

Stochastic Choice and Revealed Perturbed Utility*

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

December 21, 2013

Abstract

We argue that representations in which the agent maximizes the sum of expected utility and a non-linear perturbation function provide a simple and tractable way to model stochastic choice. In contrast to past work on such non-linear utility functions, we take a revealed preference approach: We suppose that the analyst can observe the agent's choice probabilities from every possible menu, and show that various restrictions on the probabilities correspond to particular forms of the perturbation function. For example we show how to relax Luce's IIA condition to model cases where the agent finds it harder to discriminate between items in larger menus, and how to extend the perturbation-function approach to model choice overload and nested decisions. We also show that these representations correspond to a form of variational preferences ([Maccheroni, Marinacci, and Rustichini, 2006](#)) of an ambiguity-averse agent who is uncertain about her true utility.

*We thank Sonia Jaffe, Mark Machina, Morgan McClellon, Wolfgang Pesendorfer, Drazen Prelec, Bill Sandholm, and Ricky Vohra for useful comments and suggestions, and NSF grants SES 0951462, 1258665, and CAREER grant 1255062 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. A number of explanations have been proposed in the literature, each with various advantages and drawbacks. Here we follow [Harsanyi \(1973b\)](#), [Machina \(1985\)](#), and [Mattsson and Weibull \(2002\)](#) and study representations in which the agent maximizes the sum of expected utility and a non-linear perturbation function that makes it optimal to assign positive probability to every action

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

where $P(A)$ is the probability distribution of choices from the set A , u is the utility function of the agent, and C_A is the perturbation function. As we show, this objective function corresponds to a game in which the agent has a form of variational preferences and so randomizes to guard against moves by malevolent Nature. This sort of objective function is 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 the case of repeated stochastic choice in static games.¹

One interpretation of representation (1) is that agents facing a decision problem randomize to maximize their non-EU preferences on lotteries, as in [Machina \(1985\)](#);² recent experimental evidence ([Agranov and Ortoleva, 2013](#); [Dwenger, Kubler, and Weizsacker, 2013](#)) indicates that some people may indeed have an intrinsic preference for randomization. Another interpretation is that stochastic choice arises due to inattention: 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

¹[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](#)).

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, but their non-EU preferences are more specific than in Machina and fall within the class of [Cerreia-Vioglio, Dillenberger, and Ortoleva \(2013\)](#).

avoiding them (van Damme, 1991).³

In contrast to past work on such non-linear utility functions, we take a revealed preference approach: we suppose that the analyst observes the agent’s choice probabilities from each menu, 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)⁴ 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.

It is easy to see that the most general form of perturbation function can accommodate any observed behavior that assigns strictly positive (albeit arbitrarily small) probability to every choice in every menu. Moreover if all choices have positive probability in any menu, this is also true for costs of the form

$$C_A(p) = \sum_{z \in A} c_{A,z}(p(z)).$$

These additive cost functions are the starting point for our analysis. Section 3 characterizes the *invariant* case, where the cost function has the form

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

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 *acyclicity* axiom, which ensures consistency of choice probabilities across different subsets. Acyclicity implies (and in fact is strictly stronger than) strong stochastic transitivity: 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 $\{x, z\}$ is at least the maximum of those two other probabilities. Here the utility function u aligns in the

³See also Weibull, Mattsson, and Voorneveld (2007) for an alternative approach in which the agent pays costs to improve signal precision.

⁴Rieskamp, Busemeyer, and Mellers (2006) survey the extent to which these conditions are satisfied, and the situations in which they are violated.

natural way with the choice probabilities: in any menu the higher utility objects are chosen with higher probability. Invariant perturbed utility implies that choice is *regular*: the probability of choosing a particular item cannot increase when more alternatives are added to the choice set. This 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).

The most commonly used cost function 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. Acyclicity is a more revealing 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 5 we explore cost functions that weaken IIA by allowing the odds ratios to become flatter 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 8. We also 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.

Acyclicity can be weakened to *menu-acyclicity*, which rules out only certain kinds of cycles; roughly speaking this condition means that menus can be ordered by *weakness*, with the probability of choosing any given item being higher in a weaker menu. Section 7 shows that menu-acyclicity characterizes *menu-invariant* cost functions, where the cost term c can be specified to be independent of the menu, that is

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

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 $\{x, z\}$. 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 acyclicity is the condition that *items are ordered*: If x is chosen more often than y from some menu A , then it is chosen more often from any menu that contains them both. Section 6 shows that this axiom is equivalent to a *item-invariant* costs that can depend on the menu but not on the item z

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

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; Harsanyi, 1973a; McFadden, 1973) which supposes that the agent’s choice maximizes a utility function that is subject to random shocks.⁵ As noted above, under a positivity axiom, 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.⁶ Their intersection is non-empty as it includes logistic choice. We note that, in contrast to RU, which is characterized by axioms that involve subtracting probabilities (Falmagne, 1978), or dividing them (Luce, 1959), perturbed utility is characterized by axioms that rely only on ordinal information about the choice probabilities (what we call the *ordinal stochastic preference*).

RU implies that the agent is never made worse off when items are added to a choice set, which

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) studies agents who only pay attention to a random subset of each menu. Their main model is a special case of RU.

⁶Menu-invariant and item-invariant costs do not nest RU either.

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 (2013) 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.

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 all menus. 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 will maintain the following positivity assumption throughout the paper.

Axiom 0 (Positivity). $P(z|A)$ is strictly positive for each 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).

The first, most general, representation is a useful starting point for the analysis, though as it imposes no restrictions at all it has no effective behavioral content. We say that a function c is a *cost function* if $c : (0, 1) \rightarrow \mathbb{R}$ is C^1 , strictly convex, and $\lim_{q \rightarrow 0} c'(q) = -\infty$.

Definition 1 (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

$$P(A) = \arg \max_{p \in \Delta(A)} \sum_{z \in A} u(z)p(z) - c_{z,A}(p(z)). \quad (2)$$

Note that this definition requires that the chosen probability distribution is the unique maximizer of the objective function, i.e. the argmax is assumed to be a single point.

Under positivity, APU does not restrict choice behavior.⁷ To understand why, consider the first order condition (FOC) in Lagrangian of the maximization problem (2)

$$u(z) - c'_{z,A}(p(z)) + \lambda(A) = 0, \quad (3)$$

where $\lambda(A)$ is the multiplier associated with the constraint $\sum_{z \in A} p(z) = 1$. To see how these conditions can be satisfied, set all of the utilities identically equal to 0, and choose convex cost functions such that $c'_{z,A}(P(z|A)) = 0$.⁸ Then the the first order conditions (3) for the agent's maximization are all satisfied with Lagrange multipliers $\lambda(A)$ of 0.

Kalai, Rubinstein, and Spiegel (2002) show that in the context of deterministic choice any data set can be rationalized by a menu-dependent objective function. The result above shows that in the context of purely stochastic choice a similar conclusion follows.

We now define the ordinal stochastic preference associated with each stochastic choice rule; it will play an important role in our analysis.

Definition 2 (Ordinal Stochastic Preference). Let $\mathcal{D} := \{(z, A) \in Z \times A : z \in A\}$. For $((x, A), (y, B)) \in \mathcal{D} \times \mathcal{D}$

$$(x, A) \succsim (y, B) \text{ iff } P(x|A) \geq P(y|B).$$

By definition, the relation $\succsim$ is complete and transitive (since it has a numerical representation); we denote its strict part by $\succ$. This relation contains all the ordinal information about the stochastic choice rule. To relate the observed choice probabilities to the form of the cost

⁷For this reason, allowing for menu dependent utility would provide no additional generality, nor would dropping the restriction to convex cost functions.

⁸The cost functions can be taken to be convex because for each z and A choice data only restricts $c_{z,A}$ at the single point $P(z|A)$.

function, we will impose various sorts of independence conditions on $\succsim$. 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 3 (Regularity). P satisfies *regularity* if and only if $P(x|B) \leq P(x|A) \forall x \in A \subseteq B$.

In Proposition 10 we show that regularity is a necessary condition for menu-invariant APU (and hence invariant APU), but it is not an implication of item-invariant APU. Moreover, regularity is not sufficient for any of them.

Our 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.

Definition 4 (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

$$P(z|A) = \text{Prob}\{u(z) + \epsilon_z \geq \max_{y \in A} u(y) + \epsilon_y\} > 0 \quad (4)$$

for each $A \in \mathcal{A}$ and $z \in A$.⁹

Perhaps the most familiar RU is logit/logistic choice, which is given by

$$P(z|A) = \exp(u(z)/\eta) / \sum_{z' \in A} \exp(u(z')/\eta).$$

As is well known, this corresponds to a RU where the shocks ε are i.i.d. Gumbel with variance η ; it also corresponds to additive perturbed utility with cost $c(p) = \eta p \log p$. As Luce (1959) showed, logit choice is characterized by Axiom 0 (positivity) and the "IIA" condition that $P(z|A)/P(z'|A) = P(z|B)/P(z'|B)$ for all z, z', A, B with $z, z' \in A \cap B$.¹⁰

The entropy function is an example of what we call an *invariant* cost function, as c does

⁹We include positivity as part of the definition of random utility because we will maintain the positivity assumption throughout this paper. A sufficient condition for positivity is that ε has full support on $\mathbb{R}^Z$.

¹⁰Echenique, 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.

not depend on z or A . Another class of invariant cost functions is the logarithmic form used by Harsanyi (1973b) $c(q) = -\eta \log(q)$, where $\eta > 0$. As the example below shows, some invariant APU is not consistent with any RU, even those with correlated or non-identical shocks. Conversely, we show in Example 5 that some RU with i.i.d. shocks do not correspond to an invariant APU.

Example 1. If $|Z| = 3$, then any choice rule that satisfies regularity has a RU representation (Block and Marschak, 1960), so any menu independent APU has a RU representation (see Proposition 10). When $|Z| > 3$ regularity does not characterize RU. For example if $Z = \{w, x, y, z\}$, then the following supermodularity condition is a necessary consequence of RU (Block and Marschak, 1960).

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

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,

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

3 Invariant Perturbations

The most commonly used perturbation functions are invariant, i.e., such that c does not depend directly on z nor on A .

Definition 5 (Invariant APU). An *invariant* APU has the form

$$P(\cdot|A) = \arg \max_{p \in \text{int}\Delta(A)} \sum_{z \in A} (u(z)p(z) - c(p(z)))$$

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

Since the first order necessary and sufficient conditions for this maximization problem are

$u(z) + \lambda(A) - c'(P(z|A)) = 0$, monotonicity of c' implies that

$$P(x|A) \geq P(y|B) \text{ if and only if } u(x) + \lambda(A) \geq u(y) + \lambda(B). \quad (5)$$

Thus, a necessary implication of invariant APU is that the ordinal stochastic preference $\succsim$ introduced in Definition 2 has an additive representation: the probability of choosing an item from a menu depends on the utility u of the item and on a function λ of the menu. This function λ measures the weakness of the menu in the following sense:

Definition 6. For any menus A and B such that $A \cap B \neq \emptyset$ we say that A is *weaker* than B if there exists $z \in A \cap B$ with $(z, A) \succsim (z, B)$. We denote this $A \succsim^* B$.

Intuitively, menu A is weaker than B if its elements compete against z less heavily than elements of B do. Note that if (5) holds, then the weakness relation $\succsim^*$ on menus is complete and transitive and moreover it is represented by the function λ . Thus, under invariant APU, the probability of choosing an item from a menu is a function of the sum of the utility of the item and the weakness of the menu. Section 3.1 uses this representation to characterize invariant APU. Section 3.2 presents an alternate characterization that relies on a generalization of Luce's IIA axiom.

3.1 Acyclicity

Axiom 1 (Acyclicity). There does not exist a sequence of preference comparisons

$$(z_1, A_1) \succsim (z'_1, A'_1), \quad (z_2, A_2) \succsim (z'_2, A'_2), \quad \dots, \quad (z_m, A_m) \succsim (z'_m, A'_m),$$

with at least one of them is strict, such that $(z_1, \dots, z_m)$ is a permutation of $(z'_1, \dots, z'_m)$, and likewise $(A_1, \dots, A_m)$ is a permutation of $(A'_1, \dots, A'_m)$, i.e., for each object or menu, the number of times it occurs on the left hand side is the same as on the right hand side.

Acyclicity means that the utility differences between items and the weakness differences between menus are comparable and summable. To see this, observe that the axiom can be written as $P(z|A) \geq P(z'|A')$ if there exists a sequence of inequalities of the form

$$P(z_1|A_1) \geq P(z'_1|A'_1), P(z_2|A_2) \geq P(z'_2|A'_2), \dots, P(z_m|A_m) \geq P(z'_m|A'_m) \quad (6)$$

such that

$$\sum_{i=1}^m (\mathbb{1}\{z_i = x\} - \mathbb{1}\{z'_i = x\}) = \begin{cases} 1 & \text{if } x = z \\ -1 & \text{if } x = z' \\ 0 & \text{if } x \neq z, z' \end{cases} \quad (7)$$

$$\sum_{i=1}^m (\mathbb{1}\{z_i = B\} - \mathbb{1}\{z'_i = B\}) = \begin{cases} 1 & \text{if } B = A \\ -1 & \text{if } B = A' \\ 0 & \text{if } B \neq A, A' \end{cases} \quad (8)$$

Under this condition, each of the inequalities $P(z_i|A_i) \geq P(z'_i|A'_i)$ can be interpreted as saying that the utility difference between z_i and z'_i is greater than the weakness difference between A_i and A'_i .

Proposition 1. *Axioms 0 and 1 are satisfied if and only if P is represented by invariant APU.*

The proofs of this and all other propositions are in the Appendix. Note that invariant APU is fully characterized by the ordinal stochastic preference $\succsim$, the exact cardinal values are inessential. The proof of necessity shows that invariant APU implies there is an additive representation of the ordinal stochastic preference $\succsim$, and then shows this implies acyclicity. The proof of sufficiency uses the fact that acyclicity is known to be sufficient for an additive representation of $\succsim$ (Fishburn, 1970), and then shows that additivity is sufficient for invariant APU.

3.2 Ordinal IIA

The entropy model is characterized by IIA. We show that a generalized version of this condition characterizes invariant APU.

Axiom 2 (Ordinal IIA). P satisfies *ordinal IIA* if there is a continuous monotone transfor-

mation $f : (0, 1) \rightarrow \mathbb{R}_{++}$ 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 and $x, y \in A \cap B$.

Proposition 2. *Axioms 0 and 2 are satisfied if and only if P is represented by invariant APU.*

Axiom 2 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 of the proof 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 .

3.3 Uniqueness

Under a fixed perturbation function, the following uniqueness result about utility function obtains.

Proposition 3. *u and u' represent the same invariant APU with the same c if and only if there exists constant α such that $u = u' + \alpha$.*

This result holds since with a fixed perturbation cost c under invariant APU, the incentive of an agent depends only on the payoff differences $u(x) - u(y)$ between items in the menu, so adding constant α to a utility function is irrelevant to his behavior. On the other hand, any other transformation of utility would affect his behavior, because it will change the strengths of utilities u relative to the perturbation c .

Moreover, under a richness condition that requires Z be infinite¹¹, we can show that c and $\hat{c}$ represent the same invariant APU with the same u iff there exist constants a and b such that $c(q) = \hat{c}(q) + at + b$ for all $q \in (0, 1)$.

¹¹The axiomatization of invariant APU in Proposition 2 holds also for infinite Z . On the other hand, the axiomatization in Proposition 1 requires Z to be finite.

4 Perturbed Utility as a Variational Preference

In this section we show that invariant APU corresponds to variational preferences (Maccheroni, Marinacci, and Rustichini, 2006) in a non-Bayesian model where an ambiguity-averse agent is faced with uncertainty about her true utility.¹² It 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 menu A the agent solves

$$\max_{p \in \Delta(A)} \min_{\epsilon \in \mathbb{R}^{|A|}} \sum_{x \in A} p(x)[u(x) + \epsilon_x] + \Phi(\epsilon), \quad (9)$$

where Φ is convex and C^1 . The interpretation of this objective function is analogous to that of the smooth variational preferences of Maccheroni, Marinacci, and Rustichini (2006): The agent worries that after she chooses x Nature may choose ϵ that leads to low utility of x ; the function Φ models the idea that the agent views some choices of ϵ to be more plausible than others.¹³ The agent has an incentive to choose non-degenerate probability distributions p , as this will force Nature to distort the utility of many prizes, making it more costly for Nature to depart from the baseline utility u .

To relate invariant APU to variational preferences, we specialize to the additive form $\Phi(\epsilon) = \sum_{x \in A} \phi(\epsilon_x)$.

Definition 7 (Invariant Additive Variational Utility). A stochastic choice rule P has an invariant *additive variational utility* (AVU) representation if and only if there exists a utility function $u : Z \rightarrow \mathbb{R}$ and a strictly convex and C^1 function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ such that $(-1, 0) \subseteq \text{range}(\phi') \subseteq (-\infty, 0)$ and

$$P(A) = \arg \max_{p \in \Delta(A)} \min_{\epsilon \in \mathbb{R}^{|A|}} \sum_{x \in A} p(x)[u(x) + \epsilon_x] + \sum_{x \in A} \phi(\epsilon_x). \quad (10)$$

¹²The 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 .

¹³In 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 set $\mathbb{R}^{|A|}$ of possible realizations of the values of ϵ .

Proposition 4. *If P has an invariant AVU representation with u and ϕ , then it has an invariant APU representation with the same u and $c(q) = \max_{\epsilon} \{q\epsilon - \phi(-\epsilon)\}$. Conversely, if P has an invariant APU representation with u and c , then it has an invariant AVU representation with the same u and $\phi(\epsilon) = \max_{q>0} \{-\epsilon q - c(q)\}$.*

The proof is based on a convex duality argument. The first direction constructs the cost function 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 $\hat{\phi}$ be the convex conjugate of c and then $\phi(\epsilon) := \hat{\phi}(-\epsilon)$.

Example 2. 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 $\phi(\epsilon) = -\eta \log(\epsilon)$. In this case, the optimal choice of Nature is $\epsilon_x = \frac{\eta}{p(x)}$. ▲

5 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 in 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.¹⁴ In other words, we expect the following axiom

Axiom 3 (Limited Discrimination). For all $x, y \in A \cap B$ if $A \succsim^* B$ and $P(x|A) > P(y|A)$, then

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

This axiom means that choices from stronger menus are more uniform than choices from weaker menus. As we show in Section 7, $A \subseteq B$ implies $A \succsim^* B$ under invariant APU, so that the axiom suggests that choice probabilities become flatter as we expand a menu. Note that

¹⁴In 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.

given the FOCs of invariant APU, we can express the choice probability ratios as

$$\frac{P(x|A)}{P(y|A)} = \frac{c'^{-1}(u(x) + \lambda(A))}{c'^{-1}(u(y) + \lambda(A))}. \quad (11)$$

If $\log c'^{-1}$ is convex then the right hand side of (11) is increasing in λ . As the λ of the weaker menu A is higher than that of B , Axiom 3 holds, leading to the following result.

Proposition 5. *Suppose that P is an invariant APU with a utility function u and cost function c . Let $h = \log(c'^{-1})$. If h is convex,¹⁵ then P satisfies Limited Discrimination.*

We now strengthen limited discrimination to analyze the choice ratio of 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 5 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 2 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 $P(x|A_n)/P(y|A_n) \rightarrow P(x|\{x, y\})/P(y|\{x, y\})$.¹⁶ 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 4 (Asymptotic Non-Discrimination). 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 (11), asymptotic non-discrimination can be expressed as

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

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

¹⁵Alternatively, when $c'(x^\lambda y^{1-\lambda}) \geq \lambda c'(x) + (1-\lambda)c'(y)$

¹⁶Regularity 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 .

Proposition 6. Suppose that P is an invariant APU with a utility function u and cost function c . Let $h = \log(c'^{-1})$.

1. If for all t the function h satisfies $\lim_{s \rightarrow \infty} [h(t-s) - h(-s)] = 0$, then P satisfies Axiom 4.
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 $P(x|A) = p$.¹⁷

Example 3. 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 6, 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 $n \rightarrow \infty$. ▲

We also have an analogous result for RU with i.i.d. noise. A cumulative distribution F of an $\mathbb{R}$ -valued random variable has a *long right tail*¹⁸ if for any $t > 0$

$$\lim_{s \rightarrow \infty} \frac{1 - F(s+t)}{1 - F(s)} = 1.$$

Under this assumption, when the noise exceeds a high threshold, i.e., $\epsilon_z \geq s$, it is also likely that the noise is a bit higher than that, i.e., $\epsilon_z \geq s+t$. When $P(x|A_n)$ and $P(y|A_n)$ are small, choosing x or y is mostly driven by very high realizations of ϵ_x or ϵ_y . Therefore, with a long right tail, the difference in utilities $u(x)$ and $u(y)$ should become almost negligible. The next proposition formalizes this intuition.

¹⁷This condition implies that the utility function convex-ranged, but not necessarily unbounded. It is not implied and does not imply the richness condition in Fudenberg and Strzalecki (2013).

¹⁸See Asmussen (2003). Examples include the Pareto distribution, the log-normal distribution, and the Weibull distribution with shape parameter less than 1.

Proposition 7. *Suppose that P has a RU representation with i.i.d. noise $(\epsilon_z)_{z \in Z}$ where each ϵ_z follows distribution F . If F has a long right tail, then P satisfies Axiom 4.*

6 Item-Invariant Perturbations

In this section and the next one we study two ways of generalizing invariant APU. Here we consider a model where the cost function is allowed to depend on the menu, but is invariant with respect to the items.

Definition 8. An *item-invariant* APU has the form

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

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 $A \in \mathcal{A}$.

Item-invariant APU satisfies weak stochastic transitivity; it needs not satisfy strong stochastic transitivity.¹⁹ Moreover, item-invariant APU need not satisfy regularity. Since many experimental studies have found violations of regularity, this representation may be a useful step in understanding these departures.²⁰

One sort of item-invariant APU arises when the cost function depends on the cardinality of the menu, perhaps because 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 $\eta = 1$. More generally, item-dependent APU can violate regularity whenever the marginal cost $c'_A(p)$ is decreasing in $|A|$.

We now present an axiom that characterizes item-invariant APU. As with invariant APU, the axiom depends only on the ordinal stochastic preference associated with P .

¹⁹To see this, consider a menu-dependent logit of the form $P(z|A) = \exp[\eta_A^{-1}u(z)] / \sum_{z' \in A} \exp[\eta_A^{-1}u(z')]$ and suppose $u(x) > u(y) > u(z)$. When $\eta_{\{x,z\}}$ is sufficiently large, it is possible to have $P(x|\{x,z\}) < \max\{P(x|\{x,y\}), P(y|\{y,z\})\}$, which violates strong stochastic transitivity. Also, when $\eta_{\{x,y,z\}}$ is large and $\eta_{\{x,y\}}$ is small, $P(y|\{x,y,z\})$ is close to $\frac{1}{3}$ while $P(y|\{x,y\})$ is close to 0, which violates regularity.

²⁰See, for instance, [Huber, Payne, and Puto \(1982\)](#), [Huber and Puto \(1983\)](#), and Section 5 of [Rieskamp, Busemeyer, and Mellers \(2006\)](#).

Axiom 5 (Items Are Ordered).

$$(x, A) \succsim (y, A) \Leftrightarrow (x, B) \succsim (y, B) \quad \forall A, B \in \mathcal{A}, \forall x, y \in A \cap B.$$

Axiom 5 says that the order on items given by $x \succsim^* y$ iff $(x, A) \succsim (y, A)$ for some $A \ni x, y$ is complete and transitive.²¹

Proposition 8. *Axioms 0 and 5 are satisfied if and only if P is represented by an item-invariant APU.*

Necessity follows immediately. To prove sufficiency, we note that items are ordered implies that there is a utility function u and a collection of strictly increasing functions W_A such that $P(x|A) = W_A(u(x))$. We then use the functions W_A to construct cost functions c_A such that the resulting first-order conditions are satisfied by the observed probabilities.

A RU with different shock distributions for each item need not be equivalent to an item-invariant APU. Conversely, 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 . Observe that any symmetric RU satisfies Axiom 5; this implies that item-invariant APU nests symmetric RU.

Example 4. An important class of RU is that of probit choice, where $\{\epsilon_z\}$ follows a zero-mean joint normal distribution; this was originally proposed by [Thurstone \(1927\)](#) as a model of psychological stimuli. When the variance and covariances of $\{\epsilon_z\}$ do not depend on the indices of the items, this model becomes symmetric. We can thus construct item-invariant costs by setting

$$u(x) = P(x|Z) = \int 1_{\{\epsilon \in K_{x,Z}\}} d\Phi(\epsilon)$$

where $K_{x,Z} = \{\epsilon \in R^Z : u(x) + \epsilon_x \geq u(y) + \epsilon_y, \forall y \in Z\}$ and Φ is the joint-normal distribution of $\{\epsilon_z\}$, and then constructing the cost functions c_A as in Proposition 8.

²¹Completeness is straightforward. For transitivity, suppose that $x \succsim^* y$ and that $y \succsim^* z$. Then there exist A and B such that $(x, A) \succsim^* (y, A)$ and $(y, B) \succsim^* (z, B)$; then Axiom 5 implies that $(x, A \cup B) \succsim^* (y, A \cup B)$ and $(y, B \cup A) \succsim^* (z, B \cup A)$, which yields the conclusion by the transitivity of $\succsim^*$ on X .

On the other hand, (Example 5) below shows that probit choices can generate menu cycles so need not correspond to an invariant APU. ▲

RU with a distinct vector of i.i.d. shocks $\{\epsilon_z\}_{z \in A}$ that follow a menu-dependent distribution F_A is analogous to item-invariant APU. This model satisfies Axiom 5, so it is nested by item-invariant APU under positivity.

7 Menu-Invariant Perturbations

An alternative way to generalize invariant APU is to require that the cost of controlling the probability of each item may depend on the item, but it's invariant with respect to the menu.

Definition 9 (Menu-invariant APU). A menu-invariant APU has the form

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

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

Axiom 6 (Menu Acyclicity). There does not exist a sequence of preference comparisons of the form

$$(z_1, A_1) \succsim (z_1, A_2), (z_2, A_2) \succsim (z_2, A_3) \dots, (z_m, A_m) \succsim (z_m, A_1)$$

where at least one of them is strict.

Proposition 9. *Axioms 0 and 6 are satisfied if and only if P is represented by a menu-invariant APU.*

The result that menu-invariant APU rules out “menu cycles” is straightforward from the first-order conditions $P(z|A) = c'_z{}^{-1}(u(z) + \lambda(A))$ and the strict convexity of each c_z . To show that Axiom 6 implies invariant APU we first use a result of Clark (1990) to conclude that Menu Acyclicity implies that the weakness order on menus can be represented by a function $\lambda : \mathcal{A} \rightarrow \mathbb{R}$. We then use a similar construction as in the proof of Proposition 8 to construct a utility function u and cost functions c_z so that the first-order conditions are necessary and

sufficient for maximization and are satisfied at the observed choices.²²

Note that the utility function u in a menu-invariant representation is not unique. That is, if P is defined by some u and $\{c_z\}_{z \in Z}$, then for any utility function u' there exist cost functions $\{c'_z\}_{z \in Z}$ such that choice data are rationalized by $(u', \{c'_z\}_{z \in Z})$.

Note also that menu acyclicity is stronger than the following “menus are ordered” condition, whose form parallels that of Axiom 5.

Axiom 7 (Menus are Ordered).

$$(x, A) \succsim (x, B) \text{ iff } (y, A) \succsim (y, B) \quad \forall A, B \in \mathcal{A}, \forall x, y \in A \cap B.$$

This axiom is not sufficient for the existence of a menu-invariant APU, because it does not imply that the weakness order $\succsim^*$ is transitive, as in the following stochastic choice rule on $Z = \{x, y, z\}$: $P(x|Z) = P(y|Z) = P(z|Z) = 1/3$, $P(x|\{x, y\}) = P(y|\{y, z\}) = P(z|\{x, z\}) = 3/5$. The reason that items are ordered is a sufficient condition for item-invariant APU’s but menus are ordered is not enough for menu-dependent APU is that the domain X of the ordinal preferences is not symmetric. In particular, for any three items there is a menu that contains them all, but given three pairwise nondisjoint menus there need not be an item contained in all of them.

We now analyze several properties of menu-invariant APU. First we show that it satisfies regularity.

Proposition 10. *Menus are Ordered implies Regularity.*

Next we note that menu-invariant APU implies *weak stochastic transitivity*:

$$P(x|\{x, y\}) \geq 0.5 \text{ and } P(y|\{y, z\}) \geq 0.5 \text{ imply } P(x|\{x, z\}) \geq 0.5.$$

On the other hand, APU can violate strong stochastic transitivity, so they can also violate the equivalent independence condition of [Tversky and Russo \(1969\)](#). Other violations of our ‘items

²²Here we use the positivity condition to show that the constructed cost functions can be chosen to satisfy $\lim_{p \rightarrow 0} c'_z(p) = -\infty$.

are ordered' axiom, such as

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

may obtain as well. This particular violation can arise from the “compromise effect.” (Simonson, 1989); it and other such probability reversals have been observed in many experiments.²³ Though such choice reversals are excluded by item-invariant APU, menu-invariant APU allows for them: These reversals can occur when $c'_x(p)$ increases faster than $c'_y(p)$, so that the increase in menu strength from $\{x, y\}$ to $\{x, y, z\}$ reduces the choice probability of x more than that of y .

We now turn to a comparison with random utility models. RU with a distinct vector of independent shocks $\{\epsilon_z\}_{z \in A}$ where each ϵ_z follow an item-dependent distribution F_z is analogous to menu-invariant APU. Note that RU with item-dependent shocks satisfies the Block-Marschak inequalities, so, as Example 2 shows, it does not nest menu-invariant APU. Also, such RU is not nested by menu-invariant APU: as the following example shows even some RU with i.i.d. shocks is ruled out.

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

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

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. ▲

²³See e.g. Tversky and Simonson (1993), and Section 4 of Rieskamp, Busemeyer, and Mellers (2006).

Finally, note that items are ordered and no menu cycles together imply that there are functions $u : Z \rightarrow \mathbb{R}$ and $\lambda : \mathcal{A} \rightarrow \mathbb{R}$ and a function W strictly increasing in the first variable and strictly decreasing in the second variable such that $P(z|A) = W(u(z), \lambda(A))$. However, these conditions are too weak to imply invariant APU, as the following example demonstrates.

Example 6. There are three items $Z = \{x, y, z\}$ and three menus $A = \{y, z\}, B = \{x, z\}, C = \{x, y\}$, with the following choice probabilities

$$P(x|Z) = 0.475, P(y|Z) = 0.425, P(z|Z) = 0.1,$$

$$P(y|A) = 0.525, P(z|A) = 0.475,$$

$$P(x|B) = 0.575, P(z|B) = 0.425,$$

$$P(x|C) = 0.525, P(y|C) = 0.475.$$

Notice that there are no menu cycles (A is weaker than B is weaker than C is weaker than Z), and items are ranked (x is better than both y and z , y is better than z). However, suppose that there exists an invariant APU representation (u, λ) . Then we have

$$P(x|B) = 0.575 > P(y|A) = 0.525; \text{ therefore, } u(x) + \lambda(B) > u(y) + \lambda(A)$$

$$P(y|Z) = 0.425 \geq P(z|B) = 0.425; \text{ therefore, } u(y) + \lambda(Z) \geq u(z) + \lambda(B)$$

$$P(z|A) = 0.475 \geq P(x|Z) = 0.475; \text{ therefore, } u(z) + \lambda(A) \geq u(x) + \lambda(Z).$$

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

We can also show this by noting that the choice probabilities violate Axiom 2. First, the axiom implies that

$$f(P(x|B)) = \frac{f(P(z|B))f(P(x|Z))}{f(P(z|Z))}, \quad f(P(y|A)) = \frac{f(P(z|A))f(P(y|Z))}{f(P(z|Z))}. \quad (12)$$

holds for some strictly increasing f . Note that we have $P(z|B) = P(y|Z)$ and $P(x|Z) = P(z|A)$ in Example 6, which imply $f(P(z|B)) = f(P(y|Z))$ and $f(P(x|Z)) = f(P(z|A))$. Substituting

them into (12), it follows that $f(P(x|B)) = f(P(y|A))$, which rules out $P(x|B) > P(y|A)$. $\blacktriangle$

8 Nested Choice

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 develop two applications, one that replicates the choices of the nested logit model used in empirical work (see, e.g., Train, 2009), and a second that models the idea of "choice overload".²⁴

Let $\{B_k\}_{k=1}^K$ be a partition of Z ; we will refer to each B_k as a nest. 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 , $P_2(z|B_k, A) := P(z|A)/P_1(B_k|A)$.

We do not develop the most general version of nested choice here, but instead restrict attention to a form that is sufficient for our two applications.

Let $\alpha \geq 0$ be the degree of menu-size penalty. 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.²⁵

Definition 10. P is called an α -nested entropy model if there exist utility function $u : Z \rightarrow \mathbb{R}$,

²⁴Gul, Natenzon, and Pesendorfer (2012) introduce "attribute rules" to study the problem of nesting. Their model is a special case of a RU with endogenous attributes.

²⁵Fudenberg and Strzalecki (2013) provide an axiomatic foundation for the relative-entropy cost function in a setting of recursive stochastic choice.

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

$$(P_1(B_1|A), \dots, P_1(B_K|A)) = \arg \max_{p \in \Delta(\{B_1, \dots, B_K\})} \sum_k p(B_k)U(B_k \cap A) - C_0(p),$$

$$P_2(B_k, A) = \arg \max_{p \in \Delta(B_k \cap A)} \sum_{z \in B_k \cap A} p(z)u(z) - \eta_k C_\alpha(p)$$

$$U(B_k \cap A) = \max_{p \in \Delta(B_k \cap A)} \sum_{z \in B_k \cap A} p(z)u(z) - \eta_k C_\alpha(p)$$

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; 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 .

8.1 Nested Logit

The special case of $\alpha = 0$, the nested entropy model, replicates the nested-logit model used in empirical work. This choice specification was developed in response to the “red bus/blue problem” of [Debreu \(1960\)](#), who pointed out that under logit demand, adding a duplicate item (a different color of bus) increases the probability of choosing a bus and decreased the probabilities of all other modes of transport. With nested logit the choice probability of item

x in nest B_k takes the form²⁶

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

where $\eta_k > 0$ is an exogenous parameter defined for each nest B_k . From Lemma 1, this model can be represented as the nested entropy model by setting

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

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

In particular, when $\lambda_k \in (0, 1]$ for each k , 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 λ_k measures the degree of independence among $\{\epsilon_z\}_{z \in B_k}$.²⁷ A 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.

8.2 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

²⁶See p.80 in Train (2009).

²⁷Train (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.

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

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

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 \(2000\)](#).)

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.

8.3 Axiomatization

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

Axiom 8 (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 9 (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 10 (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 [Tversky \(1972\)](#)'s order independence.²⁸

Axiom 11 (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 10, allowing for a rescaling of the choice probabilities of items D, D' by their cardinalities to power α .

Axiom 12 (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 \(2013\)](#) with the same name; it connects the first stage choice and the indirect payoffs at the second stage. In [Fudenberg and Strzalecki \(2013\)](#), 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

²⁸ *Order independence* holds if for all $x, y \in A \setminus B$ and $z \in B$, $P(z|B \cap \{x\}) \leq P(z|B \cap \{y\})$ iff $P(x|A) \geq P(y|A)$.

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: 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 0, 8, 9, 11, 12 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 9 implies that there is a utility function u such that the choice within each nest is logit. Second, we show that Axiom 8 implies that for some value function U choice between nests is also logit. Third, using Axioms 11 and 12 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 0, 8, 9, 10, 12 are satisfied if and only if P is represented as a nested entropy model.*

9 Conclusion

As we have shown, perturbed utility functions are relatively tractable and have an easily understood axiomatic characterization. They 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: Proofs

A.1 Proof of Proposition 1

“**If**”: Invariant APU satisfies the following FOC

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

for each menu A and $z \in A$, where $\lambda(A) \in \mathbb{R}$ is the Lagrange multiplier associated with the constraint $\sum_{z \in A} p(z) = 1$. Since c' admits a strictly increasing inverse function, we can write $P(z|A) = c'^{-1}(u(z) + \lambda(A))$. Then $P(z|A) \geq P(z'|A')$ if and only if $u(z) + \lambda(A) \geq u(z') + \lambda(A')$, for any $A, A' \in \mathcal{A}$ and $z \in A, z' \in A'$. Suppose there exists a cycle: $P(z_k|A_k) \geq P(z'_k|A'_k)$, $k = 1, \dots, m$, with at least one inequality strict. This implies a sequence of inequalities

$$u(z_k) + \lambda(A_k) \geq u(z'_k) + \lambda(A'_k), \quad k = 1, \dots, m,$$

with at least one of them strict. By definition of a cycle, the sums of the LHSs and RHSs are equal: $\sum_{k=1}^m (u(z_k) + \lambda(A_k)) = \sum_{k=1}^m (u(z'_k) + \lambda(A'_k))$, which is a contradiction.

“**Only if**”: By Theorem 4.1 of [Fishburn \(1970\)](#) acyclicity implies that there exist functions $u : Z \rightarrow \mathbb{R}$ and $\lambda : \mathcal{A} \rightarrow \mathbb{R}$ such that $P(z|A) \geq P(z'|A')$ iff $u(z) + \lambda(A) \geq u(z') + \lambda(A')$. Then we can find a strictly increasing function $h : (0, 1) \rightarrow \mathbb{R}$ such that $h(P(z|A)) = u(z) + \lambda(A)$ for each $(z, A) \in X$. Since X is finite we can take h to be a continuous function that satisfies $\lim_{p \rightarrow 0} h(p) = -\infty$. Construct a C^1 strictly convex function $c : (0, 1) \rightarrow \mathbb{R}$ by $c'(p) = h(p)$. For each $A \in \mathcal{A}$ consider the convex maximization problem $\max_{p \in \Delta(A)} \sum_{z \in A} u(z)p(z) - c(p(z))$. It's unique solution satisfies $u(z) + \mu(A) = c'(p(z))$ for some $\mu(A) \in \mathbb{R}$. Since $\sum_{z \in Z} p(z) = \sum_{z \in A} P(z|A) = 1$ and c^{-1} is a strictly increasing function, it follows that $\mu(A) = \lambda(A)$; hence $p(z) = P(z|A)$. **Q.E.D.**

A.2 Proof of Proposition 2

“If”: Axiom 0 is clearly satisfied by the assumption $\lim_{q \rightarrow 0} c'(q) = -\infty$. Note that invariant additive perturbed utility implies a FOC of the form

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

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 Axiom 2 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 Axiom 2. Because f is strictly increasing, c is well-defined by $c(q) = \int_0^q c'(t)dt$ as c' is Riemann integrable. 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): $x \notin \{y, z\}$.

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

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. **Q.E.D.**

A.3 Proof of Proposition 3

Since the “if” direction is obvious, we only prove the “only if” direction.

Fix any $x \in Z$ and set $\alpha := u(x) - u'(x)$. If u and u' represent the same choice data P under invariant APU with the same cost function c , then for any item $y \neq x$ there exist Lagrange multipliers $\lambda(\{x, y\})$ and $\lambda'(\{x, y\})$ such that

$$u(x) + \lambda(\{x, y\}) = c'(P(x|\{x, y\})), \quad u(y) + \lambda(\{x, y\}) = c'(P(y|\{x, y\})),$$

$$u(x) + \lambda(\{x, y\}) = c'(P(x|\{x, y\})), \quad u'(y) + \lambda'(\{x, y\}) = c'(P(y|\{x, y\})),$$

hold. This implies $u(y) - u'(y) = -\lambda(\{x, y\}) + \lambda'(\{x, y\}) = u(x) - u'(x) = \alpha$. **Q.E.D.**

A.4 Proof of Proposition 4

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

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

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

Note that the minimizer in the definition V_ϕ exists and is unique. Note also that $V_\phi(p) - V_c(p) = \sum_{z \in A} \left\{ c(p(z)) + \min_{\epsilon_z \in \mathbb{R}} [p(z)\epsilon_z + \phi(\epsilon_z)] \right\}$. For any function ϕ define the function $\hat{\phi}$ as $\hat{\phi}(\epsilon) = \phi(-\epsilon)$. Then $c(p(z)) + \min_{\epsilon_z \in \mathbb{R}} [p(z)\epsilon_z + \phi(\epsilon_z)] = c(p(z)) + \min_{\epsilon_z \in \mathbb{R}} -[p(z)\epsilon_z - \hat{\phi}(\epsilon_z)] = c(p(z)) - \max_{\epsilon_z \in \mathbb{R}} [p(z)\epsilon_z - \hat{\phi}(\epsilon_z)]$

To prove the first part of the proposition, 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 $c'(q) = -\phi'^{-1}(-q)$ and so $\lim_{q \rightarrow 0} c'(q) = \lim_{q \rightarrow 0} -\phi'^{-1}(-q) = -\infty$, from the assumption $(-1, 0) \subseteq \text{range}(\phi') \subseteq (-\infty, 0)$. Thus, c is a cost function and $V_\phi(p) - V_c(p) = 0$ for all p .

To prove the second part of the proposition, 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 $(-1, 0) = \text{range}(\phi')$. 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 .
Q.E.D.

A.5 Proof of Proposition 5

Take any A, B with $A \succsim^* B$ such that $x, y \in A \cap B$ and $P(x|A) > P(y|A)$. Because $A \succsim^* B$, 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, it follows that $\lambda(A) \geq \lambda(B)$ and $u(x) > u(z)$. Using these FOCs, we can express the log-ratio of choice

probabilities as

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

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.6 Proof of Proposition 6

1):

In order to show Axiom 4, 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

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

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

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

as $n \rightarrow \infty$. Because of $\tilde{b} > 0$, $\lim_{n \rightarrow \infty} \frac{P(x|A_n)}{P(y|A_n)} > 1$ follows, which contradicts Axiom 4. **Q.E.D.**

A.7 Proof of Proposition 7

Take items x, y such that $x, y \in A_n$ for each n in a sequence of menus $\{A_n\}_{n=1}^\infty$ with $P(x|A_n) \rightarrow 0$ and $P(y|A_n) \rightarrow 0$. Assume $u(x) \geq u(y) \geq 0$ without loss of generality. We use the notation $A_n = \{x, y, z_1^n, \dots, z_{|A_n|-2}^n\}$ to describe each menu A_n .

Let $(\epsilon_k)_{k=1}^\infty : \Omega \rightarrow \mathbb{R}^\infty$ be a random variable defined on the probability space $(\Omega, \mathbf{P})$ such that each ϵ_k independently follows distribution F over $\mathbb{R}$. Let ϵ_1 and ϵ_2 be the noise terms associated with x and y , respectively.

Define $B^n \subseteq \Omega$ as the set of states under which x is chosen from A_n , so that $P(x|A_n) = \mathbf{P}(B^n)$. We can write it as

$$B^n = \{\omega \in \Omega | \epsilon_1(\omega) \geq v_n(\omega) - u(x)\}$$

where v_n is a random variable defined by $v_n(\omega) := \max\{u(y) + \epsilon_2(\omega), \max_{k=1, \dots, |A_n|-2} \{u(z_k^n) + \epsilon_{k+2}(\omega)\}\}$.

Next we construct a set of states $D^n \subseteq B^n$ such that $\mathbf{P}(D^n)$ bounds $P(y|A_n)$ from below. This is defined by the last line of

$$\begin{aligned} P(y|A_n) &= \mathbf{P}(\{\omega \in \Omega | u(y) + \epsilon_2(\omega) \geq \max\{u(x) + \epsilon_1(\omega), \max_{k=1, \dots, |A_n|-2} \{u(z_k^n) + \epsilon_{k+2}(\omega)\}\}\}) \\ &= \mathbf{P}(\{\omega \in \Omega | u(y) + \epsilon_1(\omega) \geq \max\{u(x) + \epsilon_2(\omega), \max_{k=1, \dots, |A_n|-2} \{u(z_k^n) + \epsilon_{k+2}(\omega)\}\}\}) \\ &\geq \mathbf{P}(\{\omega \in \Omega | u(y) + \epsilon_1(\omega) \geq \max\{u(x) + u(y) + \epsilon_2(\omega), \max_{k=1, \dots, |A_n|-2} \{u(x) + u(z_k^n) + \epsilon_{k+2}(\omega)\}\}\}) \\ &= \mathbf{P}(\{\omega \in \Omega | \epsilon_1(\omega) \geq v_n(\omega) + u(x) - u(y)\}) \\ &=: \mathbf{P}(D^n) \end{aligned}$$

where in the second line we used the assumption of (ϵ_k) being i.i.d.

Therefore

$$\frac{P(y|A_n)}{P(x|A_n)} \geq \frac{\mathbf{P}(D^n)}{\mathbf{P}(B^n)} = \mathbf{P}(D^n|B^n)$$

where $\mathbf{P}(\cdot|B^n)$ is the conditional probability measure on B^n . Because ϵ_1 and v_n are independent, the conditional probability takes the form

$$\mathbf{P}(D^n|B^n) = E \left[\frac{1 - F(v_n + u(x) - u(y))}{1 - F(v_n - u(x))} \right]$$

where the expectation is taken over v_n .

Take any $\delta \in (0, 1)$. Because of the long-tail assumption, there exists $\bar{v} > 0$ such that $\frac{1-F(v+u(x)-u(y))}{1-F(v-u(x))} \geq \sqrt{1-\delta}$ holds for all $v \geq \bar{v}$.

Next, we have $\lim_{n \rightarrow \infty} \mathbf{P}(\{\omega \in \Omega | v_n(\omega) \geq \bar{v}\}) = 1$. If not, the right-hand side of

$$P(x|A_n) = \mathbf{P}(\{\omega \in \Omega | \epsilon_1(\omega) \geq v_n(\omega) - u(x)\}) \geq \mathbf{P}(\{\omega \in \Omega | v_n(\omega) < \bar{v}\})(1 - F(\bar{v} - u(x)))$$

does not converge to zero, which contradicts the hypothesis of Axiom 4. So there exists $\bar{n}$ such that $\mathbf{P}(\{\omega \in \Omega | v_n(\omega) \geq \bar{v}\}) \geq \sqrt{1-\delta}$ holds for all $n \geq \bar{n}$. Therefore $\mathbf{P}(D^n|B^n) \geq 1 - \delta$ holds for all $n \geq \bar{n}$. Because δ was arbitrary, we get $\mathbf{P}(D^n|B^n) \rightarrow 1$ as $n \rightarrow \infty$, implying that $\frac{P(y|A_n)}{P(x|A_n)} \rightarrow 1$. **Q.E.D.**

A.8 Proof of Proposition 8

The “If” part is obvious. To prove “only if,” note that Axiom 5 implies that $P(z|A) > P(z'|A)$ iff $P(z|Z) > P(z'|Z)$ for each $A \in \mathcal{A}$, and also that $P(z|A) = P(z'|A)$ iff $P(z|Z) = P(z'|Z)$. Thus for each A we can construct a strictly increasing, continuous, and bounded function $W_A : \mathbb{R} \rightarrow (0, 1)$ such that $W_A(P(z|Z)) = P(z|A)$ for each $z \in \mathcal{Z}$. Because of the Positivity Axiom, we can require such W_A to satisfy $\lim_{p \rightarrow 0} W_A^{-1}(p) = -\infty$. Construct a strictly convex and C^1 function $c_A : (0, 1) \rightarrow \mathbb{R}$ by $c_A(p) = \int_0^p W_A^{-1}(p) dp$.

Set $u(z) = P(z|Z)$ for each $z \in \mathcal{Z}$. Then $u(z) - c'_A(P(z|A)) = 0$ for each $A \in \mathcal{A}$ and $z \in \mathcal{Z}$. This is the FOC of the maximization problem $\max_{p \in \text{int} \Delta(\mathcal{A})} \sum_{z \in \mathcal{Z}} (u(z)p(z) - c_A(p(z)))$ in which Lagrange multiplier associated with $\sum_{z \in \mathcal{Z}} p(z) = 1$ is zero. Because of strict convexity, $P(A)$ is the solution to the maximization problem, proving that P is represented by an item-invariant APU. **Q.E.D.**

A.9 Proof of Proposition 9

“If”: Let P be represented by a menu-invariant APU. Axiom 0 is clearly satisfied. Suppose Axiom 6 was violated, i.e., there exists a sequence of the form

$$(z_1, A_1) \succsim (z_1, A_2), (z_2, A_2) \succsim (z_2, A_3) \dots, (z_m, A_m) \succsim (z_m, A_1)$$

where at least one of these preferences is strict. Note that the FOC implies $P(z|A) = c'_z(u(z) + \lambda(A))$ for each $A, z \in A$, where $\lambda(A)$ is the Lagrange multiplier associated with $\sum_{x \in A} p(x) = 1$. Because each c'_z is strictly increasing, the sequence of preference comparisons is equivalent to the sequence of inequalities

$$u(z_1) + \lambda(A_1) \geq u(z_1) + \lambda(A_2), u(z_2) + \lambda(A_2) \geq u(z_2) + \lambda(A_3), \dots, u(z_m) + \lambda(A_m) \geq u(z_m) + \lambda(A_1),$$

where at least one of them is strict. Summing up the LHSs and RHSs, we have $\sum_{i=1}^m u(z_i) + \lambda(A_i) > \sum_{i=1}^m u(z_i) + \lambda(A_i)$, a contradiction. Therefore Axiom 6 must hold.

“Only if”: If Axiom 6 holds, by Theorem 3 of Clark (1990), there is a numerical representation of menus $\lambda : \mathcal{A} \rightarrow \mathbb{R}$ such that for any menus A, A' and item $z \in A \cap A'$, $P(z|A) \geq P(z|A')$ if and only if $\lambda(A) \geq \lambda(A')$. Then for any item z we can construct a strictly increasing and continuous function $h_z : (0, 1) \rightarrow \mathbb{R}$ such that $h_z(P(z|A)) = \lambda(A)$ holds for each menu $A \ni z$. Note that by Axiom 0 we can require $\lim_{q \rightarrow 0} h_z(q) = -\infty$. For each item z , construct a strictly convex C^1 cost function $c_z : (0, 1) \rightarrow \mathbb{R}$ by $c_z(q) = \int_0^q h_z(p) dp$ and utility $u(z) = 0$. Then we have FOC $c'_z(P(z|A)) = u(z) + \lambda(A)$ for each menu A and item $z \in A$, which ensures that P is represented as a menu-invariant APU. **Q.E.D.**

A.10 Proof of Proposition 10

To see this, assume Axiom 7, and suppose $A \subseteq B$. Toward contradiction suppose there exists $z \in A$ such that $P(z|A) < P(z|B)$. Then Axiom 7 would imply $P(z'|A) < P(z'|B)$ for all $z' \in A$, and we would have $1 = \sum_{z' \in A} P(z'|A) < \sum_{z' \in A} P(z'|B) \leq \sum_{z' \in B} P(z'|B) = 1$. **Q.E.D.**

A.11 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

$$P_1(B_k|A) = \frac{\exp[U(B_k \cap A)]}{\sum_l \exp[U(B_l \cap A)]} \quad (13)$$

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 $\tilde{D} \subseteq B_n, \bar{D} \subseteq B_m$,

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

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 (13).

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

$$P_2(x|B_k, A) = \frac{W(x)}{\sum_{z \in B_k \cap A} W(z)}. \quad (14)$$

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 (14). Note that by the richness condition and (14), $\{W(z)|z \in B_k\} = (0, \infty)$ holds at each nest B_k

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

$$U(D) = \eta_k \left(\log \sum_{z \in D} W(z) - \alpha \log |D| \right). \quad (15)$$

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 equation²⁹ and the nonzero solutions are of the form $f_k(r) = \beta_k r^{\eta_k}$ for some $\beta_k, \eta_k > 0$.

Therefore

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

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 (15).

Step 4: There are utilities $\{u(z)\}_{z \in Z}$ such that P is represented by a nested entropy model

²⁹See Theorem 4, Section 3.1 of [Aczél \(1966\)](#).

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 $z \in B_k$, (15) implies $W(z) = \exp[u(z)/\eta_k]$. Therefore (14) and (15) are written as

$$P_2(x|B_k, A) = \frac{\exp[u(x)/\eta_k]}{\sum_{z \in B_k \cap A} \exp[u(z)/\eta_k]},$$

$$U(D) = \eta_k \left(\log \sum_{z \in D} \exp[u(z)/\eta_k] - \alpha \log |D| \right).$$

These equations and (13)-(14), combined with Lemma 1, 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 1. *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.
- AGRANOV, M., AND P. ORTOLEVA (2013): “Stochastic Choice and Hedging,” *mimeo*.
- ASMUSSEN, S. (2003): *Applied probability and queues*. Springer, 2nd edn.
- 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, 161.
- 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.
- DEBREU, G. (1960): “Review of RD Luce, Individual choice behavior: A theoretical analysis,” *American Economic Review*, 50(1), 186–88.
- DWENGER, N., D. KUBLER, AND G. WEIZSACKER (2013): “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*.
- 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 (2013): “Recursive Stochastic Choice,” *mimeo*.
- FUDENBERG, D., AND S. TAKAHASHI (2011): “Heterogeneous Beliefs and Local Information in Stochastic Fictitious Play,” *Games and Economic Behavior*, 71, 100–120.
- GUL, F., P. NATENZON, AND W. PESENDORFER (2012): “Random Choice as Behavioral Optimization,” .
- 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.
- KUKSOV, D., AND J. VILLAS-BOAS (2010): “When More Alternatives Lead to Less Choice,” *Marketing Science*, 29, 507–524.
- LUCE, D. (1956): “Semiorders 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.
- ORTOLEVA, P. (2013): “The Price of Flexibility: Towards a Theory of Thinking Aversion,” *Journal of Economic Theory*, 148, 903–934.
- 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.
- SIMONSON, I. (1989): “Choice Based on Reasons: The Case of Attraction and Compromise Effects,” *Journal of Consumer Research*, 16, 158–74.
- 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, 1179–89.
- 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, 1211–51.