Stochastic Choice and Revealed Perturbed Utility*

Drew Fudenberg† Ryota Iijima‡ Tomasz Strzalecki§

December 2, 2014

Abstract

Perturbed utility functions—the sum of expected utility and a non-linear perturbation function—provide a simple and tractable way to model various sorts of stochastic choice. We provide easily understood conditions that characterize this representation by generalizing the acyclicity condition used in revealed preference theory. We show how to relax Luce’s IIA condition to model cases where the agent finds it harder to discriminate between items in larger menus, and how to extend the perturbation-function approach to model choice overload and nested decisions. We also show that these representations correspond to a form of ambiguity-averse preferences for an agent who is uncertain about her true utility.

*We thank Kim Border, Jerry Green, Sonia Jaffe, Kohei Kawaguchi, Mark Machina, Morgan McClellon, Wolfgang Pesendorfer, Drazen Prelec, Bill Sandholm, Ricky Vohra, and Peter Wakker for useful comments and suggestions, and NSF grants SES 0951462, 1258665, and CAREER grant 1255062 and Sloan foundation for financial support.

†Department of Economics, Harvard University. E-mail:dfudenberg@harvard.edu
‡Department of Economics, Harvard University. E-mail:riijima@fas.harvard.edu
§Department of Economics, Harvard University. E-mail:tomasz_strzalecki@harvard.edu.
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

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

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. Similar perturbed utility functions have been previously used by e.g. Harsanyi (1973b), Machina (1985), Rosenthal (1989), Clark (1990), Mattsson and Weibull (2002), and Swait and Marley (2013).\(^1\) Because we want to apply the perturbed-utility representation to choice sets \( A \) of varying size, we adopt an additive form for the perturbation function, as opposed to allowing a general functions on \( \Delta(A) \). (The additive specification is vacuous if we allow the function \( c \) 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 non-linear 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

\(^1\)Perturbed utility has also been used in the theory of learning in games. Fudenberg and Levine (1995) show 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) use perturbed utility to construct Lyapunov functions for stochastic fictitious play, and van Damme and Weibull (2002) study perturbed utility in an evolutionary model.
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 (1959)’s IIA condition.

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. 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. As an example, in Section 3 we explore cost functions that weaken IIA by allowing the odds ratios to become closer to one as menus become large, which reflects the idea that it is harder to discriminate between objects in larger menus.

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);\(^2\) Recent experimental evidence (Agranov and Ortoleva, 2014; Dwenger, Kubler, and Weizsacker, 2014) indicates that stochastic choice may arise as 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 and so randomizes to guard against moves by a malevolent Nature. Another interpretation is that stochastic choice arises due to inattention or implemen-

\(^2\)Formally, Machina (1985) considers a form of the menu invariant cost functions that we define below. Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella (2013) also study randomization generated by nonlinear preferences over lotteries; they use a subclass of the non-EU preferences studied by Cerreia-Vioglio, Dillenberger, and Ortoleva (forthcoming).
tation 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).³

Acyclicity can be weakened to Menu Acyclicity, which rules out menu cycles. This condition implies that menus can be ordered by weakness, with the probability of choosing any given item being at least as high in a weaker menu. An alternative relaxation of is Item Acyclicity, which implies that items can be ordered by their desirability. In the setting of determinsitic choice, these two conditions are each equivalent to Acyclicity, and are also equivalent to Richter’s (1966) congruence axiom, which extends the Strong Axiom of Revealed Preference to settings where data is 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.⁴ 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 classes of stochastic choice rules are not nested, though their intersection is non-empty 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 is what we consider here) is not enough to reveal whether the agent prefers larger or smaller menus. Fudenberg and Strzalecki (forthcoming) use cost functions to address this in the special case in which choice

³See 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.
⁴Hofbauer and Sandholm (2002) show 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.
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 (forthcoming) study
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 (2013) consider an agent with a deterministic
priority order who uses logistic choice on the perceived items. Gul, Natzenzon, and Pesendorfer
(forthcoming) introduce “attribute rules” which are related to nested logit.

Many papers on stochastic choice assume that choice data is complete (i.e., available for
every possible menu), or at least sufficiently “rich.” Most of our results do not require this,
and apply when choice is observed for a subset of the possible menus, as in the work of e.g.
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.\footnote{See also Reny (forthcoming), who extends Afriat’s result to infinite data sets, and Kubler and Wei (2014)
and Echenique and Saito (2014) who study revealed preference in the demand for financial assets.} 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 $A$ would lead to a violation of those axioms. Our results imply that this problem
does not arise.

## 2 Additive Perturbed Utility

Let $Z$ be a finite set of items (consequences or prizes). A menu is a nonempty subset of $Z$.
Let $\mathcal{A}$ be the set of menus for which the choice probabilities of an 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, i.e., the collection $\mathcal{A}$ need not include every
non-empty subset of $Z$. We consider a stochastic choice rule $P$ that maps each menu $A \in \mathcal{A}$
to a probability distribution on its elements. Formally, a stochastic choice rule is a mapping
$P : \mathcal{A} \to \Delta(Z)$ with the property that for any $A \in \mathcal{A}$ the support of $P(A)$ is a subset of $A$. For
any given $A \in \mathcal{A}$ and $z \in A$ 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 generalize 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 \( A \in A \) 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).

Perhaps the most familiar stochastic choice rule is logit/logistic choice, also known as the Luce rule, which is given by \( P(z|A) = \frac{\exp(\eta u(z))}{\sum_{z' \in A} \exp(\eta u(z'))} \); this corresponds to additive perturbed utility with cost \( c(p) = \eta^{-1} p \log p \).\(^6\) As Luce (1959) showed, whenever all choice probabilities are positive logit choice is characterized by the “IIA” condition.

**Definition 2.** \( P \) satisfies IIA if for all \( A, B \in A \) with \( x, y \in A \cap B \)

\[
\frac{P(x|A)}{P(y|A)} = \frac{P(x|B)}{P(y|B)}
\]

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 Rosenthal (1989).

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

**Definition 3 (APU).** An APU has the form

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

for some utility function \( u : Z \rightarrow \mathbb{R} \) and cost function \( c \).

\(^6\)As is well known, this corresponds to a random utility model where the shocks \( \varepsilon \) are i.i.d. Gumbel with variance \( \eta \); we discuss the relationship between APU and random utility in Section 5.
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.

### 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 in the stochastic setting (as we will see in Section 4), and the second condition generalizes IIA.

**Definition 4.** $P$ satisfies Acyclicity if for any integer $n$ and any bijections $f, g : \{1, \ldots, n\} \to \{1, \ldots, n\}$,

$$P(x_1|A_1) > P(x_{f(1)}|A_{g(1)}), \quad P(x_k|A_k) \geq P(x_{f(k)}|A_{g(k)}) \quad \text{for } 1 < k < n$$

implies $P(x_n|A_n) < P(x_{f(n)}|A_{g(n)})$

**Remark 1.** Although Acyclicity rules out cycles of any length, a consequence of Theorem 2 below is that 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 it 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.”

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 $A$. 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. 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: in particular, data needs to fit the restriction \( \sum_{z \in A} P(z|A) = 1 \). We discuss these issues further in Section 4.

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 \( f : [0,1) \rightarrow \mathbb{R}_+ \) with \( f(0) = 0 \) such that

\[
\frac{f(P(x|A))}{f(P(y|A))} = \frac{f(P(x|B))}{f(P(y|B))}
\]

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 \( f(q) = \exp(-\frac{1}{\eta}q) \), the cost is proportional to \( -\log q \) as in Harsanyi (1973b).

**Theorem 1.** Suppose that Positivity holds and \( A \) 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 equivalence between (1) and (3) is a consequence of Theorem 2, which relaxes both the positivity condition and the requirement that \( A \) contains all menus with size 2 and 3. The next subsection provides some intuition and a proof sketch of both equivalences.
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 \( \lambda : A \rightarrow \mathbb{R} \) satisfy the *first order conditions* (FOC) for \( P \) iff

\[
    u(x) + \lambda(A) = c'(P(x|A))
\]  
(2)

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

**Definition 7.** \( u : Z \rightarrow \mathbb{R} \) and \( \lambda : A \rightarrow \mathbb{R} \) are a separable representation of \( P \) if and only if

\[
    u(x) + \lambda(A) > u(y) + \lambda(B) \text{ iff } P(x|A) > P(y|B)
\]  
(3)

If Positivity holds, the following conditions are equivalent (Lemma 1 in the Appendix):

(a) There exists \( (u, c) \) such that \( P \) has an APU representation with \( (u, c) \)

(b) There exists \( (u, c, \lambda) \) such that \( P \) satisfies the FOC with \( (u, c, \lambda) \)

(c) There exists \( (u, c) \) such that \( P \) has a separable representation with \( (u, \lambda) \).

In addition, the same \( (u, c, \lambda) \) can be used in each condition, although we will not make use of that fact. The equivalence of APU and the FOC follows from the Kuhn-Tucker theorem. That (c) follows from (b) is straightforward from the strict monotonicity of \( c' \). To show that (b) follows from (c), use a variant of the usual ordinal uniqueness argument to show that if \( P \) has a separable representation \( (u, \lambda) \), then there is a strictly increasing and continuous function \( g : [0,1] \rightarrow \mathbb{R} \) that satisfies \( g(P(x|A)) = u(x) + \lambda(A) \) whenever \( P(x|A) \in (0,1) \). We then define \( c(p) := \int_0^p g(q)dq \), and it is immediate that \( (u, c) \) satisfies the first order conditions.

To prove the equivalence of Acyclicity and APU in Theorem 1, we show 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 \((u, \lambda)\) 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’ lemma.\(^7\)

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 \(x, y \in A \cap B\)

\[
c'(P(x|A)) - c'(P(y|A)) = c'(P(x|B)) - c'(P(y|B))
\]

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 \(c(q) = \int_0^q c'(t)dt\) 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.\(^8\) 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.\(^9\) Here we consider an infinite set \(Z\), but we assume that each menu is a finite set. When \(Z\) is infinite, APU is characterized by Positivity and Ordinal IIA. Positivity and Acyclicity are still necessary, but we do not know whether it is sufficient. The following richness condition

---

\(^7\)Echenique and Saito (2014) use a related result in their analysis of deterministic portfolio choice. A key technical difference is that their analog of our acyclicity condition imposes constrains on the products of various prices, while our conditions are purely ordinal.

\(^8\)This is also the case for other models of stochastic choice, such as random utility, see, e.g., Fishburn (1998). Stronger uniqueness results can be obtained when items are lotteries, see, e.g., Gul and Pesendorfer (2006).

\(^9\)A similar situation arises for variational preferences of Maccheroni, Marinacci, and Rustichini (2006); to obtain uniqueness, they impose an additional axiom that guarantees that the range of \(u\) is rich enough.
implies that the range of $u$ equals $\mathbb{R}$.

**Definition 8.** $P$ satisfies Richness if

(i) For any $x \in Z$, $p \in (0, 1)$ there exist $y \in Z$ such that $\{x, y\} \in A$ and $P(x|\{x, y\}) = p$.

(ii) There exists $\bar{p}$ such that for any $p \in (0, \frac{1}{2}]$, there exist $A \in A$ and $z, z' \in A$ such that $P(z|A) = \bar{p}$ and $P(z'|A) = p$.

Note that Richness implies that the collection $A$ contains many binary menus and also many menus with at least three elements.

**Proposition 1.** Under Richness if $(u, c)$ and $(\hat{u}, \hat{c})$ represent the same APU $P$, 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. Note that $u$ and $c$ are expressed in the same units; that is why multiplying $u$ by a constant $\alpha$ requires multiplying $c$ by the same $\alpha$. Since the absolute level of the cost function does not matter, we are free to shift it by a constant $\delta$ without changing behavior. Finally, since on each menu the probabilities sum to 1 the term $\gamma p$ becomes a constant and similarly does not affect choice.

### 2.4 Comparative Statics

In this section we consider a pair $P_1$ and $P_2$ that are represented by APUs with common utility function $u$, and compare their perturbations. This kind of comparison is relevant in some applications: for example, in learning/evolutionary games, an analyst varies perturbation levels with a fixed utility function to analyze dynamic stability of equilibria, and different cost functions can give rise to different equilibrium selections.

We begin by comparing $P_1$ and $P_2$ at binary menus.

**Definition 9.** $P_1$ 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 each $P_1$ 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 $q \in (1/2, 1)$.

---

10 Add example here
Note that the condition in Proposition 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 non-binary menus, as in the following.

**Definition 10.** $P_1$ is more selective than $P_2$ if, for any $x, x', y_i, y_i', A_i, A'_i, i = 1, 2$, $P_1(x'|A'_1) \geq P_2(x'|A'_2)$ holds whenever

$$P_i(y_i|A_i) = P_i(y_i'|A'_i), \quad P_1(x|A_1) = P_2(x|A_2), \quad P_i(x'|\{x, x'} > \frac{1}{2}$$

for each $i = 1, 2$.

To understand this condition, here we focus on a situation where (i) $A_i$ and $A'_i$ 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 $P_1(x'|A'_1) - P_1(x|A_1) \geq P_2(x'|A'_2) - P_2(x|A_2)$.

**Assumption 1 (Joint Richness).** $A$ includes all binary menus, and for any $\bar{q} > q$, there exist menus $A_1, A'_1, A_2, A'_2$ and items $x, x', y_1, y_2$ such that

$$\bar{q} \geq P_1(x'|A'_1) > P_1(x|A_1) = P_2(x|A_2) \geq q$$

$$P_1(y_1|A_1) = P_1(y_1'|A'_1), \quad P_2(y_2|A_2) = P_2(y_2|A'_2)$$

For example, this is satisfied when $Z$ is an interval in $\mathbb{R}$, $A$ includes all menus with size 2 and 3, and $P_1, P_2$ are represented by APUs $(c_1, u), (c_2, u)$ respectively, where $u$ is continuous such that $\inf u = -\infty$ or $\sup u = \infty$.

**Proposition 3.** Assume that Joint Richness holds. 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 $q \in (0, 1)$. 

12
3 Limited Discrimination

We now use APU to model the idea that the agent has limited discrimination. Recall that the IIA axiom (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 discrimination between $x$ and $y$ will be harder in larger menus.\footnote{In a deterministic choice setting Frick (2013) extends Luce’s (1956) model of utility discrimination to capture the idea that items of similar utility are harder to distinguish in larger menus.} In other words, we expect the following axiom

**Definition 11.** $P$ satisfies *Limited Discrimination* 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

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

Limited Discrimination 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

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

If $\log c^{-1}$ is convex then the right hand side of (4) is increasing in $\lambda$. As the $\lambda$ of the weaker menu $A$ is higher than that of $B$, Limited Discrimination holds, leading 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 Limited Discrimination.

We now analyze how the agent discriminates between two rarely chosen items in large menus. 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. In order to do so, we assume that \( Z \) is infinite, and note that the equivalence of APU and Ordinal IIA holds for infinite \( Z \). On the one hand, it follows from regularity (and thus holds for an arbitrary \( h \)) that whenever \( p_n \to 1 \), we have that \( P(x|A_n)/P(y|A_n) \to P(x\{x,y\})/P(y\{x,y\}) \).\(^{12}\) On the other hand, it is not always true (even for \( h \) convex) that whenever \( p_n \to 0 \), we have \( P(x|A_n)/P(y|A_n) \to 1 \); for example this fails under IIA. The following result characterizes the class of \( h \) for which this ratio does converge to 1.

**Definition 12.** \( P \) satisfies Asymptotic Non-Discrimination 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 \( P(x|A_n)/P(y|A_n) \to 1 \).

From formula (4), asymptotic non-discrimination can be expressed as

\[
\frac{c^{-1}(u(x) + \lambda(A_n))}{c^{-1}(u(y) + \lambda(A_n))} \to 1.
\]

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 \( h = \log(c^{-1}) \).

1. If for all \( t \) the function \( h \) satisfies \( \lim_{s \to \infty} [h(t-s) - h(-s)] = 0 \), then \( P \) satisfies Asymptotic Non-Discrimination.

2. The converse is true if there exists \( \delta \in (0, \frac{1}{2}) \) such that for any \( q \in \left[ \frac{1}{2}, \frac{1}{2} + \delta \right) \) there exist \( x, y \in Z \) that satisfy the following conditions: (i) \( P(x\{x,y\}) = q \) (ii) for all sufficiently small \( p > 0 \) there exists \( A \supseteq \{x,y\} \) such that \( P(x|A) = p \).\(^{13}\)

**Example 1.** A particular class of cost functions leading to limited discrimination and asymptotic non-discrimination is the logarithmic form \( c(q) = -\eta \log(q) \). The function \( h \) is \( h(w) = \log(-\frac{w}{\eta}) \), defined on \((-\infty, -\eta)\), which is strictly convex. This also satisfies the condition for

\(^{12}\)Regularity implies that \( P(x|A_n) \leq P(x\{x,y\}) \) and likewise for \( y \), thus when \( p_n \to 1 \) we have \( P(x|A_n) \to P(x\{x,y\}) \) and likewise for \( y \).

\(^{13}\)This condition implies that the utility function convex-ranged, but not necessarily unbounded. It is not implied by and does not imply the richness condition in Fudenberg and Strzalecki (forthcoming).
Proposition 5, because \( h(r - t) - h(-t) = \log\left(\frac{r}{t-r}\right) \rightarrow 0 \) as \( t \rightarrow \infty \). 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
\[
P(x|A_n) = \frac{1}{2} \left( -n + \sqrt{4 + n^2} \right), \quad P(y_i|A_n) = \frac{1 - P(x|A_n)}{n}.
\]
The choice probability ratio \( \frac{P(x|A_n)}{P(y_1|A_n)} \) is decreasing in \( n \) and approaches to 1 as \( n \rightarrow \infty \). ▲

To relate asymptotic non-discrimination to the ambiguity representation where Nature minimizes \( \sum_{x \in A} p(x)[u(x) + \epsilon_x] + \sum_{x \in A} \phi(\epsilon_x) \), note that as \( p(x) \rightarrow 0 \) Nature will send the corresponding \( \epsilon_x \) 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 \( \lim_{s \rightarrow \infty} [h(t - s) - h(-s)] = 0 \) in Proposition 5 is equivalent to \( \lim_{s \rightarrow \infty} \frac{\phi'(s-t)}{\phi'(s)} \rightarrow 1 \), 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.

4 Beyond Positivity

So far we have assumed Positivity to simplify the exposition; however, our characterization holds more generally. This makes it possible to accommodate models that allow zero probabilities, such as Rosenthal (1989), and to cover deterministic choice data as a special case. In this section we consider weak APU, which takes the same form as APU except that \( \lim_{p \rightarrow 0} c'(p) \) is not required.

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, i.e., \( P(x|\{x, y, z\}) = P(x|\{x, y\}) = P(x|\{x, z\}) = P(y|\{y, z\}) = 1 \). 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 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
Definition 13. $P$ satisfies Weak Acyclicity if for any bijections $f, g : \{1, \ldots, n\} \rightarrow \{1, \ldots, n\}$

$$P(x_1|A_1) > P(x_{f(1)}|A_{g(1)}), \quad P(x_k|A_k) \geq^* P(x_{f(k)}|A_{g(k)}) \text{ for } 1 < k < n$$

implies $P(x_n|A_n) \not\geq^* P(x_{f(n)}|A_{g(n)})$.

We have the following generalization of our main theorem.

Theorem 2. $P$ is represented by weak APU if and only if Weak Acyclicity is satisfied, and for any $P$ whether or not this is the case can be checked in a finite number of steps.

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

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

(5)

We also modify the definition of a separable representation by

$$u(x) + \lambda(A) > u(y) + \lambda(B) \text{ if } P(x|A) > P(y|B),$$

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

(6)

Then we can show that weak APU, the existence of $(u, c, \lambda)$ 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 $(u, c, \lambda)$ that satisfy the FOC. The fact that only a finite number of steps is needed is also based on the linear programming representation.  

\footnote{See e.g. Kraft, Pratt, and Seidenberg (1959), who also note that when a solution to a linear system with rational coefficients exists in the reals, there is also a solution in the rational number.}
Like Acyclicity, Weak Acyclicity implies an order on items.

**Definition 14.** \(P\) satisfies Item Acyclicity if

\[
P(x_1|A_1) > P(x_2|A_1), \quad P(x_k|A_k) \geq^* P(x_{k+1}|A_k) \text{ for } 1 < k < n
\]

implies \(P(x_n|A_n) \not\geq^* P(x_1|A_1)\).

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 (1950)’s 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.\(^{15}\) Also, we can show that Item Acyclicity characterizes a more general form of weak APU where cost \(c\) can depend on menus.

Weak Acyclicity also implies an order on menus.

**Definition 15.** \(P\) satisfies Menu Acyclicity if

\[
P(x_1|A_1) > P(x_1|A_2), \quad P(x_k|A_k) \geq^* P(x_{k+1}|A_{k+1}) \text{ for } 1 < k < n
\]

implies \(P(x_n|A_n) \not\geq^* P(x_1|A_1)\).

**Definition 16.** \(P\) satisfies Regularity if

\[
P(x|A) \leq P(x|B) \text{ for all } A, B \in A \text{ and } x \in A \subseteq B.
\]

It is easy to see that Menu Acyclicity implies regularity so a fortiori APU are regular as well.\(^{16}\)

As we showed in an earlier version of this paper (Fudenberg, Iijima, and Strzalecki, 2014), Menu Acyclicity characterizes a more general form of weak APU where cost \(c\) can depend

\(^{15}\)Unlike SARP, congruence is defined for general menus and not just budget sets. Richter (1966) studies deterministic choice, and takes as primitive a choice correspondence that specifies a non-empty set of chosen options \(C(A) \subseteq A\) for each menu \(A\) in some collection. The congruence axiom says that if \(x \in C(A), y \in A, x_j \in C(A_j),\) 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 \(y \in C(A)\). 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.

\(^{16}\)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.
on items.\footnote{Clark (1990)’s Theorem 3 gives an incorrect characterization: The choice data $\mathcal{A} = \{(x, y), (y, z), (x, z)\}, P(x|x, y) = P(y|y, z) = P(z|x, z) = 1$ satisfies the theorem’s assumptions but does 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.} 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 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.\footnote{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 (forthcoming).}

Weak Acyclicity provides a link between rankings over items and menus. In particular, suppose $Z \in A$ so that the ranking over items is well-defined. Then, for any $x \in A, y \not\in A$, $A$ is weaker than $A \cup \{y\} \setminus \{x\}$ iff $y$ has a better rank than $x$.

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 17.** $P$ 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** \( u(x) = \max_{z \in A} u(z) \).

Heuristically, deterministic choice in our setup corresponds to a cost function that is identically equal to 0, so that we can set the Lagrange multipliers \( \lambda \) of the FOC identically equal to 0 as well, and the two-dimensional separable representation collapses to the single dimension of utility as expressed by (5).\(^{19}\)

### 5 APU as Payoff Uncertainty vs 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 (2014) find that a large majority of subjects select different options when the same 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 didn’t know which option was best, and thus didn’t want to commit to a specific choice.”\(^{20}\) Dwenger, Kubler, and Weizsacker (2014) find that subjects deliberately choose different options even when asked to make simultaneous choices from two copies of the same menu of consumption goods, and report 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.

\(^{19}\)We have required cost functions to be strictly convex, which rules out cost functions that are identically equal to 0. However, any deterministic choice data in a finite set \( Z \) that satisfied Hyper-Acyclicity can be represented by an invariant APU with a cost function that is strictly but slightly convex.

\(^{20}\)They also find that 29 percent of the subjects choose the option to flip a costly coin to randomize over items.
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).\(^{21}\)

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

\[
\inf_{\epsilon \in \mathbb{R}^A} \sum_{x \in A} p(x) [u(x) + \epsilon_x] + \sum_{x \in A} \phi(\epsilon_x),
\]

where \(\phi\) is a convex function.

The interpretation of this objective function is that Nature picks \(\epsilon = (\epsilon_x)_{x \in A}\) to minimize the agent’s expected payoff. However, it is costly for Nature to make each component of the vector \(\epsilon\) small, so it will choose to assign higher values to items that are less likely to be chosen.\(^{22}\) This gives the agent an incentive to choose non-degenerate probability distributions \(p\).\(^{23}\)

In our setting, the objective function can also be seen as a desire to avoid feeling regret about items that weren’t chosen. Here the vector \(\epsilon\) 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 \(\Phi_A(\epsilon) = \sum_{x \in A} \phi(\epsilon_x)\), where \(\phi : \mathbb{R} \to \mathbb{R} \cup \{\infty\}\) is strictly convex, continuously differentiable where it is finite-valued, with derivative whose range includes \((-1, 0)\).\(^{24}\) We will call any such function \(\phi\) a cost for Nature function. The additive form of the \(\Phi\) function is convenient for putting joint restrictions on

\(^{21}\)The item-invariant and menu-invariant classes of APU discussed below correspond to more general variational preferences, where the function \(\phi\) depends on \(z\) or \(A\).

\(^{22}\)In Maccheroni, Marinacci, and Rustichini (2006) the agent is uncertain about the probability distribution over an objective state space. In our setting, the agent is uncertain about his true utility; thus the preferences we consider here correspond to setting the space to be the possible values of \(\epsilon\).

\(^{23}\)Saito (forthcoming) studies a random choice model of an ambiguity averse agent that allows for more general timing of nature’s move. He considers choice over menus of acts, which presupposes an objective state space.

\(^{24}\)We use these last two these conditions on \(\phi\) only to ensure that \(\arg\min_{\epsilon_x} p(x)\epsilon_x + \phi(\epsilon_x)\) exists and is continuous in \(p(x) \in (0, 1)\).
choices from different menus. It can be interpreted as Nature not knowing \( u \) and hence treating each item symmetrically.

**Definition 18 (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 \( \phi \) such that

\[
P(A) = \arg \max_{p \in \Delta(A)} \left( \inf_{\epsilon \in \mathbb{R}, x \in A} p(x)[u(x) + \epsilon x] + \sum_{x \in A} \phi(\epsilon x) \right).
\]

**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 \( (u, \phi) \), then \( P \) has a weak APU representation with \( (u, c) \), where \( c(q) = \sup_{\epsilon} \{q\epsilon - \phi(-\epsilon)\} \). Conversely, if \( P \) has a weak APU representation with \( (u, c) \), then \( P \) has an AVU representation with \( (u, \phi) \), where \( \phi(\epsilon) = \sup_{q > 0} \{-\epsilon q - c(q)\} \).

2. \( P \) has an AVU representation with \( \lim_{\epsilon \to \infty} \phi'(\epsilon) = 0 \) if and only if \( P \) 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 \( \phi \) by setting \( c \) to be the convex conjugate of the function \( \hat{\phi}(\epsilon) := \phi(-\epsilon) \). The second direction constructs \( \phi \) from the cost function \( c \), by setting \( \hat{\phi} \) to be the convex conjugate of \( c \) and then setting \( \phi(\epsilon) := \hat{\phi}(-\epsilon) \). To understand the second part of the proposition, note that, by the envelope theorem, AVU implies an APU with the marginal cost \( c'(p(x)) = -\epsilon^*_x \), where \( \epsilon^*_x = \phi'^{-1}(-p(x)) \) is Nature’s optimal choice \( \epsilon_x \) against \( p(x) \in (0, 1) \). Because \( c' \) is strictly increasing, \( \epsilon^*_x \) is strictly decreasing. A steep cost \( c \) corresponds to \( \lim_{\epsilon \to \infty} \phi'(\epsilon) = 0 \) so that \( \lim_{p(x) \to 0} \epsilon^*_x = \infty \). This generates strictly positive choice probabilities because the payoff to any \( x \) diverges to \( \infty \) as its probability goes to 0. The AVU that corresponds to logit choice has \( \phi(\epsilon) = \gamma \exp(-\frac{\epsilon}{\gamma}) \). In this case, the optimal choice of Nature is \( \epsilon^*_x = -\gamma \log(p(x)) \). The AVU that corresponds to logarithmic APU has \( \phi(\epsilon) = -\eta \log(\epsilon) \) (with \( \phi(\epsilon) = \infty \) for negative \( \epsilon \)). In this case, the optimal choice of Nature is \( \epsilon^*_x = \frac{\eta}{p(x)} \).
5.2 Comparison to random utility

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

**Definition 19 (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 $\epsilon \in \mathbb{R}^Z$ such that

$$P(z|A) = \text{Prob}\{ u(z) + \epsilon_z \geq \max_{y \in A} u(y) + \epsilon_y \}$$

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

Like APU, any RU choice rule satisfies regularity.

We say that a RU is *symmetric* if the distribution of $\{\epsilon_z\}$ is exchangeable, i.e., vectors $(\epsilon_1, \ldots, \epsilon_n)$ and $(\epsilon_{\pi(1)}, \ldots, \epsilon_{\pi(n)})$ have the same distribution for any permutation $\pi$. 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.

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

**Example 3.** When $Z = \{w, x, y, z\}$, RU implies

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

We now construct an APU that violates this condition. Let $u(w) = -1$, $u(x) = 3$ and $u(y) = u(z) = 0$, and $c(p) = -\log(p)$. 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,

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

---

25 If $|Z| = 3$, then any choice rule that satisfies regularity has a RU representation (Block and Marschak, 1960), so any menu-invariant APU has a RU representation.

26 Note, that this result is different than Proposition 2.2 of Hofbauer and Sandholm (2002), where the utility function is known.
Weak stochastic transitivity can be violated by RU (Marschak, 1959), while it is satisfied by weak APU. Conversely the next example shows that some RU with i.i.d. shocks do not correspond to weak APU.

**Example 4.** 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 \( \epsilon_z \) follows i.i.d. normal distribution \( N(0,1) \) for each \( z \in Z \). Under probit the choice probabilities are

\[
P(z|A) = \int \prod_{z' \in A \setminus z} \Phi(u(z) + \epsilon_z - u(z')) \phi(\epsilon_z) d\epsilon_z
\]

where \( \Phi \) and \( \phi \) 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.

To accommodate observed choice behavior that violates IIA, the logit model has been extended to nested logit. 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” \( B_2 \). 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 is 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’ 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 $\mathbb{Q}$.\footnote{See, e.g., Kraft, Pratt, and Seidenberg (1959).} 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 $t \leq b$ whenever this inequality holds pointwise.

**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, i.e., 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 1

The following lemma generalizes the equivalence result stated in Section 2.2 by dropping the Positivity assumption.
Lemma 1. The following conditions are equivalent.

(a) There exists \((u, c)\) such that \(P\) has a weak APU representation with \((u, c)\).

(b) There exists \((u, c, \lambda)\) such that \(P\) satisfies the FOC (5) with \((u, c, \lambda)\).

(c) There exists \((u, \lambda)\) such that \(P\) satisfies (6) with \((u, \lambda)\).

Proof of Lemma 1

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

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

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

\[
\forall z \in A, \quad u(z) - c_1^z(A) + \lambda_0^z(A) = 0,
\]

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 \(\lambda\) such that \((u, \lambda)\) is a separable representation. It is without loss to assume that both take values in \((0, 1)\). Then, define

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

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

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

### A.1.3 Proof of Theorem 1

Equivalence of Acyclicity and APU follows from Theorem 2. Because of Lemma 1, it suffices to show the equivalence of Ordinal IIA and (5)

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

\[
c'(P(x|A)) - c'(P(y|A)) = c'(P(x|B)) - c'(P(y|B))
\]

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

\[
\frac{f(P(x|A))}{f(P(y|A))} = \frac{f(P(x|B))}{f(P(y|B))}
\]

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 \( f \) is strictly increasing and \( c' \) is Riemann integrable, \( c(q) = \int_0^q c'(t) dt \) is well defined. 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 pick up two items \( y, z \) from it. There are two exclusive cases.
Case (i): \( x \in \{y, z\} \). Set \( y = x \) without loss. Then

\[
\begin{align*}
u(z) - u(x) &= c'(P(z|\{x, z\})) - c'(P(x|\{x, z\})) \\
&= \log \left( \frac{f(P(z|\{x, z\}))}{f(P(x|\{x, z\}))} \right) \\
&= \log \left( \frac{f(P(z|A))}{f(P(x|A))} \right) \quad (\because \text{Ordinal IIA}) \\
&= c'(P(z|A)) - c'(P(x|A))
\end{align*}
\]

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

\[
\begin{align*}
u(z) - u(y) &= c'(P(z|\{x, z\})) - c'(P(x|\{x, z\})) - c'(P(y|\{x, y\})) + c'(P(x|\{x, y\})) \\
&= \log \left( \frac{f(P(z|\{x, z\})) f(P(x|\{x, y\}))}{f(P(x|\{x, z\})) f(P(y|\{x, y\}))} \right) \quad (\because \text{Ordinal IIA}) \\
&= \log \left( \frac{f(P(z|\{x, y, z\})) f(P(x|\{x, y, z\}))}{f(P(x|\{x, y\})) f(P(y|\{x, y\}))} \right) \quad (\because \text{Ordinal IIA}) \\
&= \log \left( \frac{f(P(z|A))}{f(P(y|A))} \right) \quad (\because \text{Ordinal IIA}) \\
&= c'(P(z|A)) - c'(P(y|A))
\end{align*}
\]

Therefore the equalities in the above two cases imply that FOC at \( A \) is satisfied. \textbf{Q.E.D.}

\section{A.1.4 \ Proof of Theorem 2}

By Lemma 1 it suffices to show the equivalence of Weak Acyclicity and a separable representation.

Suppose that there exists a separable representation \((u, \lambda)\). 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 \( Q^{\geq} \) be the vector space over the field of rational numbers whose coordinates correspond to \( \{(z, A), (z', A')\} \) with \( P(z|A) \geq^{*} P(z'|A') \). We represent a collection of relation \( \geq^{*} \) by a point in \( w \in Q^{\geq} \), where each coordinate of \( w \) counts the number of times
the corresponding relation appears.

We call a sequence \(((x_1, A_1), (x_{f(1)}, A_{g(1)})), \ldots, ((x_n, A_n), (x_{f(n)}, B_{g(n)})))\) a cycle if \(P(x_k | A_k) \geq^* P(x_{f(k)} | A_{g(k)})\) for each \(k = 1, \ldots, n\), and \(f\) and \(g\) are bijections over \(\{1, 2, \ldots, n\}\). We will denote elements of \(\{(x, A) | x \in A\} \subset A \times A\) by \(\alpha, \beta, \ldots\); each \(\alpha\) is of the form \((x_\alpha, A_\alpha)\) for some \(A_\alpha \in A\) and \(x_\alpha \in A_\alpha\).

**Step 1:** Any collection of order comparisons (allowing for potential repetitions of some statements) can be represented by a point in \(w \in \mathbb{Q}^{\geq^*}\), where each coordinate of \(w\) counts the number of times the corresponding relation appears. According to the definition, the collection is a cycle if it (a) at least one comparison is strict and (b) each item and each menu features equal number of times on each side. We will now represent these two properties geometrically.

Define \(b \in \mathbb{Q}^{\geq^*}\) 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 \(w \in \mathbb{Q}^{\geq^*}\), \(\langle w, b \rangle < 0\) iff at least one comparison in a collection of order comparisons represented by \(w\) is strict.

For each \(z \in Z\) define \(t^z \in \mathbb{Q}^{\geq^*}\) 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 $C \in A$ define $t^C \in \mathbb{Q}^\geq \ast$ 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 $\{t^z\}_{z \in Z} \cup \{t^C\}_{C \in A}$. Thus, $w \in \mathbb{Q}^\geq \ast$ 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 $w \in \mathbb{Q}^\geq \ast$ 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 exist a separable representation $(u, \lambda)$. To see that, note that if $t \in T$, then there exist rational functions $u : Z \to \mathbb{Q}$ and $\lambda : A \to \mathbb{Q}$ such that $t = \sum_{z \in Z} u(z) t^z + \sum_{C \in A} \lambda(C) t^C$. Thus, the functions $u$ and $\lambda$ are the coordinates of $t$ in $T$. Next, observe that $t \leq b$ implies that $(u, \lambda)$ is a separable representation: if $P(x|A) > P(y|B)$, then $t((x, A), (y, B)) < 0$, so $u(x) + \lambda(A) > u(y) + \lambda(B)$. If $P(x|A) = P(y|B)$, then $t((x, A), (y, B)) \leq 0$; by symmetry, $t((x, A), (y, B)) \geq 0$; thus, $u(x) + \lambda(A) = u(y) + \lambda(B)$.

Q.E.D.

### A.2 Proof of Proposition 1

**Step 1**: Fix an arbitrary $x \in Z$ and $p \in (\frac{1}{2}, 1)$ and construct the sequence $(\ldots, x_{-2}^p, x_{-1}^p, x_0^p, x_1^p, x_2^p, \ldots)$ recursively as follows. Let $x_0^p := x$. For any $x_n^p$ by Richness (i) there exists an element $y$ such that $P(y|\{x_n^p, y\}) = p$; let $x_{-n-1}^p := y$. Likewise, for any $x_{-n}^p$ by Richness (i) there exists an element $y$ such that $P(x_{-n}^p|\{x_{-n}, y\}) = p$; let $x_{n+1}^p := y$.

Suppose that $(u, c)$ and $(\hat{u}, \hat{c})$ are APU representations of $P$. Then by FOC, it follows that $u(x_{k+1}^p) - u(x_k^p) = c'(p) - c'(1 - p)$ for all $k \in Z$ and likewise $\hat{u}(x_{k+1}^p) - \hat{u}(x_k^p) = \hat{c}'(p) - \hat{c}'(1 - p)$.
for all $k \in \mathbb{Z}$. It follows that there exist $\alpha > 0, \beta \in \mathbb{R}$ such that $\hat{u}(z) = \alpha u(z) + \beta$ for all $z$ of the form $z = x_k^p$, $k \in \mathbb{Z}$.

Note that since $c'$ is a continuous function, we can find $q < p$ such that $u(x^q_2) - u(x^q_0) = u(x^p_1) - u(x^p_0)$, thus we can make the grid twice as fine. Clearly, the constants $\alpha, \beta$ that relate $\hat{u}$ and $u$ do not depend on $q$. Construct a sequence of grids indexed by $q_m$, where $q_0 = p$, $q_1 = q$, etc., where the grid corresponding to $q_{m+1}$ is twice as fine as the grid corresponding to $q_m$, i.e., $u(x^{q_{m+1}}_2) - u(x^{q_{m+1}}_0) = u(x^{q_m}_1) - u(x^{q_m}_0)$.

Fix an arbitrary $z \in \mathbb{Z}$. For each $m$ let $k_m$ be such that $u(x^{q_m}_{k_m}) \leq u(z) \leq u(x^{q_m}_{k_m+1})$. Such $k_m$ exists because for each $q$ the set $\{u(x^q_k)\}_{k \in \mathbb{Z}}$ is unbounded. Note that $\hat{u}(x^{q_m}_{k_m}) \leq \hat{u}(z) \leq \hat{u}(x^{q_m}_{k_m+1})$ because $\hat{u}$ and $u$ represent the same order over items.

Note that, since every point in the grid $q_m$ belongs to the grid $q_{m+1}$, it follows that $u(x^{q_m}_{k_m})$ is increasing in $m$ and $u(x^{q_m}_{k_m+1})$ is decreasing in $m$. Moreover $|u(x^{q_m}_{k_m+1}) - u(x^{q_m}_{k_m})| \to 0$, so
\[
\lim_{m \to \infty} u(x^{q_m}_{k_m}) = u(z) = \lim_{m \to \infty} u(x^{q_m}_{k_m+1}).
\]
Likewise $\lim_{m \to \infty} \hat{u}(x^{q_m}_{k_m}) = \hat{u}(z) = \lim_{m \to \infty} \hat{u}(x^{q_m}_{k_m+1})$.

Thus, $\hat{u}(z) = \alpha u(z) + \delta$ for all $z \in \mathbb{Z}$.

**Step 2:** Take $\bar{p}$ from condition (ii) of Richness and let $p \in (0, \frac{1}{2}]$. Richness implies that there exist $A \in \mathcal{A}$ and $z, z' \in A$ such that $P(z|A) = \bar{p}$ and $P(z'|A) = p$. By FOC it follows that $u(z) - u(z') = c'(\bar{p}) - c'(p)$ and $\hat{u}(z) - \hat{u}(z') = c'(\bar{p}) - c'(p)$. Thus, by Step 1, it follows that $c'(p) - \alpha c'(p) = c'(\bar{p}) - \alpha c'(\bar{p})$.

**Step 3:** Let $p \in (\frac{1}{2}, 1)$. Condition (i) of Richness implies that there exist $z, z'$ such that $P(z|\{z, z'\}) = p$. By FOC it follows that $u(z) - u(z') = c'(p) - c'(1 - p)$ and $\hat{u}(z) - \hat{u}(z') = c'(p) - c'(1 - p)$. Thus, by Step 1, it follows that $c'(p) - \alpha c'(p) = c'(1 - p) - \alpha c'(1 - p)$. Since $1 - p \in (0, \frac{1}{2}]$, Step 2 implies that $c'(1 - p) - \alpha c'(1 - p) = c'(\bar{p}) - \alpha c'(\bar{p})$. Thus, $c'(p) - \alpha c'(p) = c'(\bar{p}) - \alpha c'(\bar{p})$.

**Step 4:** Let $\gamma := c'(\bar{p}) - \alpha c'(\bar{p})$. By Steps 2 and 3, $c'(p) = \alpha c'(p) + \gamma$ for all $p \in (0, 1)$. Since $c$ and $c'$ are convex functions, they are absolutely continuous; hence, $c(t) = c(\bar{p}) + \int_{\bar{p}}^t c'(p)dp$ and $\hat{c}(t) = \hat{c}(\bar{p}) + \int_{\bar{p}}^t c'(p)dp$. Substituting $\hat{c}'(p) = \alpha c'(p) + \gamma$ this implies that $c(t) = \alpha c(t) + \gamma t + \delta$, where $\delta = \gamma \bar{p} + \hat{c}(\bar{p}) - \alpha c(\bar{p})$. Q.E.D.
A.3 Proof of Proposition 2

Note first that FOC and $P_i(x\{x, y\}) = 1 - P_i(x\{x, y\})$, $i = 1, 2$, imply

$$c'_i(P_1(x\{x, y\}) - c'_i(1 - P_1(x\{x, y\})) = u(x) - u(y) = c'_2(P_2(x\{x, y\})) - c'_2(1 - P_2(x\{x, y\})) \quad (9)$$

for all $\{x, y\} \in A$.

To show “if” direction, take any $x, y$ with $P_2(x\{x, y\}) \geq P_2(y\{x, y\})$, which implies $u(x) \geq u(y)$. Because each $c'_i(p) - c'_i(1 - p)$ 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. Q.E.D.

A.4 Proof of Proposition 3

For each $i = 1, 2$, let $\lambda_i(A)$ denote the Lagrange multiplier at menu under $(u, c_i)$. To show “only if” part, suppose to the contrary, i.e., $c''_1 \leq c''_2$ does not hold at some point in $(0, 1)$. By continuity of the second derivatives there exists interval $(q, \bar{q})$ such that $c''_1(q) > c''_2(q)$ for all $q \in (q, \bar{q})$. By Joint Richness, there exist $A_1, A'_1, A_2, A'_2$ and $x, x', y_1, y_2$ such that

$$\bar{q} \geq P_1(x'|A'_1) > P_1(x|A_1) = P_2(x|A_2) \geq q$$

$$P_1(x_1|A_1) = P_1(x_1|A'_1), \quad P_2(y_2|A_2) = P_2(y_1|A'_2)$$

Note that the second line ensures $\lambda_i(A_i) = \lambda_i(A'_i)$ for $i = 1, 2$. Thus, by $P_1(x'|A'_1) > P_1(x|A_1)$, $u(x') > u(x)$ holds, which ensures $P_1(x'|\{x, x'\}) > \frac{1}{2}$. Then, because $P_1$ is more selective than $P_2$, $P_1(x'|A'_1) \geq P_2(x'|A'_2)$ follows.

By FOC

$$c'_i(P_i(x|A_i)) = u(x) + \lambda_i(A_i), \quad c'_i(P_i(x'|A'_i)) = u(x') + \lambda_i(A'_i)$$
for each $i = 1, 2$, we have

$$c'_1(P_1(x'|A'_1)) - c'_1(P_1(x|A_1)) = c'_2(P_2(x'|A'_2)) - c'_2(P_2(x|A_2)),$$

or

$$\int_{P_1(x|A_1)}^{P_1(x'|A'_1)} c''_1(p)dp = \int_{P_2(x|A_2)}^{P_2(x'|A'_2)} c''_2(p)dp$$

This contradicts $P_1(x'|A'_1) \geq P_2(x'|A'_2), P_1(x|A_1) = P_2(x|A_2)$.

To show the “if”, take any $A_1, A'_1, A_2, A'_2$ and $x, x', y_1, y_2$ such that

$$P_i(y_i|A_i) = P_i(y_i|A'_i), i = 1, 2, \quad P_1(x|A_1) = P_2(x|A_2), \quad P_1(x'|{x, x'}) > \frac{1}{2}$$

As in the “only if” above, we have (10), which ensures $P_1(x'|A'_1) \geq P_2(x'|A'_2)$. Q.E.D.

### A.5 Proof of Proposition 4

Take any $A, B$ with $A \succeq_m B$ such that $x, y \in A \cap B$ and $P(x|A) > P(y|A)$. Because $A \succeq_m B$, we have both $P(x|A) \geq P(x|B)$ and $P(y|A) \geq P(y|B)$, since otherwise there would be a menu-cycle. Thus from the FOCs $u(x) + \lambda(A) = c'(P(x|A))$ of an APU with steep cost, 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

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

$$= h(u(x) + \lambda(A)) - h(u(y) + \lambda(A))$$

$$\geq h(u(x) + \lambda(B)) - h(u(y) + \lambda(B))$$

$$= \log \left( \frac{c'^{-1}(u(x) + \lambda(B))}{c'^{-1}(u(y) + \lambda(B))} \right)$$

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

where the inequality follows by the convexity of $h$, $u(x) - u(y) > 0$, and $\lambda(A) \geq \lambda(B)$. Therefore $P(x|A)/P(y|A) \geq P(x|B)/P(y|B)$, completing the proof. Q.E.D.
A.6 Proof of Proposition 5

1):

In order 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

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

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 \( h(r + \lambda_n) - h(\lambda_n) \to b \).

Choose a natural number \( k \) and construct a sequence \( (\langle h(\frac{r}{k} + \lambda_n) - h(\frac{(j-1)r}{k} + \lambda_n) \rangle_{j=1}^{k})_{n=1}^{\infty} \).

More explicitly, this sequence takes the form

\[
h(\frac{r}{k} + \lambda_1) - h(\lambda_1), h(\frac{2r}{k} + \lambda_1) - h(\frac{r}{k} + \lambda_1), ..., h(r + \lambda_1) - h(\frac{(k-1)r}{k} + \lambda_1), h(\frac{r}{k} + \lambda_2) - h(\lambda_2), ...
\]

If this sequence converges to 0, then it follows that

\[
h(r + \lambda_n) - h(\lambda_n) = \sum_{j=1}^{k} h(\frac{jr}{k} + \lambda_n) - h(\frac{(j-1)r}{k} + \lambda_n) \to 0,
\]

a contradiction. Therefore the sequence does not converge to 0. Let \( \tilde{b} > 0 \) denote its limit superior, and consider a sequence \( \{\tilde{\lambda}_n\}_{n=1}^{\infty} \) such that \( \tilde{\lambda}_n \to \infty \) and \( h(\frac{r}{k} + \tilde{\lambda}_n) - h(\tilde{\lambda}_n) \to \tilde{b} \).

Because \( k \) was an arbitrary natural number, take it large enough such that \( q \in \left[\frac{1}{2}, \frac{1}{2} + \delta\right] \), where \( q \) is uniquely defined by equation \( \frac{r}{k} = c'(q) - c'(1-q) \). By the Richness condition there
exists $x, y$ such that $P(x|\{x, y\}) = q$. Then $\frac{r}{k} = c'(q) - c'(1 - q)$ and FOCs at $\{x, y\}$ suggest that $u(x) - u(y) = \frac{r}{k}$. We assume $u(y) = 0$ without loss of generality.

Because of $\tilde{\lambda}_n \to -\infty$, we have $c^{-1}(\frac{r}{k} + \tilde{\lambda}_n) \to 0$. Therefore by the Richness condition, for all sufficiently large $n$, there exists $A_n$ such that $x, y \in A_n$ and $c^{-1}(\frac{r}{k} + \tilde{\lambda}_n) = P(x|A_n)$ hold. FOC $u(x) + \lambda(A_n) = c'(P(x|A_n))$ and $u(x) = \frac{r}{k}$ imply that $\tilde{\lambda}_n = \lambda(A_n)$. This leads to

$$\log \left( \frac{P(x|A_n)}{P(y|A_n)} \right) = \log \left( \frac{c^{-1}(\frac{r}{k} + \tilde{\lambda}_n)}{c^{-1}(\tilde{\lambda}_n)} \right) = \left[ h \left( \frac{r}{k} + \tilde{\lambda}_n \right) - h(\tilde{\lambda}_n) \right] \to b$$

as $n \to \infty$. Because $b > 0$, $\lim P(x|A_n) / P(y|A_n) > 1$ which contradicts Limited Discrimination. Q.E.D.

### A.7 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 $C : A \to Z$ by $C(A) = x$ with $P(x|A) = 1$. Then Item Acyclicity is satisfied if and only if the deterministic choice function satisfies the Congruence axiom (Richter, 1966), i.e., there is no sequence of items $x_1, x_2, \ldots, x_n$ such that

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

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

**equivalence of (1) and (3):** (3) implies (1) by definition. To see the converse direction, 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_{j2}|B_k) > P(x_{j1}|B_k) = 0. Since n is finite, we can construct a sequence P(x_{j1}|B_k) < P(x_{j2}|B_k), P(x_{j2}|B_k) < P(x_{j3}|B_k), \ldots, P(x_{j_n}|B_k) < P(x_{j1}|B_k), which contradicts Item Acyclicity.

equivalence of (2) and (3): (3) implies (2) by definition. To see the converse direction, 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 A_{j_1} = B_{k_1} for some k_1 = 1, 2, \ldots, n, P(y_{k_1}|A_{j_1}) = P(y_{k_1}|B_{k_1}) = 0. Also, as y_{k_1} = x_{j_2} for some j_2 = 1, 2, \ldots, n, P(y_{k_1}|A_{j_2}) = P(x_{j_2}|A_{j_2}) = 1. Since n is finite, we can construct a sequence P(y_{k_1}|A_{j_1}) < P(y_{k_1}|A_{j_2}), P(y_{k_2}|A_{j_2}) < P(y_{k_2}|A_{j_3}), \ldots, P(y_{k_n}|A_{j_n}) < P(y_{k_n}|A_{j_1}), which contradicts Menu Acyclicity. Q.E.D.

A.8 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 \((-1, 0) \subseteq \text{range}(\phi')\) 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:

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

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

Note that \( V_\phi(p) - V_c(p) = \sum_{z \in A} \left\{ c(p(z)) + \inf_{\epsilon_z \in \mathbb{R}} [p(z)\epsilon_z + \phi(\epsilon_z)] \right\} \). For any function \( \phi \) define the function \( \hat{\phi} \) by \( \hat{\phi}(\epsilon) = \phi(-\epsilon) \). Then \( c(p(z)) + \inf_{\epsilon_z \in \mathbb{R}} [p(z)\epsilon_z + \phi(\epsilon_z)] = c(p(z)) + \inf_{\epsilon_z \in \mathbb{R}} -[p(z)\epsilon_z - \hat{\phi}(\epsilon_z)] = c(p(z)) - \sup_{\epsilon_z \in \mathbb{R}} [p(z)\epsilon_z - \hat{\phi}(\epsilon_z)] \).

To prove the first claim, fix \( \phi \) and define \( c \) to be the convex conjugate of \( \phi \), i.e., \( c(q) = \sup_{\epsilon \in \mathbb{R}} \{ q\epsilon - \phi(\epsilon) \} \). Then \( c'(q) = \hat{\phi}'(-q) = -\phi'^{-1}(-q) \) for each \( q \in (0, 1) \) from the assumption \((-1, 0) \subseteq \text{range}(\phi')\). 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 \( \phi \) so that \( \hat{\phi} \) is the convex conjugate of \( c \), i.e., \( \phi(-\epsilon) = \hat{\phi}(\epsilon) =: \sup_{q \geq 0} \{ eq - c(q) \} \), 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 \( \lim_{\epsilon \to \infty} \phi'(\epsilon) = 0 \). Part 1 implies that \( P \) is represented by APU with \( c(\epsilon) = \sup_{\epsilon} \{ q\epsilon - \phi(-\epsilon) \} \), and since \( c'(\epsilon) = -\phi''(-\epsilon) \) for each \( q \in (0,1) \), \( \lim_{\epsilon \to 0} c'(\epsilon) = -\infty \).

Suppose that \( P \) is represented by APU with steep cost \( c \). Part (1) implies that \( P \) is represented by AVU with \( \phi(\epsilon) = \sup_{\epsilon \in (0,1]} \{ -\epsilon q - c(q) \} \), so \( \phi'(\epsilon) = -c'(\epsilon) \) , and \( \lim_{q \to 0} c'(q) = -\infty \) implies \( \lim_{\epsilon \to \infty} \phi'(\epsilon) = 0 \). Q.E.D.

**References**


