Published in Economic Theory.
Volume 68(4) (Nov 2019) pp. 811-826
https://doi.org/10.1007/s00199-018-1145-5

General Luce Model*

Federico Echenique[†] and Kota Saito[‡]
California Institute of Technology

August 12, 2018

Abstract

We extend the Luce model of discrete choice theory to satisfactorily handle zero-probability choices. The Luce mode struggles to explain choices that are not made. The 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. 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, when grouped with a *comparable* set of alternatives.

JEL Classification: D01, D03

Keywords: The Luce model, Logit model, Independence of Irrelevant Alternatives axiom

1 Introduction

Alice likes wine better than beer, and beer better than soda. When offered a choice between wine or beer, she chooses wine most of the time, but on occasion she may

*Echenique thanks the National Science Foundation for its support through the grants SES 1558757 and CNS-1518941. Saito thanks the National Science Foundation for its support through the grant SES 1558757.
[†]email address: fede@hss.caltech.edu
[‡]email address: saito@caltech.edu

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, she may on occasion decide to drink soda despite liking beer more than 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}.

Many choice situations are similar to Alice's. They pose a problem for mainstream discrete choice theory, because Luce's model does not handle zero-probability choices well.[1] 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 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 sometimes 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 we say 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 sets of comparable alternatives. The agent chooses with probability zero those alternatives that are dominated in the choice set. Importantly, alternatives that are not chosen from one set may be chosen with positive probability from a set of comparable alternatives.

[1] A recent growing literature extends the Luce model to capture various behavioral phenomena, but typically requires strictly positive choice probabilities.

An important aspect of our model is that the comparability relation (the binary relation "is comparable to") may not be transitive. Recall Alice's situation. It is possible that wine is better, but does not dominate, beer; beer is better than, but does not dominate soda; while wine dominates soda. Hence, the pairs of 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] Introducing the theory of semiorders into 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 purpose of our paper is as a study of stochastic choice that can sometimes be deterministic. One of Luce's crucial assumptions is that choice 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 axiomatic extension of Luce's model to accommodate deterministic choices.[3]

We present three versions of our model. The first version generalizes Luce's model by allowing for an agent to never choose some elements in a choice set. In this first model, the support of the stochastic choice function does not have any structure. The support can be any subset of a choice set. The first model is called the *general Luce model*.

The second version of our model, the *two-stage Luce model*, is a special case of the general Luce model in which the support of the stochastic choice function consists of all the alternatives in a choice set that are not dominated by any other alternative. The two-stage Luce model is endowed with a dominance binary relation, and zero probability choices are accounted for using dominance, as suggested above.

The third version of our model, the *threshold Luce model*, is a special case of the

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

[3]The paper by Fudenberg et al. (2015) allows for deterministic choices, but does not focus on Luce.

two-stage Luce model in which the dominance relation is directly tied to utility. One alternative is dominated by another alternative if the utility of the former is sufficiently smaller than the utility of the latter.

The main results in the paper are axiomatic characterizations of these models; that is, a complete description of the observable stochastic choices that are consistent with the models.

The axioms are simple. Our first axiom is a weakening of Luce's *Independence of Irrelevant Alternatives* (IIA) axiom. We call the axiom *Cyclical Independence.* The axiom is weaker than Luce's IIA in that it allows zero probability choice. Cyclical independence is the key to generalizing Luce's model to allow for zero probability choices.

To characterize the two-stage Luce model, we need three more axioms. *Weak Regularity* requires that the probability of choosing an alternative cannot drop from positive to zero when one removes alternatives from a choice set. Weak regularity is a weakening of Luce's regularity axiom, which says that choice probabilities cannot go down when alternatives are removed. Our 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. As noted, we will say that x dominates y if y is never chosen when x is available. Our last axiom is a simple stochastic version of Sen's γ axiom, which we call *Probabilistic* γ.

To characterize the threshold Luce model, we need one more axiom, *Path Monotonicity*, which requires that comparisons of sequences of alternatives be consistent.

One possible interpretation of our models is that it describes an agent's behavior when her preferences are incomplete. The agent follows Luce's rule when he needs to choose from a choice set over which the agent's preferences are incomplete. Based on a similar interpretation, a recent paper by Cerreia-Vioglio et al. (2016) characterize *consistent* random choice rules and then study stochastic choice in a consumer theory setting.

To conclude this 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.

Several recent papers provide generalizations of the Luce model. Gul et al. (2014) axiomatize a generalization of the Luce model to address difficulties of the Luce model that arise when alternatives have common attributes. Fudenberg and Strzalecki (2014) axiomatize a generalization of a 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. Neither model allows zero probability choices. Echenique et al. (2013) axiomatize a generalization of the Luce model that incorporates the effects of attention. None of these three papers 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 the support of the stochastic choice as being alternatives that the agent pays no 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 alternatives considered. A more recent paper by Brady and Rehbeck (2014) axiomatizes a model that encompasses the model of Manzini and Mariotti (2012), and Cerreia-Vioglio et al. (2017) present an analysis of the multinomial Logit model that allows for zero probability choices. Horan (2014) has proposed a new model of limited consideration based on the random utility model proposed by Block and Marschak (1960). Ahumada and Ulku (2017) independently developed the most general version of our model.

Finally, it is worth mentioning a very recent paper, Flores et al. (2017), that uses our generalized Luce model to study assortment optimization problems. Assortment optimization is an important pricing problem studied in operations research.

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.

2 Models

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 any 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.

The most general version of our model is as follows.[4]

Definition 1 *A stochastic choice function p is called a general Luce function if there exists a pair (u, c) of functions $u : X \to \mathbf{R}_{++}$ and $c : 2^X \setminus \emptyset \to 2^X \setminus \emptyset$ such that $c(A) \subseteq A$ for any subset A of X, and moreover,*

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

for all $A \subseteq X$.

The function c in the general Luce model captures the support of the stochastic choice. The function c is arbitrary. For example, deterministic choice can be captured using a function c for which $c(A)$ is always a singleton. Of course, if $c(A) = A$ for all subsets A of X, then p coincides with Luce's (1959) model.

In the following, we provide two special cases of the general Luce model. The first special case is the two-stage Luce model. The model restricts c so that $c(A)$ is the set

[4] A first draft of our paper focused on the threshold Luce mode, which we define below. We thank an anonymous referee and David Ahn for suggesting that we extend the discussion to cover the general Luce model and the two-stage general Luce model. After we wrote the first version of our paper, which did not include our most general model, Ahumada and Ulku (2017) independently developed the present generalization.

of undominated elements in the subset A.

Definition 2 *A general Luce function* (u, c) *is called a two-stage function if there exists a strict partial order (i.e., transitive and irreflexive binary relation*[5]*)* $\gg$ *on* X *such that for any* $A \subseteq X$

$$c(A) = \{x \in A : \nexists y \in A \ s.t. \ y \gg x\}.$$

We call $\gg$ *the dominance relation.*

Notice that the difference between the general Luce model and two-stage Luce model is obtained by adding restrictions to the function c so that it is induced by a strict partial order. Such characterizations of strict partial orders has been obtained by Jamison and Lau (1973).

Given a dominance relation $\gg$, we define the *comparability relation* $\simeq$ as follows: for all $x, y \in X$, $x \simeq y$ if $x \not\gg y$ and $y \not\gg x$.

Given the two-stage Luce model, the dominance relation and the compatibility relation can be identified from p as follows:

Remark 1 *For all* $x, y \in X$ *such that* $x \neq y$, *(i)* $x \gg y$ *if* $p(x, xy) = 1$; *(ii)* $x \simeq y$ *if* $0 < p(x, xy) < 1$.

The two-stage Luce model is, in a sense, about intransitive indifference. Recall our example in the introduction. In the two-stage Luce model, 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 $x \gg z$. Such lack of transitivity of the comparability binary relation is analogous to the lack of transitivity of indifference.

In the two-stage Luce model, c satisfies the following property:

Remark 2 *For any* $x \in X$ *and* $A \subset X$, $x \notin c(A) \Longrightarrow c(A) = c(A \setminus \{x\})$.

This property means that removing a dominated element x in a choice set A does not affect the agent's choice. This property is satisfied by the *attention filters* of Masatlioglu

[5] Transitivity and irreflexivity imply asymmetry.

et al. (2012). Overall, their model and ours are quite different, but the object c shares a similar interpretation. That said, unlike Masatlioglu et al. (2012), in our model the function c is uniquely identified. This difference comes partly from the fact that we study stochastic choices while Masatlioglu et al. (2012) study deterministic choices.[6]

The function c in the two-stage Luce model is independent of u. It is reasonable to think that dominance may sometimes 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 the threshold Luce model.

Definition 3 *A general Luce function (u, c) is called a threshold Luce function if there exists a nonnegative number α such that, for any subset A of X,*

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

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

Notice that the function c in the two-stage Luce model is induced by a strict partial order, and thus more general than the one induced by a semiorder. Notice also that the function c in the threshold Luce model is induced by a semiorder. (By linearization, the formula can be written in a more conventional form as the choice correspondence based on a semiorder.)

[6]The function c also satisfies the following property. If $B \subset A$ and $c(A) \subset B$, then $c(B) = c(A)$. This property appears in several studies, such as Aizerman and Malishevski (1981).

3 Axioms and results

3.1 General Luce model

The general Luce model is the most general model in this paper. The model is characterized by the following axiom.

Axiom 1 *(Cyclical Independence): For any sequence $x_1, x_2, \ldots, x_n \in X$, if there exists a sequence $A_1, \ldots, A_n$ of nonempty subsets of X such that $p(x_i, A_i) > 0$ and $p(x_{i+1}, A_i) > 0$, for $i = 1, \ldots, n$ (using summation mod n)*[7]*, then*

$$\frac{p(x_1, A_n)}{p(x_n, A_n)} = \frac{p(x_1, A_1)}{p(x_2, A_1)} \frac{p(x_2, A_2)}{p(x_3, A_2)} \cdots \frac{p(x_{n-1}, A_{n-1})}{p(x_n, A_{n-1})}. \tag{1}$$

Cyclical Independence is weaker than Luce IIA. To see this notice that Luce IIA implies that equation (1) should hold for any any sequences $x_1, x_2, \ldots, x_n \in X$ and $A_1, \ldots, A_n$ such that $x_i, x_{i+1} \in A_i$ for all i.[8] Cyclical Independence requires that (1) should hold when $p(x_i, A_i) > 0$ and $p(x_{i+1}, A_i) > 0$ for all i. Under the assumption that $p(x_{i+1}, A_i) > 0$ for all i, the fractions in (1) are well defined. Our interpretation of the condition $p(x_i, A_i) > 0$ and $p(x_{i+1}, A_i) > 0$ is that x_i and x_{i+1} are not dominated in choice set A_i.

Theorem 1 *A stochastic choice function satisfies cyclical independence if and only if it is a general Luce function.*

Remark 3 *The function c is the support of p for each A, and therefore uniquely identified.*

To state the uniqueness property of u, we need a preliminary definition. For any $x, y \in X$, define xIy if there exists $A \subset X$ such that $p(x, A)p(y, A) > 0$.

[7] $n = 2$ is allowed.

[8] Luce IIA implies that for any i, $p(x_i, A_i)/p(x_{i+1}, A_i) = p(x_i, \cup_{j=1}^{n-1} A_j)/p(x_{i+1}, \cup_{j=1}^{n-1} A_j)$. So the right hand side becomes $p(x_1, \cup_{j=1}^{n-1} A_j)/p(x_n, \cup_{j=1}^{n-1} A_j)$, which is $p(x_1, A_n)/p(x_n, A_n)$ by Luce IIA again.

Remark 4 *Two general Luce functions* (u, c) *and* (u', c') *represent the same stochastic choice function if and only if* $c = c'$ *and for every* $x, y \in X$ *such that there exists* $A \subset X$ *with* $\{x, y\} \subset c(A)$*, we have* $u(x)/u(y) = u'(x)/u'(y)$*.*

3.2 Two-Stage Luce model

We need additional axioms to characterize the two-stage Luce model. Our first axiom is a weaker version of Luce's axiom of *regularity*. Luce's axiom says that the probability of choosing x cannot decrease when the choice set shrinks.

Axiom 2 *(Weak Regularity): If* $p(x, xy) = 0$ *and* $y \in A$*, then* $p(x, A) = 0$*.*

Weak Regularity can be replaced with the stronger axiom: if $A \subseteq B$ and $p(x, A) = 0$, then $p(x, B) = 0$. The stronger axiom may be more natural, and our characterization continues to hold. It is worth emphasizing that the two-stage Luce model can violate the standard regularity axiom (meaning Luce's original version of the axiom), which demands that the probability of choosing x is monotone decreasing in the set of alternatives from which it is chosen. Notice that the contraposition of the axiom can be seen as a version of Sen's α axiom. (See Sen (1971).)

Our next axiom says that *removing a dominated alternative does not affect choices.* This means that when x is dominated, it does not affect whatever considerations govern choices among the remaining alternatives.

Axiom 3 *(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 axiom says that if an alternative x is not pairwise-dominated by any other alternative y in A (i.e., $p(x, xy) > 0$), then the alternative must be chosen with positive probability.

Axiom 4 *(Probabilistic* γ*) If* $p(x, xy) > 0$ *for any* $y \in A$*, then* $p(x, A) > 0$*.*

If x is dominated in A it must be dominated by some $y \in A$, therefore we must have $p(x, xy) = 0$. Probabilistic γ is in spirit similar to Sen's γ axiom (Sen (1971)).

Theorem 2 *A stochastic choice function satisfies Cyclical Independence, Weak Regularity, Independence of Dominated Alternatives, and Probabilistic γ if and only if it is a two-stage Luce function.*

3.3 Threshold Luce Model

In the two-stage Luce model, the dominance relation may be unrelated to the utility u. In contrast, the threshold Luce model imposes a relation between $c(A)$ and the utility of the elements of A. In order to calibrate the magnitude of the parameter α, we need an additional axiom. Recall that in the two-stage model for all $x, y \in X$ such that $x \neq y$, (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 $y \not\gg x$.

Our last axiom requires definitions and some notational conventions. First, we introduce the *stochastic revealed preference* relation, denoted by $\geq$, and defined by:

Definition 4 *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.

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

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

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

The intensity captures the difference in utilities of x and y. By the definition (ii) of path, for all i, we can have $z_i \gg z_{i+1}$ but not $z_{i+1} \gg z_i$. Consequently, $p(z_i, z_i z_{i+1})/p(z_{i+1}, z_i z_{i+1})$ can be $+\infty$ but not zero. Therefore, $d((z_i)_{i=1}^s)$ is well defined.[9]

Under Luce's IIA, the intensity between two alternatives x and y must be the same along any two paths. Moreover, it must equal $p(x, xy)/p(y, xy)$. We do not assume Luce's IIA, so in our setup the intensity can be path-dependent. Our final axiom, Path Monotonicity, says that the intensity between any noncomparable pair of alternatives must be larger than the intensity 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 3 *A stochastic choice function satisfies Cyclical Independence, Weak Regularity, Independence of Dominated Alternatives, Probabilistic γ, and Path Monotonicity if and only if it is a threshold Luce function.*

There is a connection between Theorem 2 and the results on deterministic choice of Sen (1971). If one interprets c as a deterministic choice function, then Sen's axiomatization of normal choice functions (choice functions that satisfy Sen's α and γ axioms) by means of his δ axiom is similar to our characterization of c. [10] Sen's results have no connection to the stochastic choice aspects of our model.

[9]Suppose that $z_i \gg z_{i+1}$ for some i. Then $d((z_i)_{i=1}^s)$ becomes $+\infty$, even though $z_{j+1} \geq z_j$ for some $j \neq i$. This is not useful to compare the values of intensity. So, by (iii), we restrict when $z_i \gg z_{i+1}$ is allowed: it is allowed only when $z_i \geq z_{i+1}$ for all i.

[10]We appreciate a referee who pointed out the relationship.

4 Proofs

4.1 Proof of Theorem 1

Necessity of the axiom is obvious. So we will show the sufficiency.

Suppose that p satisfies Cyclical Independence. Let $I \subseteq X \times X$ be a binary relation on X defined by $x \, I \, y$ iff there exists $A \subseteq X$ such that $p(x, A)p(y, A) > 0$.

Consider a partition of X according to the connected components of (X, I). In fact we can wlog assume that X is a single connected component. The following argument can be done for each connected component otherwise.

Choose $x_0 \in X$ arbitrarily. Define u as follows. Let $u(x_0) = 1$. For each $x \in X$, there exists at least one sequence $x = x_1, x_2, \ldots, x_n = x_0$, with corresponding sequence $A_1, \ldots, A_{n-1}$ of nonempty subsets of X such that $p(x_i, A_i) > 0$ and $p(x_{i+1}, A_i) > 0$, for $i = 1, \ldots, n-1$. Define $u(x)$ by

$$u(x) = \frac{p(x_1, A_1)}{p(x_2, A_1)} \frac{p(x_2, A_2)}{p(x_3, A_2)} \cdots \frac{p(x_{n-1}, A_{n-1})}{p(x_n, A_{n-1})}.$$

For any two such sequences $x = x_1, x_2, \ldots, x_n = x_0$, and $x = x'_1, x'_2, \ldots, x'_m = x_0$, with corresponding $A_1, \ldots, A_{n-1}$ and $A'_1, \ldots, A'_{m-1}$, we can define the sequence $\{x''_i\}$ by $x''_i = x_i$ for $i = 1, \ldots, n$ and $x''_{n+i} = x'_{m-i}$ for $i = 1, \ldots, m-1$, and set $A''_i = A_i$ for $i = 1, \ldots, n-1$ and $A''_{n+i} = A'_{m-1-i}$ for $i = 0, \ldots, m-2$. Then we have that $p(x''_i, A''_i) > 0$ and $p(x''_{i+1}, A''_i) > 0$, for $i = 1, \ldots, n-1$, while $p(x''_n, A''_n) = p(x_0, A'_{m-1}) = p(x'_m, A'_{m-1}) > 0$ and $p(x''_{n+1}, A''_n) = p(x'_{m-1}, A'_{m-1}) > 0$, and, finally, that $p(x''_{n+i+1}, A''_{n+i}) = p(x'_{m-i-1}, A'_{m-1-i}) > 0$ and $p(x''_{n+i}, A''_{n+i}) = p(x'_{m-i}, A'_{m-1-i}) > 0$. We have

$$\begin{aligned} 1 &= \frac{p(x''_1, \{x''_1, x''_{n+m-1}\})}{p(x''_{n+m-1}, \{x''_1, x''_{n+m-1}\})} \\ &= \frac{p(x''_1, A''_1)}{p(x''_2, A''_1)} \cdots \frac{p(x''_{n-1}, A''_{n-1})}{p(x''_n, A''_{n-1})} \frac{p(x''_n, A''_n)}{p(x''_{n+1}, A''_n)} \cdots \frac{p(x''_{n+m-2}, A''_{n+m-2})}{p(x''_{n+m-1}, A''_{n+m-2})}, \end{aligned}$$

where the first equality holds because $x''_1 = x'_1 = x''_{n+m-1}$ and the last equality holds by

Cyclical Independence. This equation implies

$$\frac{p(x''_1, A''_1)}{p(x''_2, A''_1)} \cdots \frac{p(x''_{n-1}, A''_{n-1})}{p(x''_n, A''_{n-1})} = \frac{p(x''_{n+1}, A''_n)}{p(x''_n, A''_n)} \cdots \frac{p(x''_{n+m-1}, A''_{n+m-2})}{p(x''_{n+m-2}, A''_{n+m-2})}.$$

So by the definition of A''_i and x''_i, $\frac{p(x_1,A_1)}{p(x_2,A_1)} \cdots \frac{p(x_{n-1},A_{n-1})}{p(x_n,A_{n-1})} = \frac{p(x'_{m-1},A'_{m-1})}{p(x'_m,A'_{m-1})} \cdots \frac{p(x'_1,A'_1)}{p(x'_2,A_1)}$. Thus, the definition of u is independent of the sequence used in the definition.

For any A, define $c(A) = \text{supp}\, p(A)$. By definition of u, then, if $x, y \in c(A)$, then $p(x, A)p(y, A) > 0$, so that $u(y) = \frac{p(y,A)}{p(x,A)} u(x)$. Therefore, for any $x \in c(A)$,

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

so $p(x, A) = \frac{u(x)}{\sum_{y \in c(A)} u(y)}$. For any $x \notin c(A)$, $p(x, A) = 0$.

4.2 Proof of Theorem 2

4.2.1 Necessity of the axioms

A two-stage Luce function satisfies Cyclical Independence because it is a general Luce function.

We show that the two-stage Luce function satisfies Weak Regularity. Suppose that $p(x, xy) = 0$ and $y \in A$. Then, by definition, $y \gg x$ and $x \notin c(A)$. Therefore, $p(x, A) = 0$.

We show that the two-stage Luce function satisfies IDA. Suppose that $p(x, A) = 0$. Then, $x \notin c(A)$. So by Remark 2, $c(A) = c(A \setminus \{x\})$. 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 two-stage Luce function satisfies Probabilistic γ, supoose $p(x, xy) > 0$ for all $y \in A$, then $x \in c(A)$, so $p(x, A) > 0$.

4.2.2 Sufficiency of the axioms

To show the sufficiency, by Theorem 1, it suffices to show that (i) $\gg$ is transitive and asymmetric; (ii) $c(A) = \{x \in A | \nexists y \in A \text{ such that } y \gg x\}$ for any $A \subset X$. By the definition of $\gg$, $\gg$ is asymmetric, so we will show that $\gg$ is transitive in Lemma 1 below. Then (ii) follows from Lemma 2 below.

Axiom 6 *(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 $p(z, xz) = 0$. ■

For all $A \subset X$, define $\hat{c}(A) = \{x \in A | \nexists y \in A \text{ such that } y \gg x\}$. Note that Lemma 1 implies $\hat{c}(A) \neq \emptyset$. In the following, we will show that $c(A) = \hat{c}(A)$ for all $A \subset X$. Since $c(A) = \text{supp } p(A)$, it suffices to show the following lemma:

Lemma 2 *Suppose that Weak Regularity and IDA hold. Then for any subset A of X, $p(x, \hat{c}(A)) = p(x, A) > 0$ for any $x \in \hat{c}(A)$, and $p(x, A) = 0$ for all $x \in A \setminus \hat{c}(A)$.*

Proof: First, let $x \in A \setminus \hat{c}(A)$. Then there is $y \in A$ with $y \gg x$; meaning that $p(x, xy) = 0$. By Weak Regularity, $p(x, A) = 0$.

In second place, let $x \in \hat{c}(A)$. Let $A \setminus \hat{c}(A) = \{x_1, \ldots, x_n\}$. By our previous argument, $p(x_i, A) = 0$ 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 $p(A) = p(\hat{c}(A))$.

Finally, to show $p(x, \hat{c}(A)) = p(x, A) > 0$ for any $x \in \hat{c}(A)$, it suffices to show that $p(x, A) > 0$ for any $x \in \hat{c}(A)$. If $x \in \hat{c}(A)$, then $y \not\gg x$ for any $y \in A$. So by Probabilistic γ, $p(x, A) > 0$. ■

4.3 Proof of Theorem 3

4.3.1 Necessity of the axioms

Since a threshold Luce function is a special case of a two-stage Luce function, it suffices to show that the threshold Luce function satisfies Path Monotonicity. First note that for all $x, y \in X$, $x \gg y$ if and only if $u(x) > (1+\alpha)u(y)$. 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 $x' \simeq y'$. Then

$$\frac{u(x)}{u(y)} > 1+\alpha \geq \frac{u(x')}{u(y')}.$$

First, we will show that $z'_i \simeq z'_{i+1}$ for all i. Suppose by way of contradiction that $z'_i \gg z'_{i+1}$ for some i. Then by the definition of path, we have $z'_i \geq z'_{i+1}$ for all i. Hence

$$\frac{u(x')}{u(y')} = \frac{u(z'_1)}{u(z'_t)} = \frac{u(z'_1)}{u(z'_2)} \cdots \frac{u(z'_{t-1})}{u(z'_t)} > 1+\alpha,$$

which is a contradiction. Since $z'_i \simeq z'_{i+1}$ for all i, we must have 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,

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

where the last inequality holds because $x' \simeq y'$. So it suffices to show $d((z_i)_{i=1}^s) > 1+\alpha$. 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

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

4.3.2 Sufficiency of the axioms

Axiom 7 *(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 3 *Dominance Transitivity and Path Monotonicity imply Strong Dominance*

Transitivity.

Proof: First, we show (i). 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 because $x \gg y$ and $y \geq z$. The value of the intensity along the path is $+\infty$. On the other hand, (x, z, y) is a path from x to y because $x \simeq z$ and $z \simeq y$. The value of the intensity along the path is finite because $x \simeq z$ and $z \simeq y$. This violates Path Monotonicity.

Second, we show (ii). Assume $x \geq y$ and $y \gg z$ to show $x \gg z$. If $x \gg y$, then by Dominance Transitivity, we have $x \gg z$. So consider the case in which $x \simeq y$.

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 $y \gg z$, we have $y \gg x$. This contradicts with $x \geq y$.

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

Axiom 8 *(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 4 *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 $x \gg z$ and therefore $x > z$.

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,

$$\frac{p(x,xz)}{p(z,xz)} = \frac{p(x,xy)}{p(y,xy)}\frac{p(y,zy)}{p(z,zy)} \geq 1. \tag{3}$$

Hence, $x \geq z$. Moreover, if $x > y$ or $y > z$ hold, then (3) holds strictly. ■

Definition 7 *For any $X' \subseteq X$ and $u : X' \to \mathbf{R}_+$, we say that a pair (X', u) satisfies the 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, Cyclical Independence, and let (X, u) have the L-property. If $x \in c(A)$, then $p(x, c(A)) = \frac{u(x)}{\sum_{y\in c(A)} u(y)}$.*

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

$$\frac{p(x_j,c(A))}{p(x_1,c(A))} = \frac{p(x_j,A)}{p(x_1,A)} = \frac{p(x_j,x_1x_j)}{p(x_1,x_1x_j)} = \frac{u(x_j)}{u(x_1)},$$

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

$$p(x_j, c(A)) = \frac{u(x_j)}{u(x_1)} p(x_1, c(A)). \tag{4}$$

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

$$p(x_1, c(A)) = \frac{u(x_1)}{\sum_{j=1}^m u(x_j)}. \tag{5}$$

Therefore, by (4) and (5), $p(x_j, c(A)) = \frac{u(x_j)}{\sum_{j=1}^m u(x_j)}$ for all $j \in \{1, \ldots, m\}$. ■

Now we will finish the proof of the sufficiency. By Lemma 4 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 $\alpha = 1/2$. Then, for all i, j such that $i < j$, $u(x_i) > (1 + \alpha)u(x_j)$. Then, for all $A \subset X$, $c(A) = \{x\}$, where $x \geq y$ for all $y \in A$. Hence, this (u, α) 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 α. Notice that $\min_{j,l:x_j \gg x_{j+l}} d((x_i)_{i=j}^{j+l})$ and $\max_{j',l':x_{j'} \simeq x_{j'+l'}} d((x_i)_{i=j'}^{j'+l'})$ exist because X is finite.

Since $x_i \geq x_{i+1}$ for any i, $(x_i)_{i=j}^{j+l}$ and $(x_i)_{i=j'}^{j'+l'}$ satisfy the conditions for a path. Moreover, if $x_{j'} \simeq x_{j'+l'}$, then $x_i \simeq x_{i+1}$ for any $i \in \{j', \ldots, j' + l' - 1\}$. (To see this that $x_i \gg x_{i+1}$ for some $i \in \{j', \ldots, j' + l' - 1\}$. Then by Strong Dominance Transitivity, we must have $x_{j'} \gg x_{j'+l'}$, which is a contradiction. Therefore, $x_i \simeq x_{i+1}$ for any $i \in \{j', \ldots, j' + l' - 1\}$.)

Therefore, 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 α such that

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

To see that α 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

$$u(x_i) = \frac{p(x_i, x_i x_{i-1})}{p(x_{i-1}, x_i x_{i-1})} u(x_{i-1}). \tag{6}$$

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

$$u(x_{i-1}) > u(x_i)(1+\alpha). \tag{7}$$

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

To finish the proof, by Lemma 5, it suffices to show the following three steps.

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

Proof of Step 1: Assume that $y \simeq z$ and $y \geq z$ to show $\dfrac{u(y)}{u(z)} = \dfrac{p(y, yz)}{p(z, yz)}$. 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

$$\frac{p(x_n, x_n x_{n+1})}{p(x_{n+1}, x_n x_{n+1})} = \frac{u(x_n)}{u(x_{n+1})}. \tag{8}$$

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

$$\frac{p(x_n, \{x_s, \ldots, x_{s+t}\})}{p(x_{n+1}, \{x_s, \ldots, x_{s+t}\})} = \frac{p(x_n, x_n x_{n+1})}{p(x_{n+1}, x_n x_{n+1})}. \tag{9}$$

Hence by (8) and (9), it follows that

$$\begin{aligned}
\frac{u(y)}{u(z)} &= \frac{u(x_s)}{u(x_{s+1})} \cdots \frac{u(x_{s+t-1})}{u(x_{s+t})} && (\because y = x_s, z = x_{s+t}) \\
&= \frac{p(x_s, \{x_s, \ldots, x_{s+t}\})}{p(x_{s+1}, \{x_s, \ldots, x_{s+t}\})} \cdots \frac{p(x_{s+t-1}, \{x_s, \ldots, x_{s+t}\})}{p(x_{s+t}, \{x_s, \ldots, x_{s+t}\})} && (\because (8), (9)) \\
&= \frac{p(y, yz)}{p(z, yz)}. && (\because (9), y = x_s, z = x_{s+t})
\end{aligned}$$

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

Proof of Step 2: Suppose by way of contradiction that $u(z) > (1+\alpha)u(y)$ and $z \simeq y$. Then, by L-property, $d(z, y) = \dfrac{p(z, zy)}{p(y, zy)} = \dfrac{u(z)}{u(y)} > 1 + \alpha$. This contradicts the definition of α.

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

Proof of Step 3: Suppose that $z \gg y$. If $z = x_j$ then $y \neq x_{j+1}$ by the definition of u.[11] 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

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

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

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 α, since $z \gg y$,

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

References

Ahumada, A. and L. Ulku (2017): "Luce Rule with Limited Consideration," Forthcoming, Mathematical Social Sciences.

Aizerman, M. and A. Malishevski (1981): "General theory of best variants choice: Some aspects," *IEEE Transactions on Automatic Control*, 26, 1030–1040.

Block, H. and J. Marschak (1960): "Random orderings and stochastic theories of responses," in *Contributions to probability and statistics*, ed. by I. O. SG Ghurye W Hoeffding H Hotelling WG 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.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and Rustichini (2016): "Law of Demand and Forced Choice," Tech. rep.

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

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, A. Rustichini, et al. (2017): "Multinomial logit processes and preference discovery: inside and outside the black box," Mimeo, U. Bocconi.

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

Fishburn, P. C. (1973): "Binary choice probabilities: on the varieties of stochastic transitivity," *Journal of Mathematical psychology*, 10, 327–352.

Flores, A., G. Berbeglia, and P. Van Hentenryck (2017): "Assortment Optimization Under the General Luce Model," Tech. rep.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015): "Stochastic choice and revealed perturbed utility," *Econometrica*, 83, 2371–2409.

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.

Jamison, D. T. and L. J. Lau (1973): "Semiorders and the theory of choice," *Econometrica: Journal of the Econometric Society*, 901–912.

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.

Sen, A. K. (1971): "Choice functions and revealed preference," *The Review of Economic Studies*, 38, 307–317.