The Imbalanced Luce Model

Preliminary, comments and suggestions welcome

Matthew Kovach* Gerelt Tserenjigmid †

January 26, 2017

Click here for the most recent version.

Abstract

We develop a random choice model in which a decision maker divides the alternatives she faces into two groups, where one group is favored and thus she is more likely to choose alternatives in that group relative to alternatives in the non-favored group. We show that this seemingly specific decision making procedure in fact generalizes the Luce (1959) model, and nests versions of salience theory, reference-dependent models, and consideration set models. Moreover, the model can be characterized by two simple weakenings of *independence from irrelevant alternatives* (IIA).

Keywords: Luce's Model; Random Choice; Salience; Reference-Dependence; Consideration Set; IIA.

JEL Classification Numbers: D01, D81, D9.

1 Introduction

In many choice situations, a decision maker may attempt to simplify a decision problem by dividing the alternatives she faces into two groups: red wine and white wine, iPhone and android, healthy food and junk food, meat and fish, cheap clothes and expensive clothes, and so on. However, the way she divides alternatives may affect or bias her choice.

For example, consider a situation in which a customer in a restaurant wants to choose a bottle of wine. The customer notices that the menu has many red wines but only few white

*Department of Economics and CIE, ITAM. E-mail: matthew.kovach@itam.mx

†Department of Economics, Virginia Tech. E-mail: gerelt@vt.edu

wines. In this situation, the customer may decide to buy a red wine because she likes to have greater variety or flexibility. Alternatively, the customer may decide to buy a white wine because she believes it is easier to decide among fewer alternatives.

Consider another situation in which a customer in a restaurant wants to choose a meal. When the menu offers many different cuts of steak but only few types of fish, the consumer may be tempted to eat meat instead of fish. Alternatively, in the same setting the customer may focus on the price of the meal and thus divide the menu into dishes above or below a certain price.

In this paper, we develop a random choice model of a decision maker who divides the alternatives she faces into two groups, where one group is favored and thus she more likely to choose alternatives in that group relative to alternatives in the non-favored group. In particular, we introduce and characterize a generalization of Luce's (1959) model called the *Imbalanced Luce Model* (ILM). The ILM is a minimal departure from the Luce model that is nonetheless flexible enough to accommodate the attraction and compromise effects, both of which are well-known behavioral phenomena. Since the ILM allows for violations of *regularity* (e.g., the attraction effect), the ILM is not a special case of the random utility model.

In the ILM, the decision makes divides the menu A she faces into two submenus, $F(A)$ and $U(A) \equiv A \setminus F(A)$. The ratio of choice frequencies for alternatives *within a single submenu*, say $F(A)$, is equal to the ratio of their utilities just as in Luce's model. However, the ILM departs from Luce with respect to how the decision maker chooses between alternatives in different groups. That is, alternatives in $F(A)$ are chosen relatively more frequently than alternatives in $U(A)$ after taking account of their utilities.

Our main result is that the ILM model can be characterized by two simple weakenings of Luce's *independence from irrelevant alternatives* (IIA) axiom and that both the decision maker's utility index and categorization of alternatives can both be identified uniquely from their choice behavior. Our first axiom requires that there are a limited number of violations of IIA. Our second axiom is a transitivity property that applies to choice from binary menus.

In addition to our main result we also show that the ILM can be interpreted as a model of salience theory, reference-dependent models, and consideration set models. The first interpretation relies on the salience of different attributes of the alternatives. In this case, the decision maker focuses on (or chooses) a particular attribute of the alternatives which she finds the most salient. Then she divides the set of alternatives in such a way that alternatives in one group are more attractive than alternatives in the other group according to the salient attribute. For example, when the decision maker thinks the price dimension of alternatives is most salient she divides the set of alternatives into cheap alternatives and expensive alternatives.

The second interpretation of the ILM relies on reference-dependence. In this interpretation, the decision maker considers an endogenous reference alternative from the menu and divides the menu in the following way: alternatives that dominate the reference alternative and alternatives that do not dominate the reference alternative. After the split, she chooses the dominating alternatives more frequently than the non-dominating ones. In Section 6, we pursue the reference-dependence interpretation further and explicitly model the relationship between the reference alternative and the magnitude of the decision maker's bias towards the dominating group. In particular, we develop and characterize a special case of the ILM, called the *Reference-Dependent Luce Model* (RDLM), in which the magnitude of the bias is proportional to the utility of the reference alternative.

Lastly, when the magnitude of the bias is large enough, alternatives in the unfavored group will be chosen with probabilities near zero. Hence, the choice procedure approximates a decision maker that only considers a subset of the available alternatives — the favored group. In other words, the ILM can be interpreted as a stochastic version of limited consideration models such as Masatlioglu et al. (2012). Such a decision maker draws a random utility and then chooses rationally from a deterministic attention set.

The rest of the paper is organized as follows. In section 2, we introduce the ILM and discuss three main interpretations of the model. In section 3, we provide two simple axioms to characterize the ILM model. In section 4, we introduce and characterize a general version of the ILM in which zero probabilities are allowed. In section 5, we also introduce and characterize the RDLM. In section 6, we discuss related literature. The proofs are collected in the appendix.

2 Model

Let X be a set of finite alternatives and $\mathscr{X}$ be a collection of nonempty subsets of X (menus).

Definition 1. A function $p : X \times \mathscr{X} \to [0, 1]$ is called *random choice rule* if for any $A \in \mathscr{X}$,

$$\sum_{a \in A} p(a, A) = 1.$$

A random choice rule p is nondegenerate if $p(a, A) > 0$ for all $A \in \mathscr{X}$ and $a \in A$.

The first half of the paper focuses on nondegenerate random choice rules. For notational simplicity, let us write $r(a, b) \equiv \frac{p(a,ab)}{p(b,ab)}$ and $r_A(a, b) \equiv \frac{p(a,A)}{p(b,A)}$.

Definition 2 (IIA)**.** A random choice rule p satisfies Luce's *independence of irrelevant* *alternatives (IIA) axiom at* $a, b \in X$ if, for any A with $a, b \in A$,

$$r(a,b) = r_A(a,b).$$

Moreover, p satisfies IIA if for any $a, b \in X$, p satisfies IIA at a, b.

Luce (1959) proves that if a non-degenerate random choice rule satisfies IIA, then it can be represented by the Luce's model (also referred to as multinomial logit).

Definition 3 (Luce's Model)**.** A random choice rule p is a *Luce's Model* if there exists $u : X \to \mathbb{R}_{++}$ such that for any A and $x \in A$,

$$p(x,A) = \frac{u(x)}{\sum_{a \in A} u(a)}. \tag{1}$$

In this paper, we develop the following generalization of the Luce's model which may violate IIA. First, we need to define a filter function and a distortion function. A mapping $F : \mathscr{A} \to \mathscr{A}$ is a *filter function* if $F(A) \subseteq A$ for any A, and $F(\{a,b\}) = \{a,b\}$ for any $a, b \in X$. Let $U(A) \equiv A \setminus F(A)$. A *distortion function* is a mapping $\epsilon : \mathscr{X} \to \mathbb{R}_+$

Definition 4 (Imbalanced Luce Model)**.** A random choice rule p is an *Imbalanced Luce Model* (ILM) if there exist a utility function $u : X \to \mathbb{R}_{++}$, a filter function $F : \mathscr{A} \to \mathscr{A}$, and a distortion function $\epsilon : \mathscr{X} \to \mathbb{R}_+$, such that for any $A \in \mathscr{X}$ and for any $x \in A$,

$$p(x,A) = \frac{u(x)\big(1 + \mathbb{1}\{x \in F(A)\}\,\epsilon(A)\big)}{\big(1+\epsilon(A)\big)\big(\sum_{a \in F(A)} u(a)\big) + \sum_{b \in U(A)} u(b)}. \tag{2}$$

An ILM p has a *constant weight* if there is $\epsilon > 0$ such that for any $A \in \mathscr{X}$, $\epsilon(A) \in \{0, \epsilon\}$.

Note that we have $p(a,ab) = \frac{u(a)}{u(a)+u(b)}$ and $p(b,ab) = \frac{u(b)}{u(a)+u(b)}$ as in the Luce's model. In other words, there is no bias in menus with two alternatives. However, the ILM deviates from the Luce's model when there are at least three alternatives. The decision maker splits A into $F(A)$ and $U(A)$, and choice frequencies of alternatives in $F(A)$ relative to alternatives in $U(A)$ are distorted by $1 + \epsilon(A)$. Note that if for any A, $\epsilon(A) = 0$, then the ILM reduces to Luce's model.

As previously mentioned in the introduction, the ILM may be interpreted as a model of salience, reference-dependence, or limited consideration. Here we provide further analysis of the three interpretations.

- **Salience:** The decision maker chooses a particular attribute of the alternatives (e.g., price or quality) which she finds most salient. For example, suppose she finds the price dimension to be more salient (perhaps because the alternatives are very similar in quality but there is large variation in prices). Then the decision maker splits A into $F(A)$, the set of cheap alternatives, and $U(A)$, the set of expensive alternatives, and the utilities of alternatives of $F(A)$ will be overweighted by $1 + \epsilon(A)$.

- **Reference-Dependence:** The decision maker chooses an endogenous reference alternative $R(A)$ from the menu A, and divides A into $F(A) = \{a \in A | a \succ R(A)\}$, the set alternatives that dominate $R(A)$ by some linear order $\succ$, and $U(A)$, the set of alternatives that do not dominate $R(A)$. Then, the utilities of dominating alternatives in $F(A)$ will be overweighted by $1 + \epsilon(A)$. For example, $a \succ b$ when a is cheaper than b and a's quality is higher than b's quality.

- **Limited Consideration:** Suppose $\epsilon(A)$ is very large. Then $p(a, A) \approx \mathbb{1}\{a \in F(A)\}\, p(a, F(A))$. In other words, the decision maker focuses on the subset $F(A)$ of A and behaves as if she is a Luce decision maker choosing from $F(A)$.

In Section 5, we discuss a reference-dependent specification of the ILM in which the distortion $\epsilon(A)$ is proportional to the utility of the reference alternative $R(A)$.

2.1 Discussion of ILM

2.1.1 Behavioral Properties

In this subsection we discuss some of the behavioral implications of the model. For simplicity we will focus on the constant weight ILM and use the language of the salience interpretation. Consider a menu $\{x, y\}$. Then in our model, choice frequencies of x and y are

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

Suppose x is a cheap, low-quality good and y is an expensive, medium-quality good. Now consider the effect of adding the third alternative, z, which is a very expensive, high-quality good. Further, suppose the quality dimension is the most salient for the menu $\{x, y, z\}$, since the third alternative is high quality. In other words, $F(A) = \{y, z\}$ and $U(A) = \{x\}$. Then the choice frequencies of x, y are

$$p(x, xyz) = \frac{u(x)}{u(x) + (u(y) + u(z))(1 + \epsilon)} \text{ and } p(y, xyz) = \frac{u(y)(1 + \epsilon)}{u(x) + \big(u(y) + u(z)\big)(1 + \epsilon)}.$$

Therefore, adding the third alternative z increases the relative probability of choosing y over

x, that is,

$$r(y,x) = \frac{u(y)}{u(x)} < r_{xyz}(y,x) = \frac{u(y)(1+\epsilon)}{u(x)}.$$

The above pattern is called the *compromise effect*, and is a well-known behavioral phenomenon.[1]

The ILM is not a special case of the random utility model since it can violate the regularity property: $p(a,A) \geq p(a, A \cup b)$. To illustrate, let us show that $p(y,xy) < p(y,xyz)$ is possible. Note that $p(y,xy) < p(y,xyz)$ iff $\frac{1+\epsilon}{\epsilon}u(z) < u(x)$. In other words, if the utility of z is small enough, then we can have a violation of regularity.

On the other hand, when ϵ is small enough, regularity is satisfied. In fact, we can show that when ϵ is small enough, the ILM is a special case of the random utility model.

2.1.2 Salience Theory

In this section, we show that the ILM with a constant weight nests a version of salience theory of Bordalo et al. (2013). We assume that alternatives have n-attributes, i.e., $X \subset \mathbb{R}^n$. For a given menu A, one of n-attributes to be chosen the most salient attribute. Suppose i-th attribute is the most salient attribute for a menu $A = \{x^1, \ldots, x^m\}$ where $x^j = (x^j_1, \ldots, x^j_n)$. Then alternatives with a high i-th attribute (e.g., greater than $\bar{x}_i = \sum_{j=1}^m x^j_i/m$) will receive an extra $\delta > 0$ on their utility. In other words, the utility of x^j is

$$\sum_{k=1}^n \alpha_k\, x^j_k + \delta\, \mathbb{1}\{x^j_i \geq \bar{x}_i\} + \epsilon_j.$$

When ϵ_j follows the standard type 1 extreme value distribution (standard Gumbel distribution), the probability of choosing alternative x^j from A is

$$p(x^j, A) = \frac{\exp\left(\sum_{k=1}^n \alpha_k\, x^j_k + \delta\, \mathbb{1}\{x^j_i \geq \bar{x}_i\}\right)}{\sum_{j=1}^m \exp\left(\sum_{k=1}^n \alpha_k\, x^j_k + \delta\, \mathbb{1}\{x^j_i \geq \bar{x}_i\}\right)}.$$

Let $u(x_j) = \exp\left(\sum_{k=1}^n \alpha_k\, x^j_k\right)$, $F(A) = \{x^j : x^j_i \geq \bar{x}_i\}$, and $U(A) = \{x^j : x^j_i < \bar{x}_i\}$. Let $\exp(\delta) = 1+\epsilon$. Then we obtain an ILM with a constant weight:

$$p(x^j, A) = \frac{u(x_j)\,(1+\epsilon\, \mathbb{1}\{x^j \in F(A)\})}{\sum_{j=1}^m u(x_j)(1+\epsilon\, \mathbb{1}\{x^j \in F(A)\})}.$$

[1]The compromise effect and the attraction effect are first documented in the experimental studies of Simonson (1989) and Huber et al. (1982), respectively, and confirmed by many studies in consumer choice (e.g., Simonson and Tversky 1992, Tversky and Simonson 1993, Ariely and Wallsten 1995, Herne 1998, Doyle et al. 1999, Chernev 2004, and Sharpe et al. 2008). These effects are also demonstrated in the contexts of choice over risky alternatives (Herne 1999), choice over policy issues (Herne 1997), and choice over political candidates (Sue O'Curry and Pitts 1995), among others.

2.1.3 Approximating Consideration Sets

We now discuss more formally how the ILM nests models of limited consideration. Masatlioglu et al. (2012) consider a decision maker that may not consider all available alternatives. In particular, their decision maker is characterized by a preference $\succ$ over the set of alternatives X, and an *attention filter*, $\Gamma : \mathscr{X} \to \mathscr{X}$, such that $\Gamma(A) \subseteq A$ and $\Gamma(A) = \Gamma(A \setminus x)$ when $x \notin \Gamma(A)$. When asked to make a choice from menu A the decision maker will select the alternative that is $\succ$-best in $\Gamma(A)$. A choice function, c, generated by some pair $(\succ, \Gamma)$ is called a *choice with limited attention* (CLA). Our decision maker can be thought of in a similar manner, except that her choice from A will correspond to maximizing a stochastic preference over $\Gamma(A)$.

Formally, given an attention filter Γ, suppose that $F(A) = \Gamma(A)$ and let u be the decision maker's utility function. When $\epsilon(A)$ is very large, then

$$p(a, A) \approx \mathbb{1}\{a \in F(A)\} \frac{u(a)}{\sum_{b \in F(A)} u(b)} = \mathbb{1}\{a \in \Gamma(A)\} \frac{u(a)}{\sum_{b \in \Gamma(A)} u(b)}.$$

That is, an ILM decision maker looks like someone choosing according to a Luce rule from $\Gamma(A)$. The requirement that attention filters satisfy $\Gamma(A) = \Gamma(A \setminus x)$ when $x \notin \Gamma(A)$, while potentially appealing, is not a feature of the ILM but rather a feature of the CLA model. The ILM is therefore capable of nesting a larger variety of attention models, such as from Lleras et al. (2016), by dispensing with this requirement. Nevertheless, in Section 3.1, we characterize the case where F is an attention filter. Moreover, in Section 5, we discuss a reference-dependence specification of the ILM and we show that an axiom called *Reference-Consistency* implies that F is an attention filter.

While the ILM model naturally captures a decision maker with random utility and deterministic consideration sets, there is another way to relate the ILM to models of choice with limited attention. Consider a particular special case of the CLA model, where the attention filter Γ is *attentive in small sets*: $\Gamma(\{a, b\}) = \{a, b\}$ for all binary menus. Then, by appropriately choosing sequences, u_k and ϵ_k, we can in fact show that any such choice function c is the limit of some ILM model. Thus our decision maker can in fact approximate the deterministic model in Masatlioglu et al. (2012) whenever Γ is attentive in small sets.

Lemma 1. *Let choice function c be a choice with limited attention, represented by $(\succ, \Gamma)$, where Γ is attentive in small sets. Then c is the limit of some ILM model.*

3 Characterizing Imbalanced Luce Models

In this section, we characterize the ILM using two very simple weakenings of IIA. The first axiom requires that we cannot have too many violations of IIA. In particular, it says that if IIA is violated at a and b, and is also violated at b and c, then we cannot have a violation of IIA at a and c.

AXIOM 1 (Bounds on Violations of IIA). For any A and $a, b, c \in A$,

$$\text{if } r_A(a,b) \neq r(a,b) \text{ and } r_A(b,c) \neq r(b,c), \text{ then } r_A(a,c) = r(a,c).$$

The second axiom imposes a transitivity property. It says that the impact of b on a (i.e., $r(a,b)$) times the impact of a on c (i.e., $r(c,a)$) is equal to the impact of b on c (i.e., $r(c,b)$).

AXIOM 2 (Transitive Impact). For any $a, b, c \in X$,

$$r(a,b) \cdot r(b,c) \cdot r(c,a) = 1.$$

Transitive Impact is a necessary and sufficient condition for having the Luce's model on binary menus (e.g., see Luce 1959). It turns out that, Transitive Impact implies *strong stochastic transitivity.*[2] We now state the first characterization theorem.

Theorem 1. *A nondegenerate random choice rule p satisfies Bounds on Violations of IIA and Transitive Impact if and only if p is an Imbalanced Luce Model. Moreover, u is unique up to a linear transformation, and ϵ is unique.*

We also characterize the Constant Weight ILM with an additional axiom. The axiom requires that if IIA is violated at both a, b and a', b', then the magnitude of violations must be the same.

AXIOM 3 (Constant Impact). For any $a, b \in A$, $a', b' \in A'$, if

$$r_A(a,b) > r(a,b) \text{ and } r_{A'}(a',b') > r(a',b'), \text{ then } \frac{r_A(a,b)}{r(a,b)} = \frac{r_{A'}(a',b')}{r(a',b')}.$$

Theorem 2. *A nondegenerate random choice rule p satisfies Bounds on Violations of IIA, Transitive Impact and Constant Impact if and only if p is an ILM with a constant weight. Moreover, u is unique up to a linear transformation, and ϵ is unique.*

[2]Strong stochastic transitivity: for any $a, b, c \in X$, if $p(a, ab) \geq \frac{1}{2}$ and $p(b, bc) \geq \frac{1}{2}$, then $p(a, ac) \geq \max\{p(a, ab), p(b, bc)\}$.

3.1 Modeling $F(A)$

In this section we discuss two different approaches to modeling the favored set, $F(A)$. The first approach is in line with the interpretation of the ILM model as a stochastic model of limited consideration. We provide a simple characterization for when $F(A)$ is an attention filter as in Masatlioglu et al. (2012). A filter function F is an *attention filter* if for any $A \in \mathscr{X}$ and $x \notin F(A)$, $F(A) = F(A \setminus \{x\})$.

Axiom 4 (Independence from Unfavored Alternatives). For any $A \in \mathscr{X}$, and $a, b \in A$, if $r_A(a,b) > r(a,b)$, then for any $a', b' \in A \setminus \{b\}$, $r_A(a',b') = r_{A\setminus\{b\}}(a',b')$.

Observation 1: If an ILM p with (u, F, ϵ) satisfies axiom 4, then F is an attention filter, and $\epsilon(A) = \epsilon(A \setminus \{x\})$ whenever $F(A) \subset A \setminus \{x\}$.

Next, we discuss a special case of the ILM in which $F(A)$ is decided according to disappointment aversion rule as in Gul (1991). A filter function F is a *disappointment filter* if for any $A \in \mathscr{X}$, $F(A) = \{a \in A | u(a) \geq \sum_{x\in A} p(x, A)u(x)\}$. In other words, $F(A)$ is the set of alternatives that have higher utilities than the average utility, so $U(A)$ is the set of alternatives that have lower utilities than the average utility. The following axiom characterizes disappointment filters.

Axiom 5. For any $A \in \mathscr{X}$, and $a, b \in X$, $r_A(a,b) > r(a,b)$ iff $1 \geq \sum_{x\in A} p(x, A)\, r(x,a)$.

Observation 2: If an ILM p with (u, F, ϵ) satisfies axiom 5, then F is a disappointment filter.

4 General Imbalanced Luce Model: Allowing Zero Probabilities

An important feature of the Luce's model and the random utility model is that each alternative is chosen with strictly positive probabilities. However, in many economic choice situations, the menu includes alternatives which are obviously worse than (dominated by) some other alternatives in the menu. One such situation is a well-known the *attraction effect*. In the attraction effect, the probability of choosing y from $\{x, y\}$ increases when a third alternative d_y is added where d_y is dominated by y in every aspect while d_y is not chosen. Doyle et al. (1999) is a representative experiment with evidence in favor the attraction effect we described. The authors present customers with a choice of baked beans. The first choice is between two types of baked beans: x and y; x is Heinz bake beans, while y is a local cheap brand called Spar. In the experiment, y was chosen 19% of the time. The authors then introduced a third option, d_y, a reduced amount of the brand Spar with the price of y. The third option d_y is obviously dominated by y. After d_y is introduced, the probability of choosing y increased to 33% of the time, but nobody had bought d_y.

We now define a general version of the ILM in which dominated alternatives (e.g., d_y) are chosen with zero probabilities. In Section 4.1, we explain in a detail how a general version of the ILM is consistent with the attraction effect.

Consider a binary relation $\succ$ on X that is irreflexive, transitive, and antisymmetric, but possibly incomplete. The binary relation $\succ$ captures the dominance relation; that is, $a \succ b$ means that a dominates b. For any $A \in \mathscr{X}$ and $a \in A$, let $A_a \equiv \{b \in A | b \succ a\}$. Also, let $A^{UN} = \{a \in A | \nexists a' \in A \text{ s.t } a' \succ a\}$. For a given $\succ$, a mapping $F^{\succ} : \mathscr{A} \to \mathscr{A}$ is a *general filter function* if $F^{\succ}(A) \subseteq A^{UN}$ for any A, and $F^{\succ}(A) = A$ when $|A| = |A^{UN}| = 2$. Let $U^{\succ}(A) \equiv A^{UN} \setminus F^{\succ}(A)$.

Definition 5 (General Imbalanced Luce Model)**.** A random choice rule p is a *General Imbalanced Luce Model* (GILM) if there exist $u : X \to \mathbb{R}_{++}$, a binary relation $\succ$, a general filter function $F^{\succ} : \mathscr{A} \to \mathscr{A}$, and a distortion function $\epsilon : \mathscr{X} \to \mathbb{R}_{+}$, such that for any A and for any $x \in A^{UN}$,

$$(3) \qquad p(x, A) = \frac{u(x)\big(1 + \mathbb{1}\{x \in F(A)\}\,\epsilon(A)\big)}{\big(1 + \epsilon(A)\big)\big(\sum_{a \in F^{\succ}(A)} u(a)\big) + \sum_{b \in U^{\succ}(A)} u(b)}.$$

A GILM p has a *constant weight* if there is $\epsilon > 0$ such that for any $A \in \mathscr{X}$, $\epsilon(A) \in \{0, \epsilon\}$.

Note that when $A = A^{UN} = \{a, b\}$, we have $p(a, ab) = \frac{u(a)}{u(a)+u(b)}$ and $p(b, ab) = \frac{u(b)}{u(a)+u(b)}$ as in the Luce's model. In other words, there is no bias in menus with two alternatives. However, the GILM deviates from the Luce's model in the following two ways. First, in the GILM, the decision maker only chooses from the undominated set A^{UN}. Second, the decision makes splits A^{UN} into $F^{\succ}(A)$ and $U^{\succ}(A)$, and choice frequencies of alternatives in $F^{\succ}(A)$ are distorted by $1 + \epsilon(A)$. Note that alternatives in $A \setminus A^{UN}$ are not chosen; i.e., $p(a, A) = 0$ when $a \in A \setminus A^{UN}$. However, these alternatives may affect choice frequencies of alternatives in A^{UN} through the menu-dependent distortion $1 + \epsilon(A)$.

Note that if for any A, $\epsilon(A) = 0$, and for any $a, b \in X$, neither $a \succ b$ nor $b \succ a$, then the GILM reduces to the Luce's model. Moreover, if either $a \succ b$ or $b \succ a$ for any $a, b \in X$, then the GILM is a deterministic choice function.

As previously discussed the ILM may be interpreted as a model of salience, reference-dependence, or consideration set. The GILM particularly fits well with the reference-dependence interpretation. In the next section, we discuss a reference-dependent model in which $F(A)^{\succ} = A^{UN}_{R(A)}$ where $R(A)$ is a reference alternative, and $\epsilon(A)$ is proportional to $u(R(A))$.

Echenique and Saito (2015) generalizes the Luce model to allow for zero probabilities. The GILM nests the model of Echenique and Saito (2015) since the GILM allows for zero probabilities as well as violations of IIA.

4.1 Discussion of GILM

In this subsection, we discuss a behavioral implication of the GILM. Consider a menu $\{x, y\}$. Then in our model, choice frequencies of x and y are

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

Suppose x is a cheap, low-quality good and y is an expensive, medium-quality good. Now let us discuss the effect of adding the third alternative d_y, which is a very expensive, medium-quality good. In other words, d_y is dominated by y, but not by x. Here d_y is a natural reference alternative. So let $F^{\succ}(A) = \{y\}$ when $A = \{x, y, d_y\}$. Moreover, let $\epsilon(xyd_y) = \frac{u(d_y)}{u(y)}$ (see Section 5). Then in the new menu the choice frequencies of x and y are

$$p(x, xyd_y) = \frac{u(x)}{u(x) + u(y) + u(d_y)} \text{ and } p(y, xyd_y) = \frac{u(y) + u(d_y)}{u(x) + u(y) + u(d_y)}.$$

Therefore, adding the third alternative d_y increases the probability of choosing y, that is,

$$p(y, xy) = \frac{u(y)}{u(x) + u(y)} < p(y, xyd_y) = \frac{u(y) + u(d_y)}{u(x) + u(y) + u(d_y)}.$$

The above pattern is consistent with the attraction effect.

4.2 Characterizing General Imbalanced Luce Models

In this section, we characterize the GILM using four very simple axioms. The two main axioms will be slight modifications of Bounds on Violations of IIA (A.1) and Transitive Impact (A.2). In particular, we assume that Bounds on Violations of IIA and Transitive Impact must be satisfied whenever possible (i.e., when there are non zero probabilities). The other two axioms deal with zero probabilities. As noted in the previous section, the first deviation from the Luce's model is that the decision maker only chooses from the undominated set A^{UN}. Therefore, we want to identify the undominated set from choice data. The following notation is essential for the identification of the undominated set.

Definition 6 (Dominance Relation). For any $a, b \in X$,

1. $a > b$ if $p(a, ab) = 1$,
2. $a \simeq b$ if $p(a, ab) \in (0, 1)$.

Note that $a > b$ means that b is revealed dominated by a, and $a \simeq b$ means that b is revealed not dominated by a. For each A, let $A^{>} \equiv \{a \in A | \nexists a' \text{ s.t } a' > a\}$ and $A_a = \{a' \in A : a' > a\}$. It turns out, $A^{>}$ will be our undominated set. The first axiom requires that alternatives in the dominated set will not be chosen.

Axiom 6 (Dominance Consistency). For any A and $a \in A$, $p(a, A) = 0$ iff $a \notin A^{>}$.

The next axiom is a modification of Bounds on Violations of IIA (A.1). Bounds on Violations of IIA requires that for any alternatives a, b, c, IIA is satisfied at one of three pairs from a, b, c. General Bounds on Violations of IIA requires that IIA is satisfied at one of three pairs from a, b, c only when none of a, b, c dominates other alternatives.

Axiom 7 (General Bounds on Violations of IIA). For any A and $a, b, c \in A^{>}$,

$$\text{if } r_A(a, b) \neq r(a, b) \text{ and } r_A(b, c) \neq r(b, c), \text{ then } r_A(a, c) = r(a, c).$$

For the sake of technical simplicity, we assume that there is an undominated alternative a^* that does not dominate any alternative in $X \setminus \{a^*\}$. Formally,

Axiom 8 (Bounds on Dominance). There exists a^* such that for any $a \in X$, $a^* \simeq a$.

This condition is satisfied in many natural settings. For example, when comparing alternatives that differ in price and quality, we might say that alternative a dominates b if it is both higher quality and a lower price. If there is a single good that is of the highest quality and the greatest price, then that object will not be dominated nor will it dominate any other alternative.

The last axiom is a modification of Transitive Impact. General Transitive Impact requires Transitive Impact only for alternatives a, b, c such that none of a, b, c dominates other alternatives.

Axiom 9 (General Transitive Impact). For any $a, b, c \in X$, if $a \simeq b$, $b \simeq c$, $c \simeq a$, then

$$r(a, b) \cdot r(b, c) \cdot r(c, a) = 1.$$

We now state the characterization theorem.

Theorem 3. *A random choice rule p satisfies Dominance Consistency, General Bounds on Violations of IIA, Bounds on Dominance, and General Transitive Impact if and only if p is a GILM.*

Similar to Theorem 2, the GILM with a constant weight is also characterized by Constant Impact.

Theorem 4. *A GILM p satisfies Constant Impact if and only it p is a GILM with a constant weight.*

5 Reference-Dependent Luce Model

In this section, we pursue the reference-dependence interpretation further and explicitly model how the reference alternative affects choice frequencies. To do so, we introduce a special case of the GILM, called the *Reference-Dependent Luce Model* (RDLM), in which the distortion $\epsilon(A)$ is proportional to the utility of the reference alternative as in Section 4.1. First, let us consider the following choice procedure using the idea of reference-dependence.

Choice Procedure: Take any menu A.

i) The menu A induces some reference alternative, $R(A)$.

ii) The agent selects alternative $a \in A^{UN}$ from $A^{UN} \cup \{R(A)\}$ with Luce probability

$$\frac{u(a)}{\sum_{b \in A^{UN} \cup \{R(A)\}} u(b)}.$$

iii) With probability

$$\frac{u(R(A))}{\sum_{b \in A^{UN} \cup \{R(A)\}} u(b)},$$

she fails to select something in the first round, and therefore she focuses her attention on objects that dominate the reference $R(A)$, given by $A^{UN}_{R(A)} \equiv A_{R(A)} \cap A^{UN}$.

This choice procedure can be formalized by the following model.

Definition 7 (Reference-Dependent Luce Model)**.** A random choice rule $p\colon X \times \mathscr{X} \to [0, 1]$ is a *Reference-Dependent Luce Model* (RDLM) if there exists a binary relation $\succ$ on X that is irreflexive, transitive, and antisymmetric, a mapping $R\colon \mathscr{X} \to X$ with $R(A) \in A$ for any

A, and a function $u : X \to \mathbb{R}_{++}$ such that for any $A \in \mathscr{A}$ and $a \in A$, i) when $a \notin A^{UN}$, $p(a, A) = 0$; and ii) when $a \in A^{UN}$,

$$p(a, A) = \mu\big(a, A^{UN} \cup R(A)\big) + \mu\big(R(A), A^{UN} \cup R(A)\big)\, \mu\big(a, A^{UN}_{R(A)}\big)$$

where $\mu(a, B) = \mathbb{1}\{a \in B\}\, \frac{u(a)}{\sum_{b \in B} u(b)}$.

The second and third steps of the procedure is equivalent to saying the agent divides A^{UN} into $A^{UN}_{R(A)}$ and $A^{UN} \setminus A^{UN}_{R(A)}$, and the utilities of alternatives in $A^{UN}_{R(A)}$ are biased by $u(R(A))$. Specifically, the RDLM is a special case of the GILM when $F^{\succ}(A) = A^{UN}_{R(A)}$ and the distortion $\epsilon(A)$ is expressed as

$$\epsilon(A) = \frac{u(R(A))}{\sum_{a \in F(A)} u(a)}.$$

5.1 Characterizing Reference-Dependent Luce Model

We characterize the RDLM with additional three axioms. An important property of the model is that choice frequencies can be affected by dominated alternatives. The first axiom specifies a relation between the magnitude of the bias $r_A(a, b) - r(a, b)$ and a dominated alternative c that affects choice frequencies in A. Let $A^{>}_{c} \equiv A^{>} \cap A_c$.

Axiom 10 (Proportional Impact). For any $a, b \in A^{>}$ and $c \in A \setminus A^{>}$ such that $p(a, A) \neq p(a, A \setminus c)$,

$$\text{if } r_A(a, b) \geq r(a, b), \text{ then } r_A(a, b) = r(a, b) + r(b, c)\, p(a, A^{>}_{c}).$$

Note that Proportional Impact implies that if a is favored over b in A, i.e., $r_A(a, b) > r(a, b)$, then a dominates c, but b does not dominates c, i.e., $a \in A^{>}_{c}$ and $b \notin A^{>}_{c}$.

The second axiom considers the case in which dominated alternatives do not affect choice frequencies. In this case, IIA should be satisfied.

Axiom 11 (Undominated IIA). For any A, if for any $c \in A \setminus A^{>}$ and $x \in A$, $p(x, A \setminus c) = p(x, A)$, then for any $a, b \in A^{>}$,

$$r_A(a, b) = r(a, b).$$

This axiom also implies that if $A = A^{>}$, then $r_A(a, b) = r(a, b)$ for each $a, b \in A$.

The last axiom requires that if an alternative b in A does not affect the probability of choosing a from A, then choice frequencies in A are not affected by b.

Axiom 12 (Weak Reference-Consistency). For any A and $a, b \in A$, if $p(a, A \setminus b) = p(a, A) > 0$, then for any $a' \in A$, $p(a', A \setminus b) = p(a', A)$.

Theorem 5. *Suppose p is Generalized ILM. Then p satisfies Proportional Impact, Undominated IIA, and Weak Reference-Consistency if and only if p is a Reference-Dependent Luce Model.*

If we strengthen Weak Reference-Consistency, then we obtain some consistency between choices of reference alternatives from different menus.

Axiom 13 (Reference-Consistency). For any A and $a, b \in A$, if $p(a, A \setminus b) = p(a, A) > 0$, then if $p(a', A' \setminus b) = p(a', A')$ for any $A \subseteq A'$ and $a' \in A$.

Reference-Consistency implies that if $R(A)$ is a reference alternative of A, then $R(A)$ is a reference alternative of any subset $B \subseteq A$ such that $R(A) \in B$. Formally,

Corollary 1. If a RDLM $(u, r, \succ)$ satisfies Reference-Consistency, then for any A and $b \neq R(A)$, $R(A) = R(A \setminus b)$.

This property on R is also obtained in a deterministic setting in Ok et al. (2014) (see related literature for a more detailed discussion). Moreover, as mentioned in Section 2.1.3 this property on R implies that a filter function F is an attention filter. In particular, a filter function $F(A) = \{a \in A^{UN} : a \succ R(A)\}$ is an attention filter.

6 Related Literature

There is a non-axiomatic literature that proposes several models which can explain well-known violations of IIA. Rieskamp et al. (2006) is an excellent survey. Examples are Tversky (1972), Roe et al. (2001), Usher and McClelland (2004), and Natenzon (2010). Similar to our model, the latter three papers allow for violations of Luce's regularity axiom. We shall not discuss these papers here since the focus of our paper is to relate the procedure of ILM to models of salience theory, reference-dependence, and consider sets. We focus instead on the more narrowly related axiomatic literature that directly discusses either Luce's model, salience theory, reference-dependence, or consider set models. Thus, we separate literature into four categories.

1. Random Utility Models: The benchmark economic model of rational behavior for stochastic choice is the random utility model. The random utility model is described by a probability measure over preferences over X; $p(a, A)$ is the probability of drawing a preference that ranks a above any other alternative in A. The random utility model is famously difficult to characterize behaviorally: see the papers by Falmagne (1978), McFadden and Richter (1990), and Barbera and Pattanaik (1986).

Since Luce's model is a special case of both the ILM and random utility, the ILM and random utility intersect. Random utility models always satisfy regularity while the ILM allows for a violation of regularity (e.g., the attraction effect). Therefore, the ILM is not a special case of the random utility model. Moreover, when the magnitude of the bias is small enough, the ILM is a special case of the random utility model.[3]

Moreover, the decision making procedure of the ILM will be similar to that of the random choice model. In the random utility model, a utility of a given alternative x is $u(x)+\epsilon_x$ where $u(x)$ is the deterministic part and ϵ_x is the random part, and a decision maker chooses the alternative with the highest utility. The Luce's model is a special of the random utility in which ϵ_x follows the type-I extreme value distribution. Our interpretation of ILM is similar to the random utility model. Our main point of departure is that we allow the decision maker's utility over alternatives to exhibit a small degree of menu dependence. In particular, in the ILM, the utility of a given alternative x in menu A is $v(x, A) = u(x) + \epsilon(A)\,\mathbb{1}\{x \in F(A)\} + \epsilon_x$ where the additional term $\epsilon(A)\,\mathbb{1}\{x \in F(A)\}$ is the bias towards alternatives in $F(A)$, and a decision maker chooses the alternative with the highest utility.

The recent paper by Gul et al. (2014) presents a random utility model, called the *attribute rule model*, in which object attributes play a key role as in Tversky (1972). Their model has Luce's form, but it applies sequentially, and in terms of its empirical motivation, it seeks to address the similarity effect of Tversky (1972). The attribute rule model and the ILM have an intersection that is not Luce's model. In this intersection, for any a, b, $p(a, ab) \in \{\frac{1}{2}, \delta, 1-\delta\}$ for a some fixed $\delta > 0$, which violates the Strong Richness condition Gul et al. (2014) use to identify their model. See Appendix B for more details.

<u>2. Limited Consideration Models:</u> Our model can be interpreted as a stochastic version of consideration set models such as Masatlioglu et al. (2012). The main difference between our model and other stochastic models of consideration sets, such as Manzini and Mariotti (2014), is the source of randomness. In our model, the source of randomness is utility shocks, as in random utility models, and the consideration set for a given menu is fixed. In Manzini and Mariotti (2014) the source of randomness is random attention, and thus the consideration set is random, while tastes are fixed. Brady and Rehbeck (2016), Aguiar (2016), and Zhang (2016) provide generalizations of Manzini and Mariotti (2014) in which the probability that a given alternative is in consideration set is menu-dependent.

Echenique et al. (2013) provides a generalization of Luce's model in which a decision maker looks at alternatives according to an exogenous order, and chooses each alternative with the Luce probability. Therefore, similar to Manzini and Mariotti (2014), the decision

[3] Recall that when the magnitude of the bias is large enough the ILM can be seen as a stochastic version of limited consideration models.

maker may consider a subset of the original menu, but the subset is random.

3. Generalizations of Luce's Model: Echenique and Saito (2015) generalizes the Luce model to allow for zero probabilities. The GILM nests the model of Echenique and Saito (2015) since the GILM allows for zero probabilities as well as violations of IIA. Tserenjigmid (2014b) provides a generalization of the Luce model, called the *Order-Dependent Luce Model* (ODLM), in which the ordering of alternatives affects (e.g., the location of products in a grocery store, the order of candidates on a ballot) on choice frequencies. The ODLM and the ILM do not nest each other, but they have an nontrivial intersection. In particular, consider a special case of the ILM in which the set all alternatives X is ranked by a linear order, and $F(A)$ is the set of k-highest ranked of A where $k \geq 2$ is a given natural number. It turns out, this special of the ILM is nested by the ODLM.[4]

Fudenberg et al. (2015) considers a decision maker who chooses a probability distribution over alternatives so as to maximize expected utility, with a cost function that ensures that probabilities are non-degenerate.

4. Reference-Dependent Models: Tversky and Kahneman (1991) provided the first explicit model of reference-dependence preferences. They focused on choice over two-attribute alternatives. The model of Tversky and Kahneman (1991) is further developed by Munro and Sugden (2003) (to the n-dimensional commodity space), Sugden (2003) (subjective expected utility), and Koszegi and Rabin (2006) (expectation-based reference points). These models focus on deterministic choice, and reference points are usually exogenously given. Our paper contributes to literature that endogenize reference points (e.g., Ok et al. 2014 and Tserenjigmid 2014a).

In fact, Ok et al. (2014) is closely related to the RDLM. In their model, for a given menu A, the decision maker chooses a reference alternative $R(A)$ and chooses the most preferred alternative from the subset of alternatives that dominates $R(A)$. Therefore, the RDLM can be considered as an stochastic version of their model. In fact, two of main axioms of Ok et al. (2014) called *No-Cycle* and *Reference Consistency* are closely related to two of our axioms, Transitive Impact and Reference-Consistency. No-Cycle rules out violations of transitivity while Reference Consistency makes sure that the reference alternative of A is also the reference alternative for a subset $B \subset A$. Therefore, our axioms, Transitive Impact and Reference-Consistency, can be considered as stochastic counterparts of No-Cycle and Reference Consistency of Ok et al. (2014).

A final related paper is Kovach (2016), which studies stochastic choice with an exogenous status quo. In this paper randomness in choice is due to stochastic consideration, á la Manzini

[4]Specially, the following model is the intersection of the ODLM and the ILM: for any A and $x \in A$, $p(x, A) = \frac{u(x)(1+\epsilon\,\mathbb{1}\{r(x,A)\leq k\})}{\sum_{a\in A} u(a)(1+\epsilon\,\mathbb{1}\{r(a,A)\leq k\})}$ where $r(x, A)$ is the ranking of x in A by some fixed linear order.

and Mariotti (2014), and the attention probabilities may depend on the status quo. While this paper also allows for zero choice probabilities, the assumption of an exogenous status quo makes the model setup substantially different from the setting of the ILM.

A Proofs

A.1 Proof of Lemma 1

Proof. Let c be a CLA, given by $(\succ, \Gamma)$, such that $\Gamma(A) = A$ whenever $|A| = 2$. Suppose that $v : X \to \{1, \ldots, |X|\}$ is a representation of $\succ$. Then for $k \in \mathbb{N}$, define $u_k : X \to \mathbb{R}$ by $u_k(a) = 2^{kv(a)}$, let $F(A) = \Gamma(A)$ for every A, and $\epsilon_k(A) = k^k$ for $|A| > 2$. Then let p_k denote the ILM probabilities from u_k and ϵ_k. Then we must show that

$$\lim_{k\to\infty} p_k(a, A) = \begin{cases} 1 & \text{if } a = c(A) \\ 0 & \text{otherwise} \end{cases}$$

It should be clear that $p_k(a, A) \to 0$ if $a \notin \Gamma(A)$. Suppose then that $a = c(A)$ (is $\succ$-best in $\Gamma(A)$). Then

$$p_k(a, A) = \frac{2^{kv(a)}(1 + k^k)}{(1 + k^k)\sum_{b\in\Gamma(A)} 2^{kv(b)} + \sum_{b\in A\setminus\Gamma(A)} 2^{kv(b)}} = \frac{2^{kv(a)}}{\sum_{b\in\Gamma(A)} 2^{kv(b)} + \frac{\sum_{b\in A\setminus\Gamma(A)} 2^{kv(b)}}{1+k^k}}.$$

For $k > |X|$, then

$$\frac{\sum_{b\in A\setminus\Gamma(A)} 2^{kv(b)}}{1 + k^k} < \frac{|X|2^{|X|}2^k}{1 + k^k} \to 0$$

as $k \to \infty$, hence

$$\lim_{k\to\infty} p_k(a, A) = \frac{2^{kv(a)}}{\sum_{b\in\Gamma(A)} 2^{kv(b)}} = \frac{1}{1 + \frac{\sum_{b\in\Gamma(A)\setminus a} 2^{kv(b)}}{2^{kv(a)}}}.$$

Setting $m = \max_{b\in\Gamma(A)\setminus a} v(b) < v(a)$, we can observe that

$$\frac{\sum_{b\in\Gamma(A)\setminus a} 2^{kv(b)}}{2^{kv(a)}} \le \frac{m2^{km}}{2^{kv(a)}} = \frac{m}{2^{k(v(a)-m)}} \to 0,$$

and hence $p_k(a, A) \to 1$. □

A.2 Proof of Theorem 1

We prove Theorem 1 by three steps.

Step 1: Let us fix a some alternative a^*. We then construct $u : X \to \mathbb{R}_{++}$ in the following way:

$$u(a) \equiv r(a, a^*) \text{ for any } a \in X.$$

Now note that for any $a, b \in X \setminus a^*$, $r(a, b) = \frac{u(a)}{u(b)}$ since

$$1 = r(a, b) \cdot r(b, a^*) \cdot r(a^*, a) = r(a, b) \cdot u(b) \cdot \frac{1}{u(a)}$$

by Transitive Impact. Since $\frac{p(a,ab)}{p(b,ab)} = \frac{u(a)}{u(b)}$ and $p(a, ab) + p(b, ab) = 1$, we have $p(a, ab) = \frac{u(a)}{u(a)+u(b)}$.

Step 2: Now take any $|A| \geq 3$.

Fact 1: For any $a, a', b \in A$, if $r_A(a', a) = r(a', a)$ and $r_A(a, b) > r(a, b)$, then $r_A(a', b) > r(a', b)$.

Proof of Fact 1: Suppose $r_A(a', a) = r(a', a)$ and $r_A(a, b) > r(a, b)$. Then we have

$$r_A(a', b) = r_A(a', a) \cdot r_A(a, b) > r(a', a) \cdot r(a, b) = \frac{u(a')}{u(a)} \cdot \frac{u(a)}{u(b)} = \frac{u(a')}{u(b)} = r(a', b).$$

By Bounds on Violations of IIA and Fact 1, we can have $A = F(A) \cup U(A)$ such that for any $a, a' \in F(A)$ and $b, b' \in U(A)$,

$$r_A(a, a') = r(a, a'),\ r_A(b, b') = r(b, b'), \text{ and } r_A(a, b) > r(a, b).$$

If $UF(A) = \emptyset$, since $r_A(a, a') = r(a, a') = \frac{u(a)}{u(a')}$, we have $p(a, A) = \frac{u(a)}{\sum_{a' \in A} u(a')}$.

Step 3: Now we assume that $|F(A)|, |U(A)| \geq 1$. Let $F(A) = \{a_1, \dots, a_n\}$ and $U(A) = \{b_1, \dots, b_m\}$.

By the construction of u, $r_A(a_i, a_1) = r(a_i, a_1) = \frac{u(a_i)}{u(a_1)}$. Therefore, $p(a_i, A) = \frac{u(a_i)}{u(a_1)} \cdot p(a_1, A)$. Similarly, $p(b_j, A) = \frac{u(b_j)}{u(b_1)} \cdot p(b_1, A)$. Then we have

$$(4) \qquad 1 = \sum_{i=1}^{n} p(a_i, A) + \sum_{j=1}^{m} p(b_j, A) = \frac{\sum_{i=1}^{n} u(a_i)}{u(a_1)} \cdot p(a_1, A) + \frac{\sum_{j=1}^{m} u(b_j)}{u(b_1)} \cdot p(b_1, A).$$

Note that for any i, j,

$$\frac{r_A(a_1, b_1)}{r(a_1, b_1)} = \frac{r_A(a_i, b_j)}{r(a_i, b_j)}.$$

Let

$$1 + \epsilon(A) \equiv \frac{r_A(a_1, b_1)}{r(a_1, b_1)}. \tag{5}$$

Therefore, we have

$$\begin{aligned} p(b_1, A) &= \frac{p(b_1, A)}{1} \\ &= \frac{p(b_1, A)}{\frac{\sum_{i=1}^n u(a_i)}{u(a_1)} \cdot p(a_1, A) + \frac{\sum_{j=1}^m u(b_j)}{u(b_1)} \cdot p(b_1, A)}, \text{ by (4)}, \\ &= \frac{u(b_1)}{\big(\sum_{i=1}^n u(a_i)\big) \cdot \frac{u(b_1)}{u(a_1)} \cdot \frac{p(a_1, A)}{p(b_1, A)} + \sum_{j=1}^m u(b_j)} \\ &= \frac{u(b_1)}{\big(\sum_{i=1}^n u(a_i)\big) \cdot \big(1 + \epsilon(A)\big) + \sum_{j=1}^m u(b_j)}, \text{ by (5)}. \end{aligned}$$

Since $1 = \sum_{i=1}^n p(a_i, A) + \sum_{j=1}^m p(b_j, A)$, we also have

$$p(a_1, A) = \frac{u(a_1)\big(1 + \epsilon(A)\big)}{\big(\sum_{i=1}^n u(a_i)\big) \cdot \big(1 + \epsilon(A)\big) + \sum_{j=1}^m u(b_j)}.$$

A.3 Proof of Theorem 2

Suppose p satisfies (2). Take any two menus $A, A' \in \mathscr{X}$. Suppose IIA is violated at both A and A'. Then we shall prove that $\epsilon(A) = \epsilon(A')$. Take any $a, b \in A$ and $a', b' \in A'$ such that

$$r_A(a, b) > r(a, b) \text{ and } r_{A'}(a', b) > r(a', b').$$

By Constant Impact, we have

$$\frac{r_A(a, b)}{r(a, b)} = 1 + \epsilon(A) = \frac{r_{A'}(a', b')}{r(a', b')} = 1 + \epsilon(A').$$

A.4 Proof of Theorem 3

We prove Theorem 3 by three steps.

Step 1: By Bounds on Dominance (A.7), there exists a^* such that for any $a \in X \setminus a^*$,

$a \simeq a^*$. Let us fix a^*. We then construct $u : X \to \mathbb{R}_{++}$ in the following way:

$$u(a) \equiv r(a, a^*) \text{ for any } a \in X.$$

Now note that for any $a, b \in X \setminus a^*$ with $a \simeq b$, $r(a,b) = \frac{u(a)}{u(b)}$ since

$$1 = r(a,b) \cdot r(b,a^*) \cdot r(a^*,a) = r(a,b) \cdot u(b) \cdot \frac{1}{u(a)}$$

by General Transitive Impact (A.8).

Step 2: Now take any A. When $A^> = a$ for some a, then by Dominance Consistency (A.4), $p(a, A) = 1$. Now suppose $|A^>| \geq 2$.

Fact 2: For any $a, a', b \in A^>$, if $r_A(a', a) = r(a', a)$ and $r_A(a, b) > r(a, b)$, then $r_A(a', b) > r(a', b)$.

Proof of Fact 2: Suppose $r_A(a', a) = r(a', a)$ and $r_A(a, b) > r(a, b)$. Then we have

$$r_A(a', b) = r_A(a', a) \cdot r_A(a, b) > r(a', a) \cdot r(a, b) = \frac{u(a')}{u(a)} \cdot \frac{u(a)}{u(b)} = \frac{u(a')}{u(b)} = r(a', b).$$

By General Bounds on Violations of IIA (A.6) and Fact 2, we can have $A^> = F(A) \cup U(A)$ such that for any $a, a' \in F(A)$ and $b, b' \in U(A)$,

$$r_A(a, a') = r(a, a'),\ r_A(b, b') = r(b, b'), \text{ and } r_A(a, b) > r(a, b).$$

Let $F(A) = \{a_1, \dots, a_n\}$ and $U(A) = \{b_1, \dots, b_m\}$.

Step 3: Now we assume that $|F(A)|, |U(A)| \geq 1$ and $|A \setminus A^>| \geq 1$.

By the construction of u, $r_A(a_i, a_1) = r(a_i, a_1) = \frac{u(a_i)}{u(a_1)}$. Therefore, $p(a_i, A) = \frac{u(a_i)}{u(a_1)} \cdot p(a_1, A)$. Similarly, $p(b_j, A) = \frac{u(b_j)}{u(b_1)} \cdot p(b_1, A)$. Since $p(A^>, A) = 1$ (by Dominance Consistency), we have

$$(6) \qquad 1 = \sum_{i=1}^{n} p(a_i, A) + \sum_{j=1}^{m} p(b_j, A) = \frac{\sum_{i=1}^{n} u(a_i)}{u(a_1)} \cdot p(a_1, A) + \frac{\sum_{j=1}^{m} u(b_j)}{u(b_1)} \cdot p(b_1, A).$$

Note that for any i, j,

$$\frac{r_A(a_1, b_1)}{r(a_1, b_1)} = \frac{r_A(a_i, b_j)}{r(a_i, b_j)}.$$

Let

$$1 + \epsilon(A) \equiv \frac{r_A(a_1, b_1)}{r(a_1, b_1)}. \tag{7}$$

Therefore, we have

$$\begin{aligned}
p(b_1, A) &= \frac{p(b_1, A)}{1} \\
&= \frac{p(b_1, A)}{\frac{\sum_{i=1}^{n} u(a_i)}{u(a_1)} \cdot p(a_1, A) + \frac{\sum_{j=1}^{m} u(b_j)}{u(b_1)} \cdot p(b_1, A)}, \text{ by (6)}, \\
&= \frac{u(b_1)}{\big(\sum_{i=1}^{n} u(a_i)\big) \cdot \frac{u(b_1)}{u(a_1)} \cdot \frac{p(a_1, A)}{p(b_1, A)} + \sum_{j=1}^{m} u(b_j)} \\
&= \frac{u(b_1)}{\big(\sum_{i=1}^{n} u(a_i)\big) \cdot \big(1 + \epsilon(A)\big) + \sum_{j=1}^{m} u(b_j)}, \text{ by (7)}.
\end{aligned}$$

and

$$p(a_1, A) = \frac{u(a_1)\big(1 + \epsilon(A)\big)}{\big(\sum_{i=1}^{n} u(a_i)\big) \cdot \big(1 + \epsilon(A)\big) + \sum_{j=1}^{m} u(b_j)}.$$

A.5 Proof of Theorem 4

We prove Theorem 4 by two steps. Suppose p is a GILM.

Step 1: Note that if $A = A^{>}$, then by Undominated IIA (A.10), we have $F(A) = A$ or $U(A) = A$. Therefore, as in Luce's model,

$$p(a, A) = \frac{u(a)}{\sum_{a' \in A} u(a')}.$$

Similarly, if $U(A) = \emptyset$, then we also have

$$p(a_i, A) = \frac{u(a_i)}{\sum_{s=1}^{n} u(a_s)}.$$

Now let us assume $|F(A)|, |U(A)|, |A \setminus A^{>}| \geq 1$. Without loss of generality, we also assume $\epsilon(A) > 0$.

Step 2: Since $r_A(a_1, b_1) > r(a_1, b_1)$, by Undominated IIA (A.10), there are $c \in A \setminus A^{>}$ and $x \in A$ such that $p(x, A) \neq p(x, A \setminus c)$. Let us fix c.

Fact 3: $p(y, A) \neq p(y, A \setminus c)$ for any $y \in A^{>}$.

Proof of Fact 3: By way of contradiction, suppose there is $y \in A^{>}$ such that $p(y, A) = p(y, A \setminus c) > 0$. Then by Weak Reference-Consistency (A.11), we need to have $p(x, A) = p(x, A \setminus c)$. A contradiction.

Take any i and j. By Fact 3, we have $p(a_i, A) \neq p(a_i, A \setminus c)$. Therefore, since $r_A(a_i, b_j) > r(a_i, b_j)$, by Proportional Impact (A.9), we have

$$r_A(a_i, b_j) = r(a_i, b_j) + r(c, b_j) \cdot p(a_i, A_c^{>}).$$

Since $r_A(a_i, b_j) > r(a_i, b_j)$, we must have $r(c, b_j) \cdot p(a_i, A_c^{>}) > 0$. In other words, $b_j \notin A_c^{>}$ and $a_i \in A_c^{>}$. Therefore, we have $U(A) \cap A_c^{>} = \emptyset$ and $F(A) \subseteq A_c^{>}$. Since $A_c^{>} \subseteq F(A) \cup U(A) = A^{>}$, we have $A_c^{>} = F(A)$.

Moreover, since $b_j \simeq c$ and $a_i \simeq b_j$, we have

$$r_A(a_i, b_j) = \frac{u(a_i)}{u(b_j)} + \frac{u(c)}{u(b_j)} \cdot p(a_i, A_c^{>}). \tag{8}$$

By Step 1,

$$p(a_i, A_c^{>}) = p(a_i, F(A)) = \frac{u(a_i)}{\sum_{s=1}^{n} u(a_s)}.$$

Then (8) implies that

$$\frac{p(a_1, A)}{p(b_1, A)} = \frac{u(a_1)}{u(b_1)} + \frac{u(c)}{u(b_1)} \cdot \frac{u(a_1)}{\sum_{i=1}^{n} u(a_i)}. \tag{9}$$

If we plug Equation (9) to Equation (6), then we have

$$\begin{aligned} 1 &= \sum_{i=1}^{n} p(a_i, A) + \sum_{j=1}^{m} p(b_j, A) = \frac{\sum_{i=1}^{n} u(a_i)}{u(a_1)} \cdot p(a_1, A) + \frac{\sum_{j=1}^{m} u(b_j)}{u(b_1)} \cdot p(b_1, A) \\ &= \frac{\sum_{i=1}^{n} u(a_i)}{u(a_1)} \cdot \Big(\frac{u(a_1)}{u(b_1)} + \frac{u(c)}{u(b_1)} \cdot \frac{u(a_1)}{\sum_{i=1}^{n} u(a_i)}\Big) p(b_1, A) + \frac{\sum_{j=1}^{m} u(b_j)}{u(b_1)} \cdot p(b_1, A) \\ &= \frac{u(c) + \sum_{i=1}^{n} u(a_i) + \sum_{j=1}^{m} u(b_j)}{u(b_1)} \cdot p(b_1, A) \end{aligned}$$

Therefore, we have

$$p(b_1, A) = \frac{u(b_1)}{\sum_{i=1}^{n} u(a_i) + \sum_{j=1}^{m} u(b_j) + u(c)} = \mu(a_1, A^{>} \cup c).$$

Moreover,

$$
\begin{aligned}
p(a_1, A) &= \frac{u(a_1)}{\sum_{i=1}^n u(a_i) + \sum_{j=1}^m u(b_j) + u(c)} + \frac{u(c)}{\sum_{i=1}^n u(a_i) + \sum_{j=1}^m u(b_j) + u(c)} \cdot \frac{u(a_1)}{\sum_{i=1}^n u(a_i)} \\
&= \mu(a_1, A^{>} \cup c) + \mu(c, A^{>} \cup c) \cdot \mu(a_1, A^{>}_c).
\end{aligned}
$$

Finally, we set $c = R(A)$.

References

Aguiar, V. (2016): "Random Categorization and Bounded Rationality," .

Ariely, D. and T. S. Wallsten (1995): "Seeking subjective dominance in multidimensional space: An explanation of the asymmetric dominance effect," *Organizational Behavior and Human Decision Processes*, 63, 223–232.

Barbera, S. and P. K. Pattanaik (1986): "Falmagne and the rationalizability of stochastic choices in terms of random orderings," *Econometrica: Journal of the Econometric Society*, 707–715.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013): "Salience and Consumer choice," *Journal of Political Economy*, 121, 803–843.

Brady, R. L. and J. Rehbeck (2016): "Menu-Dependent Stochastic Feasibility," *Econometrica*, 84, 1203–1223.

Chernev, A. (2004): "Extremeness Aversion and Attribute-Balance Effects in Choice," *Journal of Consumer Research*, 31, 249–263.

Doyle, J. R., D. J. O'Connor, G. M. Reynolds, and P. A. Bottomley (1999): "The robustness of the asymmetrically dominated effect: Buying frames, phantom alternatives, and in-store purchases," *Psychology & Marketing*, 16, 225–243.

Echenique, F. and K. Saito (2015): "General Luce Model," .

Echenique, F., K. Saito, and G. Tserenjigmid (2013): "The Perception-Adjusted Luce Model," *Caltech SS Working Paper 1379*.

Falmagne, J.-C. (1978): "A representation theorem for finite random scale systems," *Journal of Mathematical Psychology*, 18, 52–72.

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

Gul, F. (1991): "A theory of disappointment aversion," *Econometrica: Journal of the Econometric Society*, 667–686.

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

Herne, K. (1997): "Decoy alternatives in policy choices: Asymmetric domination and compromise effects," *European Journal of Political Economy*, 13, 575–589.

——— (1998): "Testing the reference-dependent model: An experiment on asymmetrically dominated reference points," *Journal of Behavioral Decision Making*, 11, 181–192.

——— (1999): "The effects of decoy gambles on individual choice," *Experimental Economics*, 2, 31–40.

Huber, J., J. W. Payne, and C. Puto (1982): "Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis," *Journal of consumer research*, 90–98.

Koszegi, B. and M. Rabin (2006): "A MODEL OF REFERENCE-DEPENDENT PREFERENCES." *Quarterly journal of economics*, 121.

Kovach, M. (2016): "Thinking Inside the Box: Status Quo Bias and Stochastic Consideration," Working paper.

Lleras, J. S., Y. Masatlioglu, D. Nakajima, and E. Y. Ozbay (2016): "When More is Less: Choice by Limited Consideration," *Journal of Economic theory*, forthcoming.

Luce, R. D. (1959): *Individual choice behavior: A theoretical analysis*, John Wiley and sons.

Manzini, P. and M. Mariotti (2014): "Stochastic choice and consideration sets," *Econometrica*, 82, 1153–1176.

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

McFadden, D. and M. K. Richter (1990): "Stochastic rationality and revealed stochastic preference," *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO*, 161–186.

Munro, A. and R. Sugden (2003): "On the theory of reference-dependent preferences," *Journal of Economic Behavior & Organization*, 50, 407–428.

Natenzon, P. (2010): "Random choice and learning," Tech. rep., Working paper, Princeton University.

Ok, E. A., P. Ortoleva, and G. Riella (2014): "Revealed (p) reference theory," *The American Economic Review.*

Rieskamp, J., J. R. Busemeyer, and B. A. Mellers (2006): "Extending the bounds of rationality: evidence and theories of preferential choice," *Journal of Economic Literature*, 631–661.

Roe, R. M., J. R. Busemeyer, J. T. Townsend, et al. (2001): "Multialternative decision field theory: A dynamic connectionist model of decision making," *PSYCHOLOGICAL REVIEW-NEW YORK-*, 108, 370–392.

Sharpe, K. M., R. Staelin, and J. Huber (2008): "Using extremeness aversion to fight obesity: policy implications of context dependent demand," *Journal of Consumer Research*, 35, 406–422.

Simonson, I. (1989): "Choice based on reasons: The case of attraction and compromise effects," *Journal of consumer research*, 158–174.

Simonson, I. and A. Tversky (1992): "Choice in context: tradeoff contrast and extremeness aversion." *Journal of marketing research.*

Sue O'Curry, Y. P. and R. Pitts (1995): "The attraction effect and political choice in two elections," *Journal of Consumer Psychology*, 4, 85–101.

Sugden, R. (2003): "Reference-dependent subjective expected utility," *Journal of economic theory*, 111, 172–191.

Tserenjigmid, G. (2014a): "Choosing with the Worst in Mind: A Reference-Dependent Model," *Caltech SS Working Paper 1391.*

——— (2014b): "The Order-Dependent Luce Model," *Working Paper, Caltech.*

Tversky, A. (1972): "Choice by elimination," *Journal of mathematical psychology*, 9, 341–367.

Tversky, A. and D. Kahneman (1991): "Loss aversion in riskless choice: A reference-dependent model," *The Quarterly Journal of Economics*, 106, 1039–1061.

Tversky, A. and I. Simonson (1993): “Context-dependent preferences,” *Management science*, 39, 1179–1189.

Usher, M. and J. L. McClelland (2004): “Loss aversion and inhibition in dynamical models of multialternative choice.” *Psychological review*, 111, 757.

Zhang, J. (2016): “Stochastic Choice with Categorization,” .

B Online Appendix B: A relation to Attribute Rule Model of Gul et al. (2014)

In this section, we prove that the only intersection between the ILM and the attribute rule model of Gul et al. (2014) is Luce's model. First, we define the attribute rule model. Let Z be the collection of attributes, let α_a be the set of attributes that a has, and let $\alpha(A) = \cup_{a \in A} \alpha_a$. Hence, $\alpha(A)$ is the set of attributes represented in A.

Definition 8 (Attribute Rule)**.** A random choice choice rule p is an attribute rule if there exist an attribute value function $w : Z \to \mathbb{R}_{++}$ and an attribute intensity function $\eta : Z \times X \to \mathbb{N} \cup \{0\}$ such that

$$p(a, A) = \sum_{x \in \alpha_a} \frac{\omega_x}{\sum_{x' \in \alpha(A)} \omega_{x'}} \cdot \frac{\eta_a^x}{\sum_{b \in A} \eta_b^x}$$

where $\alpha_a = \{x \in Z | \eta_a^x > 0\}$.

Let us check in whether Bounds on Violations of IIA is satisfied in Attribute Rule. Take a, b, c. Without loss of generality, we assume that

$$\alpha_a = \{\alpha_1, \dots, \alpha_m, x_1, \dots, x_n, z_1, \dots, z_k, t_1, \dots, t_s\}, \alpha_b = \{\beta_1, \dots, \beta_p, x_1, \dots, x_n, y_1, \dots, y_q, t_1, \dots, t_s\},$$

$$\text{and } \alpha_c = \{\gamma_1, \dots, \gamma_r, y_1, \dots, y_q, z_1, \dots, z_k, t_1, \dots, t_s\}.$$

Then

$$r(a, b) = \frac{\sum_\alpha \omega_\alpha + \sum_z \omega_z + \sum_x \omega_x \cdot \frac{\eta_a^x}{\eta_a^x + \eta_b^x} + \sum_t \omega_t \cdot \frac{\eta_a^t}{\eta_a^t + \eta_b^t}}{\sum_\beta \omega_\beta + \sum_y \omega_y + \sum_x \omega_x \cdot \frac{\eta_b^x}{\eta_a^x + \eta_b^x} + \sum_t \omega_t \cdot \frac{\eta_b^t}{\eta_a^t + \eta_b^t}}$$

and

$$r_{abc}(a, b) = \frac{\sum_\alpha \omega_\alpha + \sum_z \omega_z \cdot \frac{\eta_a^z}{\eta_a^z + \eta_c^z} + \sum_x \omega_x \cdot \frac{\eta_a^x}{\eta_a^x + \eta_b^x} + \sum_t \omega_t \cdot \frac{\eta_a^t}{\eta_a^t + \eta_b^t + \eta_c^t}}{\sum_\beta \omega_\beta + \sum_y \omega_y \cdot \frac{\eta_b^y}{\eta_b^y + \eta_c^y} + \sum_x \omega_x \cdot \frac{\eta_b^x}{\eta_a^x + \eta_b^x} + \sum_t \omega_t \cdot \frac{\eta_b^t}{\eta_a^t + \eta_b^t + \eta_c^t}}.$$

In order to have $r(a,b) = r_{abc}(a,b)$, we need

$$\frac{\sum_\alpha \omega_\alpha + \sum_z \omega_z + \sum_x \omega_x \cdot \frac{\eta_a^x}{\eta_a^x+\eta_b^x} + \sum_t \omega_t \cdot \frac{\eta_a^t}{\eta_a^t+\eta_b^t}}{\sum_\beta \omega_\beta + \sum_y \omega_y + \sum_x \omega_x \cdot \frac{\eta_b^x}{\eta_a^x+\eta_b^x} + \sum_t \omega_t \cdot \frac{\eta_b^t}{\eta_a^t+\eta_b^t}} = \frac{\sum_z \omega_z \cdot \frac{\eta_c^z}{\eta_a^z+\eta_c^z} + \sum_t \omega_t \cdot (\frac{\eta_a^t}{\eta_a^t+\eta_b^t} - \frac{\eta_a^t}{\eta_a^t+\eta_b^t+\eta_c^t})}{\sum_y \omega_y \cdot \frac{\eta_c^y}{\eta_b^y+\eta_c^y} + \sum_t \omega_t \cdot (\frac{\eta_b^t}{\eta_a^t+\eta_b^t} - \frac{\eta_b^t}{\eta_a^t+\eta_b^t+\eta_c^t})}.$$

Since the lefthand depends on ω_α while the righthand side does not depend on it, unless ω_a is constant, the above equality cannot hold. Now suppose ω is constant. Then we need to have

$$\frac{m + k + \sum_x \frac{\eta_a^x}{\eta_a^x+\eta_b^x} + \sum_t \frac{\eta_a^t}{\eta_a^t+\eta_b^t}}{p + q + \sum_x \frac{\eta_b^x}{\eta_a^x+\eta_b^x} + \sum_t \frac{\eta_b^t}{\eta_a^t+\eta_b^t}} = \frac{\sum_z \frac{\eta_c^z}{\eta_a^z+\eta_c^z} + \sum_t (\frac{\eta_a^t}{\eta_a^t+\eta_b^t} - \frac{\eta_a^t}{\eta_a^t+\eta_b^t+\eta_c^t})}{\sum_y \frac{\eta_c^y}{\eta_b^y+\eta_c^y} + \sum_t (\frac{\eta_b^t}{\eta_a^t+\eta_b^t} - \frac{\eta_b^t}{\eta_a^t+\eta_b^t+\eta_c^t})}.$$

Similarly, since the righthand depends on γ_c^z while the righthand side does not depend on it, unless γ_c^z is constant, the above equality cannot hold. Now suppose γ is constant. Then we have

$$\frac{m + k + \frac{n}{2} + \frac{s}{2}}{p + q + \frac{n}{2} + \frac{s}{2}} = \frac{\frac{k}{2} + \frac{s}{6}}{\frac{q}{2} + \frac{s}{6}}.$$

If there are flexibilities in values of m, p, then $k = q = n = s = 0$ in order to have the above. However, this will lead to Luce's model. So assume that $m = p$. Moreover, if there are flexibilities in values of k, q, then $m = n = s = p = 0$ in order to have the above equality. However, Transitive Impact implies $k = q$ when $m = n = s = p = 0$. Therefore, we have $m = p$ and $k = q$. Finally, $m = p$ and $k = q$ implies that there is a some $\delta \geq \frac{1}{2}$ such that for any $a, b \in X$, $p(a, ab) \in \{\frac{1}{2}, \delta, 1 - \delta\}$. However, this case violates the Strong Richness condition Gul et al. (2014) used to identify their model.