

Stochastic Choice and Revealed Perturbed Utility*

Drew Fudenberg[†] Ryota Iijima[‡] Tomasz Strzalecki[§]

July 16, 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 show that these representations correspond to a form of ambiguity-averse preferences for an agent who is uncertain about her true utility. We then provide easily understood conditions that characterize forms of 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 how to use perturbed utility to model menu-dependent choice behavior such as the attraction effect and the compromise effect.

*We thank Jerry Green, Sonia Jaffe, Kohei Kawaguchi, Mark Machina, Morgan McClellon, Wolfgang Pendorfer, 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 can be seen as following 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 non-linear perturbation function

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

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 strictly convex perturbation function that rewards the agent for randomizing. Such 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

In contrast to past work on non-linear perturbed utility functions, 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, including regularity and stochastic transitivity ([Marschak, 1959](#); [Block and Marschak, 1960](#)) as well as some new conditions that we develop here. 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.

One interpretation of representation (1) is that agents facing a decision problem randomize

¹Non-linear perturbed utility functions are also analyzed in [van Damme \(1991\)](#), [Fudenberg and Levine \(1995\)](#), [Hart and Mas-Colell \(2001\)](#), [van Damme and Weibull \(2002\)](#), [Hofbauer and Hopkins \(2005\)](#), [Hofbauer and Sandholm \(2002\)](#), [Benaim, Hofbauer, and Hopkins \(2009\)](#), [Fudenberg and Takahashi \(2011\)](#), all of which focus on repeated stochastic choice in static games. [Fudenberg and Levine \(1995\)](#) show that this generates a choice rule that is Hannan consistent, 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](#)).

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 some people may indeed have an intrinsic preference for randomization. In Section 3.1 we show that such preference for hedging may arise due to ambiguity 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 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).³

From the agent’s choice probabilities we derive a (possibly incomplete) relation \succsim over item-menu pairs (x, A) where $x \in A$: We say that $(x, A) \succ (y, B)$ if the probability of choosing x from A is greater than the probability of choosing y from B , and that $(x, A) \sim (y, B)$ if these probabilities are equal and strictly between 0 and 1.

We then relate various consistency restrictions on this relation to characterize the perturbed-utility representations. Section 3 characterizes the *invariant* case, where the cost function has the form

$$C(p) = \sum_{z \in A} c(p(z))$$

for a cost function c that is independent of both the item z and the menu A . We show that this cost function is characterized by the *Hyper-Acyclicity* axiom. This axiom rules out “item cycles,” which are cycles in the associated relation comparing pairs of items from the same menus, such as $(x, A) \succ (y, A)$, $(y, B) \succ (z, B)$, $(z, C) \succ (x, C)$. Hyper-Acyclicity also rules out “menu cycles” in the implied ranking on menus, such as $(x, A) \succ (x, B)$, $(y, B) \succ (y, C)$, $(z, C) \succ (z, A)$, but it is strictly stronger than the union of those two conditions as it also rules out “hyper-cycles” such as $(x, A) \succ (y, B)$, $(y, C) \succ (z, A)$, and $(z, B) \succ (x, C)$.

²Formally, 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 (2013).

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.

Hyper-Acyclicity implies (and is strictly stronger than) strong stochastic transitivity, which requires that if x is picked more often than y in the binary menu $\{x, y\}$, and y is picked more from the set $\{y, z\}$, then the probability that x is picked from $\{x, z\}$ is at least the maximum of those two other probabilities. Hyper-Acyclicity also implies that choice is *regular*: the probability of choosing a particular item cannot increase when more alternatives are added to the choice set.⁴

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 condition. Hyper-Acyclicity is a more general condition than IIA, as it captures the implications of maximizing the sum of expected utility and any perturbation. Specific conditions on the cost function can be used to organize alternative classes of choice rules that may be useful in applications. For instance, in section 4 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.

We also extend invariant cost functions to model nested choice, which allows us to characterize nested logit, and to model the idea of choice overload; we explore these ideas in Section 6. We then provide a revealed-preference characterization of nested logit, along with a characterization using invariant cost functions that is arguably more tractable and easier to understand than the usual random-utility representation.

Hyper-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. Section 5.2 shows that Menu Acyclicity characterizes *menu-invariant* cost functions, where the cost term c is independent of the menu, that is

$$C(p) = \sum_{z \in A} c_z(p(z))$$

Such representations always satisfy weak stochastic transitivity: if x is picked more often than y in a binary menu, and y is picked more from the set $\{y, z\}$, then x is picked more often from

⁴This is a seemingly intuitive property, but there are experimental settings where it is robustly violated, for example those associated with the “attraction effect” (Huber, Payne, and Puto, 1982; Huber and Puto, 1983).

$\{x, z\}$. They are also regular, but they need not satisfy strong stochastic transitivity. Menu-invariant costs are consistent with the “compromise effect” (Simonson, 1989)

$$P(x|\{x, y\}) > P(y|\{x, y\}), \quad P(y|\{x, y, z\}) > P(x|\{x, y, z\})$$

which has been observed in many experiments.

An alternative relaxation of Hyper-Acyclicity is *Item Acyclicity*, which rules out item cycles in \succsim . This condition is an extension to stochastic choice of the Strong Axiom of Revealed Preference and also of Richter’s (1966) congruence axiom, and ensures that choice probabilities at each menu are ordered by their utilities. Section 5.1 shows that this axiom is equivalent to *item-invariant* cost functions that can depend on the menu but not on the item z

$$C(p) = \sum_{z \in A} c_A(p(z))$$

These representations need not satisfy regularity, so they can be used to model the attraction effect and related phenomena.

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 probabilities⁵ As noted above, additive perturbed utility leads to no choice restrictions and therefore nests RU. Invariant costs rule 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. Moreover, neither menu-invariant nor

⁵Hofbauer and Sandholm (2002) show that with known utility functions and a fixed menu of alternatives any RU 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. 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.

item-invariant costs nest RU.

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 \(2014\)](#) use 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.

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.

2 Preliminaries

Let Z be a finite set of items (consequences, 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 $P : \mathcal{A} \rightarrow \Delta(Z)$ 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 .

We now associate an ordering \succsim of item-menu pairs to each stochastic choice rule P . We will use conditions on this ranking to characterize various representations of stochastic choice.

Definition 1. Let $D := \{(z, A) \in Z \times \mathcal{A} : z \in A, A \in \mathcal{A}\}$. Define binary relations over D by

$(x, A) \succ (y, B)$ if $P(x|A) > P(y|B)$, and $(x, A) \sim (y, B)$ if $P(x|A) = P(y|B) \in (0, 1)$. Let \succsim be the union of \succ and \sim .

Looking ahead, in an invariant APU $(x, A) \succ (y, B)$ will mean either that x is better than y or that menu A is weaker than B , in the sense of containing less attractive items. For this reason we do not want to interpret $P(x|A) = P(y|B) = 0$ or $P(x|A) = P(y|B) = 1$ as indifference—for example x might be better than y and yet $P(x|A) = p(Y|A) = 0$. This means that \succsim may be incomplete unless all choice probabilities are in $(0, 1)$. To relate the observed choice probabilities to the form of the cost function, we will impose various sorts of consistency conditions on \succsim .⁶ Thus, our axioms depend only on the ordinal information about the stochastic choice. An example of an axiom that relies on ordinal information only is regularity.

Definition 2 (Regularity). P satisfies *regularity* if and only if $P(x|B) \leq P(x|A)$ for all $A, B \in \mathcal{A}$ and $x \in A \subseteq B$.

In Section 5 we show that regularity is a necessary condition for a representation with an menu-invariant cost (and hence invariant costs), but it is not an implication of item-invariant costs. Moreover, regularity is not sufficient for any of these representations.

Perhaps the most familiar stochastic choice rule is logit/logistic choice, which is given by $P(z|A) = \exp(\eta u(z)) / \sum_{z' \in A} \exp(\eta u(z'))$; this corresponds to additive perturbed utility with cost $c(p) = \eta^{-1} p \log p$.⁷ As Luce (1959) showed, whenever all choice probabilities are positive logit choice is characterized by the “IIA” condition that $P(x|A)/P(y|A) = P(x|B)/P(y|B)$ for all $A, B \in \mathcal{A}$ with $x, y \in A \cap B$.⁸

The above entropy function is an example of what we call an invariant cost function, as c does not depend on z or A . Other classes of invariant cost functions in the literature include the logarithmic form used by Harsanyi (1973b) $c(q) = -\eta \log(q)$ and the quadratic perturbation $c(q) = \eta q^2$ analyzed by Rosenthal (1989).

⁶Our necessary and sufficient conditions resemble those used in work on multiattribute decision theory, but differ in a few key ways: the set \mathcal{D} does not have a product structure, and the data needs to fit the restriction $\sum_{z \in A} P(z|A) = 1$, which leads to some restrictions on \succsim which we will point out in various places below.

⁷As is well known, this corresponds to a random utility model where the shocks ε are i.i.d. Gumbel with variance η ; we discuss the relationship between APU and random utility in Section 5.3.

⁸Echenique, Saito, and Tserenjigmid (2013) generalize the Luce model by incorporating the agent’s perception priority over items; the agent perceives items sequentially by the priority order, and chooses each item stochastically.

3 Invariant Perturbations

This section studies and characterizes invariant cost functions, in which the function c depends only on the probability being evaluated and not the identity of the item in question or the other items in the menu. Using the same perturbation function makes it convenient to apply the model and at the same time imposes cross-menu consistency restrictions on choices, which we analyze later in this section. 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 $(0, 1)$. We call it *steep* if $\lim_{q \rightarrow 0} c'(q) = -\infty$.

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

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

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

As we show next, perturbed utilities of this sort can arise from the agent's ambiguity about the true utility of the various choices. We then provide two independent axiomatic characterizations of the invariance restriction. This restriction is strong, and violated in some examples, but it is a natural baseline for evaluating the consistency of stochastic choice, and it is satisfied by all of the perturbation functions that have previously appeared in the literature.

3.1 Perturbed Utility as Ambiguity Aversion

In this section we show that invariant APU can be seen as resulting from ambiguity that the agent is facing about her true utility, using the variational preferences of [Maccheroni, Marinacci, and Rustichini \(2006\)](#).⁹ An invariant APU can be also seen as a form of robustness against model misspecification, as in [Hansen and Sargent \(2008\)](#).

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 ϵ_x is an uncertain taste shock. For each probability

⁹The item-invariant and menu-invariant classes of APU discussed below correspond to more general variational preferences, where the function ϕ depends on z or A .

distribution on items $p \in \Delta(A)$ that the agent might choose, her utility is

$$\inf_{\epsilon \in \mathbb{R}^A} \sum_{x \in A} p(x)[u(x) + \epsilon_x] + \Phi_A(\epsilon), \quad (2)$$

where $\epsilon \in \mathbb{R}^A$ and for each $A \in \mathcal{A}$ the function $\Phi_A : \mathbb{R}^A \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex.¹⁰

The interpretation of this objective function is analogous to that of the smooth variational preferences of [Maccheroni, Marinacci, and Rustichini \(2006\)](#): 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 ϵ small, so it will choose to assign higher values to items that are less likely to be chosen.¹¹ This gives the agent an incentive to choose non-degenerate probability distributions p .¹²

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 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 invariant APU corresponds to the additive form $\Phi_A(\epsilon) = \sum_{x \in A} \phi(\epsilon_x)$, where $\phi : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$ is strictly convex, continuously differentiable where it is finite-valued, with derivative whose range includes $(-1, 0)$.¹³ 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 4 (Invariant Additive Variational Utility). A stochastic choice rule P has an

¹⁰Note that in general the agent’s fears about Nature can depend on the menu; this is why the function Φ_A may depend on the identities of the items of the menu A , not just their choice probabilities.

¹¹In [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 ϵ .

¹²[Saito \(2014\)](#) studies a random choice model of an ambiguity averse agent that allows for more general timing of the nature’s move. We could also consider such a generalization in which the agent believes that with probability γ Nature moves before the agent’s choice is realized. It is easy to see that the choice behavior under this model is the same as invariant APU with a rescaled cost γc , so that the same axiomatization applies. (Note that the domain of [Saito \(2014\)](#) is preference over menus of acts, which presupposes an objective state space.)

¹³We use these last two conditions on ϕ only to ensure that $\arg \min_{\epsilon_x} p(x)\epsilon_x + \phi(\epsilon_x)$ exists and is continuous in $p(x) \in (0, 1)$.

invariant *additive variational utility* (AVU) representation if and only if there exists a utility function $u : Z \rightarrow \mathbb{R}$ and a cost for Nature function ϕ such that

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

Proposition 1.

1. *P has an invariant AVU representation if and only if P has an invariant APU representation. Moreover, if P has an AVU representation with (u, ϕ) , then P has an APU representation with (u, c) , where $c(q) = \sup_{\epsilon} \{q\epsilon - \phi(-\epsilon)\}$. Conversely, if P has an APU representation with (u, c) , then P has an AVU representation with (u, ϕ) , where $\phi(\epsilon) = \sup_{q > 0} \{-\epsilon q - c(q)\}$.*
2. *P has an invariant AVU representation with $\lim_{\epsilon \rightarrow \infty} \phi'(\epsilon) = 0$ if and only if P has an invariant APU representation with a steep cost.*

The proofs of this and all other propositions are in the Appendix. Our proof of Proposition 1 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 $\hat{\phi}(\epsilon) := \phi(-\epsilon)$. The second direction constructs ϕ 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, an invariant AVU implies an invariant APU with the marginal cost $c'(p(x)) = -\epsilon_x^*$, where $\epsilon_x^* = \phi'^{-1}(-p(x))$ is Nature's optimal choice ϵ_x against $p(x) \in (0, 1)$. Because c' is strictly increasing, ϵ_x^* is strictly decreasing. A steep cost c corresponds to $\lim_{\epsilon \rightarrow \infty} \phi'(\epsilon) = 0$ so that $\lim_{p(x) \rightarrow 0} \epsilon_x^* = \infty$. 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 $\phi(\epsilon) = \gamma \exp(-\frac{\epsilon}{\gamma})$. 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 ϵ). In this case, the optimal choice of Nature is $\epsilon_x^* = \frac{\eta}{p(x)}$.

3.2 The Maximization Problem and Induced Revealed Preference Relations

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

Definition 5. A utility function u , a cost function c , and a function $\lambda : \mathcal{A} \rightarrow \mathbb{R}$ satisfy the *first order conditions* (FOC) for P iff

$$u(x) - c'(P(x|A)) + \lambda(A) \begin{cases} \geq 0 & \text{if } P(x|A) = 1 \\ = 0 & \text{if } P(x|A) \in (0, 1) \\ \leq 0 & \text{if } P(x|A) = 0. \end{cases} \quad (3)$$

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, the FOC implies that the relation \succsim associated with P is represented by (u, λ) in the following sense:

Definition 6. $u : Z \rightarrow \mathbb{R}$ and $\lambda : \mathcal{A} \rightarrow \mathbb{R}$ are a *separable representation* of \succsim if and only if

$$\begin{aligned} u(x) + \lambda(A) &> u(y) + \lambda(B) \text{ if } (x, A) \succ (y, B) \\ u(x) + \lambda(A) &= u(y) + \lambda(B) \text{ if } (x, A) \sim (y, B). \end{aligned} \quad (4)$$

In fact, as the next lemma shows, the implication can be reversed.

Lemma 1. *The following conditions are equivalent:*

1. *There exists a utility function u and a cost function c such that P has an invariant APU representation with (u, c) .*
2. *There exists a utility function u , a cost function c , and a function $\lambda : \mathcal{A} \rightarrow \mathbb{R}$ that satisfy the FOC for P .*
3. *There exists a utility function u and a function $\lambda : \mathcal{A} \rightarrow \mathbb{R}$ such that (u, λ) is a separable representation of \succsim , the relation associated with P .*

The equivalence of (1) and (2) follows from the Kuhn-Tucker theorem. That (3) follows from (2) is straightforward from the strict monotonicity of c' . To show that (2) follows from (3), note that if both P and $u + \lambda$ represent the relation \succsim , then ordinal uniqueness, modified to account for incomplete preferences, shows there exists a strictly increasing and continuous function $g : [0, 1] \rightarrow \mathbb{R}$ that satisfies $g(P(x|A)) = u(x) + \lambda(A)$ whenever $P(x|A) \in (0, 1)$. This defines a cost function by $c(p) := \int_0^p g(q) dq$, which satisfies the first order conditions.

If a relation \succsim has a separable representation (u, λ) , then the utility function u measures the relative desirability of the items: within a given menu, items with higher utility are chosen with higher probability. This motivates us to say that item x is *revealed preferred to item* y , $x \succ_i y$, if $(x, A) \succ (y, A)$ for some $A \ni x, y$. Similarly, we say that x is *revealed indifferent to* y , $x \sim_i y$, if $(x, A) \sim (y, A)$, for some $A \ni x, y$ ¹⁴ and we define a binary relation by $\succsim_i = \succ_i \cup \sim_i$. Absent any restrictions on the choice probabilities it is possible that there is a pair (x, y) such that all of $x \succ_i y$, $y \succ_i x$, and $x \sim_i y$ hold, but if the relation has a separable representation, then the relation \succ_i must be anti-symmetric and more strongly the relation \succsim_i cannot have cycles.

Axiom 1 (Item Acyclicity). There does not exist a sequence of items x_1, \dots, x_m such that

$$x_1 \succsim_i x_2 \succsim_i \dots \succsim_i x_m \succ_i x_1$$

As we will see in Section 5.1, Item Acyclicity is equivalent to the existence of a numerical representation of \succsim_i . It can be seen as an extension of Richter's (1966) congruence axiom.¹⁵

To better understand this and some of our other axioms, it is helpful to consider the classic case of deterministic (single valued) choice functions.

Definition 7. P is *deterministic* if for all $A \in \mathcal{A}$ there exists $x \in A$ such that $P(x|A) = 1$.

As we show below in Proposition 3, when P is deterministic, Item Acyclicity is equivalent to the existence of a *strict* utility function (i.e. no two items have the same utility) that rationalizes

¹⁴Recall that $(x, A) \sim (y, A)$ only if $0 < P(x|A) = P(y|A) < 1$.

¹⁵Congruence requires that if there is a cycle x_1, \dots, 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 menu A and x_1 is chosen then x_n is chosen as well. The derived representation for these choices 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.

the data.

In a separable representation (u, λ) , the multiplier λ measures the relative “weakness” of each menu. Intuitively, menu A is weaker than B if its elements compete against other items z less heavily than elements of B do. Formally, menu A is *revealed weaker than menu B*, $A \succ_m B$, if $(x, A) \succ (x, B)$ for some $x \in A \cap B$, and similarly, A is *revealed tied with B*, $A \sim_m B$, if $(x, A) \sim (x, B)$ for some $x \in A \cap B$. We define a binary relation by $\succsim_m = \succ_m \cup \sim_m$. If \succsim has a separable representation, then \succsim_m cannot have cycles.

Axiom 2 (Menu Acyclicity). There does not exist a sequence of menus A_1, \dots, A_m such that

$$A_1 \succsim_m A_2 \succsim_m \dots \succsim_m A_m \succ_m A_1.$$

In general, a relation \succsim with a separable representation may lead to \succsim_m that ranks menus in an arbitrary way. However, relations that are derived from a stochastic choice rule satisfy additional conditions, as the choice probabilities from each menu sum up to one. For instance, as we show in Section 5.2, if all choice probabilities are positive, then \succsim_m must respect set inclusion.

As with Item Acyclicity, Menu Acyclicity is also equivalent to the existence of a strict utility function when choice is deterministic, see Proposition 3, 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.¹⁶

3.3 Hyper-Acyclicity

This section presents the first of two alternative axioms that characterize invariant APU. To motivate it, we first show that separability is more restrictive than the combination of the Item Acyclicity and Menu Acyclicity menu conditions.

¹⁶The 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 (2014).

Example 1. There are three items $Z = \{x, y, z\}$, menus $A = \{y, z\}, B = \{x, z\}, C = \{x, y\}$, with¹⁷

$$(x, B) \succ (y, A), (y, Z) \sim (z, B), (z, A) \sim (x, Z).$$

Notice that the menu ranking is acyclic (A is weaker than B is weaker than C is weaker than Z), and that the item ranking is acyclic as well (x is better than both y and z , y is better than z). However, a separable representation (u, λ) would imply

$$\begin{aligned} u(x) + \lambda(B) &> u(y) + \lambda(A), \\ u(y) + \lambda(Z) &= u(z) + \lambda(B), \\ u(z) + \lambda(A) &= u(x) + \lambda(Z). \end{aligned}$$

Summing the above inequalities we obtain $0 > 0$, which is a contradiction. ▲

As this example suggests, invariant APU implies a connection between the ordering of items and the ordering of menus that is not implied by the union of Item Acyclicity and Menu Acyclicity. This is most easily seen in (4), which implies that the utility differences between items and the weakness differences between menus are comparable and summable. To capture this, the next axiom generalizes Acyclicity to rule out “hyper-cycles”:

Axiom 3 (Hyper-Acyclicity). There does not exist a *hyper-cycle*, meaning that there is no sequence

$$(x_1, A_1) \succsim (y_1, B_1), (x_2, A_2) \succsim (y_2, B_2), \dots, (x_m, A_m) \succ (y_m, B_m),$$

such that (x_1, \dots, x_m) is a permutation of (y_1, \dots, y_m) , and likewise (A_1, \dots, A_m) is a permutation of (B_1, \dots, B_m) .

Proposition 2. *Axiom 3 is satisfied if and only if P is represented by invariant APU if and only if P is represented by invariant AVU.*

¹⁷For example, this is induced by the choice probabilities $P(x|Z) = 0.475, P(y|Z) = 0.425, P(y|A) = 0.525, P(x|B) = 0.575, P(x|C) = 0.525$.

Note that this implies that invariant APU is fully characterized by relation \succsim associated with P ; the exact cardinal values of choice probabilities are inessential.¹⁸

The proof of this proposition relies on Lemma 1 and the following result.

Lemma 2. \succsim satisfies Hyper-Acyclicity iff there exist a separable representation (u, λ) of \succsim .

The proof of this lemma is based on a version of the theorem of the alternative with the following geometric interpretation. Consider the vector space \mathbb{Q}^{\succsim} whose coordinates correspond to elements of \mathcal{D}^2 , and the vector b equal to -1 on coordinates that correspond to a strict ranking and zero otherwise. Any hyper-cycle can be represented as a point w , in \mathbb{Q}^{\succsim} by setting the number of times each comparison features in the hyper-cycle equal to the corresponding coordinate of w . The permutation property of the hyper-cycle corresponds to the requirement that w is orthogonal to a certain linear subspace T , and the requirement that at least one preference relation is strict is $\langle w, b \rangle < 0$. Hyper-Acyclicity means that no such point exists, which by the theorem of alternative is equivalent to the existence of some $t \in T$ that is pointwise dominated by b , which in turn is equivalent to the existence of an additive representation (u, λ) , which functions are obtained as coefficients of t in a fixed basis of T .¹⁹

We also consider the case where all probabilities are positive.

Axiom 4 (Positivity). $P(z|A)$ 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).

Corollary 1. P is represented by invariant APU with steep cost c if and only if Hyper-Acyclicity and Positivity are satisfied.

The proof involves only minor modifications, and is omitted; analogous results hold under more general models in Section 5.1 and 5.2.

¹⁸As 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.

¹⁹The result can be seen as an incomplete preference generalization of Theorem 4.1 of Fishburn (1970), whose proof also uses a similar argument.

Finally, we consider the opposite case where choice is deterministic. In this case, Hyper-Acyclicity is equivalent to maximization of a strict utility function, and also to item- and menu-acyclicity.

Proposition 3. *Assume that P is deterministic. Then the following conditions are equivalent:*

1. *Item Acyclicity*
2. *Menu Acyclicity*
3. *Hyper-Acyclicity*
4. *There exists an injective function $u : Z \rightarrow \mathbb{R}$ s.t. $x \succ_i y$ iff $u(x) = \max_{z \in A} u(z)$ and $x, y \in A$ for some $A \in \mathcal{A}$*
5. *There exists an injective function $u : Z \rightarrow \mathbb{R}$ s.t. $A \succ_m B$ iff $\max_{z \in A} u(z) < \max_{z \in B} u(z)$ and $A \cap B \neq \emptyset$*
6. *There exists an injective function $u : Z \rightarrow \mathbb{R}$ 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 λ 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).²⁰ Because violations of rational choice in the deterministic setting correspond to cycles in \succ_i , or equivalently, in \succ_m , modeling such behavior is equivalent to characterizing the sorts of cycles that are allowed. The revealed menu weakness ranking \succ_m offers a new and potentially fruitful way to do this.

3.4 Ordinal IIA

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

²⁰We 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.

Axiom 5 (Ordinal IIA). P satisfies *ordinal IIA* if it satisfies Positivity and there is a continuous monotone transformation $f : (0, 1) \rightarrow \mathbb{R}_{++}$ such that

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

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

Proposition 4. *Assume that \mathcal{A} includes all menus with size 2 and 3.²¹ Then Ordinal IIA is satisfied if and only if P is represented by invariant APU with a steep cost if and only if P is represented by AVU with $\phi' < 0$.*

Axiom 5 requires that probabilities can be rescaled to satisfy IIA. Intuitively, invariant APU has this property because \succsim is additively separable in utility and weakness, so there is a monotone transformation f such that $f(P(z|A)) = e^{u(z)}e^{\lambda(A)}$. The converse direction uses the rescaling function f to explicitly construct an invariant cost function.

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 $f(q) = \exp(q)$, the cost is proportional to q^2 . The case $f(q) = \exp(\beta q)$ corresponds to the quadratic cost $c(q) = \beta q^2$ implicit in Rosenthal (1989).

3.5 Uniqueness

For an arbitrary set of items and menus, the invariant APU representation may not be unique, but uniqueness obtains when the range of observed behavior is rich enough.²² Intuitively, under invariant 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.²³ The following richness condition implies that the range of u equals \mathbb{R} .

Axiom 6 (Richness). Assume that (i) for any $x \in Z$, $p \in (0, 1)$ there exist $y \in Z$ such that

²¹The proposition also holds if this assumption is replaced by the assumption that $Z \in \mathcal{A}$

²²This 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).

²³A 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.

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

Richness implies that Z is infinite. The axiomatization of invariant APU in Proposition 4 holds also for infinite Z . On the other hand, the axiomatization in Proposition 2 requires Z to be finite. Note also that Richness implies that the collection A contains many binary menus and also many menus with at least three elements.

Proposition 5. *Under Richness if (u, c) and (\hat{u}, \hat{c}) represent the same invariant 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 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 a constant δ without changing behavior. Finally, since probabilities need to sum up to 1, for any menu the term γp becomes a constant and similarly does not affect choice.

4 Limited Discrimination

We now use invariant 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.²⁴ In other words, we expect the following axiom

Axiom 7 (Limited Discrimination). Assume Positivity. For all $x, y \in A \cap B$ if $A \succsim_m B$ and $P(x|A) > P(y|A)$, then

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

²⁴In 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.

This axiom means that choices from stronger menus are more uniform than choices from weaker menus. As we show in Section 5.2, $A \subseteq B$ implies $A \succsim_m B$ under invariant APU with Positivity, so that the axiom suggests that choice probabilities become flatter as we expand a menu. Note that given the FOCs of an invariant APU with a steep cost, we can express the choice probability ratios as

$$\frac{P(x|A)}{P(y|A)} = \frac{c'^{-1}(u(x) + \lambda(A))}{c'^{-1}(u(y) + \lambda(A))}. \quad (5)$$

If $\log c'^{-1}$ is convex then the right hand side of (5) is increasing in λ . As the λ of the weaker menu A is higher than that of B , Axiom 7 holds, leading to the following result.

Proposition 6. *Suppose that P is an invariant APU with a utility function u and steep cost function c . Let $h = \log(c'^{-1})$. If h is convex, then P satisfies Limited Discrimination.*

We now strengthen limited discrimination to analyze the relative choice probabilities of two very undesirable items. 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 6 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 our characterization in Proposition 4 holds for infinite Z . On the one hand, it follows from regularity (and thus holds for an arbitrary h) that whenever $p_n \rightarrow 1$, we have that $P(x|A_n)/P(y|A_n) \rightarrow P(x|\{x, y\})/P(y|\{x, y\})$.²⁵ On the other hand, it is not always true (even for h convex) that whenever $p_n \rightarrow 0$, we have $P(x|A_n)/P(y|A_n) \rightarrow 1$; for example this fails under IIA. The following result characterizes the class of h for which this ratio does converge to 1.

Axiom 8 (Asymptotic Non-Discrimination). Assume Positivity. For any sequence A_n such that $x, y \in A_n$ if $P(x|A_n) \rightarrow 0, P(y|A_n) \rightarrow 0$, then $P(x|A_n)/P(y|A_n) \rightarrow 1$.

From formula (5), asymptotic non-discrimination can be expressed as

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

²⁵Regularity implies that $P(x|A_n) \leq P(x|\{x, y\})$ and likewise for y , thus when $p_n \rightarrow 1$ we have $P(x|A_n) \rightarrow P(x|\{x, y\})$ and likewise for y .

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 7. *Suppose that P is an invariant APU with a utility function u and a steep cost function c . Let $h = \log(c'^{-1})$.*

1. *If for all t the function h satisfies $\lim_{s \rightarrow \infty} [h(t-s) - h(-s)] = 0$, then P satisfies Asymptotic Non-Discrimination.*
2. *The converse is true if there exists $\delta \in (0, \frac{1}{2})$ such that for any $q \in [\frac{1}{2}, \frac{1}{2} + \delta)$ 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 $P(x|A) = p$.²⁶*

Example 2. 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 $h(w) = \log(-\frac{\eta}{w})$, defined on $(-\infty, -\eta)$, which is strictly convex. This also satisfies the condition for Proposition 7, because $h(r-t) - h(-t) = \log(\frac{-t}{r-t}) \rightarrow 0$ as $t \rightarrow \infty$. As an illustration, consider menus of the form $A_n = \{x, y_1, \dots, y_n\}$ 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 ϵ_x to infinity. Because $c'(q) = -\phi'^{-1}(-q)$, $c'^{-1}(s) = -\phi'(-s)$, so invariant AVU implies invariant APU with $h(s) = \log(-\phi'(-s))$. Thus the condition $\lim_{s \rightarrow \infty} [h(t-s) - h(-s)] = 0$ in Proposition 7 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.

²⁶This 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 (2014).

5 Other Forms of Perturbations

Under the most general additive perturbations, a cost function can depend on each item and menu:

Definition 8 (Additive Perturbed Utility). A stochastic choice rule P has an *additive perturbed utility (APU)* representation if and only if there exists a utility function $u : Z \rightarrow \mathbb{R}$ and a collection of cost functions $c_{z,A}$ for each z, A , such that for all $A \in \mathcal{A}$,

$$P(A) = \arg \max_{p \in \Delta(A)} \sum_{z \in A} u(z)p(z) - c_{z,A}(p(z)). \quad (6)$$

APU does not restrict choice behavior.²⁷ For this reason, allowing for menu dependent utility would provide no additional generality, nor would dropping the restriction to convex cost functions; this is analogous to the [Kalai, Rubinstein, and Spiegel \(2002\)](#) “anything goes” result on menu-dependent deterministic choice.

Next we consider two non-nested, intermediate generalizations of invariant APU: item-invariant APU (Section 5.1) and menu-invariant APU (Section 5.2). They also turn out to be consistent with some empirical findings that are excluded by invariant APU. In Section 5.3, we compare our models to random utility models.

5.1 Item-Invariant Perturbations

As we have seen, invariant APU is characterized by the ordinal relation \succsim with a separable representation: the probability of choosing an item from a menu depends on the sum of the utility u of the item and a function λ of the menu. In this section we consider a model where the cost function is allowed to depend on the menu, but is invariant with respect to the items; this will generate a relation \succsim that ranks items but does not rank menus.

²⁷To understand why, set all of the utilities identically equal to 0, and choose convex cost functions such that $c'_{z,A}(P(z|A)) = 0$. (The cost functions can be taken to be convex because for each z and A the choice data only restricts $c_{z,A}$ at the single point $P(z|A)$.) Then the optimality conditions in Lemma 1 are all satisfied with Lagrange multipliers $\lambda(A)$ of 0.

Definition 9. An *item-invariant* APU has the form

$$P(A) = \arg \max_{p \in \Delta(A)} \sum_{z \in A} (u(z)p(z) - c_A(p(z)))$$

for each menu $A \in \mathcal{A}$ where $u : Z \rightarrow \mathbb{R}$ is a utility function and c_A is a cost function for each menu $A \in \mathcal{A}$.

Proposition 8. *The following conditions are equivalent.*

1. P is represented by item-invariant APU.
2. There exists a utility function u such that $u(x) > u(y)$ if $x \succ_i y$, and $u(x) = u(y)$ if $x \sim_i y$.
3. *Item Acyclicity.*

The direction (1) \Rightarrow (2) is straightforward from the first-order conditions and the strict convexity of each c_A . To show the converse (2) \Rightarrow (1), for each menu A we construct a strictly increasing function $g_A : [0, 1] \rightarrow \mathbb{R}$ that satisfies $g_A(P(x|A)) = u(x)$ whenever $P(x|A) \in (0, 1)$. This defines a cost function by $c_A(p) := \int_0^p g_A(q) dq$, which satisfies the first order conditions for the observed probabilities under $\lambda(A) = 0$. To show the equivalence of (2) and (3), we apply the following lemma to the item order \succsim_i over Z .

Lemma 3. *Let \mathcal{X} be a finite set, and take binary relations $\succ^*, \sim^* \subset \mathcal{X} \times \mathcal{X}$ where \sim^* is symmetric. Then the following conditions are equivalent.*

1. *There is no cycle meaning that there is no sequence $(\chi_1, \chi_2), (\chi_2, \chi_3), \dots, (\chi_{m-1}, \chi_m), (\chi_m, \chi_1)$ in $\succ^* \cup \sim^*$ where at least one of them belongs to \succ^* .*
2. *There exists a function $v : \mathcal{X} \rightarrow \mathbb{R}$ such that (i) $v(\chi) > v(\chi')$ if $\chi \succ^* \chi'$, and (ii) $v(\chi) = v(\chi')$ if $\chi \sim^* \chi'$.*

This lemma is similar to Lemma 2, and the proof is likewise based on a version of Farkas lemma.²⁸

²⁸As pointed out to us by Peter Wakker, this lemma can also be proved by using Richter (1966)'s theorem.

Item-invariant APU can accommodate violations of regularity (as seen in [Huber, Payne, and Puto, 1982](#); [Huber and Puto, 1983](#)). For example it might be that $P(x|A) > P(x|B)$, $B \subseteq A$, when $c''_A(p)$ is increasing in A in the set-inclusion sense.²⁹ This would occur when $c''_A(p) > c''_B(p)$, so that choice probabilities at larger menu A are less sensitive to utility differences, as then the agent's choice at A will select inferior items x more often than at B . Such item-invariant APUs, with $c''_A(p) > c''_B(p) \forall p$, correspond to item-invariant AVUs with the property $\phi''_A(\epsilon^*_A(p)) < \phi''_B(\epsilon^*_B(p)) \forall p$, where $\epsilon^*_A(p) = \phi'^{-1}_A(-p)$ and $\epsilon^*_B(p) = \phi'^{-1}_B(-p)$ are Nature's choices. This inequality means that Nature's choice $\epsilon^*_A(p)$ is more sensitive to the agent's choice p in menu A than in B , which can be interpreted as the agent facing more ambiguity at larger menu A . Item-invariant APU also allows adding items to violate regularity in a different way, with the probability of best item increasing when the menu is enlarged. This arises under the opposite condition $c''_A(p) < c''_B(p)$, as then choice is more precise on the larger menu, perhaps because the added item makes it easier for the agent to compare the items.

Item-invariant APU satisfy weak stochastic transitivity.³⁰ But it does not imply strong stochastic transitivity (or the equivalent independence condition of [Tversky and Russo \(1969\)](#)), and for example if $c''_{\{x,y\}}, c''_{\{y,z\}} < c''_{\{x,z\}}$ it is possible that $P(x|\{x,y\}), P(y|\{y,z\}) > \frac{1}{2}$, but $P(x|\{x,z\}) \leq \max\{P(x|\{x,y\}), P(y|\{y,z\})\}$. This corresponds to an AVU where the agent faces more ambiguity at $\{x,z\}$ than $\{x,y\}$ and $\{y,z\}$.³¹

To relate item-invariant APU to ambiguity aversion, we can relax invariant AVU to item-invariant AVU by allowing the function ϕ to depend on the menu. It is then an immediate corollary that Positivity and Item Acyclicity are satisfied if and only if P is represented by an item-invariant AVU.

²⁹This might arise if it is harder to spot the best items in a large list. A special case of this is where the cost function is $c_A(p) = \eta^{|A|} p \log p$, which reduces to the usual entropy representation of logistic choice when $\eta = 1$.

³⁰The converse is true when \mathcal{A} consists of all the binary menus and no others, as in [Gilboa \(1990\)](#), [Gilboa and Monderer \(1992\)](#), and [Fishburn \(1992\)](#). To see this, suppose that weak stochastic transitivity holds. There are no cycles in the item order \succsim_i with length 2 because any pair of items has only a single menu that contains them. Weak stochastic transitivity excludes cycles with length 3. Any longer cycle is excluded, as $(z_1 \succsim_i z_2, z_2 \succsim_i z_3, \dots, z_m \succ_i z_1)$ implies a shorter cycle $(z_1 \succsim_i z_2, z_2 \succsim_i z_3, \dots, z_{m-1} \succ_i z_1)$ as weak stochastic transitivity ensures $z_{m-1} \succsim_i z_1$ under this binary domain.

³¹As noted by [Mellers and Biagini \(1994\)](#) and [Rieskamp, Busemeyer, and Mellers \(2006\)](#), violations typically occur when attributes of x and z are harder to compare than other pairs, which might be interpreted as a source of ambiguity.

5.2 Menu-Invariant Perturbations

An alternative way to generalize invariant APU is to require that the cost of controlling the probability of each item depends on the item, but is invariant with respect to the menu. This will generate a consistent “weakness ranking” \succsim_m of menus without generating a fixed utility function on items.

Definition 10 (Menu-invariant APU). A menu-invariant APU has the form

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

such that $u : Z \rightarrow \mathbb{R}$ is an utility function and c_z is a cost function for each item $z \in Z$.

Proposition 9. *The following conditions are equivalent.*

1. P is represented by menu-invariant APU.
2. There exists a function $\lambda : \mathcal{A} \rightarrow \mathbb{R}$ such that $\lambda(A) > \lambda(B)$ if $A \succ_m B$, and $\lambda(A) = \lambda(B)$ if $A \sim_m B$.
3. Menu Acyclicity.

The proof is analogous to that of Proposition 8 for item-invariant APU.³² As with the equivalence of item-invariant APU and item-invariant AVU, it is immediate that Menu Acyclicity is satisfied if and only if P is represented by a menu-invariant AVU, which allows the function ϕ to depend on each item.

Note that Menu Acyclicity implies regularity, so menu-invariant APU satisfies regularity.³³ Menu-invariant APU satisfies weak stochastic transitivity, but can violate strong stochastic transitivity.

³²Clark (1990) proposed an axiomatization of this representation but his Theorem 3 is not correct: 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 a menu-invariant representation. However, his Theorem 3 is correct under the additional assumption of positivity, as then its conditions are equivalent to Menu Acyclicity.

³³To see this, assume toward contradiction that there exists $z \in A \subseteq B$ such that $P(z|A) < P(z|B)$. Then we must have $P(z'|A) < P(z'|B)$ for all $z' \in A$ with $P(z'|A) > 0$ as otherwise there would be the menu cycle $(z', A) \succ (z', B)$ and $(z, B) \succ (z, A)$. This implies $1 = \sum_{z' \in A} P(z'|A) < \sum_{z' \in A} P(z'|B) \leq \sum_{z' \in B} P(z'|B) = 1$, which is a contradiction. Note that, under Positivity, this ensures $A \subseteq B \Rightarrow A \succsim_m B$.

Note also that menu-invariant APU can violate Item Acyclicity. In particular it is consistent with

$$P(x|\{x, y\}) > P(y|\{x, y\}), \quad P(y|\{x, y, z\}) > P(x|\{x, y, z\}),$$

which can arise from the “compromise effect” (Simonson, 1989). This and other such probability reversals have been observed in many experiments (e.g. Tversky and Simonson (1993), and Section 4 of Rieskamp, Busemeyer, and Mellers (2006)). These reversals can occur when $c''_y(p) > c''_x(p)$, so that the choice probability of y is less sensitive, which corresponds to menu-invariant AVU with $\phi'_y(\epsilon^*(p)) < \phi'_x(\epsilon^*(p))$, so that the agent faces more ambiguity at y than x .

Finally, consider a condition due to Tversky (1972) that connects the relation \succsim over items and the one over menus.

Axiom 9 (Order Independence). For all $A, B \in \mathcal{A}$, $x, y \in A \setminus B$ and $z \in B$, $P(z|B \cup \{x\}) \leq P(z|B \cup \{y\})$ iff $P(x|A) \geq P(y|A)$.

This condition is implied by the combination of Hyper-Acyclicity and Positivity, but it is weaker than Hyper-Acyclicity, as it is also satisfied by Example 1. Tversky (1972) shows that when \mathcal{A} includes every menu this axiom characterizes “simple scalability;” it also implies Item Acyclicity. On the other hand, the axiom is not implied by the combination of Item Acyclicity and Menu Acyclicity.³⁴ We now characterize the additional restriction on menu-invariant APU with steep costs imposed by Order Independence. A menu-invariant APU is *weakly item-dependent* if $u(x) - c'_x(q) \geq u(y) - c'_y(q)$ holds for all $x, y \in Z$ such that $u(x) \geq u(y)$ and $q \in (0, 1]$, where the inequality is strict if $u(x) > u(y)$.³⁵ The condition guarantees that the perturbation terms $c'_x(q), c'_y(p)$ do not offset the difference between $u(x)$ and $u(y)$.

Proposition 10. *Assume that \mathcal{A} includes all the subsets of Z . Then Positivity, Menu Acyclicity, and Order Independence are satisfied if and only if P is represented by menu-invariant APU with steep costs that is weakly item-dependent.*

³⁴To see this, consider $Z = \{x, y, z\}$ and the following data $P(x|\{x, y\}) = 0.5$, $P(x|\{x, z\}) = 0.6$, $P(y|\{y, z\}) = 0.7$, $P(x|\{x, y, z\}) = P(y|\{x, y, z\}) = 0.4$. This satisfies both Item Acyclicity and Menu Acyclicity, and generates well-defined orders $x \sim^* y \succ^* z$ and $\{x, z\} \succ^* \{y, z\} \succ^* \{x, y\} \succ^* \{x, y, z\}$. But Order Independence is violated at $P(z|\{x, z\}) > P(z|\{y, z\})$ and $P(x|\{x, y, z\}) = P(y|\{x, y, z\})$.

³⁵The condition can be written as $u(x) + \epsilon_x^*(q) \geq u(y) + \epsilon_y^*(q)$ in terms of the corresponding menu-invariant AVU.

5.3 Comparison to Random Utility

We now compare various sorts of APU to random utility models.

Definition 11 (Random Utility). A stochastic choice rule P has a *random utility (RU)* representation if and only if there exists a utility function $u : Z \rightarrow \mathbb{R}$ and a random variable $\epsilon \in \mathbb{R}^Z$ such that

$$P(z|A) = \text{Prob}\{u(z) + \epsilon_z \geq \max_{y \in A} u(y) + \epsilon_y\} \quad (7)$$

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

We say that a RU is *symmetric* if the distribution of $\{\epsilon_z\}$ is exchangeable, i.e., vectors $(\epsilon_1, \dots, \epsilon_n)$ and $(\epsilon_{\pi(1)}, \dots, \epsilon_{\pi(n)})$ have the same distribution for any permutation π . Any symmetric RU satisfies Axiom 1 and 8, so item-invariant APU 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, some invariant APU models violate the Block-Marshak conditions (Block and Marschak, 1960) that are necessary for RU.

Example 3. When $Z = \{w, x, y, z\}$, RU implies³⁶

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

We now construct an invariant 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,

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

Weak stochastic transitivity can be violated by RU (Marschak, 1959), while it is satisfied by item-invariant APU, so item-invariant APU does not nest RU. The next example shows that some RU with i.i.d. shocks do not correspond to menu-invariant APU.

³⁶If $|Z| = 3$, then any choice rule that satisfies regularity has a RU representation (Block and Marschak, 1960), so any menu-invariant APU has a RU representation.

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

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

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 menu-invariant APU. ▲

6 Nested Choice and Menu-Size Penalties

Now we develop a model of sequential or “nested” choice, in which the agent first picks a “nest” and then picks an item from the nest. We use these nests to allow the agent to have different levels of ambiguity about different sets of items, while still maintaining much of the structure of the APU/AVU representations. This model of nested choice includes nested logit as a special case; it also lets us model the idea of “choice overload.”³⁷ We develop these representations in the text; Appendix A.15.2 provides an axiomatic characterization that generalizes the IIA characterization of logit choice.

Let $\{B_k\}_{k=1}^K$ be a partition of Z ; we will refer to each B_k as a nest. Like before, a stochastic choice rule P is a choice distribution over items in each menu $A \subseteq Z$, so that $\sum_{z \in A} P(z|A) = 1$. Given a full-support distribution P , let P_1 denote the induced choice rule over nests, $P_1(B_k|A) := \sum_{z \in B_k \cap A} P(z|A)$, and let $P_2(\cdot|B_k, A)$ be the observed conditional distribution of choices from B_k when the menu is A , $P_2(z|B_k, A) := P(z|A)/P_1(B_k|A)$.

³⁷Gul, Natenzon, and Pesendorfer (forthcoming) introduce “attribute rules” which are related to nested choice. The important differences are that attributes are endogenous, whereas nests are exogenous and at the same time the probability of choosing an attribute does not depend on how many items with that attribute there are in a menu, whereas the probability of choosing a nest is menu-dependent.

Our nested representations allow the agent to feel differently about menus of different sizes; this is parameterized by the *menu size penalty* $\alpha \geq 0$. Define $C_\alpha : \text{int}\Delta(A) \rightarrow \mathbb{R}$ the *entropy cost with menu-size penalty* α by

$$C_\alpha(p) = \sum_{z \in A} p(z) \log p(z) + \alpha \log |A|.$$

Note that this reduces to entropy when $\alpha = 0$ and relative entropy with respect to the uniform distribution when $\alpha = 1$. As is well known, entropy costs (i.e. logit choice) implies the agent always prefers to add an item to a menu. In contrast, the value function for menus induced by relative entropy with respect to the uniform distribution decreases when a sufficiently undesirable item is added.³⁸

Definition 12. P is called an α -nested entropy model if there exist utility function $u : Z \rightarrow \mathbb{R}$, constant $\alpha \geq 0$, and positive coefficients $\{\eta_k\}_{k=1}^K$, such that for each B_k , $z \in B_k$ and $A \subseteq Z$

$$P(z|A) = P_1(B_k|A)P_2(z|B_k, A)$$

where

$$\begin{aligned} (P_1(B_1|A), \dots, P_1(B_K|A)) &= \arg \max_{p \in \Delta(\{B_1, \dots, B_K\})} \sum_k p(B_k)U(B_k \cap A) - C_0(p), \\ P_2(B_k, A) &= \arg \max_{p \in \Delta(B_k \cap A)} \sum_{z \in B_k \cap A} p(z)u(z) - \eta_k C_\alpha(p) \\ U(B_k \cap A) &= \max_{p \in \Delta(B_k \cap A)} \sum_{z \in B_k \cap A} p(z)u(z) - \eta_k C_\alpha(p) \end{aligned}$$

and we set $U(B_k \cap A) = -\infty$ if $B_k \cap A = \emptyset$. In particular, P is called a *nested entropy model* if $\alpha = 0$.

In this model, the agent first chooses a probability distribution on nests with associated entropy cost C_0 , and then an item from the chosen nest B_k with the cost function C_α multiplied by η_k ; $U(B_k \cap A)$ is the indirect payoff from items $B_k \cap A$. Of course this nested procedure corresponds to a one-stage choice with an APU where cost may depend on both item and menu;

³⁸Fudenberg and Strzalecki (2014) provide an axiomatic foundation for menu-size penalties in a setting of recursive stochastic choice.

the nested form is a convenient way to capture the idea that some goods are substitutes for others. Note also that we could in principle generalize α -nested entropy by replacing entropy with some other invariant APU; the main stumbling block in pursuing that route is the lack of a closed-form solution for the value function U .

As we show in Appendix A.15.1, the case $\alpha = 0$ reduces to the nested-logit model used in empirical work; the axiomatization in Appendix A.15.2 provides a characterization in terms of observed behavior, and the corresponding AVU provides its ambiguity interpretation.

6.1 Choice overload

Next we consider a special kind of nested model with only two nests, a "default option" $B_1 = \{x^*\}$ and "everything else" B_2 . As above, we suppose that choice within the "everything else" nest satisfies IIA, but we will consider more general definitions of the indirect utility that allow for the possibility that the agent is more likely to choose the default when faced with a larger choice set. (Note that the nested entropy model of the previous section always prefers larger choice sets: if any item is added to the "everything else" set the probability of the default decreases.) Once we consider nested choice processes, the value function on the nest has implications for observed choice, as changing the value of a nest changes its attractiveness without changing the conditional probabilities within the nest. For example, suppose that we specify cost function $C_\alpha(p)$ where $\alpha > 1$. Then the value of items $D \subseteq B_2$ is

$$\eta_2 \left(\log \sum_{z \in D} \exp[u(z)/\eta_2] - \alpha \log |D| \right) = \eta_2 \log \frac{\sum_{z \in D} \exp[u(z)/\eta_2]}{|D|^\alpha}.^{39}$$

Hence unlike in the nested logit/entropy case, the agent is more likely to pick the default x^* when faced with n alternatives from B_2 , each of which has the same utility, than when faced with $m < n$ of them. This is consistent with evidence that consumers are less likely to purchase when confronted with a large number of choices (e.g., Iyengar and Lepper, 2000). Here the parameter α captures the effect of choice overload: if an agent with utility function u and parameters (η_2, α) prefers the default to a menu D , and $\alpha' > \alpha$, then so does an agent with the

³⁹Note that the value of η_1 is irrelevant when nest 1 has a single item.

same utility function and parameters (η_2, α') . Moreover, for any other menu D' with $|D'| > |D|$ the α' agent also prefers the default option whenever the α agent does. The literature already provides several explanations for choice overload: it can come from search or introspection costs (Kuksov and Villas-Boas, 2010; Ortoleva, 2013), or the contextual signals provided by the menu (Kamenica, 2008). Our approach provides an alternative explanation based on the agent's ambiguity aversion about her true utility. The variational preferences that correspond to C_α have $\phi_A(x) = |A|^{-\alpha} e^{1-x}$, which means that the agent is more uncertain about the utility from larger menus, while his uncertainty about the default option is held constant, so that it becomes relatively more appealing as the menu grows.

7 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 ambiguity 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, and extensions to nested models such as nested logit and cognitive overload, which may prove useful both in dynamic choice theory and in applications to game theory and industrial organization. There is already some empirical evidence for cognitive overload, but as noted by Chernev (2012), there has been relatively little research on the related issue of the impact of menu size on the option chosen. We hope that the analytic foundations provided here will stimulate further empirical work.

Appendix

A.1 Preliminary Lemma

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} .⁴⁰ 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.2 Proof of Proposition 1

Proof of (1): For any $\phi : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$ 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 $V_\phi : \Delta(A) \rightarrow \mathbb{R} \cup \{\infty\}$ and

⁴⁰See, e.g., Dax (1997).

$V_c : \Delta(A) \rightarrow \mathbb{R} \cup \{\infty\}$ 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 ϕ 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 ϕ and define c to be the convex conjugate of $\hat{\phi}$, i.e., $c(q) = \sup_{\epsilon} \{q\epsilon - \hat{\phi}(\epsilon)\}$. 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 ϕ so that $\hat{\phi}$ is the convex conjugate of c , i.e., $\phi(-\epsilon) = \hat{\phi}(\epsilon) =: \sup_{q>0} \{\epsilon q - c(q)\}$, 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 invariant AVU with $\lim_{\epsilon \rightarrow \infty} \phi'(\epsilon) = 0$. Part 1 implies that P is represented by invariant APU with $c(q) = \sup_{\epsilon} \{q\epsilon - \phi(-\epsilon)\}$, and since $c'(q) = -\phi'^{-1}(-q)$ for each $q \in (0, 1)$, $\lim_{q \rightarrow 0} c'(q) = -\infty$.

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

A.3 Proof of Lemma 1

equivalence of (1) and (2): 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

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

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 $P(x|A) = 0$.

(2) implies (3): Condition (4) holds because c'^{-1} is strictly increasing.

(3) implies (2): Suppose that there exist u and λ such that condition (4) holds. It is without loss to assume that both take values in $(0, 1)$. Then, define

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

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

Let $g : [0, 1] \rightarrow \mathbb{R}$ 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) $g(1) = \bar{w}$. Such function exists because condition (4) and the definition of \succsim ensure that $u(x) + \lambda(A) > u(y) + \lambda(B)$ 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 the optimality condition in Lemma 1 is satisfied at each menu, so that P is represented by invariant APU. **Q.E.D.**

A.4 Proof of Lemma 2

Suppose that there exists a separable representation (u, λ) . There cannot exist a hyper-cycle, as if there is one, 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 note that since \succsim is a binary relation over \mathcal{D} , it is a subset of \mathcal{D}^2 . We will denote elements of \mathcal{D} by $\alpha, \beta, \text{etc.}$; each $\alpha \in \mathcal{D}$ is of the form (x_α, A_α) for some $A_\alpha \in \mathcal{A}$ and $x_\alpha \in A_\alpha$. Consider the vector space \mathbb{Q}^{\succsim} over the field of rational numbers.

Step 1: Any collection of preference statements (allowing for potential repetitions of some statements) can be represented by a point in $w \in \mathbb{N}^{\succsim}$ (each coordinate of this point counts the number of repetitions). According to the definition, the collection is a hyper-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 $b \in \mathbb{Q}^{\succsim}$ as follows:

$$b(\alpha, \beta) = \begin{cases} -1 & \text{if } (\alpha, \beta) \in \succ \\ 0 & \text{if } (\alpha, \beta) \in \sim . \end{cases}$$

Note that for any $w \in \mathbb{N}^{\succsim}$, $\langle w, b \rangle < 0$ iff at least one comparison in a collection of preference statements represented by w is strict.

For each $z \in Z$ define $t^z \in \mathbb{Q}^{\succsim}$ 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 for any $w \in \mathbb{N}^{\succsim}$, $\langle w, t^z \rangle = 0$ iff z features equal number of times on each side of the hyper-cycle represented by w .

For each $C \in \mathcal{A}$ define $t^C \in \mathbb{Q}^{\succsim}$ by

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

Note that for any $w \in \mathbb{N}^{\succsim}$, $\langle w, t^C \rangle = 0$ 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, $w \in \mathbb{N}^{\succsim}$ represents a hyper-cycle if and only if $w \perp T$ and $\langle w, b \rangle < 0$.

Step 2: Since Hyper-Acyclicity implies that there does not exist w that meets the conditions of Step 1, there cannot exist $w \in \mathbb{Q}^{\succsim}$ 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, λ) of \succsim . To see that, note that if $t \in T$, then there exist rational functions $u : Z \rightarrow \mathbb{Q}$ and $\lambda : \mathcal{A} \rightarrow \mathbb{Q}$ 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) represent \succsim : if $(x, A) \succ (y, B)$, then $t((x, A), (y, B)) < 0$, so $u(x) + \lambda(A) > u(y) + \lambda(B)$. If $(x, A) \sim (y, B)$, then $t((x, A), (y, B)) \leq 0$; by the symmetry of \sim , $t((x, A), (y, B)) \geq 0$; thus, $u(x) + \lambda(A) = u(y) + \lambda(B)$. **Q.E.D.**

A.5 Proof of Proposition 2

Follows from Proposition 1 and Lemmas 1 and 2. **Q.E.D.**

A.6 Proof of Lemma 3

The direction (2) \Rightarrow (1) is straightforward, so we only show (1) \Rightarrow (2). Note first that $(\xi', \xi'') \in \succ^*$ implies $(\xi', \xi'') \notin \sim^*$, because otherwise the symmetry of \sim^* leads to a cycle $\xi'^* \succ^* \xi'' \sim^* \xi'$. Let $\succsim^* := \succ^* \cup \sim^*$, which is a subset of \mathcal{X}^2 .

Step 1: Any sequence of preference statements (allowing for potential repetitions of some statements) can be represented by a point in $w \in \mathbb{N}^{\succsim^*}$ (each coordinate of this point counts the number of repetitions).

Define $b \in \mathbb{Q}^{\succsim^*}$ as follows:

$$b(\xi', \xi'') = \begin{cases} -1 & \text{if } (\xi', \xi'') \in \succ^* \\ 0 & \text{if } (\xi', \xi'') \in \sim^* . \end{cases}$$

Note that for any $w \in \mathbb{N}^{\succsim^*}$, $\langle w, b \rangle < 0$ iff at least one comparison in the sequence represented by w is strict.

For each $\xi \in \mathcal{X}$ define $t^\xi \in \mathbb{Q}^{\succsim^*}$ by

$$t^\xi(\xi', \xi'') = \begin{cases} -1 & \text{if } \xi' = \xi \text{ and } \xi'' \neq \xi \\ 1 & \text{if } \xi' \neq \xi \text{ and } \xi'' = \xi \\ 0 & \text{otherwise.} \end{cases}$$

Note that for any $w \in \mathbb{N}^{\succsim^*}$, $\langle w, t^\xi \rangle = 0$ iff ξ features equal number of times on each side of the sequence represented by w .

Let T be the linear subspace generated by the collection $\{t^\xi\}_{\xi \in \mathcal{X}}$. Thus, $w \in \mathbb{N}^{\succsim^*}$ represents a cycle if and only if $w \perp T$ and $\langle w, b \rangle < 0$.

Step 2: Since the condition (1) implies that there does not exist w that meets the conditions of Step 1, there cannot exist $w \in \mathbb{Q}^{\succsim^*}$ 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 representation v of \succsim^* . To see that, note that if $t \in T$, then there exists a rational function $v : \mathcal{X} \rightarrow \mathbb{Q}$ such that $t = \sum_{\xi \in \mathcal{X}} v(\xi)t^\xi$. Thus, the function v is the coordinates of t in T . Next, observe that $t \leq b$ implies that v represents \succsim^* : if $\xi' \succ^* \xi''$, then $t(\xi', \xi'') < 0$, so $v(\xi') > v(\xi'')$. If $\xi' \sim^* \xi''$ then $t(\xi', \xi'') \leq 0$; by the symmetry of \sim^* , $t(\xi', \xi'') \geq 0$; thus, $v(\xi') = v(\xi'')$. **Q.E.D.**

A.7 Proof of Proposition 3

Note that when P is deterministic, $(x, A) \succsim (y, B)$ iff $(x, A) \succ (y, B)$ iff $P(x|A) = 1$ and $P(y|B) = 0$.

equivalence of (2) and (6): When P is deterministic, it induces a deterministic and single-valued choice function. Then Axiom 1 is satisfied if and only if the deterministic choice function satisfies the Congruence axiom (Richter, 1966). Because Z is finite, this is equivalent to the existence of a strict utility function u that rationalizes the choice function.

equivalence of (1) and (3): (3) implies (1) by definition. To see the converse direction, suppose Item Acyclicity is satisfied and there is a hyper-cycle, i.e., $(x_1, A_1) \succ (y_1, B_1)$, $(x_2, A_2) \succ (y_2, B_2)$, \dots , $(x_m, A_m) \succ (y_m, B_m)$. Suppose that (x, A) is on the left hand side; since x must be featured on the right hand side, this means that there exists a menu B such that $P(x|B) = 0$. Since B must be featured on the left hand side associated with some item x' , i.e., $P(x'|B) = 1$, we have $x' \succ_i x$. We can continue the same argument to show that, since the set Z is finite, the relation \succ_i over Z has a cycle, a contradiction.

equivalence of (2) and (3): (3) implies (2) by definition. To see the converse direction, suppose Menu Acyclicity is satisfied and there is a hyper-cycle, i.e., $(x_1, A_1) \succ (y_1, B_1)$, $(x_2, A_2) \succ (y_2, B_2)$, \dots , $(x_m, A_m) \succ (y_m, B_m)$. Suppose that (x, A) is on the left hand side; since A must be featured on the right hand side, this means that there exists an item $y \in A$ such that $P(y|A) = 0$. Since y must be featured on the left hand side associated with some menu A' , i.e., $P(y|A') = 1$, we have $A' \succ_m A$. We can continue the same argument to show that, since the set \mathcal{A} is finite, the relation \succ_m over \mathcal{A} has a cycle, a contradiction.

equivalence of (4) and (6): This is straightforward from the definitions.

(6) implies (5): Suppose there is an injective function u such that $P(x|A) = 1$ iff $u(x) = \max_{z \in A} u(z)$. Take any $A, B \in \mathcal{A}$. Then $A \succ_m B$ iff there is $x \in A \cap B$ such that $P(x|A) = 1 > 0 = P(x|B)$, which is equivalent to $\{x\} = \arg \max_{z \in A} u(z)$ and $x \notin \arg \max_{z \in B} u(z)$.

(5) implies (3): Suppose there is an injective function u s.t. $A \succ_m B$ iff $\max_{z \in A} u(z) < \max_{z \in B} u(z)$ and $A \cap B \neq \emptyset$. If Menu Acyclicity is violated, there exists a sequence $A_1 \succ_m A_2 \succ_m \dots \succ_m A_m \succ_m A_1$. This implies $\max_{z \in A_1} u(z) < \max_{z \in A_2} u(z) < \dots < \max_{z \in A_m} u(z) < \max_{z \in A_1} u(z)$, a contradiction. **Q.E.D.**

A.8 Proof of Proposition 4

Because of Proposition 1, we only need to show that Positivity and Ordinal IIA are satisfied if and only if P is represented by invariant APU with a steep cost.

“**If**”: Positivity is clearly satisfied by the assumption $\lim_{q \rightarrow 0} c'(q) = -\infty$. Note that with a steep cost c the solution to the FOC is always interior:

$$u(x) + \lambda(A) - c'(P(x|A)) = 0$$

where $\lambda(A) \in \mathbb{R}$ is the Lagrange multiplier associated with the constraint $\sum_{x \in A} p(x) = 1$ in the optimization problem. This implies

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

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

“**Only if**”: Construct a cost function by setting $c'(q) := \log(f(q) - \lim_{r \rightarrow 0} f(r))$ 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 $\lim_{q \rightarrow 0} c'(q) = -\infty$. 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

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

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

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

Therefore the equalities in the above two cases imply that FOC at A is satisfied. Since this holds for any menu A , this proves that P is represented by an invariant APU with steep cost c . **Q.E.D.**

A.9 Proof of Proposition 5

Step 1: Fix an arbitrary $x \in Z$ and $p \in (\frac{1}{2}, 1)$ and construct the sequence $(\dots, x_{-2}^p, x_{-1}^p, x_0^p, x_1^p, x_2^p, \dots)$ 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 $P(x_{-n}^p|\{x_{-n}^p, y\}) = p$; let $x_{n+1}^p := y$.

Suppose that (u, c) and (\hat{u}, \hat{c}) are invariant 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 $k \in \mathbb{Z}$ and likewise $\hat{u}(x_{k+1}^p) - \hat{u}(x_k^p) = \check{c}'(p) - \check{c}'(1-p)$ 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 $u(x_2^q) - u(x_0^q) = u(x_1^p) - u(x_0^p)$, thus we can make the grid twice as fine. Clearly, the constants α, β 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., $u(x_2^{q_{m+1}}) - u(x_0^{q_{m+1}}) = u(x_1^{q_m}) - u(x_0^{q_m})$.

Fix an arbitrary $z \in Z$. For each m let k_m be such that $u(x_{k_m}^{q_m}) \leq u(z) \leq u(x_{k_m+1}^{q_m})$. Such k_m exists because for each q the set $\{u(x_k^q)\}_{k \in \mathbb{Z}}$ is unbounded. Note that $\hat{u}(x_{k_m}^{q_m}) \leq \hat{u}(z) \leq \hat{u}(x_{k_m+1}^{q_m})$ 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 $u(x_{k_m}^{q_m})$ is increasing in m and $u(x_{k_m+1}^{q_m})$ is decreasing in m . Moreover $|u(x_{k_m+1}^{q_m}) - u(x_{k_m}^{q_m})| \rightarrow 0$, so $\lim_{m \rightarrow \infty} u(x_{k_m}^{q_m}) = u(z) = \lim_{m \rightarrow \infty} u(x_{k_m+1}^{q_m})$. Likewise $\lim_{m \rightarrow \infty} \hat{u}(x_{k_m}^{q_m}) = \hat{u}(z) = \lim_{m \rightarrow \infty} \hat{u}(x_{k_m+1}^{q_m})$. Thus, $\hat{u}(z) = \alpha u(z) + \delta$ for all $z \in Z$.

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 $\hat{u}(z) - \hat{u}(z') = \check{c}'(\bar{p}) - \check{c}'(p)$. Thus, by Step 1, it follows that $\check{c}'(p) - \alpha c'(p) = \check{c}'(\bar{p}) - \alpha c'(\bar{p})$.

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 $\hat{u}(z) - \hat{u}(z') = \check{c}'(p) - \check{c}'(1-p)$. Thus, by Step 1, it follows that $\check{c}'(p) - \alpha c'(p) = \check{c}'(1-p) - \alpha c'(1-p)$. Since $1-p \in (0, \frac{1}{2}]$, Step 2 implies that $\check{c}'(1-p) - \alpha c'(1-p) = \check{c}'(\bar{p}) - \alpha c'(\bar{p})$. Thus, $\check{c}'(p) - \alpha c'(p) = \check{c}'(\bar{p}) - \alpha c'(\bar{p})$.

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

A.10 Proof of Proposition 6

Take any A, B with $A \succsim_m B$ such that $x, y \in A \cap B$ and $P(x|A) > P(y|A)$. Because $A \succsim_m B$, 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 invariant APU with steep cost, it follows that $\lambda(A) \geq \lambda(B)$ and $u(x) > u(y)$. Using these FOCs, we can express the log-ratio of choice probabilities as

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

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.11 Proof of Proposition 7

1):

In order to show Axiom 8, 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 $\lim_{q \rightarrow 0} c'(q) = -\infty$, $P(x|A_n) \rightarrow 0$ implies $\lambda(A_n) \rightarrow -\infty$. Therefore

$$\begin{aligned} \log \left(\frac{P(x|A_n)}{P(y|A_n)} \right) &= \log \left(\frac{c'^{-1}(u(x) + \lambda(A_n))}{c'^{-1}(u(y) + \lambda(A_n))} \right) \\ &= h(u(x) + \lambda(A_n)) - h(u(y) + \lambda(A_n)) \\ &\rightarrow 0 \end{aligned}$$

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

2):

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

Choose a natural number k and construct a sequence $((h(\frac{jr}{k} + \lambda_n) - h(\frac{(j-1)r}{k} + \lambda_n)))_{j=1}^k)_{n=1}^\infty$. More explicitly, this sequence takes the form

$$h(\frac{r}{k} + \lambda_1) - h(\lambda_1), h(\frac{2r}{k} + \lambda_1) - h(\frac{r}{k} + \lambda_1), \dots, h(r + \lambda_1) - h(\frac{(k-1)r}{k} + \lambda_1), h(\frac{r}{k} + \lambda_2) - h(\lambda_2), \dots$$

If this sequence converges to 0, then it follows that

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

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 $\tilde{\lambda}_n \rightarrow \infty$ and $h(\frac{r}{k} + \tilde{\lambda}_n) - h(\tilde{\lambda}_n) \rightarrow \tilde{b}$.

Because k was an arbitrary natural number, take it large enough such that $q \in [\frac{1}{2}, \frac{1}{2} + \delta)$, 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 $\tilde{\lambda}_n \rightarrow -\infty$, we have $c'^{-1}(\frac{r}{k} + \tilde{\lambda}_n) \rightarrow 0$. 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

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

as $n \rightarrow \infty$. Because of $\tilde{b} > 0$, $\lim \frac{P(x|A_n)}{P(y|A_n)} > 1$ follows, which contradicts Limited Discrimination.

Q.E.D..

A.12 Proof of Proposition 8

As Lemma 1 shows, the optimality condition of the maximization problem $\max_{p \in \Delta(A)} \sum_{z \in A} (u(z)p(z) - c_A(p(z)))$ is that there exists $\lambda(A)$ such that

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

holds for each $x \in A$.

(1) implies (2): Suppose that P is represented by item-invariant APU, and take any menu $A \in \mathcal{A}$ with $x, y \in A$. Then, by the above FOC and strict convexity of c_A , $u(x) > u(y)$ if $P(x|A) > P(y|A)$, and $u(x) = u(y)$ if $P(x|A) = P(y|A) \in (0, 1)$.

(2) implies (1): Suppose that there is a function $u : Z \rightarrow \mathbb{R}$ such that, for any A and $x, y \in A$, $u(x) > u(y)$ if $P(x|A) > P(y|A)$, and $u(x) = u(y)$ if $P(x|A) = P(y|A) \in (0, 1)$. Take any $A \in \mathcal{A}$ and let

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

Then we can construct a strictly increasing and continuous function $g_A : [0, 1] \rightarrow \mathbb{R}$ such that (i) $g_A(0) = \underline{w}(A)$, and (ii) $g_A(P(z|A)) = u(z)$ if $P(z|A) \in (0, 1)$. This defines a strictly convex and C^1 function $c_A : [0, 1] \rightarrow \mathbb{R}$ by $c_A(p) = \int_0^p g_A(q) dq$. Then

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

Therefore the optimality condition holds at each menu A with $\lambda(A) = 0$, so that P is represented by item-invariant APU.

equivalence of (2) and (3): This follows by applying Lemma 3 to the item order \succsim_i over

$Z = \mathcal{X}$. Q.E.D.

A.13 Proof of Proposition 9

(1) implies (2): Suppose that P is represented by menu-invariant APU, and take any menus $A, B \in \mathcal{A}$ with $x \in A \cap B$. Let $\lambda : \mathcal{A} \rightarrow \mathbb{R}$ be the Lagrangean multipliers associated with each menu. Then, by the FOC and strict convexity of c_x , $\lambda(A) > \lambda(B)$ if $P(x|A) > P(x|B)$, and $\lambda(A) = \lambda(B)$ if $P(x|A) = P(x|B) \in (0, 1)$.

(2) implies (1): Suppose that there is a function $\lambda : \mathcal{A} \rightarrow \mathbb{R}$ such that for any menus A, B and item $z \in A \cap B$, $\lambda(A) > \lambda(B)$ if $P(z|A) > P(z|B)$, and $\lambda(A) = \lambda(B)$ if $P(z|A) = P(z|B) \in (0, 1)$. It is without loss to assume that λ takes values in $(0, 1)$.

Take any item z and let

$$\bar{w}(z) := \begin{cases} 1 & \text{if } P(z|A) < 1 \forall A \ni z \\ \min\{\lambda(A) | A \in \mathcal{A}, P(x|A) = 1\} & \text{otherwise.} \end{cases}$$

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

Construct for each z a strictly increasing and continuous function $g_z : [0, 1] \rightarrow \mathbb{R}$ such that (i) $g_z(0) = \underline{w}(z)$, (ii) $g_z(P(x|A)) = \lambda(A)$ if $P(x|A) \in (0, 1)$, and (iii) $g_z(1) = \bar{w}(z)$. For each z , define a strictly convex C^1 cost function $c_z : [0, 1] \rightarrow \mathbb{R}$ by $c_z(q) = \int_0^q g_z(p) dp$ and utility $u(z) = 0$. Then the optimality condition

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

holds at each menu A and item $x \in A$, so that P is represented by menu-invariant APU.

equivalence of (2) and (3): This follows by applying Lemma 3 to the menu order \succsim_m over

$\mathcal{A} = \mathcal{X}$. Q.E.D.

A.14 Proof of Proposition 10

In the following we only consider menus with at least two elements, without loss of generality.

“**IF**”: Positivity and Menu Acyclicity are satisfied if P is represented by menu-invariant APU with steep costs. To show Order Independence we proceed as follows.

First we show that for any menu A and items $x, y \in A$, $u(x) \geq u(y)$ iff $P(x|A) \geq P(y|A)$.

Note that the FOC takes the form

$$u(x) - c'_x(P(x|A)) = u(y) - c'_y(P(y|A)). \quad (8)$$

If $u(x) \geq u(y)$ then, by weak menu-dependence $c'_x(P(x|A)) - c'_y(P(x|A)) \leq u(x) - u(y)$, and from the strict convexity of c_y , $P(x|A) \geq P(y|A)$. Conversely, suppose that $P(x|A) \geq P(y|A)$. If $u(y) > u(x)$, then the weak menu-dependence condition $c'_y(P(x|A)) - c'_x(P(x|A)) < u(y) - u(x)$ and the strict convexity of c_y lead to a contradiction to (8). Therefore $u(x) \geq u(y)$.

Now take any menus A, B and $x, y \in A \setminus B$. Observe that

$$\begin{aligned} c'_x(P(x|B \cup \{x\})) - u(x) &\leq c'_y(P(y|B \cup \{y\})) - u(y) \\ \Leftrightarrow \lambda(B \cup \{x\}) &\leq \lambda(B \cup \{y\}) \\ \Leftrightarrow \exists z \in B, P(z|B \cup \{x\}) &\leq P(z|B \cup \{y\}) \\ \Leftrightarrow \forall z \in B, P(z|B \cup \{x\}) &\leq P(z|B \cup \{y\}) \\ \Rightarrow P(x|B \cup \{x\}) &\geq P(y|B \cup \{y\}). \end{aligned}$$

Suppose that $P(z|B \cup \{x\}) \leq P(z|B \cup \{y\})$ holds for all $z \in B$. If $u(x) < u(y)$, then the weak menu-dependence condition $c'_y(P(x|B \cup \{x\})) - c'_x(P(x|B \cup \{x\})) < u(y) - u(x)$, the strict convexity of c_y , and $P(x|B \cup \{x\}) \geq P(y|B \cup \{y\})$ imply $c'_y(P(y|B \cup \{y\})) - c'_x(P(x|B \cup \{x\})) < u(y) - u(x)$, a contradiction. Therefore $u(x) \geq u(y)$. Conversely, suppose that $u(x) \geq u(y)$. If $\lambda(B \cup \{x\}) > \lambda(B \cup \{y\})$, then we have both $c'_x(P(x|B \cup \{x\})) - u(x) > c'_y(P(y|B \cup \{y\})) - u(y)$

and $P(x|B \cup \{x\}) \leq P(y|B \cup \{y\})$. The weak menu-dependence condition $u(x) - c'_x(P(x|B \cup \{x\})) \geq u(y) - c'_y(P(x|B \cup \{x\}))$ and the strict convexity of c_y imply $u(x) - c'_x(P(x|B \cup \{x\})) \geq u(y) - c'_y(P(y|B \cup \{y\}))$, a contradiction. So $\lambda(B \cup \{x\}) \leq \lambda(B \cup \{y\})$. Therefore we have shown that $P(z|B \cup \{x\}) \leq P(z|B \cup \{y\})$ holds for all $z \in B$ if and only if $u(x) \geq u(y)$, which is equivalent to $P(x|A) \geq P(y|A)$.

“Only if”: Positivity and Order Independence imply Item Acyclicity, which ensures the existence of $u : Z \rightarrow \mathbb{R}$ such that $P(x|A) \geq P(y|A)$ iff $u(x) \geq u(y)$ for any $x, y \in A$. Menu Acyclicity implies the existence of $\lambda : \mathcal{A} \rightarrow \mathbb{R}$ such that $P(x|A) \geq P(x|B)$ iff $\lambda(A) \geq \lambda(B)$ for any $x \in A \cap B$. Without loss, we take u and λ to be strictly positive valued. By Positivity, there exists $\epsilon > 0$ such that $\epsilon < P(x|A)$ for all $A \in \mathcal{A}$ and $x \in A$. Take a continuous and strictly increasing function $\bar{g} : (0, \epsilon] \rightarrow \mathbb{R}$ such that $\lim_{q \rightarrow 0} \bar{g}(q) = -\infty$ and $\bar{g}(\epsilon) = 0$.

Fix $\delta > 0$. For each $x \in Z$, we construct a function $g_x : (0, 1] \rightarrow \mathbb{R}$ in the following manner. First, we set $g_x(q) = \bar{g}(q)$ for all $q \in (0, \epsilon]$. Second, we set $g_x(P(x|A)) = u(x) + \lambda(A)$ for all menu $A \ni x$. Third, values of $g_x(q)$ for each $q \in (\epsilon, \max_{A \ni x} P(x|A)]$ are defined by linear interpolations among the previously defined points. Finally, we set $g_x(q) = \delta(q - \max_{A \ni x} P(x|A)) + u(x) + \max_{A \ni x} \lambda(A)$ for each $q \in (\max_{A \ni x} P(x|A), 1]$.

This function g_x is well-defined, and continuous by construction. Because of the property $u(x) + \lambda(A) \geq u(x) + \lambda(B)$ iff $P(x|A) \geq P(x|B)$, g_x is strictly increasing in $q \in [\min_{A \ni x} P(x|A), 1]$. Furthermore, $g_x(\min_{A \ni x} P(x|A)) = u(x) + \min_{A \ni x} \lambda(A) > 0$ ensures that g_x is strictly increasing globally.

Take any x, y such that $u(x) > u(y)$, and we show that $g_x(q) - g_y(q) < u(x) - u(y)$ for all q if δ is sufficient large. Because $g_x(q) - g_y(q)$ is equal to zero in $q \in [0, \epsilon]$, and piece-wise linear in $q \in [\epsilon, 1]$, it is sufficient to only consider q on the kink-points in $(\epsilon, 1]$. There are three cases.

1. $q = P(y|A)$ for some $A \ni y$.

First, consider the case $x \in A$. Note that $P(y|A) < P(x|A)$. Then

$$g_x(q) - g_y(q) < g_x(P(x|A)) - g_y(P(y|A)) = u(x) - u(y).$$

Second, consider the case $x \notin A$. Then Order Independence implies $P(z|A) > P(z|A \cup \{x\} \setminus \{y\})$ for all $z \in A \setminus \{y\}$, so that $\lambda(A) > \lambda(A \cup \{x\} \setminus \{y\})$ and $P(y|A) < P(y|A \cup \{x\} \setminus \{y\})$. Then

$$\begin{aligned} g_x(q) - g_y(q) &< g_x(P(x|A)) - g_y(P(y|A \cup \{x\} \setminus \{y\})) \\ &= u(x) + \lambda(A) - u(y) - \lambda(A \cup \{x\} \setminus \{y\}) \\ &< u(x) - u(y). \end{aligned}$$

2. $q = P(x|A)$ for some $A \ni x$.

This case is analogous to the first case so that omitted.

3. $q = 1$.

Let $q_x = \max_{A \ni x} P(x|A)$ and $q_y = \max_{A \ni y} P(y|A)$. As in the first case, for any $A \ni y$, there exists $A' \ni x$ such that $P(y|A) < P(x|A')$. Therefore $q_x > q_y$. Then

$$\begin{aligned} g_x(q) - g_y(q) &= \delta(1 - q_x) + g_x(q_x) - \delta(1 - q_y) - g_y(q_y) \\ &= \delta(q_y - q_x) + g_x(q_x) - g_y(q_y) \\ &< u(x) - u(y) \end{aligned}$$

when δ is sufficiently large.

By finiteness of Z , we can take δ uniformly in x, y .

Next, take any x, y such that $u(x) = u(y)$. We show that $g_x(q) = g_y(q)$ for all q . To see this, take any q such that $q = P(x|A)$ for some menu $A \ni x$. First consider the case $y \in A$. Then $P(y|A) = P(x|A)$, so that $g_y(q) = u(y) + \lambda(A) = u(x) + \lambda(A) = g_x(q)$. Second, consider the case $y \notin A$. Then by Order Independence, $P(z|A) = P(z|A \cup \{y\} \setminus \{x\})$ for all $z \in A \setminus \{x\}$, so that $\lambda(A) = \lambda(A \cup \{y\} \setminus \{x\})$ and $P(x|A) = P(y|A \cup \{y\} \setminus \{x\})$. This implies that $g_y(q) = u(y) + \lambda(A \cup \{y\} \setminus \{x\}) = u(x) + \lambda(A) = g_x(q)$. Then, by construction, $g_x(q) = g_y(q)$ for all $q \in (0, 1]$.

We construct a strictly convex and C^1 cost function $c_x : (0, 1] \rightarrow \mathbb{R}$ for each $x \in Z$ by $c'_x(q) = g_x(q)$. This cost is steep by construction: $\lim_{q \rightarrow 0} c'_x(q) = \lim_{q \rightarrow 0} \bar{g}(q) = -\infty$. Then $u(x) - c'_x(q) \geq u(y) - c'_y(q)$ for each $q \in (0, 1]$ and x, y such that $u(x) \geq u(y)$, where the inequality is strict if $u(x) > u(y)$. For any menu A and item $x \in A$, FOC $c'_x(P(x|A)) = u(x) + \lambda(A)$ holds, so that P is represented by menu-invariant APU with steep costs that is weakly item-dependent. **Q.E.D.**

A.15 The α -nested Entropy Model

A.15.1 Nested Logit

With nested logit the choice probability of item x in nest B_k takes the form⁴¹

$$P(x|A) = \frac{\exp[u(x)/\eta_k]}{\sum_{z \in B_k \cap A} \exp[u(z)/\eta_k]} \frac{(\sum_{z \in B_k \cap A} \exp[u(z)/\eta_k])^{\eta_k}}{\sum_{l=1}^K (\sum_{z \in B_l \cap A} \exp[u(z)/\eta_l])^{\eta_l}}$$

where $\eta_k > 0$ is an exogenous parameter defined for each nest B_k . From Lemma 4 in Appendix A.15.3, this model can be represented as the nested entropy model by setting

$$P_1(B_k|A) = \frac{(\sum_{z \in B_k \cap A} \exp[u(z)/\eta_k])^{\eta_k}}{\sum_{l=1}^K (\sum_{z \in B_l \cap A} \exp[u(z)/\eta_l])^{\eta_l}}, \quad P_2(x|B_k, A) = \frac{\exp[u(x)/\eta_k]}{\sum_{z \in B_k \cap A} \exp[u(z)/\eta_k]},$$

$$U(B_k \cap A) = \eta_k \left(\log \sum_{z \in B_k \cap A} \exp[u(z)/\eta_k] \right).$$

The nested-logit model is a RU where the cumulative distribution of $\{\epsilon_z\}_{z \in A}$ is given by

$$\exp \left[- \sum_k \left(\sum_{z \in B_k \cap A} \exp[-\epsilon_z/\lambda_k] \right)^{\lambda_k} \right]$$

and $\lambda_k \in (0, 1]$ measures the degree of independence among $\{\epsilon_z\}_{z \in B_k}$.⁴²

⁴¹See p.80 in Train (2009).

⁴²A higher value of λ_k means greater independence and less correlation. When $\lambda_k = 1$ for all k then there is no correlation within each nest, and the model is reduced to the standard logit model. Train (2009) allows λ_k to be greater than 1. In such a case, the model can be expressed as RU for some range of the variables but not all values.

A.15.2 Axiomatization of α -nested entropy

Here we provide the necessary and sufficient conditions for the nested α -entropy representation.

Axiom 10 (IIN). For any nests B_k, B_l and menus A, A' such that $B_k \cap A = B_k \cap A' \neq \emptyset$ and $B_l \cap A = B_l \cap A' \neq \emptyset$,

$$\frac{P_1(B_k|A)}{P_1(B_l|A)} = \frac{P_1(B_k|A')}{P_1(B_l|A')}.$$

This axiom, combined with Positivity, ensures that the choice probabilities over nests are as in the logit model.

Axiom 11 (Within-nest IIA). For any nest B_k , items $x, y \in B_k$ and menus A, A' such that $x, y \in A \cap A'$

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

This axiom weakens IIA by imposing it only between items from the same nest; it is in the spirit of nested logit and responds to the same red bus/blue bus critique.

Axiom 12 (Consistency). For each items $D, D' \subseteq B_k$ and $z \in B_l \neq B_k$,

$$P(z|D \cup \{z\}) \leq P(z|D' \cup \{z\}) \text{ iff } P(D|D \cup D') \geq P(D'|D \cup D').$$

This axiom connects the choice probabilities of menus D, D' belonging to the same nest and their weakness with respect to an item z belonging to a different nest; it is similar in spirit to Order Independence (Axiom 9).

Axiom 13 (α -Consistency). There exists $\alpha \geq 0$ such that, for each items $D, D' \subseteq B_k$ and $z \in B_l \neq B_k$,

$$P(z|D \cup \{z\}) \leq P(z|D' \cup \{z\}) \text{ iff } \frac{1}{|D|^\alpha} P(D|D \cup D') \geq \frac{1}{|D'|^\alpha} P(D'|D \cup D').$$

This axiom is a generalization of Axiom 12, allowing for a rescaling of the choice probabilities of items D, D' by their cardinalities to power α .

Axiom 14 (Separable Ratios). For any $a, b, c, d \in B_k$ and $z \in B_l \neq B_k$ if

$$P(a|\{a, b\}) = P(c|\{c, d\})$$

then

$$\frac{P(a|\{a, z\})/P(z|\{a, z\})}{P(b|\{b, z\})/P(z|\{b, z\})} = \frac{P(c|\{c, z\})/P(z|\{c, z\})}{P(d|\{d, z\})/P(z|\{d, z\})}.$$

This axiom is an adaptation of the axiom in [Fudenberg and Strzalecki \(2014\)](#) with the same name; it connects the first stage choice and the indirect payoffs at the second stage. In [Fudenberg and Strzalecki \(2014\)](#), the condition after "then" equalizes the choice probability ratio of a over b and that of c over d at the first stage. Since we can't directly observe these choice probability ratios in our set-up, we use the probability ratios of these items over another nest's item z to infer the relevant information.

We require a richness condition of the following form: \mathcal{A} includes every menu, and, for all $\gamma \in (0, \infty)$, B_k , and $x \in B_k$. there exists $y \in B_k$ such that $P(x|\{x, y\})/P(y|\{x, y\}) = \gamma$.

Proposition 11. *Assume the richness condition. Fix a nest structure $\{B_k\}_{k=1}^K$. Axioms 4, 10, 11, 13, 14 are satisfied if and only if P is represented as a nested entropy model with menu-costs.*

To prove this proposition, we first show that Axiom 11 implies that there is a utility function u such that the choice within each nest is logit. Second, we show that Axiom 10 implies that for some value function U choice between nests is also logit. Third, using Axioms 13 and 14 we relate the value function U to the utility function u .

The following results is a straightforward implication of Proposition 11.

Corollary 1. *Assume the richness condition. Fix a nest structure $\{B_k\}_{k=1}^K$. Axioms 4, 10, 11, 12, 14 are satisfied if and only if P is represented as a nested entropy model.*

A.15.3 Proof of Proposition 11

We only show the sufficiency of the axioms, because their necessity is straightforward.

Step 1: There are values $U(D)$ for any subset of each nest $D \subseteq B_k$ such that the first stage choice probabilities can be written as

$$P_1(B_k|A) = \frac{\exp[U(B_k \cap A)]}{\sum_l \exp[U(B_l \cap A)]} \quad (9)$$

Fix a nest B_k and its subset D and set $U(D) := 0$. Next, for any other nest B_l and its subset $D' \subset B_l$, set $U(D') := \log \frac{P_1(B_l|D \cup D')}{P_1(B_k|D \cup D')}$. Then, using such $D' \subseteq B_l$, we set $U(D'') := \log \frac{P_1(B_k|D' \cup D'')}{P_1(B_l|D' \cup D'')} \exp[U(D')]$ for any $D'' \subseteq B_k$, $D'' \neq D$.

For any nests B_n, B_m with $n, m \neq k$, and their subsets $\tilde{D} \subseteq B_n, \bar{D} \subseteq B_m$,

$$\begin{aligned} \frac{P_1(B_n|A)}{P_1(B_m|A)} &= \frac{P_1(B_n|D \cup \tilde{D} \cup \bar{D})}{P_1(B_m|D \cup \tilde{D} \cup \bar{D})} \quad (\because \text{IIN}) \\ &= \frac{P_1(B_n|D \cup \tilde{D} \cup \bar{D})}{P_1(B_k|D \cup \tilde{D} \cup \bar{D})} \frac{P_1(B_k|D \cup \tilde{D} \cup \bar{D})}{P_1(B_m|D \cup \tilde{D} \cup \bar{D})} \\ &= \frac{P_1(B_n|D \cup \tilde{D})}{P_1(B_k|D \cup \tilde{D})} \frac{P_1(B_k|D \cup \bar{D})}{P_1(B_m|D \cup \bar{D})} \quad (\because \text{IIN}) \\ &= \frac{\exp[U(\tilde{D})] \exp[U(D)]}{\exp[U(D)] \exp[U(\bar{D})]} \\ &= \frac{\exp[U(\tilde{D})]}{\exp[U(\bar{D})]} \end{aligned}$$

whenever $A \cap B_n = \tilde{D}$ and $A \cap B_m = \bar{D}$. The case of $n = k$ and/or $m = k$ is analogous. This leads to the desired claim (9).

Step 2: There are positive values $W(z)$ for each item z such that the second stage choice probabilities can be written as

$$P_2(x|B_k, A) = \frac{W(x)}{\sum_{z \in B_k \cap A} W(z)}. \quad (10)$$

For any nest B_k , menus A, A' and items $x, y \in A \cap A'$, we have

$$\begin{aligned} \frac{P_2(x|B_k, A)}{P_2(y|B_k, A)} &= \frac{P(x|A)}{P(y|A)} \\ &= \frac{P(x|A')}{P(y|A')} \quad (\because \text{Within-nest IIA}) \\ &= \frac{P_2(x|B_k, A')}{P_2(y|B_k, A')}. \end{aligned}$$

Then, following essentially the same argument in the previous step, we can construct values $\{W(z)\}_{z \in B_k}$ that satisfy (10). Note that by the richness condition and (10), $\{W(z)|z \in B_k\} = (0, \infty)$ holds at each nest B_k

Step 3: For each nest B_k there exists $\eta_k > 0$ such that for all $D \subseteq B_k$

$$U(D) = \eta_k \left(\log \sum_{z \in B} W(z) - \alpha \log |D| \right). \quad (11)$$

For each nest B_k , its subsets D, D' and any item $z \in B_l \neq B_k$ we have

$$\begin{aligned} \exp[U(D)] \geq \exp[U(D')] &\Leftrightarrow P_1(B_k|D \cup \{z\}) \geq P_1(B_k|D' \cup \{z\}) \quad (\because \text{Step1}) \\ &\Leftrightarrow P(D|D \cup \{z\}) \geq P(D'|D' \cup \{z\}) \\ &\Leftrightarrow \frac{1}{|D|^\alpha} P(D|D \cup D') \geq \frac{1}{|D'|^\alpha} P(D'|D \cup D') \quad (\because \alpha\text{-Consistency}) \\ &\Leftrightarrow \frac{1}{|D|^\alpha} \sum_{z \in D} W(z) \geq \frac{1}{|D'|^\alpha} \sum_{z \in D'} W(z). \quad (\because \text{Step2}) \end{aligned}$$

Therefore, for each B_k , we can find an increasing function f_k such that $\exp[U(D)] = f_k(\frac{1}{|D|^\alpha} \sum_{z \in D} W(z))$ for all $D \subseteq B_k$.

By the richness condition, for each $\gamma \in (0, \infty)$ and $a, c \in B_k$, there are $b, d \in B_k$ such that

$$\frac{P(a|\{a, b\})}{P(b|\{a, b\})} = \frac{P(c|\{c, d\})}{P(d|\{c, d\})} = \gamma.$$

By Separable Ratios, for any $z \in B_l \neq B_k$

$$\frac{P_1(B_k|\{a, z\})}{P_1(B_l|\{a, z\})} \frac{P_1(B_l|\{b, z\})}{P_1(B_k|\{b, z\})} = \frac{P_1(B_k|\{c, z\})}{P_1(B_l|\{c, z\})} \frac{P_1(B_l|\{d, z\})}{P_1(B_k|\{d, z\})},$$

which is, by Step 1, equivalent to

$$\frac{\exp[U(\{a\})] \exp[U(\{z\})]}{\exp[U(\{z\})] \exp[U(\{b\})]} = \frac{\exp[U(\{c\})] \exp[U(\{z\})]}{\exp[U(\{z\})] \exp[U(\{d\})]},$$

or,

$$\frac{f_k(W(a))}{f_k(W(b))} = \frac{f_k(W(c))}{f_k(W(d))}.$$

Because of $\gamma = \frac{P_2(a|B_k, \{a, b\})}{P_2(b|B_k, \{a, b\})} = \frac{P_2(c|B_k, \{c, d\})}{P_2(d|B_k, \{c, d\})} = \frac{W(a)}{W(b)} = \frac{W(c)}{W(d)}$ by Step 2, this leads to the functional equation of the form

$$f_k(\gamma r) = g_k(\gamma) f_k(r)$$

for all $r, \gamma > 0$. This is a special case of the Pexider functional equation⁴³ and the nonzero solutions are of the form $f_k(r) = \beta_k r^{\eta_k}$ for some $\beta_k, \eta_k > 0$.

Therefore

$$\begin{aligned} \exp[U(D)] &= f_k \left(\frac{1}{|D|^\alpha} \sum_{z \in D} W(z) \right) \\ &= \beta_k \left(\frac{1}{|D|^\alpha} \sum_{z \in B} W(z) \right)^{\eta_k} \\ &= \exp \left[\log \beta_k + \eta_k \left(\log \sum_{z \in B} W(z) - \alpha \log |D| \right) \right] \end{aligned}$$

Fix an item $x_k \in B_k$. We normalize the values $\{W(z)\}_{z \in B_k}$ by multiplying a positive constant to satisfy $W(x_k) = \exp[U(\{x_k\})/\eta_k]$. Note that this normalization does not affect any previous steps. Then the above equation implies $\exp[U(\{x_k\})] = \exp[\log \beta_k + U(\{x_k\})]$, which leads to $\beta_k = 1$. This proves (11).

Step 4: There are utilities $\{u(z)\}_{z \in Z}$ such that P is represented by a nested entropy model with cost coefficients $\{\eta_k\}_{k=1}^K$.

Set $u(z) := U(\{z\})$ for each item $z \in Z$. Then, for any nest B_k and item $z \in B_k$, (11) implies $W(z) = \exp[u(z)/\eta_k]$. Therefore (10) and (11) are written as

$$\begin{aligned} P_2(x|B_k, A) &= \frac{\exp[u(x)/\eta_k]}{\sum_{z \in B_k \cap A} \exp[u(z)/\eta_k]}, \\ U(D) &= \eta_k \left(\log \sum_{z \in D} \exp[u(z)/\eta_k] - \alpha \log |D| \right). \end{aligned}$$

⁴³See Theorem 4, Section 3.1 of Aczél (1966).

These equations and (9)-(10), combined with Lemma 4, imply that P is represented by a nested entropy model with menu-costs. **Q.E.D.**

The following lemma follows from the well-known formula for the maximized value with entropy cost (see Chapter 3 of Train (2009)) and the fact that $C_\alpha(p)$ is the entropy cost minus $\eta\alpha \log |A|$.

Lemma 4. *For any $\eta > 0$, $\alpha \geq 0$, and menu A ,*

$$\max_{p \in \Delta(A)} \sum_{z \in A} p(z)u(z) - \eta C_\alpha(p) = \eta \left(\log \sum_{z \in A} \exp[u(z)/\eta] - \alpha \log |A| \right).$$

References

- ACZÉL, J. (1966): *Lectures on Functional Equations and Their Applications*. New York: Academic Press.
- AFRIAT, S. N. (1967): “The construction of utility functions from expenditure data,” *International Economic Review*, 8(1), 67–77.
- AGRANOV, M., AND P. ORTOLEVA (2014): “Stochastic Choice and Hedging,” *mimeo*.
- BENAIM, M., J. HOFBAUER, AND E. HOPKINS (2009): “Learning in games with unstable equilibria,” *Journal of Economic Theory*, 144, 1694–1709.
- BLACKWELL, D. (1956): “Controlled Random Walks,” in *Proceedings International Congress of Mathematicians III*, pp. 336–338. Amsterdam, North Holland.
- BLOCK, D., AND J. MARSCHAK (1960): “Random Orderings And Stochastic Theories of Responses,” in *Contributions To Probability And Statistics*, ed. by I. O. et al. Stanford: Stanford University Press.
- CERREIA-VIOGLIO, S., D. DILLENBERGER, AND P. ORTOLEVA (2013): “Cautious Expected Utility and the Certainty Effect,” Discussion paper.
- CERREIA-VIOGLIO, S., D. DILLENBERGER, P. ORTOLEVA, AND G. RIELLA (2013): “Deliberately Stochastic: Random Choice and Preferences for Hedging,” *mimeo*.
- CHERNEV, A. (2012): “Product Assortment and Consumer Choice: An Interdisciplinary Review,” *Foundations and Trends in Marketing*, 6, 1–61.
- CLARK, S. A. (1990): “A concept of stochastic transitivity for the random utility model,” *Journal of Mathematical Psychology*, 34(1), 95–108.
- VAN DAMME, E. (1991): *Stability and perfection of Nash equilibria*. Springer.
- VAN DAMME, E., AND J. WEIBULL (2002): “Evolution in games with endogenous mistake probabilities,” *Journal of Economic Theory*, 106(2), 296–315.

- DAX, A. (1997): “An Elementary Proof of Farkas’ Lemma,” *SIAM Review*, 39(3), 503–507.
- DE CLIPPEL, G., AND K. ROZEN (2014): “Bounded Rationality and Limited Datasets,” *mimeo*.
- DWENGER, N., D. KUBLER, AND G. WEIZSACKER (2014): “Flipping a Coin: Theory and Evidence,” Discussion paper.
- ECHENIQUE, F., K. SAITO, AND G. TSERENJIGMID (2013): “The Perception Adjusted Luce Model,” Discussion paper.
- FALMAGNE, J. (1978): “A representation theorem for finite random scale systems,” *Journal of Mathematical Psychology*, 18(1), 52–72.
- FISHBURN, P. (1970): *Utility theory for decision making*.
- FISHBURN, P. C. (1992): “Induced binary probabilities and there linear ordering polytope: A status report,” *Mathematical Social Sciences*, 23, 67–80.
- (1998): “STOCHASTIC UTILITY,” *Handbook of Utility Theory: Volume 1: Principles*, p. 273.
- FRICK, M. (2013): “Monotone Threshold Representations,” Working Paper, Harvard University.
- FUDENBERG, D., AND D. K. LEVINE (1995): “Consistency and Cautious Fictitious Play,” *Journal of Economic Dynamics and Control*, 19, 1065–1089.
- FUDENBERG, D., AND T. STRZALECKI (2014): “Dynamic Logit with Choice Aversion,” *mimeo*.
- FUDENBERG, D., AND S. TAKAHASHI (2011): “Heterogeneous Beliefs and Local Information in Stochastic Fictitious Play,” *Games and Economic Behavior*, 71, 100–120.
- GILBOA, I. (1990): “A Necessary but Insufficient Condition for the Stochastic Binary Choice Problem,” *Journal of Mathematical Psychology*, 34, 371–392.
- GILBOA, I., AND D. MONDERER (1992): “A Game-Theoretic Approach to the Binary Stochastic Choice Problem,” *Journal of Mathematical Psychology*, 36(1), 555–572.
- GUL, F., P. NATENZON, AND W. PESENDORFER (forthcoming): “Random Choice as Behavioral Optimization,” *Econometrica*.
- GUL, F., AND W. PESENDORFER (2006): “Random expected utility,” *Econometrica*, 74(1), 121–146.
- HANNAN, J. (1957): “Approximation to bayes risk in repeated plays,” in *Contributions to the Theory of Games*, 3, ed. by M. Dresher, A. Tucker, and P. Wolfe, pp. 97–139.
- HANSEN, L. P., AND T. J. SARGENT (2008): *Robustness*. Princeton university press.
- HARSANYI, J. (1973a): “Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points,” *International Journal of Game Theory*, 2(1), 1–23.
- (1973b): “Oddness of the number of equilibrium points: A new proof,” *International Journal of Game Theory*, 2(1), 235–250.
- HART, S., AND A. MAS-COLELL (2001): “A general class of adaptive strategies,” *Journal of Economic Theory*, 98(1), 26–54.

- HOFBAUER, J., AND E. HOPKINS (2005): “Learning in perturbed asymmetric games,” *Games and Economic Behavior*, 52(1), 133–152.
- HOFBAUER, J., AND W. SANDHOLM (2002): “On the global convergence of stochastic fictitious play,” *Econometrica*, 70(6), 2265–2294.
- HUBER, J., J. W. PAYNE, AND C. PUTO (1982): “Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis,” *Journal of Consumer Research*, 9, 90–98.
- HUBER, J., AND C. PUTO (1983): “Market Boundaries and Product Choice: Illustrating Attraction and Substitution Effects,” *Journal of Consumer Research*, 10, 31–44.
- 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.
- KALAI, G., A. RUBINSTEIN, AND R. SPIEGLER (2002): “Rationalizing Choice Functions by Multiple Rationales,” *Econometrica*, 70, 2481–2488.
- KAMENICA, E. (2008): “Contextual Inference in Markets: On the Informational Content of Product Lines,” *American Economic Review*, 98, 2127–2149.
- KREPS, D. (1979): “A representation theorem for” preference for flexibility”,” *Econometrica: Journal of the Econometric Society*, pp. 565–577.
- KUKSOV, D., AND J. VILLAS-BOAS (2010): “When More Alternatives Lead to Less Choice,” *Marketing Science*, 29, 507–524.
- LUCE, D. (1956): “Semiordeers and a theory of utility discrimination,” *Econometrica*, pp. 178–191.
- LUCE, D. (1959): *Individual choice behavior*. John Wiley.
- MACCHERONI, F., M. MARINACCI, AND A. RUSTICHINI (2006): “Ambiguity Aversion, Robustness, and the Variational Representation of Preferences,” *Econometrica*, 74(6), 1447–1498.
- MACHINA, M. (1985): “Stochastic choice functions generated from deterministic preferences over lotteries,” *The Economic Journal*, 95(379), 575–594.
- MANZINI, P., AND M. MARIOTTI (forthcoming): “Stochastic Choice and Consideration Sets,” *Econometrica*.
- MARSCHAK, J. (1959): “Binary Choice Constraints on Random Utility Indicators,” Discussion paper.
- MATTSSON, L.-G., AND J. W. WEIBULL (2002): “Probabilistic choice and procedurally bounded rationality,” *Games and Economic Behavior*, 41, 61–78.
- McFADDEN, D. (1973): “Conditional logit analysis of qualitative choice behavior,” in *Frontiers in Econometrics*, ed. by P. Zarembka. Institute of Urban and Regional Development, University of California.
- MELLERS, B., AND K. BIAGINI (1994): “Similarity and Choice,” *Psychological Review*, 101, 505–18.

- ORTOLEVA, P. (2013): “The Price of Flexibility: Towards a Theory of Thinking Aversion,” *Journal of Economic Theory*, 148, 903–934.
- RICHTER, M. K. (1966): “Revealed Preference Theory,” *Econometrica*, 34(3), 635–645.
- RIESKAMP, J., J. R. BUSEMEYER, AND B. A. MELLERS (2006): “Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice,” *Journal of Economic Literature*, 44(3), 631–661.
- ROCKAFELLAR, R. T. (1970): *Convex Analysis*. Princeton University Press.
- ROSENTHAL, A. (1989): “A bounded-rationality approach to the study of noncooperative games,” *International Journal of Game Theory*, 18, 273–292.
- SAITO, K. (2014): “Preference for Flexibility and Preference for Randomization under Ambiguity,” .
- SIMONSON, I. (1989): “Choice Based on Reasons: The Case of Attraction and Compromise Effects,” *Journal of Consumer Research*, 16, 158–74.
- SWAIT, J., AND A. MARLEY (2013): “Probabilistic choice (models) as a result of balancing multiple goals,” *Journal of Mathematical Psychology*, 57(1-2), 114.
- THURSTONE, L. (1927): “A law of comparative judgment.,” *Psychological Review*, 34(4), 273.
- TRAIN, K. (2009): *Discrete choice methods with simulation*. Cambridge university press, 2nd edn.
- TVERSKY, A. (1972): “Choice by Elimination,” *Journal of Mathematical Psychology*, 9, 341–367.
- TVERSKY, A., AND J. E. RUSSO (1969): “Substitutability and Similarity in Binary Choice,” *Journal of Mathematical Psychology*, 6, 1–12.
- TVERSKY, A., AND I. SIMONSON (1993): “Context Dependent Preferences,” *Management Science*, 39, 117989.
- 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.