Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Law of Demand and Forced Choice

S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and A. Rustichini

Working Paper n. 593

This Version: December, 2016

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Law of Demand and Forced Choice*

S. Cerreia-Vioglio⋆, F. Maccheroni⋆, M. Marinacci⋆, and A. Rustichini•
⋆Università Bocconi and IGIER and •University of Minnesota

December 2016

Abstract

We characterize consistent random choice rules in terms of the optimality of the support. We then proceed to study stochastic choice in a consumer theory setting. We prove a law of demand for stochastic choice. We then move to a temporal setting where we characterize the softmax decision criterion.

1 Introduction

It takes time to decide. Decisions are fully rational only when decision makers have enough time to make up their minds and understand how their basic (unmodelled) needs are satisfied by the different alternatives at hand. Otherwise, if decision makers are, nevertheless, required to choose within a given time window,[1] randomness enters the picture even if they have a well defined preference over alternatives that would determine their choices had they enough time to decide. Choices are, therefore, stochastic.

The strands of literature that, in the different disciplines (such as economics, neuroscience and psychology), have dealt with stochasticity in choices have mostly focused on pure choice behavior. Here we study how stochastic choice might affect consumer behavior, the most basic economic choice problem. We show that for an important class of stochastic choices the law of demand for normal goods, arguably the main result of traditional consumer theory, continues to hold on average when strictly dominated alternatives are instinctively dismissed.

*We thank Federico Echenique, Ryota Iijima, and Kota Saito for helpful suggestions. We especially thank Jay Lu with whom we proved an early version of Theorem 9. Simone Cerreia-Vioglio and Massimo Marinacci gratefully acknowledge the financial support of ERC (Grant SDDM-TEA and INDIMACRO, respectively).

[1] Without the burden of choice, incompleteness might otherwise arise.

The paper starts with an atemporal analysis, based on random choice functions, that clarifies the economic rationale of the "average" normal law. We then move to a temporal forced choice setting in which psychometric functions are introduced in order to account for mental preferences. In particular, we characterize softmax psychometric functions and then establish a law of demand for forced choice.

2 Atemporal analysis

2.1 Random choice rules and optimality

Let $\mathcal{A}$ be the collection of all non-empty finite choice sets A of an all inclusive set of alternatives X. In the rest of the paper, for each (not necessarily finite) $A \subseteq X$, we denote by $\Delta(A)$ the set of all finitely supported probabilities on A.

Definition 1 *A* random choice rule *is a function* $p : \mathcal{A} \to \Delta(X)$ *such that* $p(\cdot, A) \in \Delta(A)$ *for all* $A \in \mathcal{A}$*.*

We interpret $p(a, A)$ as the probability that a decision maker chooses alternative $a \in A$ within the choice set A. In a (ergodic) long run setup, this probability can be viewed as the long run frequency with which a is chosen. We regard p as a purely behavioral notion that accounts for the decision maker's choices. In other words, p is a way to organize choice data, without any mental interpretation per se.

Definition 2 *A random choice rule* p *is* consistent *if*

$$p(a, B) = p(a, A)\, p(A, B) \qquad \forall a \in A \subseteq B \tag{1}$$

This condition is a form of the classic Luce's choice axiom (see Luce [10]). It ensures that $p(\cdot, A)$ and $p(\cdot, B)$ are linked via conditioning a la Renyi [13]. We denote by $\sigma_p : \mathcal{A} \rightrightarrows X$ the support correspondence.[2]

Example 1 (Luce) Given $\varphi : X \to (0, \infty)$, define $p : \mathcal{A} \to \Delta(X)$ by

$$p(a, A) = \frac{\varphi(a)}{\sum_{b \in A} \varphi(b)} \qquad \forall A \in \mathcal{A} \tag{2}$$

This function p is a consistent random choice rule *with full support* (i.e., $p(a, A) > 0$ for all $a \in A$). Under some mild conditions, Renyi [13] and Luce [10] proved that this is the general form of consistent random choice rules that have full support. In particular, the uniform rule $p(a, A) = 1/|A|$ is the special case with $\varphi = 1$. It is close in spirit to the analysis of Becker [5]. ▲

[2] That is, $\sigma_p(A) = \{a \in A : p(a, A) > 0\}$.

Example 2 (Optimization) A correspondence $\sigma : \mathcal{A} \rightrightarrows X$ is a choice correspondence if $\emptyset \neq \sigma(A) \subseteq A$ for all $A \in \mathcal{A}$. As proved by Arrow [2], σ is *optimal* (or *rational*) if and only if it satisfies the following version of WARP:

$$\text{If } A \subseteq B \text{ and } \sigma(B) \cap A \neq \emptyset, \text{ then } \sigma(B) \cap A = \sigma(A) \tag{C}$$

Given any optimal correspondence, define $p : \mathcal{A} \to \Delta(X)$ by

$$p(a, A) = \frac{1}{|\sigma(A)|}\delta_a(\sigma(A)) = \begin{cases} \frac{1}{|\sigma(A)|} & \text{if } a \in \sigma(A) \\ 0 & \text{else} \end{cases} \tag{3}$$

This function p is easily seen to be a consistent random choice rule (without full support and not in the Luce-Renyi form).[3] In view of Arrow [2], we can give a mental interpretation to (3) as a tie-breaking randomization rule that selects an optimal – according to some underlying utility function – alternative within the set $\sigma(A)$. When σ is an optimal *choice function*,[4] this rule takes the deterministic form $p(b, A) = \delta_a(\{b\})$.[5] Optimal choice functions can thus be viewed as a special, deterministic, rule p. ▲

Given the previous example, we will say that a choice correspondence σ is *optimal* if it satisfies (C). The previous example also seems to provide a very specific rule, (3), where optimality of σ implies consistency. The next result makes this observation formal and much more general. Indeed, it characterizes consistent random choice rules in terms of the optimality of the support correspondence.

Theorem 1 *The function $p : \mathcal{A} \to \Delta(X)$ is a consistent random choice rule if and only if it has the form*

$$p(a, A) = \begin{cases} \frac{\varphi(a)}{\sum_{b \in \sigma(A)} \varphi(b)} & \text{if } a \in \sigma(A) \\ 0 & \text{else} \end{cases} \tag{4}$$

where $\varphi : X \to (0, \infty)$ and $\sigma : \mathcal{A} \rightrightarrows X$ is an optimal choice correspondence. Moreover, σ is unique and coincides with σ_p.

Note that, compared to Luce's rule (2), $p(\cdot, A)$ is not required to have full support. In particular, Luce's rule corresponds to the special case $\sigma(A) = A$ for all $A \in \mathcal{A}$ (which is trivially optimal). The optimization rule (3) is also a special case in that it corresponds to $\varphi = 1$. So, the two previous examples are special cases of (4).

Note that, in Theorem 1, when $\sigma : \mathcal{A} \rightrightarrows X$ is a function, $\sigma(A)$ is a singleton for all $A \in \mathcal{A}$. In this case, p is a degenerate random choice rule $\delta_{\sigma(A)}$. Thus, standard

[3] Of course, provided that σ is *not* such that $\sigma(A) = A$ for all $A \in \mathcal{A}$.

[4] That is, $\sigma(A)$ is a singleton for all $A \in \mathcal{A}$.

[5] Here δ_x denotes the (Dirac) probability measure at $x \in X$, that is, $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise.

optimization can be regarded as a special case of consistent stochastic choice. This is important for our study of stochastic choice in a consumer theory framework (our main purpose), which can then be regarded as a generalization of the standard theory. Moreover, our result allows us to interpret consistency as the stochastic counterpart of optimality (that is, of WARP).

When in (4) σ is rationalized by a utility function $v : X \to \mathbb{R}$,[6] the functions v and φ can be interpreted as first and second order utility functions, respectively. In particular, v determines the optimal actions in each choice set A, while φ takes care of the tie-breaking among them. In a multicriteria spirit, we can think of v and φ as accounting for two different characteristics of each alternative a, with the second characteristic becoming relevant when the first characteristic is not able to select a unique best alternative.[7]

Theorem 1 shows that consistent random choice rules are a special case of what Echenique and Saito [7, Definition 1] call general Luce rules. These rules, axiomatically characterized by [7, Theorem 1], take the form (4) where $\varphi : X \to (0, \infty)$, but the choice correspondence σ is not necessarily optimal, despite satisfying Sen's α. Echenique and Saito achieve such a characterization with four axioms that in our case are replaced by Renyi's consistency.[8]

We conclude with a corollary. A random choice rule p is *uniform* if, given any $A \in \mathcal{A}$, all possible alternatives in A have the same probability of being chosen; that is, for each $a \in A$

$$p(a, A) = \frac{1}{|\sigma_p(A)|}\delta_a(\sigma_p(A)) = \begin{cases} \frac{1}{|\sigma(A)|} & \text{if } a \in \sigma(A) \\ 0 & \text{else} \end{cases}$$

Corollary 2 *A uniform random choice rule p is consistent if and only if σ_p is optimal.*

In this case, the binary relation defined on X by

$$a \succsim b \iff p(a, \{a, b\}) \geq p(b, \{a, b\})$$

is a weak order and $\sigma_p(A) = \{a \in A : a \succsim b \quad \forall b \in A\}$ for all $A \in \mathcal{A}$.

2.2 Mental interpretation

The tie-breaking interpretation of Example 2 is best viewed as part of a mental interpretation of random choice rules in terms of expected utility maximization over mixed

[6] That is, $\sigma(A) = \operatorname{argmax}_{a \in A} v(a)$.

[7] The "lexicographic" relation between these two utility functions can be seen as a decision theoretic counterpart of the lexicographic measure theoretic results of Renyi [13] and [14].

[8] Random choice rules that satisfy Renyi's consistency have been studied in game theory under the name conditional-probability systems (see Myerson [11]). In this terminology, our result shows that conditional-probability systems have optimal supports.

actions. Specifically, we denote by α an element of $\Delta(A)$ when interpreted as a mixed action.[9,10]

Proposition 3 *If X is countable and $p : \mathcal{A} \to \Delta(X)$ is a consistent random choice rule, then there exists $v : X \to \mathbb{R}$ such that, for each $A \in \mathcal{A}$,*

$$\sigma_p(A) = \operatorname{argmax}_{a \in A} v(a) \tag{5}$$

and

$$p(\cdot, A) \in \operatorname{argmax}_{\alpha \in \Delta(A)} \sum_{a \in A} v(a)\,\alpha(a) \tag{6}$$

A consistent random choice rule can be viewed as an optimal mixed action with respect to an expected utility criterion. Tie-breaking is then among optimal "pure" actions.

2.3 Random consumption

Our aim here is to develop a consumer theory in a random choice setting which, *inter alia*, encompasses the traditional deterministic theory as a special case. A relevant related work is Mossin [12], which outlined a stochastic theory of consumption (with a different framework and motivation). Related in spirit is also Gabaix [8], in which a consumer theory under limited attention is developed.

Let $X = \mathbb{R}^n_+$ be the space of all bundles of goods and $B : \mathbb{R}^n_{++} \times \mathbb{R}_{++} \rightrightarrows X$ the budget correspondence defined by $B(q, w) = \{x \in X : q \cdot x \leq w\}$ for each strictly positive price and wealth pair (q, w) in $\mathbb{R}^n_{++} \times \mathbb{R}_{++}$. Now $\mathcal{A}$ is replaced with the class $\mathcal{B}$ that contains $\mathcal{A}$ and all budget sets $B(q, w)$.[11]

Definition 3 *A function $d : \mathbb{R}^n_{++} \times \mathbb{R}_{++} \to \Delta(X)$ is an (individual) stochastic demand if there exists a consistent random choice rule $p : \mathcal{B} \to \Delta(X)$ such that*

$$d(q, w)(x) = p(x, B(q, w))$$

We interpret $d(q, w)(x)$ as the probability that the bundle $x \in B(q, w)$ is chosen at price q with wealth w. The *average cost function* $c : \mathbb{R}^n_{++} \times \mathbb{R}_{++} \to \mathbb{R}_+$ of the bundle demanded is then defined by

$$c(q, w) = \sum_{x \in B(q,w)} (q \cdot x)\, d(q, w)(x)$$

[9] See Battigalli, Cerreia-Vioglio, Maccheroni, and Marinacci [4] for a decision theoretic setup in which mixed actions play a key role.

[10] We omit the simple proof of the next result.

[11] In this section, random choice rules as well as the property of consistency are extended to this class of subsets of X. Similarly, σ_p is defined over $\mathcal{B}$.

We denote by $\bar{d} : \mathbb{R}^n_{++} \times \mathbb{R}_{++} \to \mathbb{R}^n_+$ the (*individual*) *average demand function* defined by

$$\bar{d}\,(q, w) = \sum_{x \in B(q,w)} x d\,(q, w)\,(x)$$

Clearly, $c\,(q, w) = q \cdot \bar{d}\,(q, w)$.

Example 3 If p is the random choice rule in (4),[12] then

$$d\,(q, w)\,(x) = \begin{cases} \frac{\varphi(x)}{\sum_{y \in \sigma(B(q,w))} \varphi(y)} & \text{if } x \in \sigma\,(B\,(q, w)) \\ 0 & \text{else} \end{cases} \tag{7}$$

and

$$\bar{d}\,(q, w) = \sum_{x \in \sigma(B(q,w))} x \frac{\varphi\,(x)}{\sum_{y \in \sigma(B(q,w))} \varphi\,(y)}$$

The uniform case $\varphi = 1$ is, as we already remarked, in the spirit of Becker [5]. ▲

This example shows that standard demand functions are included in our setup. Indeed, in the special case in which $\varphi = 1$ and σ is a choice function, the average demand $\bar{d}\,(q, w) = \sigma\,(B\,(q, w))$ is a classic Walrasian demand function.

2.3.1 Walras' law

Given a stochastic demand, in general, by construction, we only have $c\,(q, w) \leq w$. Equality holds when the random choice p is *stochastically monotone*: $p\,(x, \{x, y\}) = 0$ whenever $x \ll y$.[13] We begin with a stochastic version of Walras' law and by formalizing the previous observation.

Lemma 4 *If $w < w'$, then $c\,(q, w) \leq c\,(q, w')$. If p is stochastically monotone, then $c\,(q, w) = w$ and*

$$w < w' \implies c\,(q, w) < c\,(q, w')$$

Stochastic monotonicity is, *prima facie*, a strong assumption that requires the consumer to detect whether a bundle is dominant across all goods. Indeed, it seems reasonable to say that comparisons between two bundles x and y of goods are easier to make when one of them is strictly dominant, say $x \ll y$. Stochastic monotonicity can be, then, viewed as a way to capture this comparative easiness, which in turn implies a sharp Walras' law. Indeed, such a consumer always exhausts his budget.[14]

[12] It can be verified that if $\sigma : \mathcal{B} \rightrightarrows X$ is an optimal choice correspondence such that $\sigma\,(A) \in \mathcal{A}$ for all $A \in \mathcal{B}$, then p is indeed consistent on $\mathcal{B}$. Extending the "if" part of Theorem 1.

[13] As usual, the notation $x \ll y$ means that $x_i < y_i$ for all $i \in \{1, ..., n\}$.

[14] For instance, Gabaix [8, p. 1675] assumes that the consumer he studies "is boundedly rational, but smart enough to exhaust his budget."

Example 4 Let $B \in \mathcal{B}$ and let $\partial^+ B = \{x \in B : \nexists x' \in B,\ x' \gg x\}$ be the subset of all elements of B that are not strictly dominated. If we assume $\sigma_p(B) \subseteq \partial^+ B$ for all $B \in \mathcal{B}$ we have a simple random choice rule that satisfies stochastic monotonicity. ▲

By construction, both the stochastic demand and the average demand are homogeneous of degree zero, so there is no nominal illusion. In fact, in our analysis the consumer is always able to assess correctly whether a bundle of goods is affordable.

2.3.2 Law of demand

We can now study wealth and price effects. As to wealth effects, say that a good k is *normal* if its average demand increases as wealth increases, that is, $\bar{d}_k(q, w') \geq \bar{d}_k(q, w)$ if $w' > w$. By Lemma 4, under stochastic monotonicity, we have $q \cdot \bar{d}(q, w') > q \cdot \bar{d}(q, w)$ if $w' > w$. So, intuitively, some of the goods have to be normal, at least locally. As to price effects, we have the following preliminary result.

Lemma 5 *If $q < q'$, then $\bar{d}(q', w) \not> \bar{d}(q, w)$.*

Next we show that a classic compensated law of demand continues to hold "on average".

Lemma 6 *Let (q', w') and (q, w) be in $\mathbb{R}^n_{++} \times \mathbb{R}_{++}$. If $q' \cdot \bar{d}(q, w) = c(q', w')$ and p is stochastically monotone, then*

$$(q' - q) \cdot \left(\bar{d}(q', w') - \bar{d}(q, w)\right) \leq 0 \tag{8}$$

When a sharp Walras' law holds, the condition $q' \cdot \bar{d}(q, w) = c(q', w')$ becomes a standard Slutsky wealth compensation. In this case the law of demand for normal goods – arguably the most important result of consumer theory – continues to hold on average.

Proposition 7 (Law of Demand) *Let p satisfy stochastic monotonicity. If wealth and other prices do not change, an increase (decrease) in the price of a normal good k decreases (increases) its average demand $\bar{d}_k$.*

Under consistency, on average the behavior of the consumer continues to satisfy the normal law of demand provided he is able to select strictly dominant alternatives. The standard consumer theory result for Walrasian demand functions, first stated in Slutsky [16, p. 14], is the special case that corresponds to the deterministic demand function $d(q, w) = \delta_{\sigma(B(q,w))}$, whose average demand is indeed the Walrasian demand function, i.e., $\bar{d}(q, w) = \sigma(q, w)$. Our result thus generalizes the most important finding of classical consumer theory. At the same time, it goes well beyond that, for instance it includes the purely random choice of Becker [5] (viewed as the uniform case).

Remark 1 (i) We expect that a full-fledged "average" consumer theory can be developed along the lines of this section. (ii) If stochastic monotonicity is replaced by the following condition

$$p\left(B\left(q,w\right)\cap B\left(q',w'\right),B\left(q',w'\right)\right)=0$$

then the inequality in (8), holds strictly.

3 Forced choice and soft maximization

A *timed random choice rule* is a function $p:[0,\infty]\times\mathcal{A}\to\Delta(X)$ such that $p_t(\cdot,A)\in\Delta(A)$ for all $A\in\mathcal{A}$ and all $t\in[0,\infty]$. We interpret $p_t(a,A)$ as the probability that $a\in A$ is chosen from the choice set A if t is the amount of time that the decision maker is given to make up his mind. Before moving on, we introduce two useful pieces of notation: given any timed random choice rule p and any $a,b\in X$, we denote

$$p_t\left(a,b\right)=p_t\left(a,\{a,b\}\right)\quad\text{and}\quad r_t\left(a,b\right)=\frac{p_t\left(a,b\right)}{p_t\left(b,a\right)}$$

for all $t\in[0,\infty]$. Next we introduce a class of such rules that will play a key role in our analysis.

A timed random choice rule p is:

- *bounded* if $\sup_{t,s\in(0,\infty)}\left|r_{t+s}\left(a,b\right)-r_t\left(a,b\right)r_s\left(a,b\right)\right|<\infty$ for all $a,b\in X$;
- *strictly coherent* if, given any $t\in(0,\infty)$,

  $$p_t\left(a,b\right)>p_t\left(b,a\right)\implies p_\infty\left(a,b\right)>p_\infty\left(b,a\right)$$

  coherent if also weak inequalities are preserved;
- *consistent* if $p_t:\mathcal{A}\to\Delta(X)$ is consistent for all $t\in(0,\infty)$;
- *continuous* if $\lim_{s\to t}p_s\left(a,A\right)=p_t\left(a,A\right)$ for all $t\in[0,\infty]$, all $A\in\mathcal{A}$, and all $a\in A$;
- *separable* if there are $t>0$ and a countable $C=C_t\subseteq X$ such that if $a,b\in X$ and

  $$p_t\left(a,\{a,b\}\right)>p_t\left(b,\{a,b\}\right)$$

  then

  $$p_t\left(a,\{a,b,c\}\right)\geq p_t\left(c,\{a,b,c\}\right)\geq p_t\left(b,\{a,b,c\}\right)$$

  for some $c\in C$;
- *with full support* if $p_t:\mathcal{A}\to\Delta(X)$ has full support for all $t\in(0,\infty)$.

Definition 4 *A timed random choice rule p is a* psychometric *function if:*

(i) p_0 *is uniform and has full support;*

(ii) p_∞ *is uniform;*

(iii) p *is strictly coherent.*

Property (i) says that without time to decide ($t = 0$), all alternatives have an equal chance to be chosen. Choice is then purely random, as in Becker [5], or reasonless in the terminology of Rubinstein [15]. Property (ii) says that with no time constraints the alternatives that are chosen "must be" mutually indifferent "hence" chosen with the same probability. Property (iii) says that the choice process gains in accuracy: once it is probabilistically established that x dominates y, the probability of choosing x over y never drops to $1/2$ or below.

Proposition 8 *Let X be countable. The following conditions are equivalent:*

(i) p *is a psychometric function and* p_t *is consistent for all* $t \in [0, \infty]$*;*

(ii) there exists $u : X \to \mathbb{R}$*, a family of increasing functions* $\{g_t : u(X) \to (0, \infty)\}_{t \in [0,\infty]}$*, and a family of functions* $\{\varphi_t : X \to (0, \infty)\}_{t \in [0,\infty]}$ *such that:*

- $g_0 = 1_{u(X)}$ *and* $\varphi_0 = 1_X$*;*
- $g_\infty = \mathrm{id}_{u(X)}$ *and* $\varphi_\infty = 1_X$*;*
- *for each* $t \in [0, \infty]$*, each* $A \in \mathcal{A}$*, each* $a \in A$*,*

$$p_t(a, A) = \begin{cases} \dfrac{\varphi_t(a)}{\sum_{b \in \arg\max_A g_t \circ u} \varphi_t(b)} & \text{if } a \in \arg\max_A g_t \circ u \\ 0 & \text{else} \end{cases}$$

and $u(c) \geq u(d)$ *implies* $\varphi_t(c) \geq \varphi_t(d)$ *for all* $c, d \in \arg\max_A g_t \circ u$*.*

The interpretation is straightforward:

- without time to decide, $t = 0$, choice is purely random;
- without time constraints, $t = \infty$, the DM maximizes u;
- with time constraints, $0 < t < \infty$, the DM choices are still determined by u though imperfectly so (there may be noise or maybe u is being learned). Specifically, first $g_t \circ u$-dominated alternatives are excluded, the remaining ties are then broken on $\arg\max_A g_t \circ u$ by randomizing through φ_t which is, in turn, an increasing transformation of u once restricted to $\arg\max_A g_t \circ u$.

In sum, under time constraints, u still governs imperfectly both the selection of possible alternatives and the tie breaking among them.

Next we characterize softmax rules, a very important example of consistent psychometric functions, widely used in fields that range from machine learning to neuroscience.[15]

Theorem 9 *Let X be countable. The following conditions are equivalent:*

(i) p is a coherent, consistent, continuous, and bounded psychometric with full support;

(ii) there exists a utility function $u : X \to \mathbb{R}$ such that

$$p_t(a, A) = \begin{cases} \dfrac{1}{|A|} & \text{if } t = 0 \\ \dfrac{e^{tu(a)}}{\sum_{b \in A} e^{tu(b)}} & \text{if } t \in (0, \infty) \\ \dfrac{1}{|\arg\max_A u|} \delta_a(\arg\max_A u) & \text{if } t = \infty \end{cases} \tag{9}$$

for all $A \in \mathcal{A}$ and all $a \in A$.

In this case, u is unique up to an additive constant.

Inspection of the proof shows that, if X is uncountable, then the result remains true if and only if the assumption that p is separable is added to point 1.

A Appendix

The class $\mathcal{B}$ denotes a collection of non-empty subsets of X that includes all finite sets, that is, $\mathcal{A} \subseteq \mathcal{B}$. In Section 2.3, for example, $\mathcal{B}$ is the collection of all non-empty finite sets as well as all the budget sets. We denote by A and B generic elements of $\mathcal{B}$. Let $p : \mathcal{B} \to \Delta(X)$ be a consistent random choice rule, that is, $p(\cdot, A) \in \Delta(A)$ for all $A \in \mathcal{B}$ and

$$p(a, B) = p(a, A)\, p(A, B) \qquad \forall a \in A \subseteq B \tag{10}$$

By $\sigma_p : \mathcal{B} \rightrightarrows X$, we denote the support correspondence. Note that $\sigma_p(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$. Given $B \in \mathcal{B}$, for ease of notation, we might alternatively denote by $\tilde{B}$ the support of $p(\cdot, B)$. By (10), $p(a, B) = p\left(a, \tilde{B}\right)$ for all $a \in \tilde{B}$. In particular, given $C \in \mathcal{B}$ such that $C \subseteq B$,

$$p(C, B) = p\left(C \cap \tilde{B}, \tilde{B}\right) \tag{11}$$

[15] See, e.g., Sutton and Barto [18], Vermorel and Mohri [19], Soltani and Wang [17]. The term softmax is due to Bridle [6].

Lemma 10 *Let $A \subseteq B$ with $A, B \in \mathcal{B}$. The following statements are true:*

1. $A \cap \tilde{B} = \tilde{A} \cap \tilde{B}$ *and* $p(A, B) = p\left(\tilde{A} \cap \tilde{B}, \tilde{B}\right)$*;*

2. $p(A, B) > 0$ *if and only if* $A \cap \tilde{B} = \tilde{A}$*.*

Proof 1. By definition of $\tilde{A}$, $\tilde{A} \cap \tilde{B} \subseteq A \cap \tilde{B}$. Viceversa, consider $a \in A \cap \tilde{B}$. By contradiction, assume that $a \notin \tilde{A} \cap \tilde{B}$. This implies that $a \notin \tilde{A}$, that is, $p(a, A) = 0$. By (10), we can conclude that $p(a, B) = 0$, that is, $a \notin \tilde{B}$ a contradiction with $a \in A \cap \tilde{B}$. We conclude that $A \cap \tilde{B} \subseteq \tilde{A} \cap \tilde{B}$, proving the equality between the two sets. Moreover, by definition and since $p(a, B) = p\left(a, \tilde{B}\right)$ for all $a \in \tilde{B}$, we have that

$$p(A, B) = \sum_{a \in A \cap \tilde{B}} p(a, B) = \sum_{a \in A \cap \tilde{B}} p\left(a, \tilde{B}\right) = \sum_{a \in \tilde{A} \cap \tilde{B}} p\left(a, \tilde{B}\right) = p\left(\tilde{A} \cap \tilde{B}, \tilde{B}\right)$$

2. By (10), $p(a, B) = p(a, A)\, p(A, B)$ for all $a \in A$. This implies that if $p(A, B) > 0$ and $a \in A$, then $p(a, B) > 0$ if and only if $p(a, A) > 0$. It follows that $A \cap \tilde{B} = \tilde{A}$. As to the converse, assume that $A \cap \tilde{B} = \tilde{A}$. Viceversa, since $A, \tilde{A} \neq \emptyset$, if $A \cap \tilde{B} = \tilde{A}$, then there exists $a \in A$ which belongs to $\tilde{B}$. We can conclude that $p(a, B) > 0$, that is, $p(A, B) > 0$. ■

Let $f : X \to V$ be a function that takes values in a vector space V. It can be extended to $\mathcal{B}$ by defining $\phi : \mathcal{B} \to V$ as the average $\phi(B) = \sum_{a \in \tilde{B}} f(a)\, p(a, B)$ of f with respect to p. Since $p(a, B) = p\left(a, \tilde{B}\right)$ for all $a \in \tilde{B}$ and $\sigma_p(B) = \sigma_p\left(\tilde{B}\right)$, we have that $\phi(B) = \phi(\tilde{B})$ for all $B \in \mathcal{B}$.

Proposition 11 *If the sets $\{B_i\}_{i=1}^n \subseteq \mathcal{B}$ are pairwise disjoint and $B = \bigcup_{i=1}^n B_i \in \mathcal{B}$, then*

$$\phi(B) = \sum_{i=1}^n p(B_i, B)\, \phi(B_i)$$

Proof By Lemma 10 and since $B_i \subseteq B$, it follows $B_i \cap \tilde{B} = \tilde{B}_i \cap \tilde{B} \in \mathcal{A}$ for all $i \in \{1, ..., n\}$. This implies that:[16] (a) $\tilde{B}_i \cap \tilde{B}$ are pairwise disjoint and (b)

$$\left(\bigcup_{i=1}^n \tilde{B}_i\right) \cap \tilde{B} = \bigcup_{i=1}^n \left(\tilde{B}_i \cap \tilde{B}\right) = \bigcup_{i=1}^n \left(B_i \cap \tilde{B}\right) = \left(\bigcup_{i=1}^n B_i\right) \cap \tilde{B} = B \cap \tilde{B} = \tilde{B}$$

Let $I = \left\{i : B_i \cap \tilde{B} \neq \emptyset\right\}$. By the previous equality, I is not empty. On the one hand, by Lemma 10 and since $p(B_i, B) > 0$ for all $i \in I$, we have that $B_i \cap \tilde{B} = \tilde{B}_i$ for all $i \in I$. On the other hand, if $i \notin I$, then $B_i \cap \tilde{B} = \emptyset$, yielding that $p(B_i, B) = 0$ for all

[16] To ease notation, we write $\tilde{B}_i$ in place of $\widetilde{B_i}$.

$i \notin I$. Thus, if $i \notin I$, there is no $a \in B_i$ such that $p(a, B) > 0$. Moreover, by (11), we can conclude that for each $i \in I$

$$0 < p(B_i, B) = p\left(B_i \cap \tilde{B}, \tilde{B}\right) = p\left(\tilde{B}_i \cap \tilde{B}, \tilde{B}\right)$$

and by consistency and since $\tilde{B} \supseteq \tilde{B}_i \cap \tilde{B} = \tilde{B}_i = B_i \cap \tilde{B} \neq \emptyset$ and $\tilde{B}_i \cap \tilde{B} \in \mathcal{A}$ for all $i \in I$, we have that

$$p\left(a, \tilde{B}\right) = p\left(a, \tilde{B}_i \cap \tilde{B}\right) p\left(\tilde{B}_i \cap \tilde{B}, \tilde{B}\right) = p\left(a, \tilde{B}_i\right) p\left(\tilde{B}_i \cap \tilde{B}, \tilde{B}\right)$$

for all $a \in \tilde{B}_i \cap \tilde{B} = \tilde{B}_i$ and for all $i \in I$. By Lemma 10 and since the elements of $\left\{\tilde{B}_i \cap \tilde{B}\right\}_{i \in I}$ are non-empty, pairwise disjoint, and finite, we have

$$\begin{aligned}
\phi(B) &= \phi\left(\tilde{B}\right) = \sum_{a \in \tilde{B}} f(a)\, p\left(a, \tilde{B}\right) = \sum_{a \in \bigcup_{i=1}^{n}\left(\tilde{B}_i \cap \tilde{B}\right)} f(a)\, p\left(a, \tilde{B}\right) \\
&= \sum_{a \in \bigcup_{i \in I}\left(\tilde{B}_i \cap \tilde{B}\right)} f(a)\, p\left(a, \tilde{B}\right) \\
&= \sum_{i \in I} \sum_{a \in \tilde{B}_i \cap \tilde{B}} f(a)\, p\left(a, \tilde{B}\right) = \sum_{i \in I} \sum_{a \in \tilde{B}_i} f(a)\, p\left(a, \tilde{B}\right) \\
&= \sum_{i \in I} \sum_{a \in \tilde{B}_i} f(a)\, p\left(a, \tilde{B}_i\right) p\left(\tilde{B}_i \cap \tilde{B}, \tilde{B}\right) \\
&= \sum_{i \in I} p\left(\tilde{B}_i \cap \tilde{B}, \tilde{B}\right) \sum_{a \in \tilde{B}_i} f(a)\, p\left(a, \tilde{B}_i\right) = \sum_{i \in I} p\left(\tilde{B}_i \cap \tilde{B}, \tilde{B}\right) \sum_{a \in \tilde{B}_i} f(a)\, p(a, B_i) \\
&= \sum_{i \in I} p\left(B_i \cap \tilde{B}, \tilde{B}\right) \phi(B_i) = \sum_{i \in I} p\left(B_i \cap \tilde{B}, B\right) \phi(B_i) \\
&= \sum_{i=1}^{n} p\left(B_i \cap \tilde{B}, B\right) \phi(B_i) = \sum_{i=1}^{n} p(B_i, B)\, \phi(B_i),
\end{aligned}$$

proving the statement. ■

B Proofs

B.1 Atemporal setting

The main goal of this section is to prove Theorem 1. We start by introducing some new notation and some preparatory result. Recall that for each $a, b \in X$

$$p(a, b) = p(a, \{a, b\}) \quad \text{and} \quad r(a, b) = \frac{p(a, b)}{p(b, a)}$$

Lemma 12 *If $p : \mathcal{A} \to \Delta(X)$ is a consistent random choice rule, then σ_p is optimal.*

Proof Clearly, $\emptyset \neq \sigma_p(A) \subseteq A$ for all $A \in \mathcal{A}$. Let $A, B \in \mathcal{A}$ such that $A \subseteq B$. Assume that $\sigma_p(B) \cap A \neq \emptyset$. We want to show that $\sigma_p(B) \cap A = \sigma_p(A)$. Since p is consistent, if $a \in \sigma_p(B) \cap A$, then $0 < p(a, B) = p(a, A)\, p(A, B)$. It follows that $p(a, A) > 0$, that is, $a \in \sigma_p(A)$. Thus, $\sigma_p(B) \cap A \subseteq \sigma_p(A)$. As to the converse inclusion, let $a \in \sigma_p(A) \subseteq A$, that is, $p(a, A) > 0$. By contradiction, assume that $a \notin \sigma_p(B)$, that is, $p(a, B) = 0$. Since p is consistent, we then have $0 = p(a, B) = p(a, A)\, p(A, B)$. Since $p(a, A) > 0$, this implies that $p(A, B) = 0$, that is, $\sigma_p(B) \cap A = \emptyset$. This contradicts $\sigma_p(B) \cap A \neq \emptyset$, proving the opposite inclusion and completing the proof.■

Proof of Theorem 1 "If". Let p be given by (4) with σ optimal and $\varphi : X \to (0, \infty)$. Since φ is strictly positive and σ a choice correspondence, p is a well defined random choice rule. Let $A, B \in \mathcal{A}$ such that $A \subseteq B$ and $a \in A$. We have two cases:

1. $\sigma(B) \cap A \neq \emptyset$. Since σ is optimal, $\sigma(B) \cap A = \sigma(A)$. On the one hand, by (4), if $a \in \sigma(A)$, then $a \in \sigma(B)$ and $p(a, A) = \varphi(a) / \sum_{b \in \sigma(A)} \varphi(b)$. We can conclude that

$$p(a, B) = \frac{\varphi(a)}{\sum_{b \in \sigma(B)} \varphi(b)} = \frac{\varphi(a)}{\sum_{b \in \sigma(A)} \varphi(b)} \frac{\sum_{b \in \sigma(B) \cap A} \varphi(b)}{\sum_{b \in \sigma(B)} \varphi(b)} = p(a, A)\, p(A, B)$$

   On the other hand, if $a \notin \sigma(A)$, we have that $a \in A \backslash \sigma(B)$, so $p(a, A) = 0 = p(a, B)$. In both cases (1) holds.

2. $\sigma(B) \cap A = \emptyset$. It follows that $a \notin \sigma(B)$ and $p(A, B) = 0 = p(a, B)$. Again, (1) holds.

Cases (i) and (ii) prove that p is consistent.

"Only if". Let $p : \mathcal{A} \to \Delta(X)$ be a consistent random choice rule and set

$$a \succsim b \iff a \in \sigma_p(\{a, b\})$$

By Lemma 12, σ_p is optimal. Since σ_p is optimal, $\succsim$ is a weak order and

$$\begin{aligned} a &\succsim b \iff p(a, b) > 0 \\ a &\succ b \iff p(a, b) = 1 \\ a &\sim b \iff p(a, b) \in (0, 1) \end{aligned}$$

By [2] and since σ_p is optimal, observe that

$$\sigma_p(A) = \{a \in A : a \succsim b \quad \forall b \in A\} \qquad \forall A \in \mathcal{A} \tag{12}$$

Let $\{X_i : i \in I\}$ be the family of all equivalence classes of X with respect to $\sim$. Choose $a_i \in X_i$ for all $i \in I$. Define $\varphi : X \to (0, \infty)$ to be such that

$$\varphi(x) = r(x, a_i) = \frac{p(x, a_i)}{p(a_i, x)} \qquad \forall x \in X_i, \forall i \in I \tag{13}$$

Consider $i \in I$ and let $x \sim a_i$, we have that $p(x, a_i), p(a_i, x) \in (0, 1)$. Since i and x were arbitrarily chosen, it follows that φ is well defined. By (12), we have that

$$\sigma_p(S) = S \qquad \forall S \in \mathcal{A} \text{ such that } S \subseteq X_i \text{ for some } i \in I \tag{14}$$

This implies that for each $S \in \mathcal{A}$ such that $S \subseteq X_i$ and for each $a, b \in S$

$$\begin{aligned} p(a, S) &= p(a, b)\, p(\{a, b\}, S) > 0 \\ p(b, S) &= p(b, a)\, p(\{a, b\}, S) > 0 \end{aligned}$$

yielding that

$$\frac{p(a, S)}{p(b, S)} = \frac{p(a, b)}{p(b, a)} = r(a, b) \tag{15}$$

Next, consider $a, b \in X$ such that $a \sim b$. We have that there exists $i \in I$ such that $a \sim b \sim a_i$. By consistency and definition of φ, we can conclude that

$$\begin{aligned} p(a, \{a, b, a_i\}) &= p(a, a_i)\, p(\{a, a_i\}, \{a, b, a_i\}) = \frac{p(a, a_i)}{p(a_i, a)} p(a_i, a)\, p(\{a, a_i\}, \{a, b, a_i\}) \\ &= \frac{p(a, a_i)}{p(a_i, a)} p(a_i, \{a, b, a_i\}) = \varphi(a)\, p(a_i, \{a, b, a_i\}) \\ p(b, \{a, b, a_i\}) &= p(b, a_i)\, p(\{b, a_i\}, \{a, b, a_i\}) = \frac{p(b, a_i)}{p(a_i, b)} p(a_i, b)\, p(\{b, a_i\}, \{a, b, a_i\}) \\ &= \frac{p(b, a_i)}{p(a_i, b)} p(a_i, \{a, b, a_i\}) = \varphi(b)\, p(a_i, \{a, b, a_i\}) \end{aligned}$$

By (14), we have that $p(a_i, \{a, b, a_i\}) > 0$ and $p(b, \{a, b, a_i\}) > 0$. By applying (15) twice, we can conclude that

$$\frac{p(a, S)}{p(b, S)} = r(a, b) = \frac{p(a, \{a, b, a_i\})}{p(b, \{a, b, a_i\})} = \frac{\varphi(a)}{\varphi(b)} \quad \forall S \in \mathcal{A} \text{ such that } a, b \in S \subseteq X_i \tag{16}$$

By consistency and since $p(\sigma_p(A), A) = 1$, we also have that

$$p(a, A) = p(a, \sigma_p(A))\, p(\sigma_p(A), A) = p(a, \sigma_p(A)) \qquad \forall a \in \sigma_p(A)$$

We are ready to conclude our proof, that is, proving (4) where $\sigma = \sigma_p$. We have two cases:

1. $a \notin \sigma_p(A)$. It trivially follows that $p(a, A) = 0$.

2. $a \in \sigma_p(A)$. By (12), all the elements in $\sigma_p(A)$ are equivalent with respect to $\succsim$ and therefore they are equivalent to some a_i with $i \in I$. It follows that $\sigma_p(A) \cup \{a_i\} \in \mathcal{A}$ and it is such that $\sigma_p(A) \cup \{a_i\} \subseteq X_i$. By (14), we have that $\sigma_p(\sigma_p(A) \cup \{a_i\}) = \sigma_p(A) \cup \{a_i\}$. By consistency and (15) and since $a \in \sigma_p(A)$, we can conclude that

$$\begin{aligned} p(a, A) &= p(a, \sigma_p(A)) = \frac{p(a, \sigma_p(A) \cup \{a_i\})}{p(\sigma_p(A), \sigma_p(A) \cup \{a_i\})} = \frac{\frac{p(a,\sigma_p(A)\cup\{a_i\})}{p(a_i,\sigma_p(A)\cup\{a_i\})}}{\frac{p(\sigma_p(A),\sigma_p(A)\cup\{a_i\})}{p(a_i,\sigma_p(A)\cup\{a_i\})}} \\ &= \frac{r(a, a_i)}{\sum_{b\in\sigma_p(A)} \frac{p(b,\sigma_p(A)\cup\{a_i\})}{p(a_i,\sigma_p(A)\cup\{a_i\})}} = \frac{r(a, a_i)}{\sum_{b\in\sigma_p(A)} r(b, a_i)} = \frac{\varphi(a)}{\sum_{b\in\sigma_p(A)} \varphi(b)} \end{aligned}$$

as wanted.

As for the uniqueness part, in both cases, we can consider p as in (4). It is immediate to conclude that $\sigma_p = \sigma$. ■

Proof of Corollary 2 Before entering the proof details note that, by definition and since p is uniform,

$$\begin{aligned} a &\succ b \iff p(a,b) > p(b,a) \iff p(a,b) = 1 \\ a &\sim b \iff p(a,b) = p(b,a) \iff p(a,b) = \frac{1}{2} \\ a &\prec b \iff b \succ a \iff p(b,a) = 1 \iff p(a,b) = 0 \end{aligned}$$

In particular, $a \succsim b$ if and only if $a \in \sigma_p(\{a,b\})$.[17]

By Theorem 1, if p is consistent, then σ_p is optimal. Conversely, if p is uniform and σ_p is optimal, then $\varphi = 1_X$ is strictly positive, $\sigma = \sigma_p$ is optimal, and, for each $A \in \mathcal{A}$,

$$p(a, A) = \begin{cases} \frac{\varphi(a)}{\sum_{b\in\sigma(A)} \varphi(b)} & \text{if } a \in \sigma(A) \\ 0 & \text{else} \end{cases}$$

therefore, by Theorem 1 again, p is consistent. Since in both cases, $\sigma_p = \sigma$ is consistent, by [2], it follows that $\succsim$ is a weak order and

$$\sigma_p(A) = \{a \in A : a \succsim b \quad \forall b \in A\} \qquad \forall A \in \mathcal{A}$$

holds. ■

B.2 Consumer theory

Before starting the proofs, recall that $X = \mathbb{R}^n_{++}$ and $\mathcal{B}$ is the collection of all non-empty finite sets and all the possible budget sets. In this section, by being consistent with the standard microeconomic literature, we will tend to denote the alternatives in X (that is, bundles) by x and y instead of a and b. We need two ancillary facts:

[17] In the words of Arrow, $\succsim$ is the binary relation generated by σ_p.

Proposition 13 *Let d be a stochastic demand and let $p : \mathcal{B} \to \Delta(X)$ be a consistent random choice rule that generates d. The following statements are true:*

1. *If $\succsim$ is the weak order that rationalizes σ_p restricted to $\mathcal{A}$, then*

$$\sigma_p(B) = \{x \in B : x \succsim y \quad \forall y \in B\} \qquad \forall B \in \mathcal{B} \tag{17}$$

2. *If $x \in \sigma_p(B(q,w))$ and p is stochastically monotone, then for each $x \in \sigma_p(B(q,w))$*

$$q \cdot x = w$$

Proof 1. Consider the random choice rule p restricted to $\mathcal{A}$. Construct $\succsim$ as in the proof of Theorem 1. Let $B \in \mathcal{B}$. By the proof of Theorem 1, if $B \in \mathcal{A}$, then (17) holds. If $B \notin \mathcal{A}$, then $B = B(q,w)$ for some $(q,w) \in \mathbb{R}^n_{++} \times \mathbb{R}_{++}$. As before, define $\tilde{B} = \sigma_p(B)$. We next prove (17) by proving both inclusions. By contradiction, assume that $\sigma_p(B) \not\subseteq \{x \in B : x \succsim y \quad \forall y \in B\}$. Since $\sigma_p(B) \subseteq B$, it follows that there exists $x \in B$ such that $x \succ y$ for some $y \in \sigma_p(B)$. Since $\tilde{B} = \sigma_p(B) \in \mathcal{A}$ and $p\left(y, \tilde{B}\right) = p(y,B)$ for all $y \in \tilde{B}$, we have that $\tilde{B} = \sigma_p(B) = \sigma_p\left(\tilde{B}\right)$. By Theorem 1, we can conclude that all the elements in $\sigma_p\left(\tilde{B}\right)$ are indifferent according to the weak order $\succsim$, thus, $x \succ y$ for all $y \in \tilde{B}$, as well as $x \notin \tilde{B}$. By Theorem 1, if we define $A = \tilde{B} \cup \{x\} \subseteq B$, then $A \in \mathcal{A}$, $p(x,A) = 1$, and $p(A,B) = 1$. By consistency, this implies that

$$p(x,B) = p(x,A)\,p(A,B) = 1$$

a contradiction with $x \notin \tilde{B}$, proving the $\subseteq$ inclusion and that $\{x \in B : x \succsim y \quad \forall y \in B\}$ is non-empty. Viceversa, assume that $\bar{x} \in B$ is such that $\bar{x} \succsim y$ for all $y \in B$. Define $A = \tilde{B} \cup \{\bar{x}\} \in \mathcal{A}$. By the previous part of the proof, we have that all the elements of A are indifferent according to $\succsim$. By Theorem 1 and consistency and since $1 \leq p\left(\tilde{B}, B\right) \leq p(A,B) \leq 1$, we have that

$$p(\bar{x},B) = p(\bar{x},A)\,p(A,B) = p(\bar{x},A) > 0$$

proving that $\bar{x} \in \tilde{B}$ and the opposite inclusion.

2. By contradiction and since $x \in \sigma_p(B(q,w)) \subseteq B(q,w)$, assume that $q \cdot x < w$. It follows that there exists $y \in B(q,w)$ such that $x \ll y$ and $y \in B(q,w)$. By consistency and stochastic monotonicity and since $x \in \sigma_p(B(q,w))$ and $x \in \{x,y\} \subseteq B(q,w)$, it follows that

$$0 < p(x, B(q,w)) = p(x,\{x,y\})\,p(\{x,y\}, B(q,w)) = 0$$

a contradiction. ∎

Proof of Lemma 4 In Proposition 11, define $f : X \to \mathbb{R}$ by $f(x) = q \cdot x$ for all $x \in X$. Given a stochastic demand d, let p be such that $d(q,w)(\cdot) = p(\cdot, B(q,w))$. Set also $A = B(q,w)$ and $B = B(q,w')$. Clearly, we have that $A \subseteq B$ and $A, B \in \mathcal{B}$. By Lemma 10, it follows $A \cap \tilde{B} = \tilde{A} \cap \tilde{B}$. Note that $B_1 = \tilde{A} \cap \tilde{B}$ and $B_2 = \tilde{B} \backslash \tilde{A}$ belong to $\mathcal{A} \subseteq \mathcal{B}$ (provided they are not empty), are pairwise disjoint, and $\tilde{B} = B_1 \cup B_2 \in \mathcal{A}$. Moreover, $B_2 \subseteq A^c$. Otherwise, there would exist $x \in B_2 \cap A$. Thus, we would have that $x \in B_2$ and $x \in A$. By consistency, we could conclude that $x \in \tilde{B}$, $x \notin \tilde{A}$, and $x \in A \subseteq B$

$$0 < p(x,B) = p(x,A)\,p(A,B) = 0$$

a contradiction. Observe also that

$$\begin{aligned} c(q,w') &= \sum_{x \in B(q,w')} (q \cdot x)\, d(q,w')(x) = \sum_{x \in B(q,w')} (q \cdot x)\, p(x, B(q,w')) && (18) \\ &= \sum_{x \in \widetilde{B(q,w')}} (q \cdot x)\, p(x, B(q,w')) = \phi(B(q,w')) = \phi(B) && (19) \end{aligned}$$

Similarly, we have that $c(q,w) = \phi(A) = \phi\left(\tilde{A}\right)$. Finally, since $B_2 \subseteq A^c$, if $B_2 \neq \emptyset$, then we also have that $\tilde{B}_2 \subseteq B_2 \subseteq A^c$

$$\begin{aligned} \phi(B_2) &= \sum_{x \in \tilde{B}_2} (q \cdot x)\, p(x, B_2) = \sum_{x \in \tilde{B}_2 \cap A^c} (q \cdot x)\, p(x, B_2) \\ &> w \sum_{x \in \tilde{B}_2 \cap A^c} p(x, B_2) = w \sum_{x \in \tilde{B}_2} p(x, B_2) = w \end{aligned}$$

By Proposition 11 and since B_1 and B_2 are disjoint, if $B_1, B_2 \neq \emptyset$, then we can conclude that

$$c(q,w') = \phi(B) = \phi\left(\tilde{B}\right) = \phi(B_1 \cup B_2) = p\left(B_1, \tilde{B}\right)\phi(B_1) + p\left(B_2, \tilde{B}\right)\phi(B_2) \quad (20)$$

We have two cases:

1. $p(A,B) > 0$. By Lemma 10, it follows that $B_1 = \tilde{A} \cap \tilde{B} = \tilde{A} \neq \emptyset$. On the one hand, by (20), if $B_2 \neq \emptyset$, then we have that

$$\begin{aligned} c(q,w') &= \phi(B) = \phi\left(\tilde{B}\right) = \phi(B_1 \cup B_2) = p\left(B_1, \tilde{B}\right)\phi(B_1) + p\left(B_2, \tilde{B}\right)\phi(B_2) \\ &= p\left(B_1, \tilde{B}\right)\phi\left(\tilde{A}\right) + p\left(B_2, \tilde{B}\right)\phi(B_2) \geq p\left(B_1, \tilde{B}\right) c(q,w) + p\left(B_2, \tilde{B}\right) w \\ &\geq p\left(B_1, \tilde{B}\right) c(q,w) + p\left(B_2, \tilde{B}\right) c(q,w) = c(q,w) \end{aligned}$$

   On the other hand, if $B_2 = \emptyset$, then $\tilde{A} = \tilde{A} \cap \tilde{B} = B_1 = B_1 \cup B_2 = \tilde{B}$. This implies that

$$c(q,w') = \phi(B) = \phi\left(\tilde{B}\right) = \phi\left(\tilde{A}\right) = \phi(A) = c(q,w)$$

2. $p(A,B)=0$. By (11) and Lemma 10, it follows that

$$p(A,B)=p\left(A\cap\tilde{B},\tilde{B}\right)=p\left(\tilde{A}\cap\tilde{B},\tilde{B}\right)=p\left(B_1,\tilde{B}\right)=0$$

that is, $B_1=\emptyset$ and $\emptyset\neq\tilde{B}=B_2\subseteq A^c$, which immediately yields that

$$c(q,w')=\phi(B)=\phi\left(\tilde{B}\right)=\phi(B_2)>w\geq c(q,w)$$

Points 1 and 2 prove the main statement.

Next, let us assume that p is stochastically monotone. Consider $\partial^+B(q,w)=\{x\in B(q,w):q\cdot x=w\}$. We next show by contradiction that $\operatorname{supp}d(q,w)\subseteq\partial^+B(q,w)$. By contradiction, assume that there exists $x\in\operatorname{supp}d(q,w)$ such that $q\cdot x<w$. Then, there exists $z\in B(q,w)$ such that such that $x\ll z$. Since, by stochastic monotonicity, $p(x,\{x,z\})=0$, we have that

$$0<d(q,w)(x)=p(x,B(q,w))=p(x,\{x,z\})\,p(\{x,z\},B(q,w))=0$$

contradicting $x\in\operatorname{supp}d(q,w)$. This yields that $\operatorname{supp}d(q,w)\subseteq\partial^+B(q,w)$. By (18), this yields that $c(q,w)=w$. The rest of the statement trivially follows. ■

Proof of Lemma 5 Clearly, since $q<q'$, $B(q',w)\subseteq B(q,w)$. Define $A=B(q',w)$ and $B=B(q,w)$. In Proposition 11, let $f(x)=x$. Given a stochastic demand d, let p be such that $d(q,w)(\cdot)=p(\cdot,B(q,w))$. Set also $A=B(q',w)$ and $B=B(q,w)$. Clearly, we have that $A\subseteq B$ and $A,B\in\mathcal{B}$. By Lemma 10, it follows $A\cap\tilde{B}=\tilde{A}\cap\tilde{B}$. Note that $B_1=\tilde{A}\cap\tilde{B}$ and $B_2=\tilde{B}\backslash\tilde{A}$ belong to $\mathcal{A}\subseteq\mathcal{B}$ (provided they are not empty), are pairwise disjoint, and $\tilde{B}=B_1\cup B_2$. Moreover, $B_2\subseteq A^c$. Otherwise, there would exist $x\in B_2\cap A$. Thus, we would have that $x\in B_2$ and $x\in A$. By consistency, we could conclude that $x\in\tilde{B}$, $x\notin\tilde{A}$, and $x\in A\subseteq B$

$$0<p(x,B)=p(x,A)\,p(A,B)=0$$

a contradiction. Observe that

$$\begin{aligned}\bar{d}(q,w) &= \sum_{x\in B(q,w)} xd(q,w)(x)=\sum_{x\in B(q,w)} xp(x,B(q,w))\\ &= \sum_{x\in\widetilde{B(q,w)}} xp(x,B(q,w))=\phi(B(q,w))=\phi(B)\end{aligned}$$

Similarly, we have that $\bar{d}(q',w)=\phi(A)=\phi\left(\tilde{A}\right)$. Finally, since $B_2\subseteq A^c$, if $B_2\neq\emptyset$, then we also have that $\tilde{B}_2\subseteq B_2\subseteq A^c$

$$\phi(B_2)=\sum_{x\in\tilde{B}_2} xp(x,B_2)=\sum_{x\in\tilde{B}_2\cap A^c} xp(x,B_2)\ \text{ and }\ q'\cdot\phi(B_2)=\sum_{x\in\tilde{B}_2\cap A^c}(q'\cdot x)\,p(x,B_2)>w$$

By Proposition 11 and since B_1 and B_2 are disjoint, if $B_1, B_2 \neq \emptyset$, then we have that

$$\bar{d}(q,w) = \phi(B) = \phi\left(\tilde{B}\right) = \phi(B_1 \cup B_2) = p\left(B_1, \tilde{B}\right)\phi(B_1) + p\left(B_2, \tilde{B}\right)\phi(B_2) \quad (21)$$

By (11) and Lemma 10 and since $A \subseteq B$, recall that

$$p(A,B) = p\left(A \cap \tilde{B}, \tilde{B}\right) = p\left(\tilde{A} \cap \tilde{B}, \tilde{B}\right) = p\left(B_1, \tilde{B}\right) \quad (22)$$

By contradiction, assume that $\bar{d}(q',w) > \bar{d}(q,w)$. We have three cases:

1. $p\left(B_2, \tilde{B}\right) = 0$. This implies that $p\left(B_1, \tilde{B}\right) = 1$. By Lemma 10 and (22), it follows that $B_1 = \tilde{A} \cap \tilde{B} = \tilde{A}$ as well as $B_2 = \emptyset$, that is, $\tilde{B} = B_1 = \tilde{A}$. We can conclude that

$$\bar{d}(q,w) = \phi(B) = \phi\left(\tilde{B}\right) = \phi\left(\tilde{A}\right) = \bar{d}(q',w)$$

   a contradiction.

2. $1 > p\left(B_2, \tilde{B}\right) > 0$. This implies that $1 > p\left(B_1, \tilde{B}\right) > 0$. In particular, we have that $B_1, B_2 \neq \emptyset$. By Lemma 10 and (22), it follows that $B_1 = \tilde{A} \cap \tilde{B} = \tilde{A}$. By (21), we have that

$$\begin{aligned} \bar{d}(q,w) &= p\left(B_1, \tilde{B}\right)\phi(B_1) + p\left(B_2, \tilde{B}\right)\phi(B_2) \\ &= p\left(B_1, \tilde{B}\right)\phi\left(\tilde{A}\right) + p\left(B_2, \tilde{B}\right)\phi(B_2) \\ &= p\left(B_1, \tilde{B}\right)\bar{d}(q',w) + p\left(B_2, \tilde{B}\right)\phi(B_2) \end{aligned}$$

   This yields that

$$0 > \bar{d}(q,w) - \bar{d}(q',w) = p\left(B_2, \tilde{B}\right)\left[\sum_{x \in \tilde{B}_2} xp(x, B_2) - \bar{d}(q',w)\right]$$

   that is,

$$0 > \sum_{x \in \tilde{B}_2} xp(x,B_2) - \bar{d}(q',w) \implies \bar{d}(q',w) > \sum_{x \in \tilde{B}_2} xp(x,B_2)$$

   In turn, since $\tilde{B}_2 \subseteq B_2 \subseteq A^c$, this yields that

$$w \geq q' \cdot \bar{d}(q',w) \geq q' \cdot \left(\sum_{x \in \tilde{B}_2} xp(x,B_2)\right) = \sum_{x \in \tilde{B}_2 \cap A^c} (q' \cdot x)\, p(x,B_2) > w$$

   a contradiction.

3. $p\left(B_2,\tilde{B}\right)=1$. This implies that $p\left(B_1,\tilde{B}\right)=0$. In particular, we have that $B_1=\emptyset$ and $\emptyset\neq\tilde{B}=B_2$. This implies that
$$\bar{d}(q,w)=\phi(B)=\phi\left(\tilde{B}\right)=\phi(B_2)$$
yielding that
$$w\geq q'\cdot\bar{d}(q',w)\geq q'\cdot\bar{d}(q,w)=q'\cdot\phi(B_2)>w$$
a contradiction.

Points 1, 2, and 3 prove the statement. ■

Proof of Lemma 6 We first prove an ancillary claim:
Claim Let $\bar{B}=B(\bar{q},\bar{w})$ and $\hat{B}=B(\hat{q},\hat{w})$. If $p\left(\bar{B}\cap\hat{B},\hat{B}\right)=0$, then
$$\bar{q}\cdot\bar{d}(\hat{q},\hat{w})>\bar{w}$$

Proof of the Claim By assumption, the support of $p\left(\cdot,\hat{B}\right)$ is contained in $\hat{B}\cap\bar{B}^c$, in particular, $\widetilde{\hat{B}}\subseteq\bar{B}^c$. It follows that
$$\bar{q}\cdot\bar{d}(\hat{q},\hat{w})=\sum_{x\in\widetilde{\hat{B}}}(\bar{q}\cdot x)\,p\left(x,\hat{B}\right)>\sum_{x\in\widetilde{\hat{B}}}\bar{w}p\left(x,\hat{B}\right)=\bar{w}$$
proving the claim. □

Consider (q,w) and (q',w') in $\mathbb{R}^n_{++}\times\mathbb{R}_{++}$. Define $B=B(q,w)$ and $B'=B(q',w')$. By the previous claim and setting $\bar{B}=B'$ and $\hat{B}=B$, it follows that $p(B'\cap B,B)>0$. Otherwise, we would have that $w'<q'\cdot\bar{d}(q,w)=c(q',w')\leq w'$, a contradiction. Since $p(B'\cap B,B)>0$, denote by $\bar{x}\in\sigma_p(B)\cap B'$. By Proposition 13 and since $\bar{x}\in\sigma_p(B)$, we have that $\bar{x}\succsim y$ for all $y\in B$. We have two cases:

1. $p(B'\cap B,B')=0$. By the previous claim and setting $\bar{B}=B$ and $\hat{B}=B'$, it follows that $q\cdot\bar{d}(q',w')>w$.

2. $p(B'\cap B,B')>0$. This implies that
$$q\cdot\bar{d}(q',w')=\sum_{x\in\widetilde{B'}}(q\cdot x)\,p(x,B')=\sum_{x\in\widetilde{B'}\cap B}(q\cdot x)\,p(x,B')+\sum_{x\in\widetilde{B'}\cap B^c}(q\cdot x)\,p(x,B')\tag{23}$$
Given $x\in\widetilde{B'}$, we have two subcases:

   (a) $x\in\widetilde{B'}\cap B^c$. In this case, $q\cdot x>w$.

(b) $x \in \widetilde{B'} \cap B = \sigma_p(B') \cap B$. By Proposition 13, it follows that $x \succsim y$ for all $y \in B'$. In particular, since $\bar{x} \in B'$, this implies that $x \succsim \bar{x}$. At the same time, since $\bar{x} \in B$ is such that $\bar{x} \succsim y$ for all $y \in B$, we have that $x \succsim y$ for all $y \in B$. By Proposition 13 and since $x \in B$, this yields that $x \in \sigma_p(B)$. By stochastic monotonicity, we can conclude that $q \cdot x = w$.

To sum up, by (23), we can conclude that

$$q \cdot \bar{d}(q', w') \geq \sum_{x \in \widetilde{B'} \cap B} wp(x, B') + \sum_{x \in \widetilde{B'} \cap B^c} wp(x, B') = w$$

In both cases, we have that

$$\begin{aligned}
(q' - q) \cdot \left(\bar{d}(q', w') - \bar{d}(q, w)\right) &= q' \cdot \left(\bar{d}(q', w') - \bar{d}(q, w)\right) - q \cdot \left(\bar{d}(q', w') - \bar{d}(q, w)\right) \\
&= q' \cdot \bar{d}(q', w') - q' \cdot \bar{d}(q, w) - q \cdot \left(\bar{d}(q', w') - \bar{d}(q, w)\right) \\
&= w' - w' + w - q \cdot \bar{d}(q', w') = w - q \cdot \bar{d}(q', w')
\end{aligned}$$

By points 1 and 2, this implies the main statement. ■

Proof of the Law of Demand Consider an initial price and wealth pair (q, w). Let $q' \in \mathbb{R}^n_{++}$ be such that $q'_k > q_k$ and $q'_i = q_i$ for all $i \neq k$. Let $w' = w'(q')$ be such that $w' = q' \cdot \bar{d}(q, w) \geq q \cdot \bar{d}(q, w) = w$, since p is stochastically monotone. By Lemma 4 and since p is stochastically monotone, it follows $q' \cdot \bar{d}(q, w) = w' = c(q', w')$. In view of Lemma 6, the difference $\bar{d}(q', w') - \bar{d}(q, w)$ quantifies a substitution effect on the goods' average demand due only to the price change $q' - q$. This suggests the following decomposition:

$$\bar{d}(q', w) - \bar{d}(q, w) = \underbrace{\bar{d}(q', w) - \bar{d}(q', w')}_{\text{wealth effect}} + \underbrace{\bar{d}(q', w') - \bar{d}(q, w)}_{\text{substitution effect}} \tag{24}$$

in which the r.h.s. accounts for, respectively, the wealth and substitution effects on the goods' demand. Note that the elements in (24) are vectors. Thus, the equality holds componentwise. Since good k is normal and $w \leq w'$, we have

$$\bar{d}_k(q', w) \leq \bar{d}_k(q', w') \tag{25}$$

By Lemma 6 and the choice of q and q' and $q' \cdot \bar{d}(q, w) = w' = c(q', w')$, we have that

$$(q'_k - q_k)\left(\bar{d}_k(q', w') - \bar{d}_k(q, w)\right) = (q' - q) \cdot \left(\bar{d}(q', w') - \bar{d}(q, w)\right) \leq 0$$

Since $q'_k - q_k > 0$, it follows that $\bar{d}_k(q', w') - \bar{d}_k(q, w) \leq 0$. By (24) and (25), this implies that

$$\bar{d}_k(q', w) - \bar{d}_k(q, w) = \left[\bar{d}_k(q', w) - \bar{d}_k(q', w')\right] + \left[\bar{d}_k(q', w') - \bar{d}_k(q, w)\right] \leq 0$$

proving the statement. Indeed, $\bar{d}_k(q', w) \leq \bar{d}_k(q, w)$ where in q' only the price of k increased, while the other prices did not change and wealth remained constant. ■

B.3 Temporal setting

Proof of Proposition 8 Assume (i) holds. By Corollary 2, the binary relation defined on X by

$$x \succsim_\infty y \iff p_\infty(x,y) \geq p_\infty(y,x)$$

is a weak order and, writing σ_t instead of σ_{p_t} for all $t \in [0,\infty]$,

$$\sigma_\infty(A) = \{a \in A : a \succsim_\infty b \text{ for all } b \in A\}$$

for all $A \in \mathcal{A}$. Since X is countable, then there exists $v_\infty : X \to \mathbb{R}$ such that

$$x \succsim_\infty y \iff v_\infty(x) \geq v_\infty(y)$$

and $\sigma_\infty(A) = \arg\max_A v_\infty$ for all $A \in \mathcal{A}$.

By Theorem 1, for each $t \in [0,\infty]$, σ_t is an optimal choice correspondence, therefore, the binary relation defined on X by

$$x \succsim_t y \iff x \in \sigma_t(\{x,y\}) \iff p_t(x,y) > 0$$

is a weak order and

$$\sigma_t(A) = \{a \in A : a \succsim_t b \text{ for all } b \in A\}$$

for all $A \in \mathcal{A}$. Since X is countable, then there exists $v_t : X \to \mathbb{R}$ such that

$$x \succsim_t y \iff v_t(x) \geq v_t(y)$$

and $\sigma_t(A) = \arg\max_A v_t$ for all $A \in \mathcal{A}$. Moreover, since p_0 has full support, we can assume $v_0 = 1_X$.

By Theorem 1, again, there exists a family of functions $\{\varphi_t : X \to (0,\infty)\}_{t\in[0,\infty]}$ such that

$$p_t(a,A) = \begin{cases} \dfrac{\varphi_t(a)}{\sum_{b\in\arg\max_A v_t}\varphi_t(b)} & \text{if } a \in \arg\max_A v_t \\ 0 & \text{else} \end{cases}$$

for each $t \in [0,\infty]$, each $A \in \mathcal{A}$, and each $a \in A$. Moreover, since p_0 and p_∞ are uniform, we can assume $\varphi_0 = \varphi_\infty = 1_X$.

Next we show that, for each $t \in (0,\infty)$ and each $A \in \mathcal{A}$, $v_\infty(x) \geq v_\infty(y)$ implies $v_t(x) \geq v_t(y)$ and, if $x, y \in \arg\max_A v_t$, also $\varphi_t(x) \geq \varphi_t(y)$. Indeed, strict coherence implies that

$$p_\infty(x,y) \geq p_\infty(y,x) \implies p_t(x,y) \geq p_t(y,x)$$

therefore

$$\begin{aligned} v_\infty(x) \geq v_\infty(y) \quad &\iff \quad p_\infty(x,y) \geq p_\infty(y,x) \implies p_t(x,y) \geq p_t(y,x) \\ &\implies \quad p_t(x,y) > 0 \iff x \succsim_t y \iff v_t(x) \geq v_t(y) \end{aligned}$$

moreover, if $x, y \in \arg\max_A v_t = \sigma_t(A)$, then

$$\begin{aligned} v_\infty(x) \geq v_\infty(y) \quad &\Longrightarrow \quad p_t(x,y) \geq p_t(y,x) \iff p_t(x,y)\, p(\{x,y\}, A) \geq p_t(y,x)\, p(\{x,y\}, A) \\ &\iff \quad p(x,A) \geq p(y,A) \iff \varphi_t(x) \geq \varphi_t(y) \end{aligned}$$

But also

- for $t = 0$, $v_\infty(x) \geq v_\infty(y)$ implies $v_0(x) = 1 \geq 1 = v_0(y)$ and, if $x, y \in \arg\max_A v_0$, also $\varphi_0(x) = 1 \geq 1 = \varphi_t(y)$.
- for $t = \infty$, $v_\infty(x) \geq v_\infty(y)$ implies $v_\infty(x) \geq v_\infty(y)$ and, if $x, y \in \arg\max_A v_\infty$, also $\varphi_\infty(x) = 1 \geq 1 = \varphi_\infty(y)$.

Setting $u = v_\infty$, $g_0 = 1_{u(X)}$, and $g_\infty = \mathrm{id}_{u(X)}$, and

$$g_t(\xi) = v_t(x) \qquad \text{if } \xi = u(x) \in u(X)$$

for all $t \in (0, \infty)$, delivers (i). Indeed,

- $u = v_\infty : X \to \mathbb{R}$ is a function;
- $g_0 = 1_{u(X)}$ and $g_\infty = \mathrm{id}_{u(X)}$ are increasing and $v_0 = 1_X = g_0 \circ u$, $v_\infty = g_\infty \circ u$;
- $v_t(x) = g_t(u(x))$ for all $x \in X$ and all $t \in (0, \infty)$, and g_t is (well defined and) increasing since $v_\infty(x) \geq v_\infty(y)$ implies $v_t(x) \geq v_t(y)$;
- $\{\varphi_t : X \to (0, \infty)\}_{t \in [0,\infty]}$ is a family of functions such that $\varphi_0 = \varphi_\infty = 1_X$;
- for each $t \in [0, \infty]$, each $A \in \mathcal{A}$, each $a \in A$, since $v_t = g_t \circ u$,

$$p_t(a, A) = \begin{cases} \dfrac{\varphi_t(a)}{\sum_{b \in \arg\max_A g_t \circ u} \varphi_t(b)} & \text{if } a \in \arg\max_A g_t \circ u \\ 0 & \text{else} \end{cases}$$

  moreover, if $x, y \in \arg\max_A g_t \circ u$ and $u(x) \geq u(y)$, it follows $\varphi_t(x) \geq \varphi_t(y)$.

Conversely, assume (ii) holds.

Since, for each $t \in [0, \infty]$, $\varphi_t : X \to (0, \infty)$ and $\sigma_t = \arg\max_A g_t \circ u$ is an optimal choice correspondence, by Theorem 1, p_t is consistent for all $t \in [0, \infty]$.

Since $g_0 = 1_{u(X)}$ and $\varphi_0 = 1_X$, then p_0 is uniform and has full support.

Since $g_\infty = \mathrm{id}_{u(X)}$ and $\varphi_\infty = 1_X$, then

$$p_\infty(a, A) = \begin{cases} \dfrac{1}{|\arg\max_A u|} & \text{if } a \in \arg\max_A u \\ 0 & \text{else} \end{cases}$$

is uniform and in particular, $p_\infty(a,b) \geq p_\infty(b,a)$ if and only if $u(a) \geq u(b)$.

Given any $t \in (0, \infty)$, and any $a, b \in X$, assume $p_\infty(a,b) \geq p_\infty(b,a)$. Then $u(a) \geq u(b)$, and, since g_t is increasing, therefore $g_t(u(a)) \geq g_t(u(b))$, so that

- either $\arg\max_{\{a,b\}} g_t \circ u = \{a,b\}$, then $u(a) \geq u(b)$ implies $\varphi_t(a) \geq \varphi_t(b)$, whence $p_t(a,b) \geq p_t(b,a)$;

- or $\arg\max_{\{a,b\}} g_t \circ u = \{a\}$, then $p_t(a,b) = 1 \geq 0 = p_t(b,a)$.

This proves strict coherence and yields (i). ■

Proof of Theorem 9 Assume (i) holds. Since p is consistent and continuous, if $A \subseteq B$ then, for all $a \in A$,

$$p_\infty(a,B) = \lim_{t\to\infty} p_t(a,B) = \lim_{t\to\infty} p_t(a,A)\, p_t(A,B) = p_\infty(a,A)\, p_\infty(A,B)$$

Then p_∞ is a uniform and consistent random choice rule. By Corollary 2, the binary relation defined on X by

$$x \succsim y \iff p_\infty(x,y) \geq p_\infty(y,x)$$

is a weak order and

$$\sigma_{p_\infty}(A) = \{a \in A : a \succsim b \text{ for all } b \in A\}$$

for all $A \in \mathcal{A}$.

For each $t \in (0,\infty)$, coherence implies

$$x \succsim y \iff p_\infty(x,y) \geq p_\infty(y,x) \iff p_t(x,y) \geq p_t(y,x)$$

Moreover, p_t is consistent and has full support, therefore, for every $A \in \mathcal{A}$ such that $x, y \in A$

$$\begin{aligned} p_t(x,A) &= p_t(x,y)\, p_t(\{x,y\}, A) \\ p_t(y,A) &= p_t(y,x)\, p_t(\{x,y\}, A) \end{aligned}$$

therefore:

- $x \succsim y \iff p_t(x,y) \geq p_t(y,x) \iff p_t(x,A) \geq p_t(y,A)$;
- $r_t(x,y) = \frac{p_t(x,A)}{p_t(y,A)}$.

If X is countable, then there exists $v : X \to \mathbb{R}$ such that

$$x \succsim y \iff v(x) \geq v(y)$$

Otherwise, let t be such that there exists a countable $C_t \subseteq X$ realizing the separability of p. Given any $x, y \in X$, $x \succ y \iff p_t(x,\{x,y\}) > p_t(y,\{x,y\})$, but then, there exists $z \in C_t$ such that $p_t(x,\{x,y,z\}) \geq p_t(z,\{x,y,z\}) \geq p_t(y,\{x,y,z\})$, and so $x \succsim z \succsim y$. This shows that C_t is a countable and $\succ$-order dense subset of X. Kreps

[9, Theorem 3.5] guarantees that, also in this case, there exists $v : X \to \mathbb{R}$ such that $x \succsim y \iff v(x) \geq v(y)$.

Summing up, given any $t \in (0, \infty]$ and any $A \in \mathcal{A}$, if $x, y \in A$, then

$$x \succsim y \iff v(a) \geq v(b) \iff p_t(a, A) \geq p_t(b, A) \tag{26}$$

Now, fix any $a, b \in A \in \mathcal{A}$ and define $\varphi_{a,b} : (0, \infty) \to \mathbb{R}$ by

$$\varphi_{a,b}(t) = r_t(a, b) = \frac{p_t(a, b)}{p_t(b, a)} = \frac{p_t(a, A)}{p_t(b, A)} \qquad \forall t \in (0, \infty)$$

The full support and continuity assumptions guarantee that $\varphi_{a,b}$ is well defined and continuous. Next we show that

$$\varphi_{a,b}(t + s) = \varphi_{a,b}(t)\,\varphi_{a,b}(s) \qquad \forall t, s \in (0, \infty) \tag{27}$$

Three cases have to be considered:

- If $v(a) = v(b)$, then, by (26), $\varphi_{a,b}(t) = 1$ for all $t \in (0, \infty)$, and (27) holds.
- If $v(a) > v(b)$, then $a \succ b$ and as discussed in the proof of Corollary 2, $p_\infty(a, b) = 1$ and $p_\infty(b, a) = 0$, then
$$\lim_{t \to \infty} \varphi_{a,b}(t) = \infty$$
and $\varphi_{a,b}$ is unbounded above. Moreover, by boundedness of p, there exists $M > 0$ such that $\left|\varphi_{a,b}(t + s) - \varphi_{a,b}(t)\,\varphi_{a,b}(s)\right| < M$ for all $t, s \in (0, \infty)$. Now, $(0, \infty)$ is a semigroup with respect to the addition $+$. Therefore, by Baker [3, Theorem 1], (27) holds.[18]
- Else, $v(a) < v(b)$, then the previous point shows
$$\varphi_{b,a}(t + s) = \varphi_{b,a}(t)\,\varphi_{b,a}(s) \qquad \forall t, s \in (0, \infty)$$
but then
$$\varphi_{a,b}(t + s) = \frac{p_{t+s}(a, b)}{p_{t+s}(b, a)} = \frac{1}{\varphi_{b,a}(t + s)} = \frac{1}{\varphi_{b,a}(t)\,\varphi_{b,a}(s)} = \varphi_{a,b}(t)\,\varphi_{a,b}(s)$$
as wanted.

We conclude that the functional equation (27) holds, given any $a, b \in A$. Continuity of $\varphi_{a,b}$ implies that

$$\varphi_{a,b}(t) = e^{h(a,b)t}$$

for a unique $h(a, b) \in \mathbb{R}$ (see, e.g., Aczel [1, Theorem 2.1.2.1, p.38]).

[18] *Caveat:* Baker writes the semigroup in multiplicative form, while we write it in additive form because our operation is the usual addition of real numbers.

Now fix some $a^* \in X$ and define $u : X \to \mathbb{R}$ by $u(x) = h(x, a^*)$. Fix any $t \in (0, \infty)$, $A \in \mathcal{A}$, and $a, b \in A$,

$$\begin{aligned}\varphi_{a,b}(t) &= \frac{p_t(a, A)}{p_t(b, A)} = \frac{p_t(a, \{a, b, a^*\})}{p_t(b, \{a, b, a^*\})} = \frac{p_t(a, \{a, b, a^*\})}{p_t(a^*, \{a, b, a^*\})} \frac{p_t(a^*, \{a, b, a^*\})}{p_t(b, \{a, b, a^*\})} \\ &= \varphi_{a,a^*}(t)\, \varphi_{a^*,b}(t) = \frac{\varphi_{a,a^*}(t)}{\varphi_{b,a^*}(t)} = \frac{e^{h(a,a^*)t}}{e^{h(b,a^*)t}} = \frac{e^{u(a)t}}{e^{u(b)t}}\end{aligned}$$

hence, arbitrarily choosing $c \in A$,

$$p_t(a, A) = \frac{p_t(a, A)}{\sum_{b \in A} p_t(b, A)} = \frac{\frac{p_t(a,A)}{p_t(c,A)}}{\sum_{b \in A} \frac{p_t(b,A)}{p_t(c,A)}} = \frac{\frac{e^{u(a)t}}{e^{u(c)t}}}{\sum_{b \in A} \frac{e^{u(b)t}}{e^{u(c)t}}} = \frac{e^{tu(a)}}{\sum_{b \in A} e^{tu(b)}}$$

Moreover, for all $x, y \in X$

$$\begin{aligned}x \succsim y &\iff v(x) \geq v(y) \iff p_1(x, y) \geq p_1(y, x) \\ &\iff \frac{e^{u(x)}}{e^{u(x)} + e^{u(y)}} \geq \frac{e^{u(y)}}{e^{u(x)} + e^{u(y)}} \iff u(x) \geq u(y)\end{aligned}$$

but then, for every $A \in \mathcal{A}$,

$$\sigma_{p_\infty}(A) = \{a \in A : u(a) \geq u(b) \text{ for all } b \in A\} = \arg\max{}_A u$$

and, for every $a \in A$,

$$p_\infty(a, A) = \frac{1}{|\sigma_{p_\infty}(A)|} \delta_a(\sigma_{p_\infty}(A)) = \frac{1}{|\arg\max_A u|} \delta_a\left(\arg\max_A u\right)$$

This proves (ii) holds, while the converse is trivial.

As to uniqueness of u, notice that, if $\bar{u}$ represents p in the sense of point 2, then

$$\frac{e^{u(a)t}}{e^{u(b)t}} = r_t(a, b) = \frac{e^{\bar{u}(a)t}}{e^{\bar{u}(b)t}}$$

for all $t \in (0, \infty)$ and all $a, b \in X$. Therefore $t(u(a) - u(b)) = t(\bar{u}(a) - \bar{u}(b))$ for all for all $t \in (0, \infty)$ and all $a, b \in X$. ■

References

[1] J. Aczel, *Lectures on functional equations and their applications*, Academic Press, New York, 1966.

[2] K. J. Arrow, Rational choice functions and orderings, *Economica,* 26, 121–127, 1959.

[3] J. A. Baker, The stability of the cosine equation, *Proceedings of the American Mathematical Society*, 80, 411–416, 1980.

[4] P. Battigalli, S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci, Mixed extensions of decision problems under uncertainty, *Economic Theory*, forthcoming.

[5] G. S. Becker, Irrational behavior and economic theory, *Journal of Political Economy*, 1–13, 1962.

[6] J. S. Bridle, Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters, in *Advances in Neural Information Processing Systems* (D. S. Touretzky, ed.), Morgan Kaufmann, San Mateo, 1990.

[7] F. Echenique and K. Saito, General Luce model, mimeo, 2015.

[8] X. Gabaix, A sparsity-based model of bounded rationality, *Quarterly Journal of Economics*, 129, 1661–1710, 2014.

[9] D. M. Kreps, *Notes on the theory of choice*, Westview Press, Boulder, 1988.

[10] R. D. Luce, *Individual choice behavior: A theoretical analysis*, Wiley, New York, 1959.

[11] R. B. Myerson, *Game theory: Analysis of conflict*, Harvard University Press, Cambridge, 1991.

[12] A. Mossin, Elements of a stochastic theory of consumption, *Swedish Journal of Economics*, 70, 200–220, 1968.

[13] A. Renyi, On a new axiomatic theory of probability, *Acta Mathematica Hungarica*, 6, 285–335, 1955.

[14] A. Renyi, On conditional probability spaces generated by a dimensionally ordered set of measures, *Theory of Probability and Its Applications*, 1, 55–64, 1956.

[15] A. Rubinstein, Instinctive and cognitive reasoning: A study of response times, *Economic Journal*, 117, 1243–1259, 2007.

[16] E. E. Slutsky, Sulla teoria del bilancio del consumatore, *Giornale degli Economisti*, 51, 1–26, 1915. (English translation in *Giornale degli Economisti*, 2012)

[17] A. Soltani and X.-J. Wang, A biophysically based neural model of matching law behavior: melioration by stochastic synapses, *Journal of Neuroscience*, 26, 3731–3744, 2006.

[18] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, MIT press, Cambridge, 1998.

[19] J. Vermorel and M. Mohri, Multi-armed bandit algorithms and empirical evaluation, in *European conference on machine learning*, Springer, Berlin, 2005.