DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

**CALIFORNIA INSTITUTE OF TECHNOLOGY**

PASADENA, CALIFORNIA 91125

# GENERAL LUCE MODEL

Federico Echenique and Kota Saito

**SOCIAL SCIENCE WORKING PAPER 1407R**

November 2015

# General Luce Model

Federico Echenique and Kota Saito


## Abstract

We extend the Luce model of discrete choice theory to satisfactorily handle zero-probability choices. The Luce model (or the Logit model) is the most widely applied and used model in stochastic choice, but it struggles to explain choices that are not made. The Luce model requires that if an alternative $y$ is never chosen when $x$ is available, then there is no set of alternatives from which $y$ is chosen with positive probability: $y$ cannot be chosen, even from sets of alternatives that exclude $x$. We relax this assumption. In our model, if an alternative $y$ is never chosen when $x$ is available, then we infer that $y$ is *dominated* by $x$. While dominated by $x$, $y$ may still be chosen with positive probability—even with high probability—when grouped with a *comparable* set of alternatives.



JEL classification numbers: D01,D10

Key words: Stochastic Choice, Logit Model, Luce Model, Independence of Irrelevant Alternatives, Dominance

# General Luce Model

Federico Echenique and Kota Saito

## 1 Introduction

Alice likes wine better than beer, and beer better than soda. When offered to choose between wine or beer, she chooses wine most of the time, but on occasion she may choose beer. In contrast, when Alice is offered wine or soda, she will always drink wine and never soda. Finally, when Alice is offered a choice of beer or soda, and despite liking beer more than soda, she may on occasion decide to drink soda. Standard discrete choice theory in the form of Luce's model (Luce, 1959), also known as the *Logit model*, cannot explain Alice's behavior because never choosing soda when offered {wine, soda} means that soda has the lowest possible utility: zero. This means that soda can never be chosen from any menu; in particular Alice must never choose soda from {beer, soda}.[1]

Many choice situations are similar to Alice's, and the problems they pose for Luce's model is that it does not handle zero-probability choices well. Suppose that $x$ and $y$ are two alternatives. Luce's model postulates that the probability of choosing $x$ over $y$ depends on the relative utility of $x$ compared to that of $y$. When $x$ is chosen more frequently than $y$, one infers that the utility of $x$ is higher than that of $y$. Now consider alternative $z$, which is worse than $y$. Suppose that $x$ is so much better than $z$ that $z$ would never be chosen when $x$ is present. Luce's model now says that the utility of $z$

[1]Luce's (1959) model is the most widely used and applied model of discrete choice, and one of the most successful models in decision theory. Luce's model is arguably the only model of random choice actually implemented empirically by applied economists.

is the lowest possible: zero. This means that $z$ would never be chosen, *even when $x$ is not offered*. In other words, discrete choice theory in the form of the Luce model cannot account for a situation in which $x$ is always chosen over $z$, but $z$ is some times chosen from $\{y, z\}$ for some $y$.

We propose to capture the phenomenon of probability zero choices through the idea of *dominance*. When the presence of $x$ causes $z$ not to be chosen, we say that $x$ dominates $z$. If $x$ is not present, then $z$ may be chosen with positive probability, even when $z$ is paired with alternatives that have a higher utility than $z$. More precisely, we say that *$x$ dominates $z$* if $z$ is never chosen when $x$ is available; and that $x$ and $z$ are *comparable* if neither of them dominates the other. In our theory, an agent uses Luce's model to determine the probability of choosing each alternative among a sets of comparable alternatives. The agent chooses with probability zero those alternatives that are dominated in the choice set. Importantly, this does not mean that such alternatives are never chosen: they may be comparable to another set of alternatives, and may be chosen with positive probability from a set of comparable alternatives.

An important aspect of our model is that the comparability binary relation (the relation "is comparable to") may not be transitive. It is possible that wine is better, but does not dominate, beer; beer is better, but does not dominate soda; while wine dominates soda. Hence, wine and beer, and beer and soda, are comparable; but wine and soda are not comparable. Such lack of transitivity is related to the phenomenon of semiorders. In the theory of semiorders (originally proposed by Luce (1956)), indifference may fail to be transitive.[2] Casting the theory of semiorders in the framework of stochastic choice affords considerable simplification because one can use the cardinal magnitudes of stochastic choice to measure cardinal utility differences.

One way to think of the exercise in our paper is as a study of stochastic choice that can some times be deterministic. One of Luce's crucial assumptions is that choice

[2] A common example in the literature on semiorders is the comparison of coffee with different amounts of sugar. An additional grain of sugar produces an indifferent cup of coffee. But after adding enough grains, one obtains a noticeably sweeter cup.

probabilities are always strictly positive (see for example Theorem 3 in Luce (1959)). But it is very common to see the model being applied to environments in which some choice probabilities are zero. Ours seems to be the first extension of Luce's model to accommodate deterministic choices.

We present two versions of our model. The first version generalizes Luce's model by allowing that an agent never chooses some elements in a choice set. The support of the stochastic choice function includes all the alternatives in a choice set that are not dominated by any other alternative. The second version of our model is a special case of the first where the dominance relation is tied to utility. An alternative is dominated by another if its utility is sufficiently smaller. The main results in the paper are axiomatic characterizations of these models, meaning a complete description of the observable stochastic choices that are consistent with the models.

The axioms are simple. As written above, we determine when alternatives are dominated from the stochastic choice behavior of the agent. We will say that $x$ dominates $y$ if $y$ is never chosen when $x$ is available. Our general model is captured by three axioms.

The first axiom, *Weak Regularity*, imposes that the probability of choosing an alternative cannot drop from positive to zero when one removes alternatives from a choice set. It is a weakening of an axiom of Luce's.

The second axiom, *Independence of Dominated Alternatives*, says that removing a dominated alternative from a choice set does not affect the stochastic choice behavior of the agent. The third and fourth axioms are weaker version of Luce's axiom, *Independence of Irrelevant Alternatives*. These axioms give us that agents' choices over comparable alternatives are dictated by Luce's model.

We propose a more restrictive model, where dominance is tied to utility. In addition to the axioms we have already described, it requires an additional axiom: *Path Monotonicity*, which requires that comparisons of sequences of alternatives be consistent.

To conclude the introduction, we discuss the related literature. There are many papers on semiorders, starting from Luce (1956). In particular, Fishburn (1973) studies a stochastic preference relation as a semiorder. On the other hand, in our paper, the stochastic preference relation is not a semiorder. It is the comparable relation that has the intransitive property.

Several recent papers provide generalizations of Luce model. Gul et al. (2014) axiomatize a generalization of Luce model to address difficulties of Luce model that arise when alternatives have common attributes. Fudenberg and Strzalecki (2014) axiomatizes a generalization of discounted logistic model that incorporates a parameter to capture different views that the agent might have about the costs and benefits of larger choice sets. Both models do not allow zero probability choices. Echenique et al. (2013) axiomatizes a generalization of Luce model that incorporates the effects of attention. None of them study the issues that we focus on in the present paper.

Our model is related to the recent literature on attention and inattention. In our model, an agent chooses with positive probability only a subset of the available alternatives. So one can think of the alternatives that are outside of the support of the stochastic choice as being alternatives that the agent does not pay attention to. Masatlioglu et al. (2012) provide an elegant model of attention. In their model, the agent's choice is deterministic. Some recent studies try to incorporate the effect of attention into stochastic choice. Manzini and Mariotti (2012) axiomatize a model in which an agent considers each feasible alternative with a probability and then chooses the alternative that maximizes a preference relation within the set of considered alternatives. The more recent paper by Brady and Rehbeck (2014) axiomatizes a model that encompasses the model of Manzini and Mariotti (2012). Horan (2014) has proposed a new model of limited consideration based on the random utility model (Block and Marschak, 1960).

The rest of the paper is organized as follows. In Section 2, we propose the models. Then in Section 3, we propose the axioms, the main representation theorems and a uniqueness property of the representations. In Section 4, we present the proofs of the main results. Finally, in Section 5, we present an extension of the model.

## 2 Model

The set of all possible objects of choice, or *alternatives*, is a finite set $X$. A *stochastic choice function* is a function $p$ that for an every nonempty subset $A$ of $X$ returns a probability distribution $p(A)$ over $A$. We denote the probability of choosing an alternative $a$ from $A$ by $p(a, A)$. A stochastic choice function is the primitive observable object of our study.

We propose the following model:[3]

**Definition 1.** *$p$ is a general Luce model if there exist $u : X \to \mathbf{R}_{++}$ and a function $c : 2^X \setminus \emptyset \to 2^X \setminus \emptyset$ such that $c(A) \subseteq A$ for all subsets $A$ of $X$ and*

$$p(x, A) = \begin{cases} \dfrac{u(x)}{\sum_{y \in c(A)} u(y)} & \text{if } x \in c(A), \\ 0 & \text{if } x \notin c(A). \end{cases}$$

*Moreover $c$ satisfies the following properties:*

1. *$A \subseteq B \Longrightarrow c(B) \cap A \subseteq c(A)$.*
2. *$x \in c(A')$ for all $A' \subsetneq A$ with $x \in A' \Longrightarrow x \in c(A)$.*
3. *$x \notin c(A) \Longrightarrow c(A) = c(A \setminus \{x\})$.*

**Remark 1.** *There are two special cases worth emphasizing.*

*(i) If $c(A)$ is singleton for all finite $A$ subset of $X$, then $p(x, A)$ is deterministic choice.*

*(ii) If $c(A) = A$ for all finite subset $A$ of $X$, then $p$ coincides with Luce's (1959) model.*

Properties (1)–(3) of $c$ capture our ideas of when an alternative is chosen with positive probability. Property (1) of $c$ (a well-known property called Sen's $\alpha$) ensures that if an alternative is chosen with positive probability from a large set $B \supseteq A$, in which it faces more competition than in $A$, then it must also be chosen with positive probability from

[3]A first draft of our paper focused on the threshold general Luce mode. We thank David Ahn, who suggested that we extend the discussion to cover the general Luce model.

$A$. Property (2) reflects a kind of monotonicity in the size of $A$; it says that if $x$ is chosen from all subsets of $A$ then it must be chosen from $A$. It is important for us because it means that if $x$ is not chosen from $A$ then we can find who is "responsible" for $x$ not being chosen. In our interpretation, the responsible alternative is the alternative that dominates $x$. Property (3) says that if $x$ is chosen with probability zero then it is deemed irrelevant and cannot affect which alternatives are chosen with positive probability (it is a sort of "independence of irrelevant alternatives" property).[4]

Ultimately, the role of properties (1)–(3) of $c$ is to give $c(A)$ the role of the undominated alternatives in $A$. For that, we need to introduce a dominance binary relation.

Behaviorally, the *dominance* ($\gg$) and *comparability* ($\simeq$) binary relations are defined as follows.

**Definition 2.** *For all $x, y \in X$, (i) $x \gg y$ if $p(x, xy) = 1$; (ii) $x \not\gg y$ if $p(x, xy) < 1$; (iii) $x \simeq y$ if $x \not\gg y$ and $y \not\gg x$.*

The relations $\gg$, $\not\gg$, and $\simeq$ are the behavioral counterparts to the notion of "dominance," "non-dominance," and "comparability" implicit in the general Luce model. When $x \gg y$ we infer that $x$ is revealed to dominate $y$, when $x \not\gg y$ we infer that $x$ is revealed not to dominate $y$, and when $x \simeq y$ we infer that $x$ and $y$ are revealed to be *comparable.* Note that these notions correspond to our discussion in the introduction.

**Definition 3.** *A function $c : 2^X \setminus \emptyset \to 2^X \setminus \emptyset$ with $c(A) \subseteq A$ for all $A \subset X$ is dominance rationalizable if there is a transitive and antisymmetric relation $\gg'$ such that*

$$c(A) = \{x \in A : \nexists y \in A \ s.t. \ y \gg' x\}.$$

**Proposition 1.** *If a stochastic choice function $p$ is a general Luce model $(u, c)$, then $c$ is dominance rationalizable by a relation $\gg'$. Moreover, $x \gg' y$ if and only if $x \gg y$.*

[4]Property (3) is shared by $c$ and by *attention filters,* proposed by Masatlioglu et al. (2012). Unlike Masatlioglu et al. (2012), in our model, the function $c$ is uniquely identified. This difference partly comes from the fact that we study stochastic choices while Masatlioglu et al. (2012) study deterministic choices. In all, the models are quite different.

**Proof**: Define $\gg'$ by: $x \gg' y$ iff $y \notin c(\{x, y\})$. Let $y \in A \subset X$. If there is $x \in A$ with $x \gg y$ then Property (1) of $c$ (Sen's $\alpha$) implies that $y \notin c(A)$. Thus $c(A)$ is contained in the set $\{x \in A : \nexists y \in A \text{ s.t. } y \gg' x\}$. Conversely, if $x \notin c(A)$ then either $A$ has cardinality 2 and there is $y$ with $y \gg' x$, or Property (2) of $c$ implies that there is a set $A' \ni x$ of smaller cardinality than $A$ for which $x \notin c(A')$. By recursion then there is $y$ with $y \gg' x$.

Next, to prove that $\gg'$ is transitive, suppose that $x \gg' y$ and $y \gg' z$. Note that Property (1) implies that $y, z \notin c(\{x, y, z\})$. So $c(\{x, y, z\}) = \{x\}$. Then Property (3) implies that $\{x\} = c(\{x, z\})$.

Finally, note that $x \gg' y$ if and only if $p(x, xy) = 1$ if and only if $x \gg y$. $\square$

Proposition 1 makes it clear that the general Luce model is, in a sense, about semiorders: $x$ may be comparable to $y$, and $y$ may be comparable to $z$, but it is possible that $x$ dominates $z$: we may have $x \simeq y$, $y \simeq z$, but $x \gg z$.

The function $c$ in a general Luce model is independent of $u$. It is reasonable to think that dominance may some times be tied to utility. We introduce the idea that an alternative $z$ is dominated by $x$ if $u(x)$ is sufficiently larger than $u(z)$. The resulting model is called a threshold general Luce model.

**Definition 4.** *A general Luce model $(u, c)$ is called a threshold general Luce model if there exists a nonnegative number $\varepsilon$ such that for all finite $A$ subset of $X$,*

$$c(A) = \{y \in A | (1 + \varepsilon)u(y) \geq u(z) \text{ for all } z \in A\}.$$

In the threshold general Luce model, the function $c$ is defined by $u$ and a new parameter $\varepsilon$. So the agent considers that an alternative $x$ dominates another alternative $y$ if $u(x) > (1 + \varepsilon)u(y)$. The number $\varepsilon \geq 0$ captures the threshold beyond which alternatives become dominated. A utility ratio of more than $1 + \varepsilon$ means that the less-preferred alternative is dominated by the more-preferred alternative.

# 3 Axioms and results

Four axioms characterize the general Luce model. The first axiom is a weaker version of Luce's axiom *regularity.* Luce's axiom says that the probability of choosing $x$ cannot decrease when the choice set shrinks.

**Axiom 1.** *(Weak Regularity): If $p(x, xy) = 0$ and $y \in A$, then $p(x, A) = 0$.*

Our next axiom says that *removing a dominated alternative does not affect choices.* It means that when $x$ is dominated then it does not affect whatever consideration governs choices among the remaining alternatives.

**Axiom 2.** *(Independence of Dominated Alternatives (IDA)):*
*If $p(x, A) = 0$, then $p(y, A) = p(y, A \setminus \{x\})$ for all $y \in A \setminus \{x\}$.*

The last two axioms are weakenings of Luce's Independence of Irrelevant Alternatives axiom (IIA; see Luce (1959)). The first axiom imposes IIA among the comparable alternatives in a set; we do not impose it for dominated alternatives. The second axiom imposes a consequence of Luce's IIA.

**Axiom 3.** *(Weak IIA): Suppose that $x \simeq y$ for all $x, y \in A$. Then for all $x, y, z \in A$,*

$$\frac{p(x, A)}{p(y, A)} = \frac{p(x, A \setminus \{z\})}{p(y, A \setminus \{z\})}.$$

**Axiom 4.** *(Cyclical Independence): For any $x_1, x_2, \ldots, x_n \in X$, if $x_i \simeq x_{i+1}$ for all $i \in \{1, \ldots, n-1\}$ and $x_1 \simeq x_n$, then*

$$\frac{p(x_1, x_1x_n)}{p(x_n, x_1x_n)} = \frac{p(x_1, x_1x_2)}{p(x_2, x_1x_2)} \frac{p(x_2, x_2x_3)}{p(x_3, x_2x_3)} \cdots \frac{p(x_{n-1}, x_{n-1}x_n)}{p(x_n, x_{n-1}x_n)}. \quad (1)$$

Cyclical independence compares the relative probability of $x_n$ to $x_1$ with the sequence of relative probabilities one would obtain through any other path from $x_1$ to $x_n$. It is obvious that in Luce's model (1) holds, and that it results directly from Luce's IIA. But to prove (1) from our weak IIA would require us to consider the sets of the form $\{x_i, x_{i+1}, x_{i+2}\}$ but even if $x_i \simeq x_{i+1}$ and $x_{i+1} \simeq x_{i+2}$ it may not be the case that $x_i \simeq x_{i+2}$. So we cannot use weak IIA to prove (1).

**Theorem 1.** *A stochastic choice function satisfies Weak Regularity, IDA, Weak IIA, and Cyclical Independence if and only if it is a general Luce model* $(u, c)$*. Moreover,* $c$ *is unique.*

## 3.1 Threshold General Luce Model

In a general Luce model, the dominance relation may be unrelated to the utility $u$. In contrast, the "threshold" model imposes a relation between $c(A)$ and the utility of the elements of $A$. In order to, in a sense, calibrate the magnitude of the parameter $\varepsilon$ we need an additional axiom.

Our last axiom requires a definition and some notational conventions.

**Definition 5.** *For any* $x, y \in X$*, a sequence* $(z_i)_{i=1}^s$ *is a path from* $x$ *to* $y$ *if* $x = z_1$ *and* $y = z_s$ *and* $z_{i+1} \not\gg z_i$ *for all* $i$*.*

We use the number $+\infty$, and assume that it has the following properties: $+\infty > x$ for all $x \in \mathbf{R}$, $1/0$ is equal to $+\infty$, and $+\infty x = +\infty$ for any $x > 0$. With this notational convention, the following notion of distance is well defined.

**Definition 6.** *(Distance): For any path* $(z_i)_{i=1}^s$ *from* $x$ *to* $y$*, define*

$$d((z_i)_{i=1}^s) = \frac{p(z_1, z_1z_2)}{p(z_2, z_1z_2)} \frac{p(z_2, z_2z_3)}{p(z_3, z_2z_3)} \cdots \frac{p(z_{s-1}, z_{s-1}z_s)}{p(z_s, z_{s-1}z_s)}.$$

By the definition of path, for all $i$, we can have $z_i \gg z_{i+1}$ but not $z_{i+1} \gg z_i$. So $p(z_i, z_iz_{i+1})/p(z_{i+1}, z_iz_{i+1})$ can be $+\infty$ but not zero. So $d((z_i)_{i=1}^s)$ is well defined.

Under Luce's IIA, the distance between two alternatives $x$ and $y$ must be the same along any two paths; it must equal $p(x, xy)/p(y, xy)$. We do not assume Luce's IIA, so in our setup the distance can be path-dependent. Our final axiom, path monotonicity, says that the distance between any incomparable pair of alternatives must be larger than the distance between any comparable pair of alternatives.

**Axiom 5.** *(Path Monotonicity): For any pair of paths* $(z_i)_{i=1}^s$ *from* $x$ *to* $y$ *and* $(z'_i)_{i=1}^t$ *from* $x'$ *to* $y'$*, if* $x \gg y$ *and* $x' \simeq y'$*, then*

$$d((z_i)_{i=1}^s) > d((z'_i)_{i=1}^t). \tag{2}$$

**Theorem 2.** *A stochastic choice function satisfies Weak Regularity, IDA, Weak IIA, and Path Monotonicity if and only if it is a threshold general Luce model* $(u, \varepsilon)$*.*

In Theorem 2, we do not assume Cyclical Independence. This is because Weak IIA and Path Monotonicity imply Cyclical Independence. See Remark 2 in Section 4.2.

## 3.2 Uniqueness

We now argue that a general Luce model is uniquely identified. As stated in Theorem 1, the function $c$ is unique. In the following, we will show a uniqueness property of $u$. We should emphasize that our identification result requires a richness condition on the environment.

**Axiom 6.** *(Richness): For any* $x, y \in X$*, if* $x \gg y$*, then there exists a path* $(z_i)_{i=1}^n$ *from* $x$ *to* $y$ *such that* $z_i \simeq z_{i+1}$ *for all* $i$*.*

**Proposition 2.** *Under Richness, two general Luce models* $(u, c)$ *and* $(u', c)$ *represent the same stochastic choice function if and only if there exists a positive number* $\lambda$ *such that* $u = \lambda v$*.*

# 4 Proofs

In the following, we will say that a set $A \subset X$ is pairwise comparable if any alternatives $x, y \in A$ are comparable.

**Axiom 7.** *(Dominance Transitivity) For all* $x, y, z \in X$*, if* $x \gg y$ *and* $y \gg z$*, then* $x \gg z$*.*

**Lemma 1.** *Weak Regularity and IDA imply Dominance Transitivity.*

**Proof:** Since $x \gg y$ we have $p(y, xy) = 0$. By Weak Regularity, $p(y, xyz) = 0$.

By IDA, $p(z, xyz) = p(z, xz)$. Since $y \gg z$, $p(z, yz) = 0$. By Weak Regularity, $p(z, xyz) = 0$. So $p(z, xz) = 0$. $\square$

For all $A \subset X$, define

$$c(A) = \{x \in A | \nexists y \in A \text{ such that } y \gg x\}.$$

Note that Lemma 1 implies $c(A) \neq \emptyset$. By definition, then, $c$ satisfies Property (1) and (2) of a generalized Luce model.

**Lemma 2.** *Suppose that Weak Regularity and IDA hold. Then for any subset $A$ of $X$, $p(c(A)) = p(A)$.*

**Proof:** Let $x \in A \setminus c(A)$. Then there is $y \in A$ with $y \gg x$. In fact, by transitivity of $\gg$ and finiteness of $A$ we can take $y \in c(A)$. So Weak Regularity implies that $p(x, A) = p(x, c(A)) = 0$.

Consider the case where $x \in c(A)$. Let $A \setminus c(A) = \{x_1, \ldots, x_n\}$.Then, $p(x_i, A) = 0$ for all $i \in \{1, \ldots, n\}$. So by IDA, $p(A) = p(A \setminus x_1)$. Hence, $p(x_i, A \setminus \{x_1\}) = 0$ for all $i \in \{2, \ldots, n\}$. So by IDA, $p(A) = p(A \setminus x_1) = p(A \setminus \{x_1, x_2\})$. By recursion, we have $p(A) = p(c(A))$. $\square$

Lemma 2 implies that $p(c(A)) = 1$, but that does not mean that there are no $x \in c(A)$ with $p(x, A) = 0$. The next lemma takes care of this issue.

**Lemma 3.** *Let $p$ satisfy Weak IIA, Weak Regularity, and IDA. For any subset $A$ of $X$, $x \notin c(A)$ if and only if $p(x, A) = 0$. So, $c(A) = supp\ p(A)$.*

**Proof:** If $x \in A \setminus c(A)$ then $p(x, A) = 0$ by the previous lemma.

So suppose that $p(x, A) = 0$ and (towards a contradiction) that $x \in c(A)$. For all $y \in c(A)$, we have

$$\frac{p(x, A)}{p(y, A)} = \frac{p(x, c(A))}{p(y, c(A))} = \frac{p(x, xy)}{p(y, xy)} \in (0, 1),$$

where $=$ holds by Lemma 2 and $\in$ holds by Weak IIA and the fact that $c(A)$ is pairwise comparable. This contradicts that $p(x, A) = 0$. $\square$

**Lemma 4.** *Suppose that Weak Regularity and IDA hold. Then, $c$ satisfies Property (3).*

**Proof:** Suppose that $x \notin c(A)$ to show $c(A) = c(A \setminus \{x\})$. Then, there is $y \in A$ with $y \gg x$. By Weak Regularity, $p(x, A) \leq p(x, xy) = 0$. Hence, IDA shows $p(A) = p(A \setminus \{x\})$. Then, by Lemma 3, $c(A) = \text{supp } p(A) = \text{supp } p(A \setminus \{x\}) = c(A \setminus \{x\})$. $\square$

For any $X' \subseteq X$ and $u : X' \to \mathbf{R}_+$, we say that a pair $(X', u)$ satisfies *$L$-property* if for any $x, y \in X'$ such that $x \simeq y$, we have

$$\frac{u(x)}{u(y)} = \frac{p(x, xy)}{p(y, xy)}.$$

**Lemma 5.** *Let $p$ satisfy Weak Regularity and Weak IIA, and let $(X, u)$ have $L$-property. If $x \in c(A)$, then $p(x, c(A)) = \frac{u(x)}{\sum_{y \in c(A)} u(y)}$.*

**Proof:** We denote $c(A) = \{x_1, \ldots, x_m\}$. By the definition of $c$, for all $i, j$, $x_i \simeq x_j$. Then,

$$\frac{p(x_j, c(A))}{p(x_1, c(A))} = \frac{p(x_j, x_1 x_j)}{p(x_1, x_1 x_j)} = \frac{u(x_j)}{u(x_1)},$$

where the first equality holds by virtue of Lemma 2, and the second equality holds by $L$-property. For all $j \in \{2, \ldots, m\}$,

$$p(x_j, c(A)) = \frac{u(x_j)}{u(x_1)} p(x_1, c(A)). \tag{3}$$

Since $\sum_{j=1}^m p(x_j, c(A)) = 1$, we have $1 = \frac{\sum_{j=1}^m u(x_j)}{u(x_1)} p(x_1, c(A))$, so that

$$p(x_1, c(A)) = \frac{u(x_1)}{\sum_{j=1}^m u(x_j)}. \tag{4}$$

Therefore, by (3) and (4), $p(x_j, c(A)) = \frac{u(x_j)}{\sum_{j=1}^m u(x_j)}$ for all $j \in \{1, \ldots, m\}$. $\square$

## 4.1 Proof of Theorem 1

**Necessity:** We show that a general Luce model satisfies Weak Regularity. Suppose that $p(x, xy) = 0$ and $y \in A$. Then, by Proposition 1, $y \gg x$ and $x \notin c(A)$. Therefore, $p(x, A) = 0$.

We show that the general Luce model satisfies IDA. Suppose that $p(x, A) = 0$. Then, $x \notin c(A)$. So by Property 3, $c(A) = c(A \setminus \{x\})$. Therefore, if $y \notin c(A)$, so $y \notin c(A \setminus \{x\})$ then $p(y, A) = 0 = p(y, A \setminus \{x\})$; if $y \in c(A)$, then

$$p(y, A) = \frac{u(y)}{\sum_{z \in c(A)} u(z)} = \frac{u(y)}{\sum_{z \in c(A \setminus \{x\})} u(z)} = p(y, A \setminus \{x\}).$$

Therefore, $p(A) = p(A \setminus \{x\})$, so that IDA holds.

To show that the general Luce model satisfies Weak IIA. Suppose that $A$ is pairwise comparable. By definition, $p(x, xy) \in (0, 1)$ for all $x, y \in A$. So, $x \in c(\{x, y\})$ for all $x, y \in A$. By Proposition 1, it means that $x \not\gg y$ for all $x, y \in A$. Hence, $c(A) = A$. Since $p$ has Luce formula on $c(A)$ and $c(A) = A$, $p$ satisfies IIA on $A$. Therefore, Weak IIA holds.

To show that the general Luce model satisfies Cyclical Independence, choose any $x_1, x_2, \ldots, x_n \in X$ such that $x_i \simeq x_{i+1}$ for all $i \in \{1, \ldots, n-1\}$ and $x_1 \simeq x_n$. Then, $\frac{p(x_1, x_1x_n)}{p(x_n, x_1x_n)} = \frac{u(x_1)}{u(x_n)}$ and $\frac{p(x_i, x_ix_{i+1})}{p(x_{i+1}, x_ix_{i+1})} = \frac{u(x_i)}{u(x_{i+1})}$ for all $i$. Therefore,

$$\frac{p(x_1, x_1x_n)}{p(x_n, x_1x_n)} = \frac{u(x_1)}{u(x_n)} = \frac{p(x_1, x_1x_2)}{p(x_2, x_1x_2)} \frac{p(x_2, x_2x_3)}{p(x_3, x_2x_3)} \cdots \frac{p(x_{n-1}, x_{n-1}x_n)}{p(x_n, x_{n-1}x_n)}. \tag{5}$$

**Sufficiency:** Consider subsets $\{X_i\}_{i \in I}$ of $X$ such that $X = \bigcup_{i \in I} X_i$ such that (i) for all $i \in I$ and all $x, y \in X_i$, there exists a path $(z_j)_{j=1}^n$ from $x$ to $y$ such that $z_j \simeq z_{j+1}$ for all $j$; (ii) for all $i \neq j$ and all $x \in X_i$ and $y \in X_j$, there exists no path $(z_j)_{j=1}^n$ from $x$ to $y$ or from $y$ to $x$ such that $z_j \simeq z_{j+1}$ for all $j$. By this definition, for all $i, j \in I$, if $i \neq j$, then $X_i \cap X_j = \emptyset$. So $\{X_i\}_{i \in I}$ is a partition of $X$.

Now we define the function $u$ on $X$. For each $i \in I$, choose $x_i^* \in X_i$. Define $u(x_i^*) = 1$. Choose $x \in X$ to define $u(x)$. Since $\{X_i\}_{i \in I}$ is a partition of $X$, there exists unique $i \in I$ such that $x \in X_i$. Define

$$u(x) = \inf \left\{ d((z_j)_{j=1}^n) \middle| (z_j)_{j=1}^n \text{ is a path from } x \text{ to } x_i^* \text{ such that } z_j \simeq z_{j+1} \text{ for all } j \right\},$$

where remember that $d((z_j)_{j=1}^n) \equiv \frac{p(z_1, z_1 z_2)}{p(z_2, z_1 z_2)} \cdots \frac{p(z_{n-1}, z_{n-1} z_n)}{p(z_n, z_{n-1} z_n)}$

Now we will show that $u$ is well defined. Notice that since $X$ is finite, there are only finitely many paths $(z_j)_{j=1}^n$ in which each element in $X$ appears at most once. Therefore, among such paths, we can find a path which minimizes the distance. Now consider a path $(z_j)_{j=1}^n$ in which there exists $x' \in X$ such that $x'$ appears more than once. Then, in the path, we can find a cycle $(z_j)_{j=l}^{l+k}$ such that $z_j \simeq z_{j+1}$ for all $j \in \{l, \ldots, l+k-1\}$ and $z_l = x' = z_{l+k}$. Notice that

$$\begin{aligned}
d((z_j)_{j=l}^{l+k}) &\equiv \frac{p(z_l, z_l z_{l+1})}{p(z_{l+1}, z_l z_{l+1})} \cdots \frac{p(z_{l+k-2}, z_{l+k-2} z_{l+k-1})}{p(z_{l+k-1}, z_{l+k-2} z_{l+k-1})} \frac{p(z_{l+k-1}, z_{l+k-1} z_{l+k})}{p(z_{l+k}, z_{l+k-1} z_{l+k})} \\
&= \frac{p(z_l, z_l z_{l+k})}{p(z_{l+k}, z_l z_{l+k})} \quad (\because \text{Cyclical Independence}) \\
&= 1 \qquad (\because z_l = z_{l+k})
\end{aligned}$$

Now consider a shorter path $(z_1, z_2, \ldots, z_l, z_{l+k+1}, \ldots, z_n)$. Then, by definition,

$$d(z_1, z_2, \ldots, z_l, z_{l+k+1}, \ldots, z_n) d((z_j)_{j=l}^{l+k}) = d((z_j)_{j=1}^n).$$

Hence, $d(z_1, z_2, \ldots, z_l, z_{l+k+1}, \ldots, z_n) = d((z_j)_{j=1}^n)$. That is, the distance of the shorter path is the same as that of the original path. By this way, for any path, we can delete cycles by keeping the value of the distance the same. This means that we can focus on a path $(z_j)_{j=1}^n$ in which each element in $X$ appears at most once without loss of generality. Hence, the function $u$ is well defined.

Moreover, for all $j$, we have $z_{j+1} \not\gg z_j$, so that $p(z_j, z_j z_{j+1}) / p(z_{j+1}, z_j z_{j+1})$ is not zero. Hence, $u(x)$ is a positive real number.

To complete the proof of sufficiency, by Lemmas 3, 4, and 5, it suffices to show that $(X, u)$ has $L$-property.

Choose $x, y \in X$ such that $x \simeq y$. Without loss of generality assume that $x \in X_i$ for some $i \in I$. Since $x \simeq y$, then $y \in X_i$. By definition of $u$ we have that

$$\frac{p(x, xy)}{p(y, xy)} u(y) \geq u(x). \tag{6}$$

Suppose by way of contradiction that the above inequality (6) holds strictly. Then, there exists a path $(z_i)_{i=1}^n$ from $x$ to $x_i^*$ such that

$$\frac{p(x, xy)}{p(y, xy)} u(y) > d((z_i)_{i=1}^n). \tag{7}$$

Since $x \simeq y$, we can consider the path $(y, z_1, \ldots, z_n)$ from $y$ to $x_i^*$. Then

$$d(y, z_1, \ldots, z_n) = \frac{p(y, xy)}{p(x, xy)} d((z_i)_{i=1}^n). \tag{8}$$

By (7) and (8), we have $u(y) > d(y, z_1, \ldots, z_n)$, this is a contradiction with the definition of $u$. So we obtain the equality in (6). Hence $L$-property holds.

By Lemma 3, since $c(A) = \text{supp } p(A)$, it is easy to see that $c$ is unique. $\square$

## 4.2 Lemmas for Theorem 2

**Definition 7.** *For all $x, y \in X$, (i) $x > y$ if $p(x, xy) > p(y, xy)$; (ii) $x = y$ if $p(x, xy) = p(y, xy)$; (iii) $x \geq y$ if $x > y$ or $x = y$.*

We interpret $x > y$ as $x$ being strictly revealed preferred to $y$, because it is chosen more frequently in the simple pairwise comparison of $x$ and $y$. In the following, we will use the standard *stochastic revealed preference* relation, denoted by $\geq$.

**Axiom 8.** *(Strong Dominance Transitivity): For all $x, y, z \in X$, (i) if $x \gg y$ and $y \geq z$, then $x \gg z$; (ii) if $x \geq y$ and $y \gg z$, then $x \gg z$.*

**Lemma 6.** *Dominance Transitivity and Path Monotonicity imply Strong Dominance Transitivity.*

**Proof:** Assume $x \gg y$ and $y \geq z$ to show $x \gg z$. If $y \gg z$, then by Dominance Transitivity, we have $x \gg z$. So consider the case in which $y \simeq z$.

Suppose towards a contradiction that not $x \gg z$. Then, $z \gg x$ or $x \simeq z$.

**Case** 1: $z \gg x$. Then by Dominance Transitivity and $x \gg y$, we have $z \gg y$. This contradicts with $y \geq z$.

**Case** 2: $x \simeq z$. Then $(x, y, z)$ is a path from $x$ to $z$. The value of the distance along the path is $+\infty$. On the other hand, $(x, z, y)$ is a path from $x$ to $y$ and $x \gg z$. The value of the distance along the path is finite because $x \simeq z$ and $z \simeq y$. This violates Path Monotonicity.

The other statement of Strong Dominance Transitivity can be proved in the same way. $\square$

**Axiom 9.** *(Stochastic Transitivity) For all $x, y, z \in X$, (i) if $x \geq y$ and $y \geq z$, then $x \geq z$; (ii) if $x > y$ and $y \geq z$, then $x > z$; (iii) if $x \geq y$ and $y > z$, then $x > z$.*

**Lemma 7.** *Cyclical Independence and Strong Dominance Transitivity imply Stochastic Transitivity.*

**Proof:** Choose $x \geq y$ and $y \geq z$.

**Case** 1: $x \gg y$ or $y \gg z$. Then by Strong Dominance Transitivity, we have $x \gg z$.

**Case** 2: $x \simeq y$ and $y \simeq z$. We can rule out that $z \gg x$ as that would mean that $z \gg y$ by Strong Dominance Transitivity. So we have either $x \gg z$ or $x \simeq z$. Firstly, if $x \gg z$, then $x > z$, as desired. Secondly, let $x \simeq z$. Since $x \simeq z$, by Cyclical Independence,

$$\frac{p(x, xz)}{p(z, xz)} = \frac{p(x, xy)}{p(y, xy)} \frac{p(y, zy)}{p(z, zy)} \geq 1. \tag{9}$$

Hence, $x \geq z$. Moreover, if $x > y$ or $y > z$ hold, then (9) holds strictly. $\square$

**Remark 2.** *Weak IIA and Path Monotonicity imply Cyclical Independence.*

**Proof:** Choose any $x_1, x_2, \ldots, x_n \in X$. Suppose that $x_i \simeq x_{i+1}$ for all $i \in \{1, \ldots, n-1\}$ and $x_1 \simeq x_n$.

**Case** 1: Consider the case in which $x_i \simeq x_j$ for any pair $i, j \in \{1, \ldots, n\}$. Then by Weak IIA, we have

$$\frac{p(x_i, x_i x_j)}{p(x_j, x_i x_j)} = \frac{p(x_i, \{x_k\}_{k=1}^n)}{p(x_j, \{x_k\}_{k=1}^n)}.$$

Therefore, we will have

$$\begin{aligned}
\frac{p(x_1, x_1 x_n)}{p(x_n, x_1 x_n)} &= \frac{p(x_1, \{x_k\}_{k=1}^n)}{p(x_n, \{x_k\}_{k=1}^n)} \\
&= \frac{p(x_1, \{x_k\}_{k=1}^n)}{p(x_2, \{x_k\}_{k=1}^n)} \frac{p(x_2, \{x_k\}_{k=1}^n)}{p(x_3, \{x_k\}_{k=1}^n)} \cdots \frac{p(x_{n-1}, \{x_k\}_{k=1}^n)}{p(x_n, \{x_k\}_{k=1}^n)} \\
&= \frac{p(x_1, x_1 x_2)}{p(x_2, x_1 x_2)} \frac{p(x_2, x_2 x_3)}{p(x_3, x_2 x_3)} \cdots \frac{p(x_{n-1}, x_{n-1} x_n)}{p(x_n, x_{n-1} x_n)}.
\end{aligned}$$

**Case** 2: Consider the case in which $x_i \gg x_j$ for some pair $i, j \in \{1, \ldots, n\}$. Since $x_i \simeq x_{i+1}$ for all $i \in \{1, \ldots, n-1\}$ and $x_1 \simeq x_n$, there must be a path $(z_k)_{k=1}^s$ from from $x_i$ to $x_j$ such that $z_k \simeq z_{k+1}$ for any $k$. Therefore, $d((z_k)_{k=1}^s)$ is a finite number.

Now suppose that Cyclical Independence fails. Then

$$\frac{p(x_1, x_1 x_n)}{p(x_n, x_1 x_n)} \neq \frac{p(x_1, x_1 x_2)}{p(x_2, x_1 x_2)} \frac{p(x_2, x_2 x_3)}{p(x_3, x_2 x_3)} \cdots \frac{p(x_{n-1}, x_{n-1} x_n)}{p(x_n, x_{n-1} x_n)}.$$

**Subcase 2.1:** Suppose that

$$\frac{p(x_1, x_1 x_n)}{p(x_n, x_1 x_n)} > \frac{p(x_1, x_1 x_2)}{p(x_2, x_1 x_2)} \frac{p(x_2, x_2 x_3)}{p(x_3, x_2 x_3)} \cdots \frac{p(x_{n-1}, x_{n-1} x_n)}{p(x_n, x_{n-1} x_n)}.$$

Then

$$\frac{p(x_n, x_{n-1} x_n)}{p(x_{n-1}, x_{n-1} x_n)} \cdots \frac{p(x_3, x_2 x_3)}{p(x_2, x_2 x_3)} \frac{p(x_2, x_1 x_2)}{p(x_1, x_1 x_2)} \frac{p(x_1, x_1 x_n)}{p(x_n, x_1 x_n)} > 1.$$

So we obtain a path $(x_n, x_{n+1}, \ldots, x_1, x_n)$ from $x_n$ to $x_n$ such that $d(x_n, x_{n+1}, \ldots, x_1, x_n) > 1$. By repeating this path $M$ times, we can obtain a path $x_n$ to $x_n$ and the value of the distance from $x_n$ to $x_n$ along the path is $(d(x_n, x_{n+1}, \ldots, x_1, x_n))^M$. Since $d(x_n, x_{n+1}, \ldots, x_1, x_n) > 1$, there exists a positive integer $M$ such that

$$(d(x_n, x_{n+1}, \ldots, x_1, x_n))^M > d((z_k)_{k=1}^s).$$

This is a contradiction to Path Monotonicity because $x_n \simeq x_n$ and $(z_k)_{k=1}^s$ is a path from from $x_i$ to $x_j$ and $x_i \gg x_j$.

**Subcase 2.2:** Suppose that

$$\frac{p(x_1, x_1x_n)}{p(x_n, x_1x_n)} < \frac{p(x_1, x_1x_2)}{p(x_2, x_1x_2)}\frac{p(x_2, x_2x_3)}{p(x_3, x_2x_3)} \cdots \frac{p(x_{n-1}, x_{n-1}x_n)}{p(x_n, x_{n-1}x_n)}.$$

Then

$$1 < \frac{p(x_1, x_1x_2)}{p(x_2, x_1x_2)}\frac{p(x_2, x_2x_3)}{p(x_3, x_2x_3)} \cdots \frac{p(x_{n-1}, x_{n-1}x_n)}{p(x_n, x_{n-1}x_n)}\frac{p(x_n, x_1x_n)}{p(x_1, x_1x_n)}.$$

So we obtain a path $(x_1, x_2, \ldots, x_n, x_1)$ from $x_1$ to $x_1$ such that $d(x_1, x_{n+1}, \ldots, x_1) > 1$. The rest of the proof is the same as in Subcase 2.1. □

## 4.3 Proof of Theorem 2

**Necessity:** Since a threshold general Luce model is a special case of general Luce model, it suffices to show that the model satisfies Path Monotonicity. First note that for all $x, y \in X$, $x \gg y$ if and only if $u(x) > (1+\varepsilon)u(y)$. To show Path Monotonicity, choose any pair of paths $(z_i)_{i=1}^s$ from $x$ to $y$ and $(z'_i)_{i=1}^t$ from $x'$ to $y'$. Suppose that $x \gg y$ and $x' \simeq y'$. By Strong Dominance Transitivity, we must have $z'_i \simeq z'_{i+1}$ for all $i$. So $p(z'_i, z'_i z'_{i+1})/p(z'_{i+1}, z'_i z'_{i+1}) = u(z'_i)/u(z'_{i+1})$ for all $i$. Therefore,

$$d((z'_i)_{i=1}^t) = \frac{u(x')}{u(y')} \leq 1 + \varepsilon,$$

where the last inequality holds because $x' \simeq y'$. So it suffices to show $d((z_i)_{i=1}^s) > 1 + \varepsilon$. There are two cases. If $z_i \gg z_{i+1}$ for some $i$, then $p(z_i, z_i z_{i+1})/p(z_{i+1}, z_i z_{i+1}) = \infty$, so that $d((z_i)_{i=1}^s) = \infty$. If $z_i \simeq z_{i+1}$ for all $i$, then, we have

$$d((z_i)_{i=1}^s) = \frac{u(x)}{u(y)} > 1 + \varepsilon,$$

where the last inequality holds because $x \gg y$.

<u>**Sufficiency:**</u> By Lemma 7 and the finiteness of $X$, we can order all alternatives in $X$ as follows: $x_1 \geq x_2 \geq \cdots \geq x_N$. In the following, we will keep this notation.

First, consider the case where $x_i \gg x_{i+1}$ for all $i$. In this case, for all $A \subset X$, $p(x, A) = 1$ if $x \geq y$ for all $y \in A$ and $p(x, A) = 0$ otherwise. To construct the function $u$, set $u(x_i) = 2(N + 1 - i)$ for all $i$ and set $\varepsilon = 1/2$. Then, for all $i, j$ such that $i < j$, $u(x_i) > (1 + \varepsilon)u(x_j)$. Then, for all $A \subset X$, $c(A) = \{x\}$, where $x \geq y$ for all $y \in A$. Hence, this $(u, \varepsilon)$ represents $p$ and the proof is completed in this case.

In the following, consider the case in which $x_k \simeq x_{k+1}$ for some $k$. First we define $\varepsilon$. By Path Monotonicity, we have $\min_{j,l:x_j \gg x_{j+l}} d((x_i)_{i=j}^{j+l}) > \max_{j',l':x_{j'} \simeq x_{j'+l'}} d((x_i)_{i=j'}^{j'+l'})$. Choose a number $\varepsilon$ such that

$$\min_{j,l:x_j \gg x_{j+l}} d((x_i)_{i=j}^{j+l}) - 1 > \varepsilon > \max_{j',l':x_{j'} \simeq x_{j'+l'}} d((x_i)_{i=j'}^{j'+l'}) - 1.$$

To see that $\varepsilon$ is nonnegative, remember $x_k \simeq x_{k+1}$ for some $k$. Hence,

$$\max_{j',l':x_{j'} \simeq x_{j'+l'}} d((x_i)_{i=j'}^{j'+l'}) \geq d(x_k, x_{k+1}) = \frac{p(x_k, x_k x_{k+1})}{p(x_{k+1}, x_k x_{k+1})} \geq 1,$$

where the second inequality holds because $x_k \geq x_{k+1}$.

Now, we define the function $u$. Define $u(x_1) = 1$. We define $u(x_i)$ for all $i > 1$ sequentially as follows. If $x_{i-1} \simeq x_i$, define

$$u(x_i) = \frac{p(x_i, x_i x_{i-1})}{p(x_{i-1}, x_i x_{i-1})} u(x_{i-1}). \tag{10}$$

Since $x_{i-1} \geq x_i$, we have $p(x_{i-1}, x_i x_{i-1}) \geq p(x_i, x_i x_{i-1})$. Hence, $u(x_{i-1}) \geq u(x_i)$.

If $x_{i-1} \gg x_i$, choose a positive number $u(x_i)$ such that

$$u(x_{i-1}) > u(x_i)(1 + \varepsilon). \tag{11}$$

This definition implies $u(x_{i-1}) > u(x_i)$. By this way, we have defined a nonnegative number $\varepsilon$ and positive numbers $(u(x_i))_{i=1}^n$ such that $u(x_i) \geq u(x_{i+1})$ for all $i$.

By Lemma 5 and Proposition 1, it suffices to show the following three steps.

**Step** 1: $(X, u)$ has $L$-property.

**Proof of Step** 1: Assume that $y \simeq z$ and $y \geq z$. Remember $X = \{x_1, \ldots, x_N\}$ such that $x_i \geq x_{i+1}$ for all $i$. So there exist $x_s$ and $x_{s+t}$ such that $x_s = y$ and $x_{s+t} = z$. By Strong Dominance Transitivity, $\{x_s, \ldots, x_{s+t}\}$ is pairwise comparable. For $n \in \{s, \ldots, s+t-1\}$, the definition shows

$$\frac{p(x_n, x_n x_{n+1})}{p(x_{n+1}, x_n x_{n+1})} = \frac{u(x_n)}{u(x_{n+1})}. \tag{12}$$

Since $\{x_s, \ldots, x_{s+t}\}$ is pairwise comparable, by applying Weak IIA repeatedly, we obtain for all $n$ such that $s \leq n \leq s + t - 1$,

$$\frac{p(x_n, \{x_s, \ldots, x_{s+t}\})}{p(x_{n+1}, \{x_s, \ldots, x_{s+t}\})} = \frac{p(x_n, x_n x_{n+1})}{p(x_{n+1}, x_n x_{n+1})}. \tag{13}$$

Hence by (12) and (13), it follows that

$$\begin{aligned}
\frac{u(y)}{u(z)} &= \frac{u(x_s)}{u(x_{s+t})} \\
&= \frac{u(x_s)}{u(x_{s+1})} \cdots \frac{u(x_{s+t-1})}{u(x_{s+t})} \\
&= \frac{p(x_s, \{x_s, \dots, x_{s+t}\})}{p(x_{s+1}, \{x_s, \dots, x_{s+t}\})} \cdots \frac{p(x_{s+t-1}, \{x_s, \dots, x_{s+t}\})}{p(x_{s+t}, \{x_s, \dots, x_{s+t}\})} \quad (\because (12), (13)) \\
&= \frac{p(x_s, \{x_s, \dots, x_{s+t}\})}{p(x_{s+t}, \{x_s, \dots, x_{s+t}\})} \\
&= \frac{p(x_s, x_s x_{s+t})}{p(x_{s+t}, x_s x_{s+t})} \quad (\because (13)) \\
&= \frac{p(y, yz)}{p(z, yz)}.
\end{aligned}$$

□

**Step** 2: $u(z) > (1+\varepsilon)u(y) \Rightarrow z \gg y$.

**Proof of Step** 2: Suppose by way of contradiction that $u(z) > (1+\varepsilon)u(y)$ and $z \simeq y$. Then, by $L$-property,

$$d(z, y) = \frac{p(z, zy)}{p(y, zy)} = \frac{u(z)}{u(y)} > 1 + \varepsilon.$$

This contradicts the definition of $\varepsilon$.

**Step** 3: $z \gg y \Rightarrow u(z) > (1+\varepsilon)u(y)$.

**Proof of Step** 3: Suppose that $z \gg y$. If $z = x_j$ then $y \neq x_{j+1}$ by the definition of $u$.[5] So there exists an integer $l$ strictly larger than 1 such that $y = x_{j+l}$.

**Case** 1: There exists some $i$ such that $x_i \gg x_{i+1}$. Then

$$\frac{u(z)}{u(y)} \equiv \frac{u(x_j)}{u(x_{j+l})} = \frac{u(x_j)}{u(x_{j+1})} \cdots \frac{u(x_{j+l-1})}{u(x_{j+l})} \geq \frac{u(x_i)}{u(x_{i+1})} > 1 + \varepsilon$$

because $u(x_{i'})/u(x_{i'+1}) \geq 1$ for any $i'$.

[5] Otherwise, we have $x_j \equiv z \gg y \equiv x_{j+1}$, which shows $u(z)/u(y) \equiv u(x_j)/u(x_{j+1}) > 1 + \varepsilon$, which is a contradiction.

**Case** 2: $x_i \simeq x_{i+1}$ for all $i$. Hence, by $L$-property, $u(x_i)/u(x_{i+1}) = p(x_i, x_i x_{i+1})/p(x_{i+1}, x_i x_{i+1})$. Therefore, by the definition of $\varepsilon$, since $z \gg y$,

$$\frac{u(z)}{u(y)} \equiv \frac{u(x_j)}{u(x_{j+l})} = \frac{u(x_j)}{u(x_{j+1})} \cdots \frac{u(x_{j+l-1})}{u(x_{j+l})} = d(\{x_i\}_{i=j}^{j+l}) > 1 + \varepsilon.$$

□

□

## 4.4 Proof of Propositions 2

Suppose there exist two general Luce model $(u, c)$ and $(u', c)$ represents the same stochastic choice function $p$.

Note that the values of $u$ and $u'$ are positive numbers. Hence, we can choose $y \in X$ such that $u(y) \neq 0 \neq u'(y)$. Define

$$\lambda = \frac{u(y)}{u'(y)}.$$

Choose any $x \in X$ to show $u(x) = \lambda u'(y)$.

First, consider the case in which $x \simeq y$, then by the definition of the general Luce model,

$$\frac{u(x)}{u(y)} = \frac{1}{p(y, xy)} - 1 = \frac{u'(x)}{u'(y)}.$$

So, $u(x) = \dfrac{u(y)}{u'(y)} u'(x) = \lambda u'(x)$.

Now, consider the case in which $y \gg x$. By Richness assumption, there exists a sequence $\{z_j\}_{j=1}^n$ of $X$ such that (i) $z_1 = y$ and $z_n = x$; (ii) $z_j \simeq z_{j+1}$ for all $j$. Since for each $j$, we have $z_j \simeq z_{j+1}$. So this means that

$$\frac{u(z_{j+1})}{u(z_j)} = \frac{1}{p(z_j, z_j z_{j+1})} - 1 = \frac{u'(z_{j+1})}{u'(z_j)}.$$

Therefore,

$$\frac{u(x)}{u(y)} = \frac{u(z_n)}{u(z_{n-1})} \cdots \frac{u(z_2)}{u(z_1)} = \frac{u'(z_n)}{u'(z_{n-1})} \cdots \frac{u'(z_2)}{u'(z_1)} = \frac{u'(x)}{u'(y)}.$$

This proves that $u(x) = \dfrac{u(y)}{u'(y)} u'(x) = \lambda u'(x)$.

Finally, we consider the case in which $x \gg y$. We have

$$\frac{u(y)}{u(x)} = \frac{u(z_n)}{u(z_{n-1})} \cdots \frac{u(z_2)}{u(z_1)} = \frac{u'(z_n)}{u'(z_{n-1})} \cdots \frac{u'(z_2)}{u'(z_1)} = \frac{u'(y)}{u'(x)}.$$

This proves that $u(x) = \dfrac{u(y)}{u'(y)} u'(x) = \lambda u'(x)$.

We show that $\lambda$ is positive. If $\lambda$ is negative, there exists $x, y \in X$ such that $u(x) > u(y)$ and $u'(y) > u'(x)$, which contradicts that $(u, c)$ and $(u', c)$ represents the same stochastic choice function $p$.

# 5 Extension of Theorem 2

We present an extension of Theorem 2 to the case where $X$ is not finite. The extension requires a stronger path monotonicity axiom.

**Definition**: For all $x, y \in X$ such that $x \geq y$,

$$\begin{aligned} \underline{d}(x, y) &= \inf \Big\{ d((z_i)_{i=1}^n) : (z_i)_{i=1}^n \text{ is a path from } x \text{ to } y \Big\}, \\ \overline{d}(x, y) &= \sup \Big\{ d((z_i)_{i=1}^n) : (z_i)_{i=1}^n \text{ is a path from } x \text{ to } y \Big\}. \end{aligned}$$

**Axiom 10.** *(Strong Path Monotonicity):*

$$\inf \Big\{ \underline{d}(x, y) : x, y \in X \textit{ such that } x \gg y\} > \sup\{\overline{d}(x', y') : x', y' \in X \textit{ such that } x' \simeq y' \Big\}.$$

Strong Path Monotonicity is stronger than Path Monotonicity in that Path Monotonicity requires that the inequality holds even “in the limit”.

**Theorem 3.** *Under Richness, a stochastic choice function satisfies Weak Regularity, IDA, Weak IIA , and Strong Path Monotonicity if and only if it is a threshold general Luce model.*

In the following, we will prove the theorem. We proceed by establishing intermediate lemmas. To state the lemmas, we define a preliminary concept: The set $X'$ is *closed under intervals* if for any $x, y \in X'$ with $x \geq y$ the set $\{z \in X : x \geq z \geq y\}$ is contained in $X'$.

**Lemma 8.** *There is one pair $(X', u)$ that satisfies L-property and $X'$ is closed under intervals.*

**Proof**: Let $x \simeq y$. Assume without loss of generality that $x \geq y$. Let $X' = \{z \in X : x \geq z \geq y\}$. Let $u(x) = 1$. Define $u(z) = p(z, xz)/p(x, xz)$. By Strong Dominance Transitivity, all alternatives in $X'$ are comparable because $x \simeq y$. Then it follows from Weak IIA that $(X', u)$ satisfies $L$-property: for any $z, w \in X'$,

$$\frac{u(z)}{u(w)} = \frac{p(z, xz)/p(x, xz)}{p(w, xw)/p(x, xw)} = \frac{p(z, xzw)/p(x, xzw)}{p(w, xzw)/p(x, xzw)} = \frac{p(z, xzw)}{p(w, xzw)} = \frac{p(z, zw)}{p(w, zw)}.$$

It is also immediate that $X'$ is closed under intervals. □

**Lemma 9.** *Let $(X', u)$ be a pair with L-property in which $X'$ is closed under intervals. Suppose that there is $x \in X \setminus X'$ and $y \in X'$ such that $x \simeq y$. Then, there is a pair $(\hat{X}, \hat{u})$ that has L-property, $\hat{X}$ is closed under intervals, and where*

$$\hat{X} = \{z \in X : y \geq z \geq x\} \cup X'$$

*if $y \geq x$ and*

$$\hat{X} = \{z \in X : x \geq z \geq y\} \cup X'$$

*if $x \geq y$.*

**Proof**: Suppose that $y \geq x$. Define $\hat{X}$ as in the statement of the lemma. Let $\hat{u}|_{X'} = u$. For all $z \in X$ with $y \geq z \geq x$, let $\hat{u}(z) = u(y)(p(z, zy)/p(y, zy))$. Since $x \simeq y$, for all

$z \in X$ with $y \geq z \geq x$, it must hold that $x \simeq z \simeq y$ because of Strong Dominance Transitivity.

**Step** 1: $(\hat{X}, \hat{u})$ has $L$-property.

**Proof**: The $L$-property is immediate to verify for two alternatives of $X'$, as $\hat{u}$ is identical to $u$ on $X'$. So to check $L$-property we need to look at two cases.

**Case** 1: Firstly, let $y' \in X'$ and $z \in \hat{X} \setminus X'$ are comparable. Note that we cannot have $z \geq y'$ as $X'$ is closed under intervals.[6] Hence $y' > z$.

If $y \geq y'$ then $y \simeq y'$ because $y \simeq z$. (If $y \gg y'$, then $y \gg z$ because $y' > z$, which is a contradiction.)

If $y' \geq y$ then $y \simeq y'$ as $y' \simeq z$. (If $y' \gg y$, then $z \gg y'$ because $z \geq y'$, which is a contradiction.)

Either way we know that $u(y)/u(y') = p(y, yy')/p(y', yy')$ by $L$-property of $(X', u)$. We can then use Weak IIA as follows:

$$\frac{\hat{u}(z)}{\hat{u}(y')} = \frac{\hat{u}(z)}{u(y)} \frac{u(y)}{\hat{u}(y')} = \frac{p(z, zy)}{p(y, zy)} \frac{p(y, yy')}{p(y', yy')} = \frac{p(z, zyy')}{p(y, zyy')} \frac{p(y, zyy')}{p(y', zyy')} = \frac{p(z, zy')}{p(y', zy')},$$

which establishes $L$-property.

**Case** 2: Secondly, if $z, y' \in \hat{X} \setminus X'$, then $z$, $y'$ and $y$ are comparable because $x \simeq y$. By the definition of $\hat{u}$ we have that

$$\frac{\hat{u}(z)}{\hat{u}(y')} = \frac{p(z, zy)/p(y, zy)}{p(y', zy')/p(z, zy')} = \frac{p(z, zyy')}{p(y, zyy')} \frac{p(y, zyy')}{p(y', zyy')} = \frac{p(z, zy')}{p(y, zy')},$$

using Weak IIA again.

**Step** 2: $\hat{X}$ is closed under intervals.

[6] If $z \geq y'$, then $y \geq z \geq y'$. Since $y, y' \in X'$ and $X'$ is closed under intervals, $z$ must belong to $X'$, which is a contradiction.

**Proof**: Choose $z, x' \in \hat{X}$ and choose $w$ between $z$ and $x'$. If $z, x' \in X'$, then $w \in X' \subset \hat{X}$ because $X'$ is closed under interval. If $z, x' \notin X'$, then it must hold that $y \geq w \geq x$ or $x \geq w \geq y$ because $\geq$ is transitive. Hence, $w \in \hat{X}$. Therefore, in the following, consider the case in which only one of $z$ or $x'$ belongs to $X'$.

Without loss of generality, assume $x' \in X'$ and $y \geq z \geq x$.

**Case** 1: First, consider the case where $z \geq x'$. Choose $w$ such that $z \geq w \geq x'$. Then, $y \geq z \geq w \geq x'$. Since $\geq$ is transitive, $y \geq w \geq x'$. Since $y, x' \in X'$ and $X'$ is closed under intervals, we have $w \in X' \subset \hat{X}$.

**Case** 2: Second, consider the case where $x' \geq z$. Choose $w$ such that $x' \geq w \geq z$.

**Case** 2.1: $y \geq x'$. Then, $y \geq x' \geq w \geq z \geq x$, so that $y \geq w \geq x$. Hence, $w \in \hat{X}$.

**Case** 2.2: $x' \geq y$. If $x' \geq w \geq y$, then $w \in X' \subset \hat{X}$ because $x', y \in X'$ and $X'$ is closed under interval. $\square$

**Lemma 10.** $(X, u)$ *has L-property.*

**Proof**: Consider the class of all pairs $(X', u)$ that satisfy $L$-property and for which $X'$ is closed under intervals. This class is nonempty by Lemma 8. The class of all pairs is partially ordered by the following order. Let $(\hat{X}, \hat{u})$ be larger than $(X', u)$ if $X'$ is a subset of $\hat{X}$ and if $u$ is the restriction of $\hat{u}$ to $X'$. It is obvious that this order is a partial order.

Let $(X_b, u_b)_{b \in B}$ be a chain in the class of pairs that satisfy $L$-property and are closed under intervals with respect to the partial order defined above.

Let $\bar{X} = \cup_{b \in B} X_b$ and $\bar{u} : \bar{X} \to \mathbf{R}_{++}$ be defined by $\bar{u}(x) = u_b(x)$ for $b$ such that $x \in X_b$. This is well defined because of the chain property. Then $(\bar{X}, \bar{u})$ is an upper bound on $(X_b, u_b)_{b \in B}$. It is also easy to see that $(\bar{X}, \bar{u})$ has $L$-property and is closed under intervals.

So any chain has an upper bound. By Zorn's Lemma there is a maximal pair $(\bar{X}, \bar{u})$ with $L$-property and that is closed under intervals. We want to argue that $\bar{X} = X$. So suppose towards a contradiction that $\bar{X} \neq X$. Then there must exist $x \notin \bar{X}$. There is $y \in \bar{X}$ with either $x \geq y$ or $y \geq x$ (or both). Suppose without loss of generality that $x \geq y$. By Richness there is a sequence $(z_i)_{i=1}^n$ with $x = z_1$ and $z_n = y$, such that $z_i \simeq z_{i+1}$ for all $i$ such that $1 \leq i \leq n-1$. Then there must exist $i$ such that $z_i \notin \bar{X}$ and $z_{i+1} \in \bar{X}$. Since $z_i \simeq z_{i+1}$, Lemma 9 would imply a larger pair with $L$-property and closed under intervals. This is a contradiction □

Fix $u : X \to \mathbf{R}_{++}$ as obtained from Lemma 10. Then, by Lemma 5, we obtain the representation.

Choose $\varepsilon$ such that

$$\inf\{\underline{d}(x,y) : x \gg y\} > 1 + \varepsilon > \sup\{\overline{d}(x,y) : x \simeq y\}.$$

to show $c(A) = \{y \in A | (1+\varepsilon)u(y) \geq u(z) \text{ for all } z \in A\}$. Then, by using $L$-property and Richness, Steps 2 and 3 in the proof of Theorem 2 hold. (Richness is necessary in Step 3.)

# References


Block, H. and J. Marschak (1960): "Random orderings and stochastic theories of responses," in *Contributions to probability and statistics*, ed. by O. G. Hoeffding, Hotelling, and Madow, Stanford University Press, vol. 2, 97–132.

Brady, R. and J. Rehbeck (2014): "Menu-Dependent Stochastic Consideration," Tech. rep., mimeo, University of California at San Diego.

Echenique, F., K. Saito, and G. Tserenjigmid (2013): "The Perception-Adjusted Luce Model," Tech. rep., mimeo, California Institute of Techonology.

Fishburn, P. C. (1973): "Interval representations for interval orders and semiorders," *Journal of Mathematical Psychology*, 10, 91–105.

Fudenberg, D. and T. Strzalecki (2014): "Dynamic logit with choice aversion," Tech. rep., mimeo.

Gul, F., P. Natenzon, and W. Pesendorfer (2014): "Random choice as behavioral optimization," *Econometrica*, 82, 1873–1912.

Horan, S. (2014): "Random Consideration and Choice," Tech. rep., mimeo, Université du Québeca Montréal.

Luce, R. D. (1956): "Semiorders and a theory of utility discrimination," *Econometrica, Journal of the Econometric Society*, 178–191.

——— (1959): *Individual Choice Behavior a Theoretical Analysis*, John Wiley and sons.

Manzini, P. and M. Mariotti (2012): "Stochastic choice and consideration sets," Forthcoming, Econometrica.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012): "Revealed attention," *The American Economic Review*, 102, 2183–2205.