assumed by van Damme (1991) and Mattsson and Weibull (2002).[4]

Acyclicity can be weakened to *Menu Acyclicity*, which implies that menus can be ordered by weakness. An alternative relaxation is *Item Acyclicity*, which implies that items can be ordered by their desirability.[5] In the setting of deterministic choice, these two conditions are each equivalent to Acylicity, and are also equivalent to Richter's (1966) congruence axiom, which extends the Strong Axiom of Revealed Preference to settings where data are incomplete in the sense that only some menus are observed.

The most familiar stochastic choice model in economics is *random utility* (RU) (Thurstone (1927), Marschak (1959), Harsanyi (1973a), McFadden (1973)), which supposes that the agent's choice maximizes a utility function that is subject to random shocks. We note that, in contrast to existing characterizations of RU, which impose conditions on how adding items to a menu changes the *difference* between choice probabilities (Falmagne (1978)), or the *ratio* of choice probabilities (Luce (1959)), we characterize perturbed utility with axioms that rely only on pairwise ordinal comparisons of the choice probabilities.[6] APU rules out some RU models, even some with i.i.d. shocks, but also allows for choice rules that do not admit a RU representation, so the two

---

[3]Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella (2015) also studied randomization generated by nonlinear preferences over lotteries; they used a subclass of the non-EU preferences studied by Cerreia-Vioglio, Dillenberger, and Ortoleva (2015).

[4]See Weibull, Mattsson, and Voorneveld (2007) for an alternative approach in which the agent pays costs to improve signal precision. None of these three papers derives the functional forms from observed behavior.

[5]As we showed in an earlier version of the paper (Fudenberg, Iijima, and Strzalecki (2014)), these two conditions characterize more general versions of APU in which costs are item- or menu-dependent. These models can accommodate several empirical patterns, such as violations of regularity, that APU cannot.

[6]Hofbauer and Sandholm (2002) showed that with known utility functions and a fixed menu of alternatives, any RU that satisfies a smoothness condition has a convex perturbation representation. In our setting, the analyst does not know the utility function, and in addition, we consider choices from menus of varying size.

classes of stochastic choice rules are not nested, though their intersection is nonempty, as it includes logistic choice.

RU implies that the agent is never made worse off when items are added to a choice set, which seems counterintuitive in some situations. One advantage of the perturbed-utility approach that we take here is that it can accommodate both cases where the agent prefers larger menus and those where she does not. Of course, purely static choice data (which are what we consider here) are not enough to reveal whether the agent prefers larger or smaller menus. Fudenberg and Strzalecki (2015) used cost functions to address this in the special case in which choice satisfies Luce's IIA axiom so that choice is logistic; the results in this paper may help extend the analysis of dynamic stochastic choice to more general choice rules.

Various recent papers consider extensions of RU. Manzini and Mariotti (2014) and Aguiar (2015) studied agents who only pay attention to a random subset of each menu; their main model is a special case of RU. Echenique, Saito, and Tserenjigmid (2014) considered an agent with a deterministic priority order who uses logistic choice on the perceived items. Gul, Natenzon, and Pesendorfer (2014) introduced "attribute rules" which are related to nested logit.

Although we do use a richness condition to pin down the sense in which the APU representation is unique, our characterization results (Theorems 1 and 3) do not require it, and apply when choice is observed for a subset of the possible menus, as in the work of, for example, Afriat (1967) and Richter (1966) on revealed preference or the work of Gilboa (1990), Gilboa and Monderer (1992), and Fishburn (1992) on RU when only binary menus are observed.[7] As observed by de Clippel and Rozen (2014), in some models of choice it is possible for limited data to be consistent with the characterizing axioms even when any specification of choices outside of $\mathcal{A}$ would lead to a violation of those axioms. Our results imply that this problem does not arise here.

2. ADDITIVE PERTURBED UTILITY

Let Z be a set of items (consequences or prizes). To begin, we will assume that Z is finite; this restriction is relaxed in Section 2.3. A *menu* is a nonempty subset of Z; we assume menus are finite whether Z is or not. Let $\mathcal{A}$ be the set of menus for which the choice probabilities of the agent have been observed; without loss of generality, we assume that every $z \in Z$ appears in at least one menu. We allow for the available choice data to be limited, that is, the collection $\mathcal{A}$ need not include every nonempty subset of Z. We consider a *stochastic choice rule P* that maps each menu $A \in \mathcal{A}$ to a probability distribution on its

---

[7]See also Reny (2015), who extended Afriat's result to infinite data sets, and Kubler, Selden, and Wei (2014) and Echenique and Saito (2015), who studied revealed preference in the demand for financial assets.

elements. Formally, a stochastic choice rule is a mapping P that assigns a measure $P(A) \in \Delta(A)$ to each menu $A \in \mathcal{A}$. We write $P(z|A)$ to denote the probability that item z is chosen from the menu A. To relate the observed choice probabilities to the form of the cost function, we will impose various sorts of consistency conditions on P.

To facilitate the exposition, we first consider the case where all probabilities are positive. We relax this assumption in Section 4, where we extend our model to include deterministic choice, and show how our conditions generalize the strong axiom of revealed preference.

DEFINITION 1: P satisfies *Positivity* if $P(z|A)$ is strictly positive for each $A \in \mathcal{A}$ and $z \in A$.

As noted by McFadden (1973), a zero probability is empirically indistinguishable from a positive but small probability. In dynamic settings, Positivity can also be motivated by the fact that no deterministic rule can be Hannan (or "universally") consistent (Hannan (1957), Blackwell (1956)).

We say that a function c is a *cost function* if $c : [0, 1] \to \mathbb{R} \cup \{\infty\}$ is strictly convex and C^1 over $(0, 1)$, and $\lim_{q \to 0} c'(q) = -\infty$.

DEFINITION 2—APU: An *APU representation* has the form

$$P(A) = \underset{p \in \Delta(A)}{\arg\max} \sum_{z \in A} \big[u(z)p(z) - c\big(p(z)\big) \big]$$

for some utility function $u : Z \to \mathbb{R}$ and cost function c.

As we show in Section 5.1, perturbed utilities of this sort can arise from the agent's ambiguity about the true utility of the various choices.

Perhaps the most familiar stochastic choice rule is logit/logistic choice, also known as the Luce rule, which is given by $P(z|A) = \exp(\eta u(z)) / \sum_{z' \in A} \exp(\eta u(z'))$; as is well known, this choice rule is generated by additive perturbed utility with cost $c(p) = \eta^{-1} p \log p$.[8] More generally, Theorem 2 shows that logit choice is characterized by $c(p) = \alpha p \ln p + \gamma p + \delta$. As is also well known, this form of APU is observationally equivalent to a random utility representation where the additive shocks are i.i.d. Gumbel with variance η;[9] we discuss the relationship between APU and random utility in Section 5. Other classes of cost functions in the literature include the logarithmic form used by Harsanyi (1973b) $c(p) = -\eta \log(p)$ and the quadratic perturbation $c(p) = \eta p^2$ implicitly assumed by Ben-Akiva and Lerman (1985) and Rosenthal (1989).[10]

---

[8]See Section 3.6 in Anderson, de Palma, and Thisse (1992).

[9]This was first shown by E. Holman and A. A. J. Marley (Luce and Suppes (1965)).

[10]See Voorneveld (2006) for a more detailed analysis on this concept in games.

As far as we know, logistic choice is the only sort of APU whose revealed choice implications have already been characterized: The Luce rule is equivalent to the following "IIA" condition if choice probabilities are strictly positive and a sufficiently rich set of menus is observed.[11]

DEFINITION 3: P satisfies *IIA* if, for all $A, B \in \mathcal{A}$ with $x, y \in A \cap B$,

$$\frac{P(x|A)}{P(y|A)} = \frac{P(x|B)}{P(y|B)}.$$

2.1. *Characterization of the Model*

In this section, we discuss two conditions, each of which characterizes APU. The first condition, called Acyclicity, is an extension of SARP to the stochastic setting, as we will see in Section 4. The second condition generalizes IIA to allow the ratio of choice probabilities to vary with the menu in a particular way.

We say that a finite sequence of quadruples $\{(x_k, A_k), (y_k, B_k)\}_{k=1}^n$ is admissible if (i) $x_k \in A_k$ and $y_k \in B_k$ for all k, (ii) $(y_k)_{k=1}^n$ is a permutation of the $(x_k)_{k=1}^n$, and (iii) $(B_k)_{k=1}^n$ is a permutation of the $(A_k)_{k=1}^n$.[12]

In reading the next definition, remember that $P(x|A)$ is only defined for $x \in A$.

DEFINITION 4: P satisfies *Acyclicity* if there is no admissible sequence such that

$$P(x_1|A_1) > P(y_1|B_1), \quad P(x_k|A_k) \geq P(y_k|A_k), \quad \text{and}$$
$$P(x_n|A_n) \geq P(y_n|B_n).$$

Although Acyclicity rules out cycles of any length, the condition can be checked on any given Z and choice data P in a finite number of steps.

To understand Acyclicity, consider a few of its implications. First, note that Acyclicity implies that, for all $x, y \in A \cap B$, we have $P(x|A) \geq P(y|A)$ if and only if $P(x|B) \geq P(y|B)$. Thus, P induces an ordinal ranking of all items in Z; and moreover, that ranking is preserved in every menu (though the particular numerical values of the likelihoods may change and their ratios do not have to be preserved). As will become apparent shortly, this ranking is represented by the utility function u; it implies that the agent choice probabilities do not reverse due to "menu effects."

---

[11]Luce (1959) assumed choice data at every menu are observable, but it is enough to have all menus with size 2 or 3, or have $Z \in \mathcal{A}$.

[12]That is, $y_k = x_{f(k)}$ for some permutation $f : \{1, \ldots, n\} \to \{1, \ldots, n\}$ and $B_k = A_{g(k)}$ for some permutation $g : \{1, \ldots, n\} \to \{1, \ldots, n\}$.

Second, note that Acyclicity implies that, for all $x, y \in A \cap B$, we have $P(x|A) \geq P(x|B)$ if and only if $P(y|A) \geq P(y|B)$. Thus, P induces an ordinal ranking of all menus in $\mathcal{A}$. One interpretation of $P(x|A) \geq P(x|B)$ is that menu A is *weaker* than menu B in the sense that its items compete less with x than items in B. As we will show, this ranking is represented by the Lagrange multiplier in the maximization problem of the agent.

Acyclicity has more bite than the two implications noted above. Intuitively, it ensures that the rankings on items and on menus "agree" with each other. For example, take any menus $A, B \in \mathcal{A}$ and items $x, y \in A \setminus B$. Then Acyclicity implies that $P(x|A) \geq P(y|A)$ iff $B \cup \{y\}$ is weaker than $B \cup \{x\}$.[13] Thus, Acyclicity implies the Order Independence condition of Tversky (1972), which leads to a more general but more complicated and less tractable "simply scalable" representation. Acyclicity is related to the cancellation condition used in work on multiattribute decision theory (Scott (1964), Tversky (1964)), but differs in a few key ways. Most notably, the choice domains in that literature have a product structure, while in our case $P(x|A)$ is only defined if $x \in A$, and our choice data need to fit the restriction $\sum_{z \in A} P(z|A) = 1$.

Our second characterization of APU generalizes the IIA condition that is known to characterize the entropy model.

DEFINITION 5: P satisfies *Ordinal IIA* if for some continuous and monotone $f : [0, 1] \to \mathbb{R}_+ \cup \{\infty\}$ with $f(0) = 0$ such that

$$\frac{f\big(P(x|A)\big)}{f\big(P(y|A)\big)} = \frac{f\big(P(x|B)\big)}{f\big(P(y|B)\big)}$$

for each menu $A, B \in A$ and $x, y \in A \cap B$.

Ordinal IIA requires that probabilities can be rescaled so that the rescaled choice probability ratios are the same in every menu. Ordinal IIA reduces to IIA under $f(q) = q$, which implies that the cost function is $\eta q \log(q)$ for some $\eta > 0$, and thus that cost is proportional to the negative of the entropy function. If instead $f(q) = \exp(-\frac{1}{q})$, the cost is proportional to $-\log q$, as in Harsanyi (1973b). When Z is finite, it is without loss of generality to suppose that f takes the form of a "probability distortion" function, that is, that $f(1) = 1$ so that f maps $[0, 1]$ to $[0, 1]$.[14]

---

[13]To see "if," note that $P(z|B \cup \{y\}) \geq P(z|B \cup \{x\})$ for all $z \in B$, which implies $P(x|B \cup \{x\}) \geq P(y|B \cup \{y\})$. Thus $P(y|A) > P(x|A)$ violates Acyclicity. To prove the "only if," suppose to the contrary that $P(z|B \cup \{x\}) > P(z|B \cup \{y\})$ for some $z \in B$. By Acyclicity, $P(z'|B \cup \{x\}) > P(z'|B \cup \{y\})$ for all $z' \in B$, which implies $P(y|B \cup \{y\}) > P(x|B \cup \{x\})$. These and $P(x|A) \geq P(y|A)$ lead to a cycle.

[14]This follows from the fact that the data only pin down the value of f at a finite number of points in $(0, 1)$.

THEOREM 1: *Suppose that Positivity holds and $\mathcal{A}$ contains all menus with size 2 and 3. Then the following conditions are equivalent*:
1. *P satisfies Acyclicity.*
2. *P satisfies Ordinal IIA.*
3. *P is represented by APU.*

The next section provides some intuition and a proof sketch of both equivalences. The assumption that $\mathcal{A}$ includes all menus of sizes 2 and 3 is not needed for the equivalence of acyclicity and APU; we include it here for expositional simplicity.

2.2. *Proof Sketch*

To study the restrictions that APU places on observed choice probabilities, we first analyze the agent's maximization problem.

DEFINITION 6: A utility function u, a cost function c, and a function $\lambda : \mathcal{A} \to \mathbb{R}$ satisfy the *first order conditions* (FOC) for P iff

$$(2) \qquad u(x) + \lambda(A) = c'\big(P(x|A)\big).$$

Here $\lambda(A)$ is the Lagrange multiplier on the constraint that the choice probabilities from menu A sum up to 1. Since c' is monotone, FOC holds if and only if P has a separable representation in the following sense:

DEFINITION 7: $u : Z \to \mathbb{R}$ and $\lambda : \mathcal{A} \to \mathbb{R}$ are a *separable representation* of P if and only if

$$(3) \qquad u(x) + \lambda(A) > u(y) + \lambda(B) \quad \text{iff} \quad P(x|A) > P(y|B).$$

LEMMA 1: *Assume Positivity holds. Then the following conditions are equivalent*:
  (a) *There exists (u, c) such that P has an APU representation with (u, c).*
  (b) *There exists (u, c, λ) such that P satisfies the FOC with (u, c, λ).*
  (c) *There exists (u, λ) such that P has a separable representation with (u, λ).*

We omit the proof of this lemma, which follows the same lines as that of Lemma 2 in the Appendix, but the intuition for the result is easy to explain: The equivalence of APU and the FOC follows from the Kuhn–Tucker theorem. That the FOC implies a separable representation is straightforward from the strict monotonicity of c'. To show the converse, we use a variant of the usual ordinal uniqueness argument to show that if P has a separable representation (u, λ), then there is a strictly increasing and continuous function $g : (0, 1) \to \mathbb{R}$ that satisfies $g(P(x|A)) = u(x) + \lambda(A)$, and $\lim_{q \to 0} g(q) = -\infty$. We then

define $c(p) := \int_{1/2}^{p} g(q)\,dq$, and it is immediate that (u, c) satisfies the first order conditions.

With the lemma in hand, we prove the equivalence of Acyclicity and APU in Theorem 1 by demonstrating that Acyclicity is equivalent to the separability property (property (c)). It is easy to show that Acyclicity is necessary: if there were both a separable representation (u, λ) and cycle, then along the cycle we would have $u(x_i) + \lambda(A_i) \geq u(y_i) + \lambda(B_i)$ for all i with at least one strict. Summing over i yields a contradiction because of the permutation property. The proof that Acyclicity is sufficient for (c) formulates the existence of a cycle as a system of linear inequalities, and obtains the desired conclusion from a version of Farkas's lemma.[15] This linear programming argument also shows that Acyclicity can be checked in a finite number of steps.[16]

To prove the equivalence of Ordinal IIA and APU in Theorem 1, we show that Ordinal IIA is equivalent to the FOC (property (b)). To see why this is true, note that the FOC implies that, for each $x, y \in A \cap B$,

$$c'\big(P(x|A)\big) - c'\big(P(y|A)\big) = c'\big(P(x|B)\big) - c'\big(P(y|B)\big).$$

Define the strictly increasing function $f : (0, 1) \to \mathbb{R}_+$ by $f(q) := \exp(c'(q))$ to obtain Ordinal IIA. For the converse, we define $c'(q) := \log(f(q))$ and show that $c(q) = \int_{1/2}^{q} c'(t)\,dt$ is indeed a cost function. We then define $u(z) := c'(P(z|\{x, z\})) - c'(P(x|\{x, z\}))$, where $x \in Z$ is an arbitrary fixed element with $u(x) := 0$. Proving that (b) holds is now a matter of substituting these definitions into the Ordinal IIA condition.

2.3. *Uniqueness*

For an arbitrary set of items and menus, the APU representation may not be unique, but uniqueness obtains when the range of observed behavior is rich enough.[17] Intuitively, under APU, the incentive of an agent depends only on the payoff differences $u(x) - u(y)$ between items in the menu; to identify the cost function, we need to be able to vary this utility difference freely.[18] To allow for arbitrary utility differences, we consider an infinite set Z, but we maintain

[15]Echenique and Saito (2015) used a related result in their analysis of deterministic portfolio choice. A key technical difference is that their analog of our acyclicity condition imposes constraints on the products of various prices, while our conditions are purely ordinal.

[16]See, for example, Kraft, Pratt, and Seidenberg (1959), who also noted that when a solution to a linear system with rational coefficients exists in the reals, there is also a solution in the rational number.

[17]This is also the case for other models of stochastic choice, such as random utility; see, for example, Fishburn (1998). Stronger uniqueness results can be obtained when items are lotteries; see, for example, Gul and Pesendorfer (2006).

[18]A similar situation arises for variational preferences of Maccheroni, Marinacci, and Rustichini (2006); to obtain uniqueness, they imposed an additional axiom that guarantees that the range of u is rich enough.

the assumption that each menu is a finite set; formally, let $\mathcal{A}$ be the collection of all finite, nonempty subsets of Z. We observe a stochastic choice function P defined on $\mathcal{A}$. When Z is infinite, APU is still characterized by Positivity and Ordinal IIA. Positivity and Acyclicity characterize APU under additional technical assumptions of Richness and Continuity.

DEFINITION 8: P satisfies *Richness* if, for any $x \in Z$ and any $p, q \in (0, 1)$ s.t. $p + q \leq 1$, there exist $y, z \in Z$ (not necessarily distinct) such that $P(x|\{x, y, z\}) = p$ and $P(y|\{x, y, z\}) = q$.

Richness implies that the range of the utility function u equals $\mathbb{R}$ and that there are at least three items with each utility level.[19]

DEFINITION 9: P satisfies *Continuity* if Z is a separable and connected metric space and, for any menu $\{x^1, \ldots, x^m\}$ and sequences of items with $\lim_{n \to \infty} x_n^i = x^i$ for each $i = 1, 2, \ldots, m$,

$$\lim_{n \to \infty} P(x_n^i | \{x_n^1, \ldots, x_n^m\}) = P(x^i | \{x^1, \ldots, x^m\}), \quad \forall i = 1, \ldots, m.$$

Continuity implies that the utility function u is continuous.

THEOREM 2: *Suppose that Positivity and Richness hold*.
  1. *P satisfies Ordinal IIA if and only if there is an APU representation of P.*
  2. *P satisfies Acyclicity and Continuity if and only if there is an APU representation of P with continuous u.*
  Moreover, if (u, c) and $(\hat{u}, \hat{c})$ represent the same P, then there exist constants $\alpha > 0, \beta, \gamma, \delta \in \mathbb{R}$ such that $\hat{u} = \alpha u + \beta$ and $\hat{c}(p) = \alpha c(p) + \gamma p + \delta$ for all $p \in (0, 1)$.

The utility function u is unique up to positive affine transformations; its scale matters because higher utility differences between items lead to higher differences in their choice probabilities, while items with about the same utility are chosen with about the same probability. Note that u and c are expressed in the same units; that is why multiplying u by a constant α requires multiplying c by the same α. Since the absolute level of the cost function does not matter, we are free to shift it by constants β, δ without changing behavior. Finally, since

---

[19]These should be thought of as different items, and not duplicates of the same item. Although duplicates and similarities between items are important in many contexts, we do not think they are important in all of them. Like the assumption of an infinite data set, the existence of items with exactly the same utility is a purely technical condition that is needed to pin down a unique representation; neither assumption can be satisfied in any real-world application of the model.

on each menu the probabilities sum to 1, the term γp becomes a constant and similarly does not affect choice.[20]

To prove this result, we exploit the fact that under Positivity, Additive Perturbed Utility can be seen as an extension of the "Fechnerian" model of stochastic choice from binary to general menus. This model, which is also called the "constant utility model," was studied and axiomatized by Davidson and Marschak (1959), Debreu (1958), Block and Marschak (1960), and Scott (1964), among others.

DEFINITION 10: A stochastic choice rule P on binary menus has a *Fechnerian representation* if there exist a utility function $u : Z \to \mathbb{R}$ and a strictly increasing transformation function g such that $P(a|\{a, b\}) = g(u(a) - u(b))$.

PROPOSITION 1: *Suppose that P is defined on binary menus and satisfies Positivity, and either Z is finite or Richness and Continuity hold. Then P satisfies Acyclicity iff P has a Fechnerian representation.*[21]

3. MONOTONE DISCRIMINATION

With the uniqueness result of Theorem 2 in hand, we can relate the discrimination or selectivity of choice rules to properties of their associated cost functions. Just as with measures of risk aversion, we can do this both by comparing the behavior of different agents and by examining how the behavior of a given agent varies across decision problems.

3.1. *Comparing Choice Rules*

In this section, we consider a pair of choice rules P_1 and P_2 that are represented by APUs with common utility function u, so that differences in choice probabilities come only from differences in the cost function, and define rankings of their selectiveness in the sense of choosing better options more often. (The reason for fixing u is that the agent also becomes more selective when the utility function is multiplied by any $\lambda > 1$ holding the cost function fixed.) These rankings generalize the effect of multiplying the cost function by a positive number.[22] We begin by comparing P_1 and P_2 at binary menus. We say

---

[20]The uniqueness result shows that APU is not nested in the "generalized entropy" representation of Fosgerau and de Palma (2015) because many cost functions violate their homogeneity condition (1). Conversely, their representation is not nested in APU, as they showed.

[21]When P is defined on binary menus only and Z is infinite, Acyclicity may not be strong enough to imply an APU representation, though the converse is always true and both directions are true on any finite Z.

[22]For example, in the literature on learning and evolutionary games, an analyst varies perturbation levels with a fixed utility function to analyze long-run payoffs and the dynamic stability of equilibria (Fudenberg and Levine (1995), Benaim and Hirsch (1999), Kreindler and Young (2013)), and in the empirical application of quantal response equilibrium (McKelvey and Palfrey (1995)) the logit functional form is held fixed and the scale of the perturbation is estimated.

that agent 1 is more pairwise-selective than agent 2 if she has the same ordinal ranking over items, but her choices are less random.

DEFINITION 11: P_1 is *more pairwise-selective than* P_2 if $P_1(x|\{x, y\}) \geq P_2(x|\{x, y\})$ whenever $P_2(x|\{x, y\}) \geq P_2(y|\{x, y\})$.

PROPOSITION 2: *Assume that Richness holds for both* P_1 *and* P_2. *For each* $i = 1, 2$, *let* P_i *be represented by APU* (u, c_i); *then* P_1 *is more pairwise-selective than* P_2 *if and only if* $c'_1(q) - c'_1(1 - q) \leq c'_2(q) - c'_2(1 - q)$ *for all* $q \in (1/2, 1)$.

Note that when costs are twice differentiable, the condition in Proposition 2 is equivalent to $\int_{1-q}^{q} c''_1(p) \, dp \leq \int_{1-q}^{q} c''_2(p) \, dp$ for all $q \in (1/2, 1)$, which is implied by $c''_1(q) \leq c''_2(q)$ for all $q \in (0, 1)$. For the logit case, $c_i(q) = \eta_i q \log q$, this is equivalent to $\eta_1 \leq \eta_2$. To characterize all of the implications of $c''_1 \leq c''_2$, we need to look at nonbinary menus, as in the following.

DEFINITION 12: P_1 *is more selective than* P_2 *if, for any* $x, x', y, z_i, z'_i, i = 1, 2$, $P_1(x'|\{x', y, z'_1\}) \geq P_2(x'|\{x', y, z'_2\})$ *holds whenever*

$$P_i\big(y|\{x, y, z_i\}\big) = P_i\big(y|\{x', y, z'_i\}\big),$$

$$P_1\big(x|\{x, y, z_1\}\big) = P_2\big(x|\{x, y, z_2\}\big),$$

$$P_i\big(x'|\{x, x'\}\big) > \frac{1}{2}$$

for each $i = 1, 2$.

To understand this condition, here we focus on a situation where (i) $\{x, y, z_i\}$ and $\{x', y, z'_i\}$ have the same menu strength under P_i, $i = 1, 2$, (ii) x' is better than x. If P_1 is more selective than P_2, then $P_1(x'|\{x', y, z'_1\}) - P_1(x|\{x, y, z_1\}) \geq P_2(x'|\{x', y, z'_2\}) - P_2(x|\{x, y, z_2\})$.

PROPOSITION 3: *Assume Continuity and Richness. For each* $i = 1, 2$, *let* P_i *be represented by APU* (u, c_i), *where* c_i *is* C^2. *Then* P_1 *is more selective than* P_2 *if and only if* $c''_1(q) \leq c''_2(q)$ *for all* $q \in (0, 1)$.

3.2. *Decreasing Selectivity*

We now use APU to model the idea that the agent is less attentive to small probabilities and in this sense displays "decreasing selectivity." Recall that IIA (corresponding to logit choice and the entropy cost function) implies that the choice ratio of x and y in the pairwise choice problem $\{x, y\}$ is the same as it is in the grand set Z. If the agent has limited cognitive resources to implement her choices, we might expect that the agent will be more careful in setting the

probabilities of x and y when they are both more likely to be chosen, so that there will be more discrimination in weaker menus.[23] In other words, we expect the following condition:

DEFINITION 13: *P* satisfies *Decreasing Selectivity* if it satisfies Positivity, and for all $x, y \in A \cap B$, if $P(x|A) \geq P(x|B)$ and $P(x|A) > P(y|A)$, then

$$\frac{P(x|A)}{P(y|A)} \geq \frac{P(x|B)}{P(y|B)}.$$

Decreasing Selectivity says that choices from stronger menus are more uniform than choices from weaker menus. As we show in Section 4, $A \subseteq B$ implies $P(x|A) \geq P(x|B)$ under APU, so that the axiom suggests that choice probabilities become flatter as we expand a menu. Note that given the FOCs of an APU, we can express the choice probability ratios as

$$(4) \qquad \frac{P(x|A)}{P(y|A)} = \frac{c'^{-1}\big(u(x) + \lambda(A)\big)}{c'^{-1}\big(u(y) + \lambda(A)\big)}.$$

If $\log c'^{-1}$ is convex, then the right-hand side of (4) is increasing in λ. As the λ of the weaker menu A is higher than that of B, Decreasing Selectivity, leads to the following result.

PROPOSITION 4: *Suppose that P is an APU with a utility function u and cost function c. Let $h = \log(c'^{-1})$. If h is convex, then P satisfies Decreasing Selectivity.*

In many settings, there are many duplicates or near-duplicate choices, so that menu sizes do not have a natural bound. This motivates the study of choice in large menus. In such menus, many items will be chosen with small probability, which leads us to study the ratio of choice probabilities of two items that are each rarely chosen. Specifically, consider a collection of menus A_n such that $x, y \in A_n$ for each n and let $p_n := P(x|A_n) + P(y|A_n)$. Proposition 4 implies that, for a convex h, the ratio $P(x|A_n)/P(y|A_n)$ is monotone in p_n; that is, the worse the items x and y are compared to the remainder of A_n, the flatter their choice ratio. We now investigate what happens in the limit.

DEFINITION 14: *P* satisfies *Asymptotic Non-Selectivity* if it satisfies Positivity, and for any sequence A_n such that $x, y \in A_n$ if $P(x|A_n) \to 0, P(y|A_n) \to 0$, then $P(x|A_n)/P(y|A_n) \to 1$.

---

[23]In a deterministic choice setting, Frick (2013) extended Luce's (1956) model of utility discrimination to capture the idea that items of similar utility are harder to distinguish in larger menus.

From formula (4), asymptotic nonselectivity can be expressed as

$$\frac{c'^{-1}\big(u(x) + \lambda(A_n)\big)}{c'^{-1}\big(u(y) + \lambda(A_n)\big)} \to 1.$$

For this to hold, the function h must flatten out asymptotically as its argument $u(x) - \lambda(A_n)$ becomes extremely low. This is formalized by the next proposition.

PROPOSITION 5: *Suppose that P is an APU with a utility function u and cost function c. Let $h = \log(c'^{-1})$. If, for all t, the function h satisfies $\lim_{s\to\infty}[h(t-s) - h(-s)] = 0$, then P satisfies Asymptotic Non-Selectivity. The converse is true under Richness and Continuity.*

EXAMPLE 1: A particular class of cost functions leading to limited discrimination and asymptotic nonselectivity is the logarithmic form $c(q) = -\eta \log(q)$. The function h is $h(w) = \log(-\frac{\eta}{w})$, defined on $(-\infty, -\eta)$, which is strictly convex. This also satisfies the condition for Proposition 5, because $h(t - s) - h(-s) = \log(\frac{s}{s-t}) \to 0$ as $s \to \infty$. As an illustration, consider menus of the form $A_n = \{x, y_1, \ldots, y_n\}$, where $u(x) = 1$ and $u(y_i) = 0$ for each i. Choice probabilities under $\eta = 1$ are $P(x|A_n) = \frac{1}{2}(-n + \sqrt{4 + n^2})$, $P(y_i|A_n) = n^{-1}[1 - P(x|A_n)]$. The choice probability ratio $\frac{P(x|A_n)}{P(y_1|A_n)}$ is decreasing in n and approaches to 1 as $n \to \infty$.

4. BEYOND POSITIVITY

So far we have assumed Positivity to simplify the exposition. This rules out situations in which some items are clearly inferior to others so that they are never chosen. However, our characterization holds more generally, which makes it possible to accommodate models that allow both zero and nonzero choice probabilities, such as the tractable linear choice probabilities used in Ben-Akiva and Lerman (1985) and Rosenthal (1989). Relaxing positivity also lets us cover deterministic choice as a special case and thus demonstrate how our acyclicity condition is a natural extension of past work.

To represent choice without imposing positivity, we use weak APU, which takes the same form as APU except that the steepness condition $\lim_{p\to 0} c'(p) = -\infty$ is not required.[24] Formally, a function c is a *weak cost function* if $c : [0, 1] \to \mathbb{R} \cup \{\infty\}$ is strictly convex and C^1 over $(0, 1)$.

[24]Iijima (2014) and Mertikopoulos and Sandholm (2015) clarified the importance of using nonsteep cost functions in the study of learning and evolution in games.

DEFINITION 15—Weak APU: A *weak APU representation* has the form

$$P(A) = \arg\max_{p \in \Delta(A)} \sum_{z \in A} \left[u(z) p(z) - c\big(p(z)\big) \right],$$

for some utility function $u : Z \to R$ and weak cost function c.

We first note that Acyclicity no longer characterizes weak APU without Positivity. Let $u(x) > u(y) > u(z)$, and suppose that c is sufficiently small so that the agent always picks the best item with probability 1, that is, $P(x|\{x, y, z\}) = P(x|\{x, y\}) = P(x|\{x, z\}) = P(y|\{y, z\}) = 1$. This violates Acyclicity because $P(y|\{y, z\}) > P(z|\{y, z\})$ and $P(z|\{x, y, z\}) \geq P(y|\{x, y, z\})$. We modify the Acyclicity condition by replacing $\geq$ with the relation $\geq^*$ on $[0, 1] \times [0, 1]$ defined by $p \geq^* q$ iff $p > q$ or $p = q \in (0, 1)$. The following condition is weaker than Acyclicity but equivalent to it when Positivity holds.

DEFINITION 16: P satisfies Weak *Acyclicity* if there is no admissible sequence such that

$$P(x_1|A_1) > P(y_1|B_1), \quad P(x_k|A_k) \geq^* P(y_k|A_k), \quad \text{and}$$
$$P(x_n|A_n) \geq^* P(y_n|B_n).$$

We have the following generalization of our main theorem.

THEOREM 3: *Suppose Z is finite. Then P is represented by weak APU if and only if Weak Acyclicity is satisfied.*

The proof idea is essentially the same as in Section 2.2, except that we need additional care in dealing with choice probabilities that are 0 or 1. FOC now takes the Kuhn–Tucker form

$$(5) \qquad u(x) - c'\big(P(x|A)\big) + \lambda(A) \begin{cases} \geq 0 & \text{if } P(x|A) = 1, \\ = 0 & \text{if } P(x|A) \in (0, 1), \\ \leq 0 & \text{if } P(x|A) = 0. \end{cases}$$

We also modify the definition of a separable representation by

$$(6) \qquad u(x) + \lambda(A) > u(y) + \lambda(B) \quad \text{if} \quad P(x|A) > P(y|B),$$
$$u(x) + \lambda(A) = u(y) + \lambda(B) \quad \text{if} \quad 1 > P(x|A) = P(y, B) > 0.$$

Then we can show that weak APU, the existence of (u, c, λ) that satisfy (5), and the existence of a separable representation (6) are all equivalent. And we again formulate the existence of a cycle as a system of linear equalities and inequalities, and use a version of Farkas's lemma to show that Weak Acyclicity is equivalent to the existence of (u, c, λ) that satisfy the FOC.

Like Acyclicity, Weak Acyclicity implies an order on items.

DEFINITION 17: P satisfies *Item Acyclicity* if

$$P(x_1|A_1) > P(x_2|A_1), \quad P(x_k|A_k) \geq^* P(x_{k+1}|A_k) \quad \text{for} \quad 1 < k < n$$
$$\text{implies} \quad P(x_n|A_n) \not\geq^* P(x_1|A_n).$$

Item Acyclicity is equivalent to the existence of an ordinal ranking of items. It can be seen as an extension of Richter's (1966) congruence axiom, which is itself a generalization of Houthakker's (1950) Strong Axiom of Revealed Preference, and requires that if there is a cycle $x_1, \ldots, x_n$ where each x_i is chosen from a menu that contains x_{i+1}, then if x_1 and x_n are both in a menu and x_1 is chosen, then x_n is chosen as well.[25] Also, as we showed in an earlier version of this paper (Fudenberg, Iijima, and Strzalecki (2014)), Item Acyclicity characterizes the more general item-invariant representation, where the cost c can be different on each menu.

Weak Acyclicity also implies an order on menus.

DEFINITION 18: P satisfies *Menu Acyclicity* if

$$P(x_1|A_1) > P(x_1|A_2), \quad P(x_k|A_k) \geq^* P(x_k|A_{k+1}) \quad \text{for} \quad 1 < k < n$$
$$\text{implies} \quad P(x_n|A_n) \not\geq^* P(x_n|A_1).$$

DEFINITION 19: P satisfies *Regularity* if $P(x|B) \leq P(x|A)$ for all $A, B \in \mathcal{A}$ and $x \in A \subseteq B$.

It is easy to see that Menu Acyclicity implies regularity, so a fortiori, APU are regular.[26] Moreover, Menu Acyclicity characterizes a more general form of weak APU where cost c can depend on items.[27] As with Item Acyclicity, Menu Acyclicity is also equivalent to the existence of a strict utility function when choice is deterministic (see Proposition 6), so in this case it is also equivalent to

---

[25]Unlike SARP, congruence is defined for general menus and not just budget sets. Richter (1966) studied deterministic choice, and took as primitive a choice correspondence that specifies a nonempty set of chosen options $\mathcal{C}(A) \subseteq A$ for each menu A in some collection. The congruence axiom says that if $x \in \mathcal{C}(A)$, $y \in A$, $x_j \in \mathcal{C}(A_j)$, and $x_{j+1} \in A_j$ hold for $j = 1, 2, \ldots, n-1$ at some menus $A, A_1, A_2, \ldots, A_n$ and items $y = x_1, \ldots, x_n = x$, then $y \in \mathcal{C}(A)$. The derived representation sets the utilities of x_1 and x_n to be equal, which in our setting corresponds to the case where the choice probabilities of x_1 and x_n are equal.

[26]To see this, take any $A \subseteq B$ and suppose that $P(x|A) < P(x|B)$ for some x. Then $P(y|A) < P(y|B)$ holds for any $y \in A$ such that $P(y|A) > 0$; otherwise $P(y|A) \geq^* P(y|B)$, which violates Menu Acyclicity. Thus $1 = \sum_{y \in A} P(x|A) < \sum_{y \in B} P(y|B) = 1$, a contradiction.

[27]See Fudenberg, Iijima, and Strzalecki (2014). Clark (1990)'s Theorem 1 gives an incorrect characterization: The choice data $\mathcal{A} = \{\{x, y\}, \{y, z\}, \{x, z\}\}$, $P(x|\{x, y\}) = P(y|\{y, z\}) = P(z|\{x, z\}) = 1$ satisfy the theorem's assumptions but do not have the asserted representation. Clark's characterization is correct under the additional assumption of Positivity, as then its conditions are equivalent to Menu Acyclicity.

congruence. Perhaps for this reason, the notion of a revealed weakness ranking of menus has not been used in the literature on deterministic choice, but it is a natural counterpart to the revealed attractiveness of items, and is potentially useful in other models of stochastic choice.[28]

Finally, we note that though Item Acyclicity and Menu Acyclicity are both necessary consequences of Weak Acyclicity, they are not sufficient.

EXAMPLE 2: There are three items $Z = \{x, y, z\}$, menus $A = \{y, z\}$, $B = \{x, z\}$, $C = \{x, y\}$, with the choice probabilities $P(x|Z) = 0.475$, $P(y|Z) = 0.425$, $P(y|A) = 0.525$, $P(x|B) = 0.575$, $P(x|C) = 0.525$. Notice that the menu ranking is acyclic (A is weaker than B is weaker than C is weaker than Z) and the item ranking is acyclic (x is better than both y and z, y is better than z). However, Weak Acyclicity fails because $P(x|B) > P(y|A)$, $P(y|Z) \geq^* P(z|B)$, $P(z|A) \geq^* P(x|Z)$.

However, when choice is deterministic (and single valued), all three conditions are equivalent.

DEFINITION 20: P is *deterministic* if, for all $A \in A$, there exists $x \in A$ such that $P(x|A) = 1$.

PROPOSITION 6: *Assume that P is deterministic. Then the following conditions are equivalent*:
  1. *Item Acyclicity*;
  2. *Menu Acyclicity*;
  3. *Weak Acyclicity*;
  4. *There exists an injective function $u : Z \to \mathbb{R}$ s.t. $P(x|A) = 1$ iff $u(x) = \max_{z \in A} u(z)$.*

The equivalence of (1) and (4) follows from Richter (1966), and it is easy to see that (4) implies (3) which implies (1) and (2). Finally, with deterministic choice, the menu ranking boils down to that of the best element, which is why (1) and (2) are equivalent.

5. APU AS PAYOFF UNCERTAINTY VERSUS RANDOM UTILITY

Recent experimental papers show that stochastic choice can arise as deliberate randomization by subjects, rather than random variation in their expected utility functions. In lottery choice experiments, Agranov and Ortoleva (2015) found that a large majority of subjects select different options when the same

---

[28]The literature following Kreps (1979) generates rankings of menus from data on menu choice, but we do not use such data here, and two representations that are equivalent in our setting can have different implications for menu choice—see Fudenberg and Strzalecki (2015).

menus are offered several times in a row even when they are told that the menus will be repeated. This tendency is most relevant for "hard" questions where there is no item that is "clearly" better than others. According to the ex post questionnaire, "subjects' typical answer was that they did so because they did not know which option was best, and thus did not want to commit to a specific choice."[29] Dwenger, Kubler, and Weizsacker (2014) found that subjects deliberately choose different options even when asked to make simultaneous choices from two copies of the same menu of consumption goods, and reported analogous findings in a field study of students applying to German universities. This behavior is consistent with the "payoff uncertainty" formulation of APU that we present below.

5.1. *Perturbed Utility Arising From Payoff Uncertainty*

There are many possible ways to model the impact of the agent's uncertainty about the payoffs of various choices including robustness to model misspecification, as in Hansen and Sargent (2008). Here we develop a specification that generalizes this idea along the lines of the variational preferences of Maccheroni, Marinacci, and Rustichini (2006).

Suppose that when the agent chooses x, she receives total utility $u(x) + \varepsilon_x$, where $u(x)$ is a baseline utility that she knows, and ε_x is an uncertain taste shock. For each probability distribution on items $p \in \Delta(A)$ that the agent might choose, her utility is

$$
(7) \qquad \inf_{\varepsilon \in \mathbb{R}^A} \sum_{x \in A} p(x)\big[u(x) + \varepsilon_x\big] + \sum_{x \in A} \phi(\varepsilon_x),
$$

where ϕ is a convex function.

The interpretation of this objective function is that Nature picks $\varepsilon = (\varepsilon_x)_{x \in A}$ to minimize the agent's expected payoff. However, it is costly for Nature to make each component of the vector ε small, so it will choose to assign higher values to items that are less likely to be chosen. This gives the agent an incentive to choose nondegenerate probability distributions p.

Similar forces are at play in various other models. Maccheroni, Marinacci, and Rustichini (2006) studied preferences over acts when the agent is uncertain about the probability distribution over an objective state space and Nature minimizes the agent's payoff subject to a cost. Here minimax behavior in the resulting game leads the agent to prefer to randomize, as it does in other models of ambiguity-aversion such as Gilboa and Schmeidler (1989). Saito (2015) studied the preference for randomization of an ambiguity-averse agent in a model that allows for more general timing of Nature's move. He took preferences

---

[29]They also found that 29% of the subjects choose the option to flip a costly coin to randomize over items.

over menus of acts as primitives, while we look directly at choice probabilities. Note that these papers assume there is an objective state space. In our setting, the agent is uncertain about his true utility and the analyst only observes the items the agent chooses, so our state space is subjective, and the states are the possible values of ε. Epstein, Marinacci, and Seo (2007) studied a related subjective state space model, where the agent considers a set of utility functions and Nature picks a utility function in that set to minimize the agent's payoff, thereby making randomization valuable to the agent. They studied preferences on lotteries over menus.

In our setting, the objective function can also be seen as a desire to avoid feeling regret about items that were not chosen. Here the vector ε specifies the "extra utility" of each item, and the agent worries that Nature will choose the largest bonus on items he selects with low probability.

We now show that weak APU corresponds to the additive form $\Phi_A(\varepsilon) = \sum_{x \in A} \phi(\varepsilon_x)$, where $\phi : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is strictly convex, continuously differentiable where it is finite-valued, with derivative whose range includes $(-1, 0)$.[30] We will call any such function ϕ a *cost for Nature function*. The additive form of the Φ function is convenient for putting joint restrictions on choices from different menus. It can be interpreted as Nature not knowing u and hence treating each item symmetrically.

DEFINITION 21—Additive Variational Utility: A stochastic choice rule P has an *additive variational utility (AVU)* representation if and only if there exists a utility function $u : Z \to \mathbb{R}$ and a cost for Nature function ϕ such that

$$P(A) = \arg\max_{p \in \Delta(A)} \left(\inf_{\varepsilon \in \mathbb{R}^A} \sum_{x \in A} p(x)\big[u(x) + \varepsilon_x\big] + \sum_{x \in A} \phi(\varepsilon_x) \right).$$

PROPOSITION 7:
1. *P has an AVU representation if and only if P has a weak APU representation. Moreover, if P has an AVU representation with (u, ϕ), then P has a weak APU representation with (u, c), where $c(q) = \sup_\varepsilon \{q\varepsilon - \phi(-\varepsilon)\}$. Conversely, if P has a weak APU representation with (u, c), then P has an AVU representation with (u, ϕ), where $\phi(\varepsilon) = \sup_{q>0} \{-\varepsilon q - c(q)\}$.*
2. *P has an AVU representation with $\lim_{\varepsilon \to \infty} \phi'(\varepsilon) = 0$ iff P has an APU representation.*

Our proof of Proposition 7 uses convex duality. The first direction of the proof of part 1 constructs the cost function c from ϕ by setting c to be the convex conjugate of the function $\hat{\phi}(\varepsilon) := \phi(-\varepsilon)$. The second direction constructs

---

[30]We use these last two conditions on ϕ only to ensure that $\arg\min_{\varepsilon_x} p(x)\varepsilon_x + \phi(\varepsilon_x)$ exists and is continuous in $p(x) \in (0, 1)$.

ϕ from the cost function c, by setting $\hat{\phi}$ to be the convex conjugate of c and then setting $\phi(\varepsilon) := \hat{\phi}(-\varepsilon)$. To understand the second part of the proposition, note that, by the envelope theorem, AVU implies an APU with the marginal cost $c'(p(x)) = -\varepsilon_x^*$, where $\varepsilon_x^* = \phi'^{-1}(-p(x))$ is Nature's optimal choice ε_x against $p(x) \in (0, 1)$. Because c' is strictly increasing, ε_x^* is strictly decreasing. The Inada condition $\lim_{p \to 0} c'(p) = -\infty$ corresponds to $\lim_{\varepsilon \to \infty} \phi'(\varepsilon) = 0$ so that $\lim_{p(x) \to 0} \varepsilon_x^* = \infty$. This generates strictly positive choice probabilities because the payoff to any x diverges to ∞ as its probability goes to 0. The AVU that corresponds to logit choice has $\phi(\varepsilon) = \gamma \exp(-\frac{\varepsilon}{\gamma})$. In this case, the optimal choice of Nature is $\varepsilon_x^* = -\gamma \log(p(x))$. The AVU corresponding to logarithmic APU has $\phi(\varepsilon) = -\eta \log(\varepsilon)$, with $\phi(\varepsilon) = \infty$ for negative ε. In this case, Nature's optimal choice is $\varepsilon_x^* = \frac{\eta}{p(x)}$.

To relate asymptotic nondiscrimination (Proposition 5) to AVU, note that as $p(x) \to 0$, Nature will send the corresponding ε_x to infinity. Because $c'(q) = -\phi'^{-1}(-q)$, $c'^{-1}(s) = -\phi'(-s)$, so AVU implies APU with $h(s) = \log(-\phi'(-s))$. Thus the condition $\lim_{s \to \infty}[h(t - s) - h(-s)] = 0$ in Proposition 5 is equivalent to $\lim_{s \to \infty} \frac{\phi'(s-t)}{\phi'(s)} \to 1$, so Nature's marginal cost for rarely chosen items becomes flat. In this limit, Nature's choice depends on p but is insensitive to the differences in utilities, so it is optimal for the agent to assign about the same probability to all of the rarely chosen items.

5.2. *Comparison to Random Utility*

We now compare the revealed-preference implications of APU/AVU to those of random utility models.

DEFINITION 22—Random Utility: A stochastic choice rule P has a *random utility (RU)* representation if and only if there exists a utility function $u : Z \to \mathbb{R}$ and a random variable $\varepsilon \in \mathbb{R}^Z$ such that, for each $A \in \mathcal{A}$ and $z \in A$,

$$(8) \qquad P(z|A) = \text{Prob}\left\{u(z) + \varepsilon_z \geq \max_{y \in A} u(y) + \varepsilon_y\right\}.$$

Like APU, any RU choice rule satisfies regularity. We say that a RU is *symmetric* if the distribution of $\{\varepsilon_z\}$ is exchangeable, that is, vectors $(\varepsilon_1, \ldots, \varepsilon_n)$ and $(\varepsilon_{\pi(1)}, \ldots, \varepsilon_{\pi(n)})$ have the same distribution for any permutation π. Under positivity, symmetric RU and Fechnerian are observationally equivalent if the data consist only of binary menus; the additional structure imposed by APU only matters when choice is observed from some larger menus. More generally, any symmetric RU satisfies Item Acyclicity, so weak APU with menu-dependent costs nests symmetric RU. This applies in particular to any RU with i.i.d. shocks, as in the standard specification of the probit model. As the following example illustrates, such RU with i.i.d. shocks excludes some choice data that are consistent with weak APU.

EXAMPLE 3: Let $Z = \{x, y, z\}$ with $P(x|\{x, y\}) > P(y|\{x, y\}) > 0$ and $P(x|\{x, z\}) > P(z|\{x, z\}) > 0$. Then RU with i.i.d. shocks predicts $P(z|\{x, y, z\}) > 0$, because each of $u(x) - u(z) < \varepsilon_z - \varepsilon_x$ and $u(y) - u(z) < \varepsilon_z - \varepsilon_y$ holds with positive probability. On the other hand, weak APU can accommodate $P(z|\{x, y, z\}) = 0$ when y is preferred to z.

Moreover, as the example below shows, APU can violate the Block–Marshak conditions (Block and Marschak (1960)) that are necessary for RU.

EXAMPLE 4: When $Z = \{w, x, y, z\}$, RU implies[31]

$$P\big(w|\{w, x\}\big) + P\big(w|\{w, x, y, z\}\big) \geq P\big(w|\{w, x, y\}\big) + P\big(w|\{w, x, z\}\big).$$

We now construct an APU that violates this condition. Let $u(w) = -1$, $u(x) = 3$, and $u(y) = u(z) = 0$, and $c(p) = -\log(p)$.[32] Then $P(w|\{w, x\}) \approx 0.191$, $P(w|\{w, x, y\}) = P(w|\{w, x, z\}) \approx 0.177$, and $P(w|\{w, x, y, z\}) \approx 0.161$; thus,

$$P\big(w|\{w, x\}\big) + P\big(w|\{w, x, y, z\}\big) < P\big(w|\{w, x, y\}\big) + P\big(w|\{w, x, z\}\big).$$

Weak stochastic transitivity, the condition that $P(x|\{x, y\}) \geq \frac{1}{2}$ and $P(y|\{y, z\}) \geq \frac{1}{2}$ imply $P(x|\{x, z\}) \geq \frac{1}{2}$, can be violated by RU (Marschak (1959)), while it is satisfied by weak APU. More strongly, even with i.i.d. shocks RU need not correspond to weak APU.

EXAMPLE 5: Let $Z = \{x_1, x_2, y_1, y_2, y_3\}$. Let the utility function be $u(x_1) = u(x_2) = w$ and $u(y_1) = u(y_2) = u(y_3) = 0$. Let $A = \{x_1, x_2, y_1\}$ and $B = \{x_1, y_1, y_2, y_3, y_4, y_5, y_6\}$. Consider the probit model in which ε_z follows i.i.d. normal distribution $N(0, 1)$ for each $z \in Z$. Under probit, the choice probabilities are

$$P(z|A) = \int \prod_{z' \in A \setminus z} \Phi\big(u(z) + \varepsilon_z - u(z')\big)\phi(\varepsilon_z)\,d\varepsilon_z,$$

where Φ and ϕ are the cumulative distribution and the density under $N(0, 1)$. Then we have $P(x_1|B) \approx 0.4574 > P(x_1|A) \approx 0.4526$ and $P(y_1|A) \approx 0.0949 > P(y_1|B) \approx 0.0904$ when w is near 1.13. That is, there exists a menu cycle, as B is weaker than A for x_1 but A is weaker than B for y_1. This implies that this choice behavior cannot be rationalized by any weak APU.

---

[31]If $|Z| = 3$, then any choice rule that satisfies regularity has a RU representation (Block and Marschak (1960)), so any APU has a RU representation.

[32]Note that this result is different than Proposition 2.2 of Hofbauer and Sandholm (2002), where the utility function is known.

To accommodate observed choice behavior that violates IIA, the logit model has been extended to nested logit.[33] Our working paper (Fudenberg, Iijima, and Strzalecki (2014)) gives a revealed-preference characterization of nested logit and of an extension that allows for menu-size penalties. It also considers a particular form of nested model with only two nests, a "default option" $B_1 = \{x^*\}$ and "everything else" B_2. Here we suppose that when deciding whether to take the default or choose from B_2, the agent uses a cost function that depends on the size of B_2, such as $\sum_{z \in B_2} p(z) \log p(z) + \alpha \log |B_2|$. We use this to capture the phenomenon of "choice overload" as seen in Iyengar and Lepper (2000), where consumers are less likely to purchase when faced with a superset of a smaller menu.

6. CONCLUSION

As we have shown, perturbed utility functions are relatively tractable and have an easily understood axiomatic characterization that applies even when choice data are only observed for a subset of the possible menus. Moreover, these utility functions can be understood as describing choices of an agent who faces uncertainty about his true utility, modeled as smooth variational preferences. These features made it easy to develop further refinements, such as limited discrimination, which relaxes the IIA assumption implicit in the entropy cost function. As noted by Chernev (2012), there has been relatively empirical work on how menu size changes choice probabilities; we hope that the analytic foundations provided here may stimulate further empirical work. Our results may also prove helpful in designing more careful empirical analyses of just what sorts of randomization devices people prefer to use.

APPENDIX

A.1. *Proofs of Main Results*

A.1.1. *Rational Farkas*

The following result, called the theorem of the alternative, or Farkas's lemma, is usually applied to vector spaces over the field of real numbers $\mathbb{R}$, but also applies to vector spaces over the field of rational numbers $\mathbb{Q}$.[34] Let S be a finite set and treat $\mathbb{Q}^S$ as a vector space over the field of rational numbers $\mathbb{Q}$. Let $\langle \cdot, \cdot \rangle$ denote the inner product in $\mathbb{Q}^S$. For any vector $w \in \mathbb{Q}^S$ and subset $T \subseteq \mathbb{Q}^S$, we write $w \perp T$ if $\langle w, t \rangle = 0$ for all $t \in T$. For any $t, b \in \mathbb{Q}^S$, we write $t \le b$ whenever this inequality holds pointwise.

---

[33]Mattsson, Weibull, and Lindberg (2014) proposed a random utility model that generalizes nested logit and yields a closed form for the induced choice probabilities.

[34]See, for example, Kraft, Pratt, and Seidenberg (1959).

LEMMA A.1.1: *Let $b \in \mathbb{Q}^S$ and T be a linear subspace of $\mathbb{Q}^S$. Exactly one of the following conditions holds*:
1. *There exists $t \in T$ such that $t \leq b$.*
2. *There exists $w \in \mathbb{Q}_+^S$ such that $w \perp T$ and $\langle w, b \rangle < 0$.*

To understand the geometric interpretation of this lemma, consider first the case when T is a hyperplane, that is, is of dimension $|S| - 1$, and let B be the set of all points weakly dominated by b. The set $B \cap T$ is nonempty whenever Condition (1) holds. The set $B \cap T$ is empty whenever there exists a hyperplane that separates B from T, namely T itself; because of the shape of B, this hyperplane is generated by a vector $w \in \mathbb{Q}_+^S$. This is equivalent to Condition (2). To obtain the separating hyperplane in the case when T is lower dimensional, a superspace of T is used.

A.1.2. *Lemma 2*

LEMMA 2: *The following conditions are equivalent*:
(a) *There exists (u, c) such that P has a weak APU representation with (u, c).*
(b) *There exists (u, c, λ) such that P satisfies the FOC (5) with (u, c, λ).*
(c) *There exists (u, λ) such that P satisfies (6) with (u, λ).*

PROOF OF LEMMA 2: *Equivalence of (a) and (b)*: By the strict convexity of the objective function, a necessary and sufficient condition for $P(A) = \arg\max_{p \in \Delta(A)} V_c^u(p)$ is that $P(A)$ solves

$$
\max_{p \in \mathbb{R}^{|A|}} \sum_{z \in A} \left[u(z) p(z) - c\big(p(z)\big) \right] + \lambda(A) \left(\sum_z p(z) - 1 \right)
$$
$$
+ \sum_z \left[\lambda_0^z(A) p(z) + \lambda_1^z(A)\big(p(z) - 1\big) \right]
$$

such that $\lambda_0^z(A), \lambda_1^z(A) \geq 0$ and $\lambda_0^z(A) p(z) = \lambda_0^z(A)(p(z) - 1) = 0$ for each $z \in A$, where multipliers $\lambda(A)$, $\lambda_0^z(A)$, and $\lambda_1^z(A)$ are associated with $\sum_z p(z) = 1$, $p(z) \geq 0$, and $p(z) \leq 1$, respectively. This is equivalent to the conditions

$$
\forall z \in A, \quad u(z) - c_1^{\prime z}(A) + \lambda_0^z(A) = 0,
$$

where $\lambda_1^z(A) \geq 0 = \lambda_0^z(A)$ if $P(x|A) = 1$, $\lambda_0^z(A) = \lambda_1^z(A) = 0$ if $P(x|A) \in (0, 1)$, and $\lambda_0^z(A) \geq 0 = \lambda_1^z(A)$ if $P(x|A) = 0$.

(b) implies (c): The separability condition holds because c'^{-1} is strictly increasing.

(c) implies (b): Suppose that there exist u and λ such that (u, λ) is a separable representation. It is without loss to assume that both take values in $(0, 1)$.

Then, define

$$\overline{w} := \begin{cases} 2 & \text{if } P(z|A) < 1 \; \forall (z, A) \in \mathcal{D}, \\ \min\{u(x) + \lambda(A)|(x, A) \in \mathcal{D}, P(x|A) = 1\} \\ \quad \text{otherwise,} \end{cases}$$

$$\underline{w} := \begin{cases} 0 & \text{if } P(z|A) > 0 \; \forall (z, A) \in \mathcal{D}, \\ \max\{u(x) + \lambda(A)|(x, A) \in \mathcal{D}, P(x|A) = 0\} \\ \quad \text{otherwise.} \end{cases}$$

Let $g : [0, 1] \to \mathbb{R}$ be a strictly increasing and continuous function such that (i) $g(0) = \underline{w}$, (ii) $g(P(x|A)) = u(x) + \lambda(A)$ if $P(x|A) \in (0, 1)$, and (iii) $g(1) = \overline{w}$. Such function exists because $u(x) + \lambda(A) > u(y) + \lambda(B)$ if $P(x|A) > P(y|B)$, and $u(x) + \lambda(A) = u(y) + \lambda(B)$ if $P(x, A) = P(y|B) \in (0, 1)$. Define $c : [0, 1] \to \mathbb{R}$ by $c(p) = \int_0^p g(q) \, dq$. Then FOC (2) is satisfied at each menu. *Q.E.D.*

The proof of Lemma 1 follows the same lines, with (5) in part (b) replaced by (2), so that the FOC holds with equality, and the definition of separability in (c) tightened from (6) to (3).

A.1.3. *Proof of Theorem 1*

To show the equivalence of Acyclicity and APU, modify the proof of Theorem 3 to show the equivalence of Acyclicity and (6), and then use Lemma 1. Because of Lemma 2, it suffices to show the equivalence of Ordinal IIA and (5).

To show that (5) implies Ordinal IIA, we have

$$c'\big(P(x|A)\big) - c'\big(P(y|A)\big) = c'\big(P(x|B)\big) - c'\big(P(y|B)\big)$$

for each A, B and $x, y \in A \cap B$. Setting the strictly increasing function $f : [0, 1) \to \mathbb{R}_+$ by $f(0) = 0$ and $f(q) = \exp[c'(q)]$ for each $q > 0$, we obtain

$$\frac{f\big(P(x|A)\big)}{f\big(P(y|A)\big)} = \frac{f\big(P(x|B)\big)}{f\big(P(y|B)\big)}$$

for each A, B and $x, y \in A \cap B$, so that Ordinal IIA is satisfied.

To show that Ordinal IIA implies (5), construct a cost function by setting $c'(q) := \log(f(q))$ where f is taken from the Ordinal IIA property. Because c' is continuous, $c(q) = \int_{1/2}^q c'(t) \, dt$ is well defined for each $q \in (0, 1)$. Note that c is C^1 and strictly convex, and that $\lim_{q \to 0} c'(q) = -\infty$. Fix any item x and set $u(x) := 0$. For any other item $z \neq x$, set $u(z) := c'(P(z|\{x, z\})) - c'(P(x|\{x, z\}))$. Take an arbitrary menu A and $y, z \in A$. There are two exclusive cases.

Case (i): $x \in \{y, z\}$. Set $y = x$ without loss of generality. Then

$$u(z) - u(x) = c'\big(P(z|\{x, z\})\big) - c'\big(P(x|\{x, z\})\big)$$

$$= \log\left(\frac{f\big(P(z|\{x, z\})\big)}{f\big(P(x|\{x, z\})\big)}\right)$$

$$= \log\left(\frac{f\big(P(z|A)\big)}{f\big(P(x|A)\big)}\right) \quad (\because \text{Ordinal IIA})$$

$$= c'\big(P(z|A)\big) - c'\big(P(x|A)\big).$$

Case (ii): $x \notin \{y, z\}$:

$$u(z) - u(y) = c'\big(P(z|\{x, z\})\big) - c'\big(P(x|\{x, z\})\big)$$

$$\qquad - c'\big(P(y|\{x, y\})\big) + c'\big(P(x|\{x, y\})\big)$$

$$= \log\left(\frac{f\big(P(z|\{x, z\})\big)}{f\big(P(x|\{x, z\})\big)} \frac{f\big(P(x|\{x, y\})\big)}{f\big(P(y|\{x, y\})\big)}\right)$$

$$= \log\left(\frac{f\big(P(z|\{x, y, z\})\big)}{f\big(P(x|\{x, y, z\})\big)} \frac{f\big(P(x|\{x, y, z\})\big)}{f\big(P(y|\{x, y, z\})\big)}\right)$$

$$\quad (\because \text{Ordinal IIA})$$

$$= \log\left(\frac{f\big(P(z|\{x, y, z\})\big)}{f\big(P(y|\{x, y, z\})\big)}\right)$$

$$= \log\left(\frac{f\big(P(z|A)\big)}{f\big(P(y|A)\big)}\right) \quad (\because \text{Ordinal IIA})$$

$$= c'\big(P(z|A)\big) - c'\big(P(y|A)\big).$$

Therefore, the equalities in the above two cases imply that FOC at A is satisfied.

A.1.4. *Proof of Theorem 3*

By Lemma 2, it suffices to show the equivalence of Weak Acyclicity and a separable representation. Suppose that there exists a separable representation (u, λ). Weak Acyclicity is satisfied, as otherwise then $u(x_i) + \lambda(A_i) \geq u(y_i) + \lambda(B_i)$ for all i with at least one strict. Summing over i yields a contradiction because of the permutation property.

For the converse, let $\mathbb{Q}^*$ be the vector space over the field of rational numbers whose coordinates correspond to ordered pairs $(\alpha, \beta) = \{(x_\alpha, A_\alpha), (x_\beta, A_\beta)\}$ with $P(x_\alpha|A_\alpha) \geq^* P(x_\beta|A_\beta)$. Represent a collection of

these pairs by the vector w whose coordinates count the number of times the corresponding relation appears.

A collection is a cycle if (a) at least one comparison is strict and (b) each item and each menu appear the same number of times on each side.

Step 1: We will now represent a cycle as a collection that satisfies certain linear inequalities. Define $b \in \mathbb{Q}^*$ as follows:

$$b(\alpha, \beta) = \begin{cases} -1 & \text{if } P(x_\alpha | A_\alpha) > P(x_\beta | A_\beta), \\ 0 & \text{if } P(x_\alpha | A_\alpha) = P(x_\beta | A_\beta) \in (0, 1). \end{cases}$$

Note that for any $w \in \mathbb{Q}^*$, $\langle w, b \rangle < 0$ iff at least one comparison in a collection of order comparisons represented by w is strict. For each $z \in Z$, define $t^z \in \mathbb{Q}^*$ by

$$t^z(\alpha, \beta) = \begin{cases} -1 & \text{if } x_\alpha = z \text{ and } x_\beta \neq z, \\ 1 & \text{if } x_\alpha \neq z \text{ and } x_\beta = z, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\langle w, t^z \rangle = 0$ iff z features equal number of times on each side of the cycle represented by w. For each $C \in \mathcal{A}$, define $t^C \in \mathbb{Q}^*$ by

$$t^C(\alpha, \beta) = \begin{cases} -1 & \text{if } A_\alpha = C \text{ and } A_\beta \neq C, \\ 1 & \text{if } A_\alpha \neq C \text{ and } A_\beta = C, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, $\langle w, t^C \rangle = 0$ iff C features equal number of times on each side of the cycle represented by w. Let T be the linear subspace generated by the collection $\{t^z\}_{z \in Z} \cup \{t^C\}_{C \in \mathcal{A}}$. Thus, $w \in \mathbb{Q}^*$ represents a cycle if and only if $w \perp T$ and $\langle w, b \rangle < 0$.

Step 2: Since Weak Acyclicity implies that there does not exist w that meets the conditions of Step 1, there cannot exist $w \in \mathbb{Q}^*$ such that $w \perp T$ and $\langle w, b \rangle < 0$. Lemma A.1.1 implies that there exists $t \in T$ such that $t \leq b$.

Step 3: The existence of such t implies that there exists a separable representation (u, λ). To see that, note that since $t \in T$, there are functions $u : Z \to \mathbb{Q}$ and $\lambda : \mathcal{A} \to \mathbb{Q}$ such that $t = \sum_{z \in Z} u(z) t^z + \sum_{C \in \mathcal{A}} \lambda(C) t^C$. Thus, the functions u and λ are the coordinates of t in T. Next, observe that $t \leq b$ implies that (u, λ) is a separable representation: if $P(x|A) > P(y|B)$, then $t((x, A), (y, B)) = -u(x) - \lambda(A) + u(y) + \lambda(B) \leq b((x, A), (y, B)) = -1$, so $u(x) + \lambda(A) > u(y) + \lambda(B)$. If $P(x|A) = P(y|B) \in (0, 1)$, then $t((x, A), (y, B)) \leq 0$; by symmetry, $t((x, A), (y, B)) \geq 0$; thus, $u(x) + \lambda(A) = u(y) + \lambda(B)$.

A.2. *Proof of Theorem 2*

To prove the first part of the theorem we rely on the same argument as in the proof of Theorem 1. We only prove the "only if" direction of the second part.

For this purpose, we construct an increasing sequence of finite subsets Z_n of Z with corresponding APU representations (u_n, c_n). Given this sequence, we show that the limit cost function c is finite, strictly increasing, and continuous on $(0, 1)$. We use this limit cost function to construct a function f so that Ordinal IIA holds, and then the first part of the theorem implies there is an APU representation.

A.2.1. *Intermediate Results*

We first show that Continuity, combined with other conditions, implies several useful richness properties of P. The first one is just a form of Richness for binary menus.

LEMMA 3: *For any $x \in Z$ and any $p \in (0, 1)$, there exists $y \in Z$ such that $P(x|\{x, y\}) = p$.*

PROOF: A direct consequence of Richness applied to $p + q = 1$, in which $y = z$ necessarily holds.   *Q.E.D.*

The second property guarantees the existence of a *midpoint* between any two elements of Z.

LEMMA 4: *For any $x, y \in Z$, there exists $z \in Z$ such that $P(x|\{x, z\}) = P(z|\{y, z\})$.*

PROOF: If $P(x|\{x, y\}) = 0.5$, then the existence of z follows from Richness (with $p = q = \frac{1}{3}$) and Acyclicity. Otherwise, let $f(z) := P(x|\{x, z\}) - P(z|\{y, z\})$ and note that by Continuity, f is a continuous function. By Lemma 3, there exist items $x', y' \in Z$ such that $P(x'|\{x, x'\}) = P(y'|\{y, y'\}) = 0.5$. Thus, $f(x') = 0.5 - P(x'|\{x', y\}) = 0.5 - P(x|\{x, y\}) = -f(y')$, where the equality of the two probabilities follows from Acyclicity. Since Z is a connected space, the intermediate value theorem implies that there exists $z \in Z$ with $f(z) = 0$.   *Q.E.D.*

Acyclicity implies the *quadruple condition* of Davidson and Marschak (1959).

LEMMA 5: *For any $a, b, c, d \in Z$, we have $P(a|\{a, b\}) \geq P(c|\{c, d\})$ if and only if $P(a|\{a, c\}) \geq P(b|\{b, d\})$.*

PROOF: Suppose that $P(a|\{a, b\}) \geq P(c|\{c, d\})$ but $P(a|\{a, c\}) < P(b|\{b, d\})$. Then since the choice probabilities sum up to 1, $P(c|\{a, c\}) > P(d|\{b, d\})$ and $P(b|\{a, b\}) \leq P(d|\{c, d\})$. These four inequalities are a cycle.   *Q.E.D.*

The conditions derived in Lemmas 3 and 5 let us invoke the theorem of Debreu (1958) to conclude that P on binary menus has a "Fechnerian utility" representation.

COROLLARY 1: *Under Acyclicity, Richness, and Continuity, there exists* u : $Z \to \mathbb{R}$ *such that* $P(a|\{a, b\}) \geq P(c|\{c, d\})$ *iff* $u(a) - u(b) \geq u(c) - u(d)$. *Moreover,* u *is unique up to positive affine transformations.*

For any finite subset of Z, the function u induces the same order on elements as any of the possible APU representations. We use this fact in the proof of Claims 1 and 3 below.

A.2.2. *Constructing a Grid Approximation*

Since Z is a separable space, there exists a countable dense set $Y = \{y_1, y_2, \ldots\}$. Let $Y_n := \{y_1, \ldots, y_n\}$. The finite sets Z_n will be constructed recursively. Each of them will be the union of Y_n, a *standard sequence* $\tilde{Z}_n$ of points with utility differences equal to 2^{-n}, and additional points with choice probabilities $p = \frac{k}{2^{n+1}}$, $k = 1, \ldots, 2^{n+1}$.

A.2.2.1. *The Construction of* Z_1. Fix an arbitrary element $z_0 \in Z$ and let $p^* \in (0, \frac{1}{2})$. By Richness, there exist elements $z_{-1}, z_1 \in Z$ s.t. $P(z_{-1}|\{z_{-1}, z_0\}) = p^* = P(z_0|\{z_0, z_1\})$. By Lemma 4, there exist midpoints $z_{-0.5}, z_{0.5}$ such that $P(z_{-1}|\{z_{-1}, z_{-0.5}\}) = P(z_{-0.5}|\{z_{-0.5}, z_0\})$ and likewise for $z_{0.5}$. Let $\tilde{Z}_1 := \{z_{-1}, z_{-0.5}, z_0, z_{0.5}, z_1\}$.

By Richness, there exist $A_{1/4} \in \mathcal{A}$ and $x_{1/4} \in A$ and such that $P(x_{1/4}|A_{1/4}) = \frac{1}{4}$ and $P(z_0|A_{1/4}) = \frac{1}{2}$, and there is $x_{3/4} \in Z$ such that $P(x_{1/4}|\{x_{1/4}, x_{3/4}\}) = \frac{1}{4}$. Let $Z_1 := \tilde{Z}_1 \cup A_{1/4} \cup \{x_{3/4}\} \cup Y_1$.

By Theorem 1 on Z_0, there is an APU representation $(\hat{u}_1, \hat{c}_1)$ of P. Note that by the FOC, we have that $\hat{u}_1(z_1) - \hat{u}_1(z_0) = \hat{c}_1'(1 - p^*) - \hat{c}_1'(p^*) = \hat{u}_1(z_0) - \hat{u}_1(z_{-1})$. Note that (u_1, c_1) is also an APU representation of P, where

$$u_1(z) := \left[\hat{u}_1(z_1) - \hat{u}_1(z_0)\right]^{-1}\left[\hat{u}_1(z) - \hat{u}_1(z_0)\right] \quad \text{for all } z \in Z_1,$$

$$c_1(p) := \left[\hat{u}_1(z_1) - \hat{u}_1(z_0)\right]^{-1}\left[\hat{c}_1(p) - p\hat{c}_1'\left(\frac{1}{2}\right) - \hat{c}_1\left(\frac{1}{2}\right)\right]$$

$$\text{for all } p \in (0, 1).$$

Note that $u_1(z_r) = r$ for $r \in \{-1, -0.5, 0, 0.5, 1\}$ and $c_1(\frac{1}{2}) = 0 = c_1'(\frac{1}{2})$. Note also that by FOC, we have $c_1'(\frac{1}{2}) - c_1'(\frac{1}{4}) = u_1(z_0) - u_1(x_{1/4})$, so $c_1'(\frac{1}{4}) = u_1(x_{1/4})$. We also have $c_1'(\frac{3}{2}) - c_1'(\frac{1}{4}) = u_1(x_{3/4}) - u_1(x_{1/4})$, so $c_1'(\frac{3}{4}) = u_1(x_{3/4})$.

A.2.2.2. *The Construction of* Z_{n+1}. By Richness, there are $z_{-n-1}, z_{n+1} \in Z$ s.t. $P(z_{-n-1}|\{z_{-n-1}, z_{-n}\}) = p^* = P(z_n|\{z_n, z_{n+1}\})$. By Lemma 4, for any $z_r, z_{r+2^{-n}} \in \tilde{Z}_n$, there exists a midpoint $z_{r+2^{-n-1}}$. Likewise, there is a midpoint $z_{n+0.5}$ between z_n and z_{n+1}, and there are further midpoints between these, such as $z_{n+0.25}$, etc. Let $\tilde{Z}_{n+1} := \{z_r : r = \frac{k}{2^{n+1}}, k = -(n+1)2^{n+1}, \ldots, (n+1)2^{n+1}\}$.

By Richness, for $p = \frac{k}{2^{n+2}}$, $k = 1, \ldots, 2^{n+1}$, there exist $x_p \in A_p \in \mathcal{A}$ such that $P(x_p|A_p) = p$ and $P(z_0|A_p) = \frac{1}{2}$. By Richness, there exist $x_{1-p} \in Z$ such that $P(x_p|\{x_p, x_{1-p}\}) = p$. Let

$$Z_2 := \tilde{Z}_2 \cup Y_2 \cup \bigcup_{p=k/2^{n+2}, k=1, \ldots, 2^{n+1}} A_p \cup \bigcup_{p=k/2^{n+2}, k=2^{n+1}+1, \ldots, 2^{n+2}-1} \{x_p\}.$$

By Theorem 1 on Z_{n+1}, there exists an APU representation $(\hat{u}_{n+1}, \hat{c}_{n+1})$. Use the rescaling as above so that the representation (u_{n+1}, c_{n+1}) has the following properties:

- $u_{n+1}(z_r) = r$ for $r = \frac{k}{2^{n+1}}$, $k = -(n+1)2^{n+1}, \ldots, (n+1)2^{n+1}$.
- $c'_{n+1}(p) = u_{n+1}(x_p)$ for $p = \frac{k}{2^{n+2}}$, $k = 1, \ldots, 2^{n+2} - 1$; $c'_n(\frac{1}{2}) = c_n(\frac{1}{2}) = 0$.

A.2.3. *Properties of the Limit Cost Function*

Define $c'(q) := \limsup_m c'_m(q)$ for each $q \in (0, 1)$. FOC implies $c'_n(1 - p^*) - c'_n(p^*) = u_n(z_1) - u_n(z_0) = 1$ for each n, so that $c'(1 - p^*) - c'(p^*) = 1$. Also note that $c'(\frac{1}{2}) = 0$.

CLAIM 1: *$c'(q)$ is strictly increasing and finite at all $q \in (0, 1)$.*

PROOF: First, note that Corollary 1 implies that $u(z_{i+1}) - u(z_i) = u(z_i) - u(z_{i-1})$ for all i, which implies that the range of u is unbounded from both above and below. Moreover, $u(z_{r+1/2^n}) - u(z_r) = 2^{-n}[u(z_{r+1}) - u(z_r)] \to_{n \to \infty} 0$, which follows since the sequence $\tilde{Z}_n$ was constructed by taking midpoints and from the fact that the Fechnerian utility of a midpoint equals the average of the two endpoints.

Take any $q = \frac{k}{2^{m+1}}$, $q' = \frac{k-1}{2^{m+1}}$ for some $m \in \mathbb{N}$, $k \in \{2, 3, \ldots, 2^{m+1} - 1\}$. By the above observation, for n' large enough, there exist $z_r, z_{r'} \in \tilde{Z}_{n'}$ such that $u(z_{-n'}) < u(x_{q'}) < u(z_{r'}) < u(z_r) < u(x_q) < u(z_{n'})$. This implies that $-n' = u_n(z_{-n'}) < u_n(x_{q'}) < u_n(z_{r'}) < u_n(z_r) < u_n(x_q) < u_n(z_{n'}) = n'$ for all $n > n'$.

Since $c'(q) = \limsup_n u_n(x_q)$, it follows that $c'(q)$ is finite (and likewise for $c'(q')$). To show monotonicity, note that, by construction, $u_n(z_r) - u_n(z_{r'})$ is independent of n; denote this value $\delta > 0$. By construction, we have $c'_n(q) = u_n(x_q)$ and $c'_n(q') = u_n(x_{q'})$, thus, $c'_n(q) - c'_n(q') > \delta$, so $c'(q) - c'(q') > \delta > 0$.

To show that c' is finite on $(0, 1)$, note that, for any $q'' \in (0, 1)$, there exist q, q' of the form above such that $q' < q'' < q$, so the result follows from weak monotonicity of c', which holds by its definition. To show that c' is strictly increasing on $(0, 1)$, note that, for any $q > q'$ in $(0, 1)$, there are k, m such that $q > \frac{k}{2^m} > \frac{k-1}{2^m} > q'$, which ensures $c'(q) > c'(q')$.   Q.E.D.

Let C denote the set of points $q \in (0, 1)$ at which $c'(q)$ is continuous. By the previous claim, $(0, 1) \setminus C$ is at most countable.

CLAIM 2: *For any* $A, B, x, y \in A \cap B$ *such that* $P(x|A), P(x|B), P(y|A),$ $P(y|B) \in C$, $c'(P(x|A)) - c'(P(y|A)) = c'(P(x|B)) - c'(P(y|B))$ *holds.*

PROOF: Denote $A \cup B = \{z^1, \ldots, z^m\}$. By construction of Z_n and separability of Z, we can choose sequences $(z_n^j)_{n \in \mathbb{N}, j=1,\ldots,m}$ of items such that $\lim_{n\to\infty} z_n^j = z^j$ for each $j = 1, \ldots, m$ and $\{z_n^1, \ldots, z_n^m\} \subseteq Z_n$ for each n. We can choose a subsequence so that $\lim_{k\to\infty} c'_{n_k}(p) = c'(p)$ for each $p = P(x|A), P(x|B), P(y|A), P(y|B)$. By Continuity, $P(x_{n_k}|A_{n_k}) \to P(x|A)$, $P(y_{n_k}|A_{n_k}) \to P(y|A)$, $P(x_{n_k}|B_{n_k}) \to P(x|B)$, and $P(y_{n_k}|B_{n_k}) \to P(y|B)$, for some $A_n, B_n \subseteq \{z_n^1, \ldots, z_n^m\}$ and $x_n, y_n \in \{z_n^1, \ldots, z_n^m\}$. It remains to show that, along a subsequence, $c'_n(P(x_n|A_n)) \to c'(P(x|A))$ and likewise for the other three elements, which together with the FOC $c'_n(P(x_n|A_n)) - c'_n(P(y_n|A_n)) = c'_n(P(x_n|B_n)) - c'_n(P(y_n|B_n))$ implies the conclusion.

Denote $q^* := P(x|A)$ and $q_n := P(x_n|A_n)$ for each n. Fix $\varepsilon > 0$; because c' is continuous at q^*, there is $\delta > 0$ such that $|c'(q^*) - c'(q)| < \frac{\varepsilon}{2}$ for all $q \in [q^* - \delta, q^* + \delta]$. By taking a subsequence, there is $\bar{n}$ such that $|c'_{n_{k_l}}(q^* - \delta) - c'(q^* - \delta)|, |c'_{n_{k_l}}(q^* + \delta) - c'(q^* + \delta)| < \frac{\varepsilon}{2}$ for all $n_{k_l} \geq \bar{n}$. As $\lim_l q_{n_{k_l}} = q^*$, there is $\bar{n}'$ such that $q_{n_{k_l}} \in [q^* - \delta, q^* + \delta]$ for all $n_{k_l} \geq \bar{n}'$. Therefore, for any $n_{k_l} \geq \max\{\bar{n}, \bar{n}'\}$, we have

$$\left| c'(q^*) - c'_{n_{k_l}}(q_{n_{k_l}}) \right|$$
$$\leq \max\left\{ \left| c'(q^*) - c'_{n_{k_l}}(q^* - \delta) \right|, \left| c'(q^*) - c'_{n_{k_l}}(q^* + \delta) \right| \right\}$$
$$\leq \max\left\{ \left| c'(q^*) - c'(q^* - \delta) \right| + \left| c'(q^* - \delta) - c'_{n_{k_l}}(q^* - \delta) \right|, \right.$$
$$\left. \left| c'(q^*) - c'(q^* + \delta) \right| + \left| c'(q^* + \delta) - c'_{n_{k_l}}(q^* + \delta) \right| \right\}$$
$$< \varepsilon,$$

where the first inequality follows because $c'_{n_{k_l}}$ is increasing. *Q.E.D.*

CLAIM 3: c' *is continuous at all* $q \in (0, 1)$.

PROOF: Suppose to the contrary that c' is discontinuous at some $q^* \in (0, 1)$. Take any $q^{**} \in C$ strictly less than $1 - q^*$. Fix $x \in Z$; by Richness 1.1, there are y, z such that $P(x|\{x, y, z\}) = q^*, P(y|\{x, y, z\}) = q^{**}$. Define a subset H of $(0, 1)^2$ by $H := \{(P(x|\{x, y, z'\}), P(y|\{x, y, z'\})) : z' \in Z\}$. By continuity of P and connectedness of Z, the set H is connected. Let $H_+ := \{(p_x, p_y) \in H : p_x \geq q^*, p_y \geq q^{**}\}$ and $H_- := \{(p_x, p_y) \in H : p_x \leq q^*, p_y \leq q^{**}\}$. Note also that by Acyclicity for all $z' \in Z$,

$$(*) \qquad P(x|\{x, y, z'\}) > q^* \quad \text{iff} \quad P(y|\{x, y, z'\}) > q^{**};$$

thus, $H = H_+ \cup H_-$. By Richness, there exist $z_-, z_+ \in Z$ such that $P(z|\{z_-, z\}) > P(z_-|\{z_-, z\})$ and $P(z|\{z_+, z\}) < P(z_+|\{z_+, z\})$. By Acyclicity, this implies that the sets $H_+ \setminus \{(q^*, q^{**})\}$ and $H_- \setminus \{(q^*, q^{**})\}$ are nonempty.[35]

Since $(0, 1) \setminus C$ is at most a countable set and H is connected, there exist sequences $z_+^n, z_-^n \in Z$ such that $C^2 \cap H_- \ni (P(x|\{x, y, z_+^n\}), P(y|\{x, y, z_+^n\})) \to (q^*, q^{**})$ and $C^2 \cap H_+ \ni (P(x|\{x, y, z_-^n\}), P(y|\{x, y, z_-^n\})) \to (q^*, q^{**})$. By Claim 2, $c'(P(x|\{x, y, z_-^n\})) - c'(P(y|\{x, y, z_-^n\})) = c'(P(x|\{x, y, z_+^n\})) - c'(P(y|\{x, y, z_+^n\}))$ for every n, and thus $\lim_{p \nearrow q^*} c'(p) - \lim_{q \nearrow q^{**}} c'(q) = \lim_{p \searrow q^*} c'(p) - \lim_{q \searrow q^{**}} c'(q)$, which leads to a contradiction. *Q.E.D.*

Thus the equality in Claim 2 holds for all menus A, B and items $x, y \in A \cap B$, so Ordinal IIA holds with $f := \exp[c']$. By the first part, there exists an APU representation (u, c) of P. By Continuity of P and FOC of the form $u(x) - u(y) = c'(P(x|\{x, y\})) - c'(1 - P(x|\{x, y\}))$, it follows that u is continuous.

Finally, to show uniqueness, let (u, c) and $(\hat{u}, \hat{c})$ be APU representations of P. Note that $P(a|\{a, b\}) \ge P(c|\{c, d\})$ iff $u(a) - u(b) \ge u(c) - u(d)$ iff $\hat{u}(a) - \hat{u}(b) \ge \hat{u}(c) - \hat{u}(d)$. Thus from Corollary 1, $\hat{u} = \alpha u + \beta$ for some $\alpha > 0, \beta \in \mathbb{R}$. Take any $p, p' \in (0, 1)$. Then for any $q < 1 - p, 1 - p'$, by Richness, $P(x|A) = p, P(y|A) = q, P(x'|A') = p', P(y'|A') = q$ for some menus and items. By FOC,

$$\hat{c}'(p) - \hat{c}'(p') = \hat{c}'(p) - \hat{c}'(q) + \hat{c}'(q) - \hat{c}'(p')$$
$$= \hat{u}(x) - \hat{u}(y) + \hat{u}(y') - \hat{u}(x')$$
$$= \alpha(u(x) - u(y) + u(y') - u(x'))$$
$$= \alpha(c'(p) - c'(q) + c'(q) - c'(p'))$$
$$= \alpha(c'(p) - c'(p')).$$

Let $\gamma := \hat{c}'(\frac{1}{2}) - \alpha c'(\frac{1}{2})$. Then, for any $p \in (0, 1)$, $\hat{c}'(p) - \hat{c}'(\frac{1}{2}) = \alpha(c'(p) - c'(\frac{1}{2}))$ and thus $\hat{c}'(p) = \alpha c'(p) + \gamma$. Define $\delta := \hat{c}(\frac{1}{2}) - \alpha c(\frac{1}{2}) - \frac{\gamma}{2}$. Then, for any $p \in (0, 1)$, $\hat{c}(p) - \hat{c}(\frac{1}{2}) = \int_{1/2}^p \hat{c}'(q) \, dq = \int_{1/2}^p (\alpha c'(q) + \gamma) \, dq = \alpha(c(p) - c(\frac{1}{2})) + (p - \frac{1}{2})\gamma$ and thus $\hat{c}(p) = \alpha c(p) + \delta + \gamma p$.

A.3. *Proof of Proposition 1*

First, suppose that P has a Fechnerian representation. Thus there exists u such that $P(x|\{x, y\}) \ge P(x'|\{x', y'\})$ iff $u(x) - u(y) \ge u(x') - u(y')$. Define $\lambda(\{x, y\}) := -\frac{u(x) + u(y)}{2}$ for each binary menu $\{x, y\}$. Then $P(x|\{x, y\}) \ge$

---

[35]For example, we show that $H_+ \setminus \{(q^*, q^{**})\}$ is nonempty. Suppose not. Then by (*) we have that $q^{**} \ge P(y|\{x, y, z_-\})$, and since probabilities sum up to 1, we have that $P(z_-|\{x, y, z_-\}) \ge P(z|\{x, y, z\})$. The last two inequalities, together with $P(z|\{z_-, z\}) > P(z_-|\{z_-, z\})$, form a cycle; contradiction.

$P(x'|\{x', y'\})$ iff $u(x) + \lambda(\{x, y\}) \geq u(x') + \lambda(\{x', y'\})$. This separable representation ensures that Acyclicity is satisfied.

Second, suppose that Acyclicity is satisfied. Then P has an APU representation by either Theorem 1 or 2. By FOC, $u(x) - u(y) = c'(P(x|\{x, y\})) - c'(1 - P(x|\{x, y\}))$ holds for any x, y. Because c' is strictly increasing, $P(x|\{x, y\}) \geq P(x'|\{x', y'\})$ iff $u(x) - u(y) \geq u(x') - u(y')$, which implies the existence of a Fechnerian representation.

A.4. *Proof of Proposition 2*

Note first that FOC and $P_i(x|\{x, y\}) = 1 - P_i(x|\{x, y\})$, $i = 1, 2$, imply that, for all $\{x, y\} \in \mathcal{A}$,

$$(9) \qquad c_1'\big(P_1\big(x|\{x, y\}\big)\big) - c_1'\big(1 - P_1\big(x|\{x, y\}\big)\big)$$
$$= u(x) - u(y) = c_2'\big(P_2\big(x|\{x, y\}\big)\big) - c_2'\big(1 - P_2\big(x|\{x, y\}\big)\big).$$

To show "if" direction, take any x, y with $P_2(x|\{x, y\}) \geq P_2(y|\{x, y\})$, which implies $u(x) \geq u(y)$. Since $c_i'(p) - c_i'(1 - p)$ is strictly increasing in p, $P_1(x|\{x, y\}) \geq P_2(x|\{x, y\})$ follows by (9).

To show "only if," suppose to the contrary that there is $q \in (\frac{1}{2}, 1)$ such that $c_1'(q) - c_1'(1 - q) > c_2'(q) - c_2'(1 - q)$. By Richness, there exist x, y such that $P_1(x|\{x, y\}) = q$. Since $c_2'(p) - c_2'(1 - p)$ is strictly increasing in p, (9) implies $P_2(x|\{x, y\}) > q$, a contradiction.

A.5. *Proof of Proposition 3*

For each $i = 1, 2$, let $\lambda_i(A)$ denote the Lagrange multiplier at menu A under (u, c_i). To show the "only if" part, suppose to the contrary, that $c_1'' \leq c_2''$ does not hold at some point in $(0, 1)$. By continuity of the second derivatives, there exists interval $(\underline{q}, \overline{q})$ such that $c_1''(q) > c_2''(q)$ for all $q \in (\underline{q}, \overline{q})$.

Note that Richness ensures that the range of u is unbounded from below and above. Thus there are x, y such that $P_i(x|\{x, y\}) > \overline{q}$ for each $i = 1, 2$. By Continuity, u is continuous, and $\{P_i(x|\{x, y, z\})|z \in Z\}$ is connected for each $i = 1, 2$. Note that $P_i(x|\{x, y, z\}) \to 0$ as $u(z) \to \infty$ and $P_i(x|\{x, y, z\}) \to P_i(x|\{x, y\})$ as $u(z) \to -\infty$. Thus by the intermediate theorem, we can take z_1 such that $P_1(x|\{x, y, z_1\}) \in (\underline{q}, \overline{q})$. And likewise we can take z_2 such that $P_2(x|\{x, y, z_2\}) = P_1(x|\{x, y, z_1\})$. Note that, for fixed y, the value $\lambda_i(\{x', y, z_i'\})$ depends only on utilities $(u(x'), u(z_i'))$, and thus can be written as $\lambda_i(\{x', y, z_i'\}) = g_i(u(x'), u(z_i'))$. This function g_i is continuous and strictly decreasing in each argument. Therefore, for a strictly decreasing sequence $\varepsilon_k \searrow 0$, we can find an increasing sequence $\varepsilon_k' \nearrow 0$ such that $g_1(u(x) + \varepsilon_k, u(z) + \varepsilon_k') = g_1(u(x), u(z_1))$ for all large enough k. Pick such k so that $c_1'^{-1}(u(x) + \varepsilon_k + g(u(x), u(z))) < \overline{q}$. By connectedness and continuity, we can take x' and z_1' such that $u(x') = u(x) + \varepsilon_k$ and $u(z_1') = u(z) + \varepsilon_k'$. Note that we

have $P_1(x'|\{x', y, z_1'\}) < \bar{q}$ and $P_1(y|\{x, y, z_1\}) = P_1(y|\{x', y, z_1'\})$ by construction. Using the same argument, we can take z_2' such that $g_2(u(x'), u(z_2')) = g_2(u(x), u(z_2))$, or $P_2(y|\{x, y, z_2\}) = P_2(y|\{x', y, z_2'\})$.

In the above construction, $u(x') > u(x)$ holds, which ensures $P_1(x'|\{x, x'\}) > \frac{1}{2}$. Then, because P_1 is more selective than P_2, $P_1(x'|\{x', y, z_1'\}) \geq P_2(x'|\{x', y, z_2'\})$ follows.

By FOC, for each $i = 1, 2$,

$$c_i'\big(P_i\big(x|\{x, y, z_i\}\big)\big) = u(x) + \lambda_i\big(\{x, y, z_i\}\big),$$
$$c_i'\big(P_i\big(x'|\{x', y, z_i'\}\big)\big) = u(x') + \lambda_i\big(\{x', y, z_i'\}\big),$$

so

$$c_1'\big(P_1\big(x'|\{x', y, z_i'\}\big)\big) - c_1'\big(P_1\big(x|\{x, y, z_i\}\big)\big)$$
$$= c_2'\big(P_2\big(x'|\{x', y, z_2'\}\big)\big) - c_2'\big(P_2\big(x|\{x, y, z_2\}\big)\big),$$

or

$$(10) \qquad \int_{P_1(x|\{x,y,z_1\})}^{P_1(x'|\{x',y,z_1'\})} c_1''(p)\, dp = \int_{P_2(x|\{x,y,z_2\})}^{P_2(x'|\{x',y,z_2'\})} c_2''(p)\, dp.$$

By $P_1(x'|\{x', y, z_1'\}) \geq P_2(x'|\{x', y, z_2'\})$ and $P_1(x|\{x, y, z_1\}) = P_2(x|\{x, y, z_2\})$, it leads to a contradiction.

To show the "if," take any x, x', y, z_1, z_2 such that $P_i(y|\{x, y, z_i\}) = P_i(y|\{x', y, z_i'\})$, $i = 1, 2$ and $P_1(x|\{x, y, z_1\}) = P_2(x|\{x, y, z_2\})$, $P_1(x|\{x, x'\}) > P_1(x|\{x, x'\})$. As in the "only if" above, FOC implies (10), which ensures $P_1(x'|\{x', y, z_1'\}) \geq P_2(x'|\{x', y, z_2'\})$.

A.6. *Proof of Proposition 4*

To show the first direction, take any A, B with $x, y \in A \cap B$ such that $P(x|A) \geq P(x|B)$ and $P(x|A) > P(y|A)$. Acyclicity implies that $P(y|A) \geq P(y|B)$ holds. Thus from the FOCs $u(x) + \lambda(A) = c'(P(x|A))$, it follows that $\lambda(A) \geq \lambda(B)$ and $u(x) > u(z)$. Using these FOCs, we can express the log-ratio of choice probabilities as

$$\log\left(\frac{P(x|A)}{P(y|A)}\right) = \log\left(\frac{c'^{-1}\big(u(x) + \lambda(A)\big)}{c'^{-1}\big(u(y) + \lambda(A)\big)}\right)$$
$$= h\big(u(x) + \lambda(A)\big) - h\big(u(y) + \lambda(A)\big)$$
$$\geq h\big(u(x) + \lambda(B)\big) - h\big(u(y) + \lambda(B)\big)$$
$$= \log\left(\frac{c'^{-1}\big(u(x) + \lambda(B)\big)}{c'^{-1}\big(u(y) + \lambda(B)\big)}\right) = \log\left(\frac{P(x|B)}{P(y|B)}\right),$$

where the inequality follows by the convexity of h, $u(x) - u(y) > 0$, and $\lambda(A) \geq \lambda(B)$. Therefore, $P(x|A)/P(y|A) \geq P(x|B)/P(y|B)$.

A.7. *Proof of Proposition 5*

(1): To show that Asymptotic Non-Discrimination holds, take any sequence of menus A_n such that $x, y \in A_n$ for each n and $\lim_n P(x|A_n) = \lim_n P(y|A_n) = 0$. From the FOC $u(x) + \lambda(A_n) = c'(P(x|A_n))$ and $\lim_{q \to 0} c'(q) = -\infty$, $P(x|A_n) \to 0$ implies $\lambda(A_n) \to -\infty$. Therefore,

$$\log\left(\frac{P(x|A_n)}{P(y|A_n)}\right) = \log\left(\frac{c'^{-1}\big(u(x) + \lambda(A_n)\big)}{c'^{-1}\big(u(y) + \lambda(A_n)\big)}\right)$$
$$= h\big(u(x) + \lambda(A_n)\big) - h\big(u(y) + \lambda(A_n)\big) \to 0$$

as $n \to \infty$. Therefore, $\frac{P(x|A_n)}{P(y|A_n)} \to 1$.

(2): Suppose to the contrary that there exists $r > 0$ such that $h(r - t) - h(-t)$ does not converge to 0 as $t \to \infty$. Let $b \in (0, \infty]$ denote the limit superior of the sequence. Then there exists a sequence $\{\lambda_n\}_{n=1}^{\infty}$ such that $\lambda_n \to -\infty$ and $h(r + \lambda_n) - h(\lambda_n) \to b$.

By Continuity, u is continuous, and Richness ensures that its range is unbounded from below and above. Define q as the unique solution to $r = c'(q) - c'(1 - q)$. By Richness, there exist x, y such that $P(x|\{x, y\}) = q$. Then FOC implies $u(x) - u(y) = r$. We assume $u(y) = 0$ without loss of generality.

Because of $\lambda_n \to -\infty$, we have $c'^{-1}(r + \lambda_n) \to 0$. By connectedness of Z, using the intermediate value theorem argument, for each n we can find z_n such that $c'^{-1}(r + \lambda_n) = P(x|\{x, y, z_n\})$. FOC $u(x) + \lambda(\{x, y, z_n\}) = c'(P(x|\{x, y, z_n\}))$ and $u(x) = r$ imply $\lambda_n = \lambda(\{x, y, z_n\})$. This leads to

$$\log\left(\frac{P\big(x|\{x, y, z_n\}\big)}{P\big(y|\{x, y, z_n\}\big)}\right) = \log\left(\frac{c'^{-1}(r + \lambda_n)}{c'^{-1}(\lambda_n)}\right)$$
$$= \big[h(r + \lambda_n) - h(\lambda_n)\big] \to b$$

as $n \to \infty$. Because $b > 0$, $\lim \frac{P(x|\{x,y,z_n\})}{P(y|\{x,y,z_n\})} > 1$, which contradicts Asymptotic Non-Discrimination.

A.8. *Proof of Proposition 6*

Note that when P is deterministic, $P(x|A) \geq^* P(y|B)$ iff $P(x|A) > P(y|B)$ iff $P(x|A) = 1$ and $P(y|B) = 0$.

Equivalence of (1) and (4): When P is deterministic, it induces a deterministic and single-valued choice function $\mathcal{C} : \mathcal{A} \to Z$ by $\mathcal{C}(A) = x$ with $P(x|A) = 1$. Then Item Acyclicity is satisfied if and only if the deterministic choice function

satisfies the Congruence axiom (Richter (1966)), that is, there is no sequence of items $x_1, x_2, \ldots, x_n$ such that

$$x_1 = \mathcal{C}(A_1) \neq x_2 \in A_1, \quad x_2 = \mathcal{C}(A_2) \neq x_3 \in A_2, \quad \ldots$$
$$x_n = \mathcal{C}(A_n) \neq x_1 \in A_n.$$

As shown by Richter (1966), this is equivalent to the existence of a preference over Z such that for each A, $\mathcal{C}(A)$ is the set of the most preferred elements of A. Because Z is finite and $\mathcal{C}$ is single valued, this is equivalent to the existence of a strict utility function that rationalizes $\mathcal{C}$.

(4) ⇒ (3): Let u be an injective function such that $P(x|A) = 1$ iff $u(x) = \max_{z \in A} u(z)$. Suppose that Weak Acyclicity is violated by a sequence $P(x_1|A_1) > P(y_1|B_1), P(x_2|A_2) > P(y_2|B_2), \ldots, P(x_n|A_n) > P(y_n|B_n)$. Pick any $j_1 = 1, 2, \ldots, n$. Because $x_{j_1} = y_{k_1}$ for some $k_1 = 1, 2, \ldots, n$, $P(x_{j_1}|B_k) = P(y_k|B_k) = 0$. Also, as $B_k = A_{j_2}$ for some $j_2 = 1, 2, \ldots, n$, $1 = P(x_{j_2}|B_k) > P(x_{j_1}|B_k) = 0$. Since n is finite, we can construct a sequence $P(x_{j_1}|B_{k_1}) < P(x_{j_2}|B_{k_1}), P(x_{j_2}|B_{k_2}) < P(x_{j_3}|B_{k_2}), \ldots, P(x_{j_n}|B_{k_n}) < P(x_{j_1}|B_{k_n})$, which leads to a contradiction $u(x_{j_1}) < u(x_{j_2}) < \cdots < u(x_{j_n}) < u(x_{j_1})$.

(3) ⇒ (2): This immediately follows by definition.

(2) ⇒ (1): Suppose that Item Acyclicity is violated by a sequence $P(x_1|A_1) > P(x_2|A_1), P(x_2|A_2) > P(x_3|A_2), \ldots, P(x_n|A_n) > P(x_1|A_n)$. Then, $0 = P(x_{k+1}|A_k) < P(x_{k+1}|A_{k+1}) = 1$ for each $k = 1, \ldots, n-1$, and $0 = P(x_1|A_n) < P(x_1|A_1) = 1$ hold. Thus Menu Ayclicity is also violated, a contradiction.

A.9. *Proof of Proposition 7*

Proof of (1): For any $\phi : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ that is C^1 over $\phi^{-1}(\mathbb{R})$, strictly convex, and satisfies $(-1, 0) \subseteq \mathrm{range}(\phi')$ and any cost function c, define the functions $V_\phi : \Delta(A) \to \mathbb{R} \cup \{\infty\}$ and $V_c : \Delta(A) \to \mathbb{R} \cup \{\infty\}$ as follows:

$$V_\phi(p) = \inf_{\varepsilon \in \mathbb{R}^{|A|}} \sum_{x \in A} p(x)[u(x) + \varepsilon_x] + \sum_{x \in A} \phi(\varepsilon_x),$$

$$V_c(p) = \sum_{z \in A} (u(z)p(z) - c(p(z))).$$

Note that $V_\phi(p) - V_c(p) = \sum_{z \in A} \{c(p(z)) + \inf_{\varepsilon_z \in \mathbb{R}} [p(z)\varepsilon_z + \phi(\varepsilon_z)]\}$. For any function ϕ, define the function $\hat{\phi}$ by $\hat{\phi}(\varepsilon) = \phi(-\varepsilon)$. Then $c(p(z)) + \inf_{\varepsilon_z \in \mathbb{R}} [p(z)\varepsilon_z + \phi(\varepsilon_z)] = c(p(z)) + \inf_{\varepsilon_z \in \mathbb{R}} -[p(z)\varepsilon_z - \hat{\phi}(\varepsilon_z)] = c(p(z)) - \sup_{\varepsilon_z \in \mathbb{R}} [p(z)\varepsilon_z - \hat{\phi}(\varepsilon_z)]$.

To prove the first claim, fix ϕ and define c to be the convex conjugate of $\hat{\phi}$, that is, $c(q) = \sup_\varepsilon \{q\varepsilon - \hat{\phi}(\varepsilon)\}$. Then $c'(q) = \hat{\phi}'(-q) = -\phi'^{-1}(-q)$ for each $q \in (0, 1)$ from the assumption $(-1, 0) \subseteq \mathrm{range}(\phi')$. Thus, c is a cost function and $V_\phi(p) - V_c(p) = 0$ for all p.

To prove the second claim, note that if we choose ϕ so that $\hat{\phi}$ is the convex conjugate of c, that is, $\phi(-\varepsilon) = \hat{\phi}(\varepsilon) =: \sup_{q>0}\{\varepsilon q - c(q)\}$, then $\hat{\phi}'(w) = c'^{-1}(-w)$, so it is strictly convex and satisfies $(-1, 0) \subseteq \text{range}(\phi')$. By the Fenchel biconjugation theorem (Theorem 12.2 of Rockafellar (1970)) c is the convex conjugate of $\hat{\phi}$, so likewise $V_\phi(p) - V_c(p) = 0$ for all p.

Proof of (2): Suppose that P is represented by AVU with $\lim_{\varepsilon\to\infty} \phi'(\varepsilon) = 0$. Part 1 implies that P is represented by APU with $c(q) = \sup_\varepsilon\{q\varepsilon - \phi(-\varepsilon)\}$, and since $c'(q) = -\phi'^{-1}(-q)$ for each $q \in (0, 1)$, $\lim_{q\to 0} c'(q) = -\infty$.

Suppose that P is represented by APU with cost c. Part (1) implies that P is represented by AVU with $\phi(\varepsilon) = \sup_{q\in(0,1]}\{-\varepsilon q - c(q)\}$, so $\phi'(\varepsilon) = -c'^{-1}(w)$, and $\lim_{q\to 0} c'(q) = -\infty$ implies $\lim_{\varepsilon\to\infty} \phi'(\varepsilon) = 0$.

REFERENCES

AFRIAT, S. N. (1967): "The Construction of Utility Functions From Expenditure Data," *International Economic Review*, 8 (1), 67–77. [2374]

AGRANOV, M., AND P. ORTOLEVA (2015): "Stochastic Choice and Preferences for Randomization," *Journal of Political Economy* (forthcoming). [2373,2387]

AGUIAR, V. H. (2015): "Random Consideration Sets: Inferring Preferences From Psychological Biases," Report. [2374]

ANDERSON, S., A. DE PALMA, AND J. F. THISSE (1992): *Discrete Choice Theory of Product Differentiation*. Cambridge: MIT Press. [2375]

BEN-AKIVA, M., AND S. R. LERMAN (1985): *Discrete Choice Analysis*. Cambridge: MIT Press. [2375,2384]

BENAIM, M., AND M. HIRSCH (1999): "Mixed Equilibria and Dynamical Systems Arising From Fictitious Play in Perturbed Games," *Games and Economic Behavior*, 29 (1–2), 36–72. [2381]

BENAIM, M., J. HOFBAUER, AND E. HOPKINS (2009): "Learning in Games With Unstable Equilibria," *Journal of Economic Theory*, 144, 1694–1709. [2371]

BLACKWELL, D. (1956): "Controlled Random Walks," in *Proceedings International Congress of Mathematicians*, Vol. III. Amsterdam: North-Holland, 336–338. [2375]

BLOCK, D., AND J. MARSCHAK (1960): "Random Orderings and Stochastic Theories of Responses," in *Contributions to Probability and Statistics*, ed. by I. Olkin et al. Stanford: Stanford University Press. [2381,2391]

CERREIA-VIOGLIO, S., D. DILLENBERGER, AND P. ORTOLEVA (2015): "Cautious Expected Utility and the Certainty Effect," *Econometrica*, 83 (2), 693–728. [2373]

CERREIA-VIOGLIO, S., D. DILLENBERGER, P. ORTOLEVA, AND G. RIELLA (2015): "Deliberately Stochastic," Report. [2373]

CHERNEV, A. (2012): "Product Assortment and Consumer Choice: An Interdisciplinary Review," *Foundations and Trends in Marketing*, 6, 1–61. [2392]

CLARK, S. A. (1990): "A Concept of Stochastic Transitivity for the Random Utility Model," *Journal of Mathematical Psychology*, 34 (1), 95–108. [2371,2386]

DAVIDSON, D., AND J. MARSCHAK (1959): "Experimental Tests of Stochastic Decision Theory," in *Measurement: Definitions and Theories*, ed. by C. W. Churchman. New York: Wiley. [2381, 2397]

DE CLIPPEL, G., AND K. ROZEN (2014): "Bounded Rationality and Limited Datasets," Report. [2374]

DEBREU, G. (1958): "Stochastic Choice and Cardinal Utility," *Econometrica*, 26 (3), 440–444. [2381,2397]

DWENGER, N., D. KUBLER, AND G. WEIZSACKER (2014): "Flipping a Coin: Theory and Evidence," Technical Report. [2373,2388]

ECHENIQUE, F., AND K. SAITO (2015): "Savage in the Market," *Econometrica*, 82 (4), 1467–1497. [2374,2379]

ECHENIQUE, F., K. SAITO, AND G. TSERENJIGMID (2014): "The Perception Adjusted Luce Model," Technical Report. [2374]

EPSTEIN, L. G., M. MARINACCI, AND K. SEO (2007): "Coarse Contingencies and Ambiguity," *Theoretical Economics*, 2 (4), 355–394. [2389]

FALMAGNE, J. C. (1978): "A Representation Theorem for Finite Random Scale Systems," *Journal of Mathematical Psychology*, 18 (1), 52–72. [2373]

FISHBURN, P. C. (1992): "Induced Binary Probabilities and the Linear Ordering Polytope: A Status Report," *Mathematical Social Sciences*, 23, 67–80. [2374]

——— (1998): "Stochastic Utility," in *Handbook of Utility Theory*, Vol. 1. Dordrecht: Kluwer Academic, 273–319. [2379]

FOSGERAU, M., AND A. DE PALMA (2015): "Demand Systems for Market Shares," Report. [2381]

FRICK, M. (2013): "Monotone Threshold Representations," Working Paper, Harvard University. [2383]

FUDENBERG, D., AND D. K. LEVINE (1995): "Consistency and Cautious Fictitious Play," *Journal of Economic Dynamics and Control*, 19, 1065–1089. [2371,2381]

FUDENBERG, D., AND T. STRZALECKI (2015): "Dynamic Logit With Choice Aversion," *Econometrica*, 83 (2), 651–691. [2374,2387]

FUDENBERG, D., AND S. TAKAHASHI (2011): "Heterogeneous Beliefs and Local Information in Stochastic Fictitious Play," *Games and Economic Behavior*, 71, 100–120. [2371]

FUDENBERG, D., R. IIJIMA, AND T. STRZALECKI (2014): "Stochastic Choice and Revealed Perturbed Utility," Report. [2373,2386,2392]

GILBOA, I. (1990): "A Necessary but Insufficient Condition for the Stochastic Binary Choice Problem," *Journal of Mathematical Psychology*, 34, 371–392. [2374]

GILBOA, I., AND D. MONDERER (1992): "A Game-Theoretic Approach to the Binary Stochastic Choice Problem," *Journal of Mathematical Psychology*, 36 (1), 555–572. [2374]

GILBOA, I., AND D. SCHMEIDLER (1989): "Maxmin Expected Utility With Non-Unique Prior," *Journal of Mathematical Economics*, 18 (2), 141–153. [2388]

GUL, F., AND W. PESENDORFER (2006): "Random Expected Utility," *Econometrica*, 74 (1), 121–146. [2379]

GUL, F., P. NATENZON, AND W. PESENDORFER (2014): "Random Choice as Behavioral Optimization," *Econometrica*, 82 (5), 1873–1912. [2374]

HANNAN, J. (1957): "Approximation to Bayes Risk in Repeated Plays," in *Contributions to the Theory of Games*, Vol. 3, ed. by M. Dresher, A. Tucker, and P. Wolfe. Princeton: Princeton University Press, 97–139. [2371,2375]

HANSEN, L. P., AND T. J. SARGENT (2008): *Robustness*. Princeton: Princeton University Press. [2388]

HARSANYI, J. C. (1973a): "Games With Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points," *International Journal of Game Theory*, 2 (1), 1–23. [2373]

——— (1973b): "Oddness of the Number of Equilibrium Points: A New Proof," *International Journal of Game Theory*, 2 (1), 235–250. [2371,2375,2377]

HOFBAUER, J., AND W. H. SANDHOLM (2002): "On the Global Convergence of Stochastic Fictitious Play," *Econometrica*, 70 (6), 2265–2294. [2371,2373,2391]

HOUTHAKKER, H. S. (1950): "Revealed Preference and the Utility Function," *Economica*, 17, 159–174. [2386]

IIJIMA, R. (2014): "Deterministic Equilibrium Selection Under Payoff-Perturbed Dynamics," Report. [2371,2384]

IYENGAR, S. S., AND M. R. LEPPER (2000): "When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology*, 79 (6), 995–1006. [2392]

KRAFT, C. H., J. W. PRATT, AND A. SEIDENBERG (1959): "Intuitive Probability on Finite Sets," *The Annals of Mathematical Statistics*, 30 (2), 408–419. [2379,2392]

KREINDLER, G. E., AND H. P. YOUNG (2013): "Fast Convergence in Evolutionary Equilibrium Selection," *Games and Economic Behavior*, 80, 39–67. [2381]

KREPS, D. M. (1979): "A Representation Theorem for 'Preference for Flexibility'," *Econometrica*, 47, 565–577. [2387]

KUBLER F., L. SELDEN, AND X. WEI (2014): "Asset Demand Based Tests of Expected Utility Maximization," *American Economic Review*, 104 (11), 3459–3480. [2374]

LUCE, D. (1956): "Semiorders and a Theory of Utility Discrimination," *Econometrica*, 24, 178–191. [2383]

——— (1959): *Individual Choice Behavior*. New York: Wiley. [2372,2373,2376]

LUCE, R. D., AND P. SUPPES (1965): "Preference, Utility, and Subjective Probability," in *Handbook of Mathematical Psychology*, Vol. III, ed. by R. D. Luce, R. R. Bush, and E. Galanter. New York: Wiley. [2375]

MACCHERONI, F., M. MARINACCI, AND A. RUSTICHINI (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74 (6), 1447–1498. [2373, 2379,2388]

MACHINA, M. J. (1985): "Stochastic Choice Functions Generated From Deterministic Preferences Over Lotteries," *The Economic Journal*, 95 (379), 575–594. [2371,2373]

MANZINI, P., AND M. MARIOTTI (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82 (3), 1153–1176. [2374]

MARSCHAK, J. (1959): "Binary Choice Constraints on Random Utility Indicators," Technical Report. [2373,2391]

MATTSSON, L.-G., AND J. W. WEIBULL (2002): "Probabilistic Choice and Procedurally Bounded Rationality," *Games and Economic Behavior*, 41, 61–78. [2371,2373]

MATTSSON, L.-G., J. W. WEIBULL, AND P. O. LINDBERG (2014): "Extreme Values, Invariance and Choice Probabilities," *Transportation Research Part B: Methodological*, 59, 81–95. [2392]

MCFADDEN, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka. New York: Academic Press. [2373,2375]

MCKELVEY, R., AND T. PALFREY (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, 10, 6–38. [2381]

MERTIKOPOULOS, P., AND W. H. SANDHOLM (2015): "Learning in Games via Reinforcement and Regularization," Report. [2384]

NOGUCHI, Y. (2015): "Bayesian Learning, Smooth Approximate Optimal Behavior, and Convergence to ε-Nash Equilibrium," *Econometrica*, 83 (1), 353–373. [2371]

RENY, P. J. (2015): "A Characterization of Rationalizable Consumer Behavior," *Econometrica*, 83 (1), 175–192. [2374]

RICHTER, M. K. (1966): "Revealed Preference Theory," *Econometrica*, 34 (3), 635–645. [2373, 2374,2386,2387,2405]

ROCKAFELLAR, R. T. (1970): *Convex Analysis*. Princeton: Princeton University Press. [2406]

ROSENTHAL, A. R. W. (1989): "A Bounded-Rationality Approach to the Study of Noncooperative Games," *International Journal of Game Theory*, 18, 273–292. [2371,2375,2384]

SAITO, K. (2015): "Preferences for Flexibility and Randomization Under Uncertainty," *American Economic Review*, 105 (3), 1246–1271. [2388]

SCOTT, D. (1964): "Measurement Structures and Linear Inequalities," *Journal of Mathematical Psychology*, 1 (2), 233–247. [2377,2381]

SWAIT, J., AND A. MARLEY (2013): "Probabilistic Choice (Models) as a Result of Balancing Multiple Goals," *Journal of Mathematical Psychology*, 57 (1–2), 1–14. [2371]

THURSTONE, L. L. (1927): "A Law of Comparative Judgment," *Psychological Review*, 34 (4), 273–286. [2373]

TVERSKY, A. (1964): "Finite Additive Structures," Technical Report, Michigan Mathematical Psycholoy Program, University of Michigan. [2377]

——— (1972): "Choice by Elimination," *Journal of Mathematical Psychology*, 9, 341–367. [2377]

VAN DAMME, E. (1991): *Stability and Perfection of Nash Equilibria*. Berlin: Springer. [2373]

VAN DAMME, E., AND J. W. WEIBULL (2002): "Evolution in Games With Endogenous Mistake Probabilities," *Journal of Economic Theory*, 106 (2), 296–315. [2371]

VOORNEVELD, M. (2006): "Probabilistic Choice in Games: Properties of Rosenthal's *t*-Solutions," *International Journal of Game Theory*, 34 (1), 105–121. [2375]

WEIBULL, J., L.-G. MATTSSON, AND M. VOORNEVELD (2007): "Better May Be Worse: Some Monotonicity Results and Paradoxes in Discrete Choice Under Uncertainty," *Theory and Decision*, 63, 121–151. [2373]

Dept. of Economics, Harvard University, Littauer Center, 1805 Cambridge Street, Cambridge, MA 02138, U.S.A.; dfudenberg@harvard.edu,

Dept. of Economics, Harvard University, Littauer Center, 1805 Cambridge Street, Cambridge, MA 02138, U.S.A.; riijima@fas.harvard.edu,

and

Dept. of Economics, Harvard University, Littauer Center, 1805 Cambridge Street, Cambridge, MA 02138, U.S.A.; tomasz_strzalecki@harvard.edu.