9-2016

Associationistic luce rule

Jiangtao LI
Singapore Management University, jtli@smu.edu.sg

Rui TANG

Associationistic Luce Rule*

Jiangtao Li† Rui Tang‡

September 25, 2016

Abstract

We develop an extension of Luce's random choice model that incorporates a role for the association of alternatives. Each alternative is characterized by a salience value, a Luce value, and its associated alternatives. The salience value captures the alternative's ability to attract the decision maker's attention, and the Luce value measures the alternative's desirability. The decision maker is first attracted by some alternative according to a salience-based Luce-type formula, and then chooses among its associated alternatives according to another desirability-based Luce-type formula. While retaining the simplicity of the Luce rule, the theory accommodates some well-known behavioral phenomena in individual choice, such as the attraction effect (violations of regularity), and violations of stochastic transitivity.

*We are grateful to John Quah for his continuous and invaluable encouragement, support, and guidance. We thank Soo Hong Chew, Faruk Gul, Shaowei Ke, Matthew Kovach, Yusufcan Masatlioglu, and Satoru Takahashi for helpful discussions. Part of this paper was written while Li was visiting the Economics Department at the University of Michigan, and he would like to thank the institution for its hospitality and support. All remaining errors are our own.

†Department of Economics, National University of Singapore, jasonli1017@gmail.com

‡Department of Economics, National University of Singapore, rui.tang@u.nus.edu

1 Introduction

In both experimental and market settings, individual choice responses typically exhibit variability; see for example, Sippel (1997), McFadden (2001), and Manzini, Mariotti, and Mittone (2010). This paper studies a random choice model, in which we assume the choice responses to be given by a probability distribution ρ that indicates the probability $\rho(A', A)$ that when the possible alternatives faced by the decision maker are the alternatives in A, some alternative in A' is chosen, as in the seminal work of Luce (1959).[1]

In the celebrated Luce model, each alternative x is characterized by a *Luce value* v_x that measures the alternative's desirability. The probability of choosing x from an alternative set A containing x is

$$\rho(\{x\}, A) := \frac{v_x}{\sum_{y \in A} v_y}.$$

Despite many attractive features, there are well-documented violations of the Luce model. In this paper, we develop an extension of Luce's random choice model that incorporates a role for the *association of alternatives*. In an associationistic Luce rule, each alternative x is characterized by a *salience value*, a *Luce value*, and the set of alternatives that are associated with x. The salience value captures the alternative's ability to attract the decision maker's attention. As in the Luce model, the Luce value measures the alternative's desirability. The decision maker is first attracted by some alternative according to a salience-based Luce-type formula, and then chooses among its associated alternatives according to another desirability-based Luce-type formula. Our model reduces to the Luce model when all alternatives are associated with each other.

We borrow the term association from the philosophy and psychology literature. As noted in Mandelbaum (2016), "associationism" is a certain arationality of thought: a creature's mental states are associated because of some facts about its causal history, and having these mental states associated entails that *bringing one of a pair of associates to mind will, ceteris paribus, ensure that the other also becomes activated*. In our setting, association of alternatives means that, even if the decision maker is not first attracted by alternative x, as long as the decision maker is attracted by some alternative that is associated with x, the decision maker still considers x when she eventually makes a choice.

[1] Also see Gul, Natenzon, and Pesendorfer (2014), Manzini and Mariotti (2014), Echenique, Saito, and Tserenjigmid (2014), Brady and Rehbeck (2016), Yildiz (2016), and Kovach (2016), etc.

This type of decision procedure is ubiquitous in daily life. An investor is attracted by a particular automobile manufacturer (possibly due to prior knowledge). But when she compares within the automobile manufacturing industry, she might eventually invest in another automobile manufacturer, which has stronger fundamentals. A shopper is attracted by a recently launched handbag of a luxury brand (possibly through advertisement). But when she visits the store and compares the various handbags, she might eventually purchase a different handbag, which fits her even better. A foodie is attracted by a spicy dish at a Michelin restaurant (possibly recommended by a friend). But when she dines there and compares the various choices from the menu, she might eventually order a non-spicy dish, which suits her taste even better.

The examples above suggest specific choice of association. It seems natural that all automobile manufacturers (resp. all handbags in the store, all dishes in the restaurant) are associated with each other. The difficulty is that the designation of association is rarely clear-cut. Consider an alternative set with three alternatives: red bus, yellow bus, and red train. The decision maker may associate the yellow bus with the red bus (same means of transportation), or associate the red train with the red bus (same color), or associate both the yellow bus and the red train with the red bus (all three alternatives are means of transportation). Therefore, how the decision maker associates one alternative with another cannot be decided based on the physical characteristics of the alternatives. Whether two alternatives are associated with each other is subjective, and must be derived from the choice data. We formally define the association of alternatives in Section 4.

While retaining the simplicity of the Luce rule, the theory can explain some well-known violations of the Luce model, such as the attraction effect (violations of regularity), and violations of stochastic transitivity. In what follows, we illustrate how the theory accommodates violations of regularity.[2] Regularity asserts that when an alternative is added to an alternative set, the choice probability of the original members of the set cannot increase.[3] Although regularity seems innocuous, this property conflicts with a substantial body of evidence, most notably the attraction effect. The following is a modification of an example in Simonson and Tversky (1992).

[2] A fuller discussion is contained in Section 5.

[3] Regularity is necessarily satisfied in many random choice models including the Luce rule (Luce (1959)), the elimination-by-aspects rule (Tversky (1972)), the attribute rule (Gul, Natenzon, and Pesendorfer (2014)), and the random consideration set rule (Manzini and Mariotti (2014)).

Example 1. *Let* $X = \{x, y, z\}$*. Alternatives* x *and* z *are different variants of the same good where* x *is high quality and expensive good, and* z *is a more expensive version of* x*. Alternative* y *is of a different brand (low quality and cheap good). We might have* $\rho(\{x\}, \{x, y\}) < \rho(\{x\}, \{x, y, z\})$*, exhibiting attraction effect (violations of regularity).*

In Example 1, since x and z are different variants of the same good, we may assume that x and z are associated with each other. Also assume that y and z are not associated with each other. Furthermore, it is plausible that x is more desirable (higher Luce value) than z, since z is a more expensive version of x. On the one hand, the addition of z to $\{x, y\}$ decreases the probability that the decision maker is first attracted by x and y. On the other hand, when the decision maker is attracted by z, she still considers x, but not y, when she eventually makes a choice. If x has a sufficiently high Luce value relative to z, then the addition of z increases the probability of x being chosen. Hence, our model allows for violations of the regularity property.

This paper joins a growing literature that studies random choice. The closest to our model in terms of representation is the attribute rule (Gul, Natenzon, and Pesendorfer (2014)). They interpret alternatives as bundles of attributes. An attribute value maps attributes to positive reals, and measures the desirability of the attributes.[4] The decision maker is first attracted by some attribute from all perceived attributes according to an attribute-value-based Luce-type formula, and then randomly chooses an alternative containing the selected attribute. Consider an investor who is contemplating an investment in automobile manufacturer BMW and several other companies. If we add another automobile manufacturer Porsche in the alternative set, this never increases the probability of BMW being considered when the investor eventually makes a choice. At best, Porsche does not introduce attributes that are non-existent in the original alternative set, in which case the probability of BMW being considered when the investor eventually makes a choice remains unchanged.

In our model, the decision maker is randomly attracted by some alternative rather than some attribute, and then randomly chooses from its associated alternatives. In the BMW-Porsche example, an investor is attracted by each company with certain probability (possibly due to prior knowledge of the companies, transaction history of the company stocks,

[4]More formally, an attribute set Z is an arbitrary index set, and each element of Z is an attribute. An attribute value is a function $\omega : Z \to \mathbb{R}_{++}$. Note that the attribute value only depends on the attribute itself, and is independent of the alternatives that contain the attribute.

or recommendations by a financial advisor). We hasten to stress that the addition of Porsche increases the probability of BMW being considered when the investor eventually makes a choice, if BMW and Porsche are associated with each other. When she is first attracted by Porsche, she still considers BMW when she eventually makes the investment decision.

Section 2 presents the basics of the model. Section 3 introduces the axiom of *additivity of relative probabilities*, and shows that the Luce rule is the only choice rule that satisfies additivity of relative probabilities. Section 4 presents the associationistic Luce rule. We formally define the association of alternatives, using a revealed stochastic preference approach. We weaken the axiom of additivity of relative probabilities. Theorem 2, our main result, shows that the associationistic Luce rule is the only choice rule that satisfies the weak version of additivity of relative probabilities and an additional transitivity axiom. Section 5 applies the theory to explain the attraction effect (violations of regularity), and violations of stochastic transitivity. We also show that the associationistic Luce rule satisfies a property that we call *attraction transitivity*. Section 6 compares the associationistic Luce rule to other models in the literature. The appendix contains proofs omitted from the main body of the paper.

2 Preliminaries

We work with any arbitrarily fixed nonempty finite set X, which can be viewed as the universal set of available alternatives. Let $\mathcal{A}$ be a domain of alternative sets which are subsets of X. We assume that the domain satisfies the following richness condition: $A \in \mathcal{A}$ for all A containing four alternatives, and $B \in \mathcal{A}$ whenever $B \subset A$ and $A \in \mathcal{A}$. Let $\mathcal{A}_+ = \mathcal{A} \backslash \emptyset$. The elements of $\mathcal{A}_+$ are viewed as feasible sets that a decision maker may need to choose an alternative from. We use $A, B, C, \ldots$ to denote alternative sets, and $x, y, z, \ldots$ to denote alternatives. Throughout the rest of the paper, unless it leads to confusion, we abuse the notation by suppressing the set delimiters, e.g., writing x rather than $\{x\}$. To simplify the statements below, we use the following notational convention:

$$AB := A \cup B,$$

$$xA := x \cup A,$$

$$xy := x \cup y.$$

Definition 1. *A* random choice rule *is a map* $\rho : \mathcal{A} \times \mathcal{A}_+ \to [0, 1]$ *such that:*

i) *for all* $A \in \mathcal{A}_+$*,* $\rho(\cdot, A)$ *is additive, and* $\rho(A, A) = 1$*;*

ii) *for all* $A \in \mathcal{A}_+$ *with* $|A| \geq 2$ *and* $x \in A$*,* $\rho(x, A) \in (0, 1)$*.*

The interpretation is that $\rho(A', A)$ denotes the probability that when the possible alternatives faced by the decision maker are the alternatives in A, some alternative in A' is chosen. Additivity is the requirement that $\rho(\cdot, A)$ is a probability. Equation $\rho(A, A) = 1$ is the feasibility constraint; ρ must choose among alternatives available in A. The constraint $\rho(x, A) \in (0, 1)$ says that the random choice rule is non-degenerate; that is, each alternative x in the feasible set A is chosen with positive probability.

We write $\rho(x|A)$ rather than $\rho(x, A)$, and write $\rho(A'|A)$ rather than $\rho(A', A)$. In an alternative set C that contains A and B, we denote the *relative probability* of choosing alternatives in A and B conditional on C as follows:

$$\beta(A, B|C) = \frac{\rho(A|C)}{\rho(B|C)}.$$

For all pairwise disjoint alternative sets $A, B, D \in \mathcal{A}_+$, we call $\beta(A, B|ABD) - \beta(A, B|AB)$ the *impact* of D on the relative probability of choosing A and B.

3 Luce Rule

The Luce rule is a well-known behavioral optimization model that can be described as follows. Let $v : \mathcal{A}_+ \to R_{++}$, and denote $v(x)$ by v_x. Such a function v is a *Luce value* if it is additive. That is, for all $A \in \mathcal{A}_+$,

$$v(A) = \sum_{x \in A} v_x.$$

Call the choice rule ρ a *Luce rule* if there exists a Luce value v such that

$$\rho(x|A) = \frac{v_x}{v(A)} \tag{1}$$

whenever $x \in A \in \mathcal{A}_+$. We say that the Luce value v represents ρ if equation (1) holds for all such x, A.

Luce (1959) proves that a choice rule ρ is a Luce rule if and only if it satisfies *Luce's independence of irrelevant alternatives (IIA) axiom*: for all pairwise disjoint alternative sets $A, B, C \in \mathcal{A}_+$,

$$\beta(A, B|ABC) = \beta(A, B|AB).$$

In words, Luce's IIA axiom says that for all disjoint alternative sets A, B and C, C has no impact on the relative probability of choosing A and B. It is a simple form of menu independence.

In a Luce rule, the relative probability of choosing AB and C conditional on ABC hinges upon *the desirability of the alternatives in AB relative to that of the alternatives in C*.[5] Since the desirability of the alternatives in AB can be decomposed into the desirability of the alternatives in A and the desirability of the alternatives in B, the relative probability of choosing AB and C conditional on ABC is the sum of two components: the relative probability of choosing A and C conditional on AC, and the relative probability of choosing alternatives in B and C conditional on BC. We introduce the following axiom of *additivity of relative probabilities (ARP)*.

Axiom ARP - Additivity of Relative Probabilities: For all pairwise disjoint alternative sets $A, B, C \in \mathcal{A}_+$, $\beta(AB, C|ABC) = \beta(A, C|AC) + \beta(B, C|BC)$.

Axiom Weak ARP does not claim that B has no impact on the relative probability of choosing A and C *per se* (or that A has no impact on the relative probability of choosing B and C). Rather, the axiom says that these two impacts cancel out. Having said that, Axiom ARP places virtually no restrictions on the alternative sets A, B and C. The following theorem shows that Axiom ARP is another probabilistic form of Luce's IIA axiom, and the Luce rule is the only choice rule that satisfies Axiom ARP.

Theorem 1. *A choice rule is a Luce rule if and only if it satisfies Axiom ARP.*

4 Associationistic Luce rule

Despite many attractive features, there are well-documented violations of the Luce model. In this section, we develop an extension of Luce's random choice model that incorporates a role for the association of alternatives. In the associationistic Luce rule, each alternative is characterized by a salience value, a Luce value, and its associated alternatives. The salience value captures the alternative's ability to attract the decision maker's attention. As in the Luce model, the Luce value measures the alternative's desirability. The decision maker is

[5] Similarly, the relative probability of choosing A (resp. B) and C hinges upon the desirability of the alternatives in A (resp. B) relative to that of the alternatives in C.

first attracted by some alternative according to a salience-based Luce-type formula, and then chooses among its associated alternatives according to another desirability-based Luce-type formula.

Salience value: A *salience value* is a function $\gamma : \mathcal{A}_+ \to R_{++}$, and captures the alternative's ability to attract the decision maker's attention. Again, we write γ_x rather than $\gamma(x)$. Like Luce value, we require γ to be additive. That is, for all $A \in \mathcal{A}_+$,

$$\gamma(A) := \sum_{x \in A} \gamma_x.$$

Association of alternatives: We say that two alternatives x and y are associated with each other if whenever the decision maker is first attracted by y, she also considers x (given that x is in the feasible set) when she eventually makes a choice. In other words, even if the decision maker is not first attracted by alternative x, as long as the decision maker is attracted by some alternative that is associated with x, the decision maker still considers x when she eventually makes a choice.

Whether two alternatives are associated with each other is subjective, and must be derived from behavior. It cannot be decided based on the physical characteristics of the alternatives, and therefore, it is a property of the choice rule and not of the alternatives. We use the choice data to reveal whether two alternatives are associated in the following way. We say that two alternatives x and y are (revealed to be) associated if the relative probability of choosing x and y is menu-independent; that is, the relative probability of choosing x and y is the same in any alternative set A that contains x and y. More formally,

Definition 2. *Two alternatives x and y are* associated *with each other if in any alternative set A that contains x and y, $\beta(x, y|A) = \beta(x, y|xy)$.*

We write $x \sim y$ if x and y are associated with each other, and $x \nsim y$ if otherwise. It is easy to see that the relation $\sim$ is symmetric and reflexive. We write $A \approx B$ if for any $x \in A$ and $y \in B$, we have $x \sim y$. We say that an alternative set A is a collection of associated alternatives if $x, y \in A$ implies $x \sim y$. In this case, we write $A \in \mathcal{L}$. As the following example demonstrates, the relation $\sim$ is not necessarily transitive, and therefore, is not an equivalence relation.

Example 2. *Let $X = \{x, y, z\}$. The random choice rule ρ is given by*

$$\rho(x|x) = \rho(y|y) = \rho(z|z) = 1,$$

$$\rho(x|xy) = \rho(y|xy) = \rho(y|yz) = \rho(z|yz) = \frac{1}{2},$$
$$\rho(x|xyz) = \rho(y|xyz) = \rho(z|xyz) = \frac{1}{3},$$
$$\rho(x|xz) = \frac{1}{3}, \rho(z|xz) = \frac{2}{3}.$$

It is straightforward to verify that $x \sim y$, $y \sim z$, and yet $x \nsim z$.

Next we define the notion of *connectedness* of two alternatives sets: when A and B have elements in common, they are connected. Even if A and B have no elements in common, they are connected if there are elements of A and B that are associated with each other.

Definition 3. *A and B are* disconnected *if whenever $x \in A, y \in B$, we have $x \nsim y$.*

We write $A \perp B$ if A and B are disconnected.

Our first axiom asserts that the relation $\sim$ is transitive; that is, if x and y are associated with each other, and y and z are associated with each other, then x and z are associated with each other.

Axiom T - Transitivity : The relation $\sim$ is transitive.

Axiom T implies that we can partition the universal set of alternatives X in the following way: $\{P_i\}_{i=1}^n$, such that i) $P_i \in \mathcal{L}$, for any $i = 1, 2, \ldots, n$; and ii) $P_i \perp P_j$ for any $i, j = 1, 2, \ldots, n$ and $i \neq j$.

Suppose that AB and C are disconnected. If the decision maker is first attracted by the alternatives in C, she does not consider the alternatives in AB when she eventually makes a choice. Similarly, if the decision maker is first attracted by the alternatives in AB, she does not consider the alternatives in C when she eventually makes a choice. In other words, the relative probability of choosing AB and C conditional on ABC hinges upon *the ability of the alternatives in AB to attract the decision maker's attention relative to that of the alternatives in C.*[6] Since the ability of the alternatives in AB to attract the decision maker's attention can be decomposed into the ability of the alternatives in A to attract the decision maker's attention and that of the alternatives in B, the relative probability of choosing AB and C conditional on ABC is the sum of two components: the relative probability of choosing A and C conditional on AC, and the relative probability of choosing alternatives in B and C conditional on BC. More formally,

[6] Similarly, the relative probability of choosing A (resp. B) and C hinges upon the ability of the alternatives in A (resp. B) to attract the decision maker's attention relative to that of the alternatives in C.

Axiom Weak ARP - Weak Additivity of Relative Probabilities : If $A \cap B = \emptyset$ and $AB \perp C$, then $\beta(AB, C|ABC) = \beta(A, C|AC) + \beta(B, C|BC)$.

Mathematically, Axiom Weak ARP is a weakening of Axiom ARP. However, the rationale for the two axioms is different. In a Luce rule, the relative probability is solely driven by the desirability of the alternatives. In contrast, in an associationistic Luce rule, in the case that AB and C are disconnected, the relative probability is solely driven by the ability of the alternatives to attract the decision maker's attention. Consequently, this weak version requires additivity of relative probability only for the case in which AB and C are disconnected. We hasten to stress that Axiom Weak ARP does not claim that, for C disconnected with AB, B has no impact on the relative probability of choosing A and C (or that A has no impact on the relative probability of choosing B and C). Rather, the axiom says that these two impacts cancel out.

Associationistic Luce rule: For a partition of X and $x \in X$, we denote by P_x the block that contains x. Call the choice rule ρ an *associationistic Luce rule* if there exists a partition $(P_i)_{i=1}^n$ of X, a salience value γ, and a Luce value v such that

$$\rho(x|A) = \frac{\gamma(A \cap P_x)}{\gamma(A)} \frac{v(x)}{v(A \cap P_x)} \tag{a}$$

whenever $x \in A \in \mathcal{A}_+$. We say that $((P_i)_{i=1}^n, \gamma, v)$ represents ρ if equation (a) holds for all such x, A. Figure 1 illustrates the associationistic Luce rule.

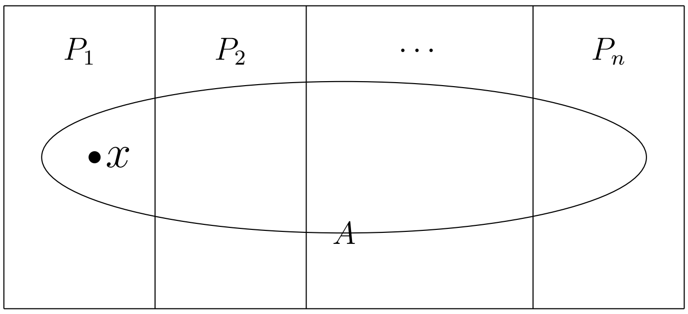

Figure 1: The universal set of alternatives X is depicted as the outermost rectangle. The feasible set A that the decision maker needs to choose an alternative from is depicted as the ellipse. The partition $(P_i)_{i=1}^n$ of X are depicted as the smaller rectangles marked by $P_1, P_2, \ldots, P_n$. We have $\rho(x|A) = \frac{\gamma(A \cap P_1)}{\gamma(A)} \frac{v(x)}{v(A \cap P_1)}$.

We show that the associationistic Luce rule is the only choice rule that satisfies Axiom T and Axiom Weak ARP. Formally, we prove a stronger result. Consider the following axiom.

Axiom Weak ARP* : If $A \cap B = \emptyset$ and $AB \perp C$ where $C \in \mathcal{L}$, then $\beta(AB, C|ABC) = \beta(A, C|AC) + \beta(B, C|BC)$.

It is easy to see that Axiom ARP implies Axiom ARP*. The following theorem shows that the associationistic Luce rule is the only choice rule that satisfies Axiom T and Axiom Weak ARP*.

Theorem 2. *A choice rule is an associationistic Luce rule if and only if it satisfies Axiom T and Axiom Weak ARP*.*

In what follows, we discuss the uniqueness of the representation. Consider the following example.

Example 3. *Let $X = \{x, y, z, w\}$. Suppose that it is revealed from the choice data that x and y are associated with each other, z and w are associated with each other, and that $(\{\{x, y\}, \{z, w\}\}, \gamma, v)$ represents choice rule ρ, where $\gamma_x = \gamma_z = \gamma_w = 1$, $\gamma_y = 2$, $v_x = v_w = 1$, and $v_y = v_z = 2$. In this example, although x and y are associated with each other, it is easy to verify that $\beta(x, z|xyz) = \beta(x, z|xz)$, and $\beta(x, w|xyw) = \beta(x, w|xw)$. In other words, y has no impact on the relative probability of choosing x and z (or w). Note that this is also the case if y is in a different block from x and z. Indeed, one can construct $(\{\{x\}, \{y\}, \{z, w\}\}, \gamma', v')$ that also represents ρ. In particular, $\gamma'_x = \gamma'_z = \gamma'_w = 1$, $\gamma'_y = 2$, $v'_x = v'_y = v'_w = 1$, and $v'_z = 2$.*

We say that $x \in X$ is an isolated alternative if i) for any $y \in X$, $y \nsim x$; or ii) for any $y \sim x$ where $y \neq x$, and $z \nsim x$, we have $\beta(x, z|xyz) = \beta(x, z|xz)$. Denote the set of isolated alternatives under choice rule ρ by S_ρ. The following theorem shows that the partition structure is unique in $X \backslash S_\rho$. In Example 3, $S_\rho = \{x, y\}$. This theorem implies that, for any $((P_i)_{i=1}^n, \gamma, v)$ that represents ρ, $\{z, w\}$ is a block of $(P_i)_{i=1}^n$.

Theorem 3. *If $((P_i)_{i=1}^n, \gamma, v)$ represents choice rule ρ, then either $P_i \cap S_\rho = \emptyset$ or $P_i \subset S_\rho$ for any $i \in \{1, 2, ..., n\}$. Furthermore, if both $((P_i)_{i=1}^n, \gamma, v)$ and $((P'_j)_{j=1}^m, \gamma', v')$ represent ρ, $P'_i \subset X \backslash S_\rho$ implies that P'_i is a block of $(P_i)_{i=1}^n$ (and vice versa).*

Fix an partition of X that represents ρ, it is easy to see that the salience value γ is unique up to a positive multiplicative constant, and the Luce value v is unique up a positive multiplicative constant within each block.

5 Anomalies

This section illustrates how the associationistic Luce rule accommodates some well-known violations of the Luce model, such as the attraction effect (violations of regularity), and violations of stochastic transitivity. Throughout this section, we assume that the decision maker follows the associationistic Luce rule. Note that when all alternatives are associated with each other, the associationistic Luce rule reduces to the Luce rule. Consequently, the data can not exhibit any violations of the Luce model. For simplicity of exposition, we shall exclude the discussion of this case.[7]

5.1 Violation of Regularity - Attraction Effect

The associationistic Luce rule can accommodate violations of regularity. We focus on the attraction effect, a well-known violation of regularity. As in Rieskamp, Busemeyer, and Mellers (2006), the attraction effect (violations of regularity) can be formalized as follows:

$$\rho(x|xyz) > \rho(x|xy).$$

In words, the addition of a third alternative z to the alternative set xy increases the probability of x being chosen. Strong evidence of this phenomenon has been found in many studies in psychology and marketing, in different contexts, and in both laboratory and field experiments; see for example, Huber, Payne, and Puto (1982), Simonson and Tversky (1992), and Doyle, O'Connor, Reynolds, and Bottomley (1999).

In this subsection, we show that the associationistic Luce rule captures the attraction effect. Consider the following example.

Example 4. *Let $\gamma_x = 1, \gamma_y = \gamma_z = 2$, and $v_x = v_y = 2, v_z = 1$, where x and z are in the same block, y is in a different block. We have*

$$\rho(x|xyz) = \frac{\gamma_x + \gamma_z}{\gamma_x + \gamma_y + \gamma_z} \frac{v_x}{v_x + v_z} = \frac{2}{5},$$
$$\text{and } \rho(x|xy) = \frac{\gamma_x}{\gamma_x + \gamma_y} = \frac{1}{3},$$

[7] Note that it is never revealed from the choice data that $x \perp y$ for any x and y. Suppose to the contrary that it is revealed from the choice data that $x \perp y$ for any x and y. For any alternative set A that contains alternatives z and w, we have $\beta(z, w|A) = \beta(z, w|zw) = \frac{\gamma_z}{\gamma_w}$. Therefore, z and w are (revealed to be) associated with each other. We arrive at a contradiction.

exhibiting attraction effect.

The following proposition provides a necessary and sufficient condition for the attraction effect.

Proposition 1. *The following two statements are equivalent:*

i) $\rho(x|xyz) > \rho(x|xy)$*;*

ii) x *and* z *are in the same block,* y *is in a different block, and* $\frac{v_x}{v_z} > \frac{\gamma_x}{\gamma_z}\frac{\gamma_x+\gamma_y+\gamma_z}{\gamma_y}$*.*

Proposition 1 summarizes the underlying logic for how the associationistic Luce rule can account for the attraction effect. Suppose that x and z are associated with each other, but x and y are not associated with each other. On the one hand, the addition of z decreases the probability that the decision maker is first attracted by x and y. On the other hand, when the decision maker is first attracted by z, she still considers x, but not y, when she eventually makes a choice. Therefore, the probability of y being chosen necessarily decreases. If x has a sufficiently high Luce value relative to z, then the addition of z increases the probability of x being chosen, exhibiting attraction effect.

The next example builds upon Example 1 from the introduction.

Example 5 (Continued from Example 1)**.** *Recall that alternatives* x, z *and* w *are different variants of the same good where* x *is high quality and expensive good,* z *is a more expensive version of* x*. Alternative* y *is of a different brand (low quality and cheap good). We might have* $\rho(x|xyz) > \rho(x|xy)$*, exhibiting attraction effect. Now we introduce another alternative* w*, which is an even more expensive version of* z*. Again, we might have* $\rho(z|yzw) > \rho(z|yz)$*, exhibiting attraction effect.*

In Example 5, the addition of z to xy "helps" increase the probability of x being chosen, and the addition of w to yz "helps" increase the probability of z being chosen. In this case, it seems natural that the addition of w to xy should "help" increase the probability of x being chosen. We introduce the following property that we call *attraction transitivity.*

Definition 4 (Attraction transitivity)**.** *For any* $x, y, z, w \in X, \rho(x|xyz) > \rho(x|xy)$ *and* $\rho(z|yzw) > \rho(z|yz)$ *imply* $\rho(x|xyw) > \rho(x|xy)$*.*

It is easy to see that the our model satisfies attraction transitivity.

Proposition 2. *The associationistic Luce rule satisfies attraction transitivity.*

5.2 Stochastic Intransitivity

Several psychologists, starting from Tversky (1969), have argued that choices may fail to be transitive. When choice is stochastic, there are many ways to define analogues of transitive behavior in deterministic models. A weak such analogue is the following:

Definition 5 (Weak Stochastic Transitivity)**.** *For any $x, y, z \in X$, $\rho(x|xy) \geq \frac{1}{2}$ and $\rho(y|yz) \geq \frac{1}{2}$ imply $\rho(x|xz) \geq \frac{1}{2}$.*

Rieskamp, Busemeyer, and Mellers (2006) provide a detailed review of evidence suggesting that choice rules may violate this property. Example 6 below illustrates how the associationistic Luce rule can account for violations of weak stochastic transitivity, and thus of the stronger version:

Definition 6 (Strong Stochastic Transitivity)**.** *For any $x, y, z \in X$, $\rho(x|xy) \geq \frac{1}{2}$ and $\rho(y|yz) \geq \frac{1}{2}$ imply $\rho(x|xz) \geq \max\{\rho(x|xy), \rho(y|yz)\}$.*

Example 6. *Let $v_x > v_y > v_z$ and $\gamma_y > \gamma_z > \gamma_x$, where x and y are in the same block, z is in a different block. We have*

$$\rho(x|xy) = \frac{v_x}{v_x + v_y} > \frac{1}{2},$$
$$\rho(y|yz) = \frac{\gamma_y}{\gamma_y + \gamma_z} > \frac{1}{2},$$

but also

$$\rho(x|xz) = \frac{\gamma_x}{\gamma_x + \gamma_z} < \frac{1}{2},$$

violating weak stochastic transitivity.

In Example 6, since x and y are associated with each other, in the binary choice between x and y, no matter which alternative the decision maker is attracted by in the first stage, she considers both alternatives in the second stage. Since x is more desirable (higher Luce value), she chooses x more frequently than y. Since y and z are not associated with each other, in the binary choice between y and z, the decision maker chooses the alternative that attracts her in the first stage. Since y is more salient, she chooses y more frequently than z. By similar reasoning, in the binary choice between x and z, the decision maker chooses z more frequently than x.

The key for the violations of weak stochastic transitivity is that the ordering of the salience value γ and the ordering of the Luce value v are reversed for some pair of alternatives.

It is easy to check that if the ordering of the salience value γ weakly agrees with the ordering of the Luce value v,[8] the associationistic Luce rule satisfies weak stochastic transitivity.

6 Comparison to Related Models

In this section, we consider how the associationistic Luce rule compares to other models in the literature. Block and Marschak (1960) consider a class of stochastic choice functions known as random utility models. A random utility model is described by a probability measure over preference orderings, where the agent selects the maximal alternative available according to the randomly assigned preference ordering. The Luce rule is a special case of random utility models. Random utility models necessarily obey the regularity condition. Therefore, the associationistic Luce rule is not nested in random utility models.

Gul, Natenzon, and Pesendorfer (2014) introduce the attribute rule where they interpret alternatives as bundles of attributes. The decision maker first randomly chooses an attribute from all perceived attributes, and then randomly chooses an alternative containing the selected attribute. The attribute rule is a random utility model (they further show that the set of attribute rules and random utility maximizers are essentially the same). Therefore, the associationistic Luce rule and the attribute rule are not equivalent from the discussion of random utility models.

Manzini and Mariotti (2014) axiomatize the random consideration set rule, where the source of choice errors is the decision maker's failure to consider all feasible alternatives. While the preference over the alternatives is deterministic, each alternative is considered with a certain probability. Brady and Rehbeck (2016) examine the role of stochastic feasibility in consumer choice using a random conditional choice set rule (RCCSR).[9] They show that RCCSR generalizes the random consideration set rule of Manzini and Mariotti (2014). Both models are incompatible with the Luce model, as they necessarily violate Luce's IIA axiom when $A \in \mathcal{A}$ and $|A| \geq 3$, since the ratio of the probability of choosing the most preferred alternative over the probability of choosing the least preferred alternative necessarily decreases

[8] More formally, $(\gamma_x - \gamma_y)(v_x - v_y) \geq 0$ for any x and y. In words, this condition says that a more salient alternative is also more desirable.

[9] Manzini and Mariotti (2014) assume that the probabilities of each alternative being considered are independent, whereas Brady and Rehbeck (2016) model feasibility to permit correlation in availability of alternatives.

once the middle ranked alternative is removed from the alternative set. The associationistic Luce rule, in contrast, includes the Luce rule as a special case.

Echenique, Saito, and Tserenjigmid (2014) develop the perception-adjusted Luce model (PALM). In this model, a decision maker is described by a weak order and a utility function. She perceives each element of the alternative set sequentially according to the perception priority. Each perceived alternative is chosen with probability described by a function that depends on utility according to a Luce-type formula. The perception priority partitions the alternative set, where each block contains alternatives with the same perception priority. Note that two alternatives are associated with each other in our model, if and only if the two alternatives have the same perception priority in the PALM. Consider any alternative set xy, where x and y are not associated with each other. Adding an alternative z which is neither associated with x nor y will not change the relative probability of choosing x and y in our model. However, the perception priority of z is not well defined in the PALM; see Echenique, Saito, and Tserenjigmid (2014, Section 3). Therefore, the associationistic Luce rule is not nested in the PALM. Lastly, we note that there are models which are both a PALM and an associationistic Luce rule, such as the Luce model.

A Appendix

A.1 Preparatory Lemmas

This subsection contains several lemmas that are used to prove Theorem 1 and Theorem 2. Lemmas 1 - 4 are purely arithmetic.

Lemma 1. *If* $a, b, c, d, e, f \in (0, 1)$ *and*

$$\begin{aligned} a + b + c &= 1 \\ \frac{1-a}{a} &= \frac{d}{1-d} + \frac{1-e}{e}; \\ \frac{1-b}{b} &= \frac{e}{1-e} + \frac{1-f}{f}; \\ \frac{1-c}{c} &= \frac{f}{1-f} + \frac{1-d}{d}, \end{aligned}$$

then

$$\frac{a}{b} = \frac{e}{1-e}.$$

Proof. Rearranging the three equations, we have

$$\frac{a+b}{1-a-b} = \frac{1}{\frac{1-a}{a} - \frac{1-e}{c}} + \frac{1}{\frac{1-b}{b} - \frac{e}{1-e}},$$

and

$$\begin{aligned} & a^2e^2 + 2abe^2 + b^2e^2 - 2a^2e - 2abe + a^2 \\ & = (ae+be)^2 - 2a(ae+be) + a^2 \\ & = (ae+be-a)^2 \\ & = 0. \end{aligned}$$

Therefore,

$$\frac{a}{b} = \frac{e}{1-e}.$$

□

Lemma 2. *If* $a, b, c, d > 0$, *and*

$$(\frac{1}{a+b} + \frac{1}{c+d})(\frac{1}{\frac{1}{a}+\frac{1}{c}} + \frac{1}{\frac{1}{b}+\frac{1}{d}}) = 1,$$

then

$$ad = bc.$$

Proof. Rearranging gives

$$\begin{aligned} & (a+b+c+d)(abc+acd+abd+bcd) \\ & = (a+b)(c+d)(a+c)(b+d). \end{aligned}$$

Simplifying further, we have

$$a^2d^2 - 2abcd + b^2c^2 = (ad-bc)^2 = 0$$

and $ad = bc$. □

We say that an $n \times n$ matrix M is symmetrically reciprocal, denoted by SR, if $a_{ii} = 1$ for all $i = 1, 2, \ldots, n$ and $a_{ij} = \frac{1}{a_{ji}}$ for all $i, j = 1, 2, \ldots, n$ and $i \neq j$. We say that an $n \times n$ matrix M is transitive if the matrix M is entrywise positive and $a_{ij} = a_{ik}a_{kj}$ for all $i, j, k = 1, 2, \ldots, n$.

Lemma 3. *For any SR and transitive matrix $M = (a_{ij})_{n\times n}$, there exists $w_i, i = 1, 2, \ldots, n$ such that*

$$a_{ij} = \frac{w_i}{w_j} \text{ and } \sum_{i=1}^{n} w_i = 1.$$

Proof. We explicitly construct w_i for $i = 1, 2, \ldots, n$. Let

$$w_i = \frac{a_{i1}}{\sum_{j=1}^{n} a_{j1}}, \forall i = 1, 2, \ldots, n.$$

It is easy to see that $a_{i1} = \frac{w_i}{w_1}$. Since the matrix M is SR, we have $a_{1j} = \frac{w_1}{w_j}$. Since the matrix M is transitive, we have that

$$a_{ij} = a_{i1}a_{1j} = \frac{w_i}{w_1}\frac{w_1}{w_j} = \frac{w_i}{w_j}.$$

□

Let $N = (1, 1, \ldots, 1)^T$ denote the $n \times 1$ vector of $1'$s.

Lemma 4. *For any SR and transitive matrices $M = (a_{ij})_{n\times n}$ and $\bar{M} = (\bar{a}_{ij})_{n\times n}$, if $N'(M - \bar{M}) = 0$, we have $a_{ij} = \bar{a}_{ij}$ for all $i, j = 1, 2, \ldots, n$.*

Proof. By Lemma 3, we can choose w_i for M and $\bar{w}_i$ for $\bar{M}$ such that

$$a_{ij} = \frac{w_i}{w_j}, \bar{a}_{ij} = \frac{\bar{w}_i}{\bar{w}_j}, \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} \bar{w}_i = 1.$$

Since $N'(M_1 - M_2) = 0$, we have

$$\begin{aligned}
(N'(M - \bar{M}))_j &= (N'M)_j - (N'\bar{M})_j \\
&= \sum_{k=1}^{n} (a_{kj} - \bar{a}_{kj}) \\
&= \sum_{k=1}^{n} (\frac{w_k}{w_j} - \frac{\bar{w}_k}{\bar{w}_j}) \\
&= \frac{1}{w_j} - \frac{1}{\bar{w}_j} \\
&= 0.
\end{aligned}$$

Therefore, $w_j = \bar{w}_j$ and $a_{ij} = \bar{a}_{ij}$ for all $i, j = 1, 2, \ldots, n$. □

A.2 Proof of Theorem 1

It is easy to see that the Luce rule satisfies Axiom ARP. For the other direction, suppose that a choice rule ρ satisfies Axiom ARP. For pairwise disjoint alternative sets $A, B, C \in \mathcal{A}_+$, we have

$$\beta(AB, C|ABC) = \beta(A, C|AC) + \beta(B, C|BC);$$
$$\beta(AC, B|ABC) = \beta(A, B|AB) + \beta(C, B|CB);$$
$$\beta(BC, A|ABC) = \beta(B, A|AB) + \beta(C, A|AC).$$

By Lemma 1, we have[10]

$$\beta(A, B|ABC) = \beta(A, B|AB). \tag{1}$$

Since (1) holds for all pairwise disjoint alternative sets $A, B, C \in \mathcal{A}_+$, we conclude that the choice rule ρ is a Luce rule.

A.3 Proof of Theorem 2

We first show that if a choice rule ρ satisfies Axiom T and Axiom Weak ARP*, then it is an associationistic Luce rule. Axiom T implies that we can partition the set X in the following way: $\{P_i\}_{i=1}^n$, such that i) $P_i \in \mathcal{L}$, for any $i = 1, 2, \ldots, n$; and ii) $P_i \perp P_j$ for any $i, j = 1, 2, \ldots, n$ and $i \neq j$. Without loss of generality, let us assume that there are at least two blocks in the partition.[11] Consider any alternative set $A \in \mathcal{A}_+$ and alternative $x \in A$. Let $A_i = A \cap X_i$, we have $A = \cup_{i=1}^n (A \cap X_i) = A_1 A_2 \cdots A_n$. It is easy to see that i) $A_i \in \mathcal{L}$, for any $i = 1, 2, \ldots, n$; and ii) $A_i \perp A_j$ for any $i, j = 1, 2, \ldots, n$ and $i \neq j$. Without loss of generality, we assume that $x \in A_1$.

The logic of the proof for this direction is summarized as follows. In Step i), we show that the choice data satisfies *independence of irrelevant blocks*; that is $\beta(A_i, A_1|A) = \beta(A_i, A_1|A_1 A_i)$. Step ii) constructs the salience value γ, and Step iii) constructs the Luce value v. Lastly, Step iv) verifies that the choice rule is an associationistic Luce rule.

Step i) Independence of Irrelevant Blocks: We first show that $\beta(A_i, A_1|A) = \beta(A_i, A_1|A_1 A_i)$ in a series of lemmas.

[10] Let $\rho(A|ABC) = a, \rho(B|ABC) = b, \rho(C|ABC) = c, \rho(C|AC) = d, \rho(A|AB) = e$, and $\rho(B|BC) = f$.

[11] If $X \in \mathcal{L}$, our model reduces to the Luce rule.

Lemma 5. *For distinct blocks A_i, A_j, A_k, we have $\beta(A_i, A_j|A_iA_jA_k) = \beta(A_i, A_j|A_iA_j)$.*

Proof. Axiom Weak ARP* implies that

$$\beta(A_iA_j, A_k|A_iA_jA_k) = \beta(A_i, A_k|A_iA_k) + \beta(A_j, A_k|A_jA_k);$$
$$\beta(A_iA_k, A_j|A_iA_jA_k) = \beta(A_i, A_j|A_iA_j) + \beta(A_k, A_j|A_jA_k);$$
$$\beta(A_jA_k, A_i|A_iA_jA_k) = \beta(A_j, A_i|A_iA_j) + \beta(A_k, A_i|A_iA_k).$$

By Lemma 1, we have[12]

$$\beta(A_i, A_j|A_iA_jA_k) = \beta(A_i, A_j|A_iA_j).$$

□

Lemma 6. *For distinct blocks A_i, A_j, A_k, we have $\beta(A_i, A_j|A_iA_j)\beta(A_j, A_k|A_jA_k) = \beta(A_i, A_k|A_iA_k)$.*

Proof. By the definition of β, we have

$$\beta(A_i, A_k|A_iA_jA_k) = \beta(A_i, A_j|A_iA_jA_k)\beta(A_j, A_k|A_iA_jA_k).$$

By Lemma 5, we have

$$\beta(A_i, A_j|A_iA_jA_k) = \beta(A_i, A_j|A_iA_j),$$
$$\beta(A_j, A_k|A_iA_jA_k) = \beta(A_j, A_k|A_jA_k),$$
$$\beta(A_i, A_k|A_iA_jA_k) = \beta(A_i, A_k|A_iA_k).$$

Therefore,

$$\beta(A_i, A_k|A_iA_k) = \beta(A_i, A_j|A_iA_j)\beta(A_j, A_k|A_jA_k).$$

□

Lemma 7. $\beta(A_i, A_j|A_iA_j) = \beta(A_i, A_j|A)$.

[12] Let $\rho(A_i|A_iA_jA_k) = a, \rho(A_j|A_iA_jA_k) = b, \rho(A_k|A_iA_jA_k) = c, \rho(A_k|A_iA_k) = d, \rho(A_i|A_iA_j) = e$, and $\rho(A_j|A_jA_k) = f$.

Proof. Repeatedly applying Axiom Weak ARP*, we have

$$\beta(A_{-i}, A_i|A) = \sum_{j:j\neq i} \beta(A_j, A_i|A_iA_j). \tag{2}$$

By the definition of β,

$$\beta(A_{-i}, A_i|A) = \sum_{j:j\neq i} \beta(A_j, A_i|A). \tag{3}$$

Therefore, it follows from (2) and (3) that

$$\sum_{j:j\neq i} \beta(A_j, A_i|A_iA_j) = \sum_{j:j\neq i} \beta(A_j, A_i|A). \tag{4}$$

Now let $a_{ji} = \beta(A_j, A_i|A_iA_j)$, and $\bar{a}_{ji} = \beta(A_j, A_i|A)$ if $j \neq i$. Let $a_{ii} = \bar{a}_{ii} = 1$. Therefore, it follows from the (4) that

$$\sum_j a_{ji} = \sum_j \bar{a}_{ji}. \tag{5}$$

Furthermore, by Lemma 6, $a_{ji}a_{ik} = a_{jk}$. By the definition of β, $\bar{a}_{ji}\bar{a}_{ik} = \bar{a}_{jk}$. By Lemma 4, we have $a_{ij} = \bar{a}_{ij}$. That is, $\beta(A_i, A_j|A_iA_j) = \beta(A_i, A_j|A)$. □

Step ii) Construction of the Salience Value γ: Fix an arbitrary $\bar{x} \in P_1$, and let $\gamma(\bar{x}) = 1$. Define $\gamma(y) = \gamma(\bar{x})\beta(y, \bar{x}|\bar{x}y) = \beta(y, \bar{x}|\bar{x}y)$ for any $y \notin P_1$. Fix an arbitrary $\bar{y} \in X_2$. Define $\gamma(x) = \gamma(\bar{y})\beta(x, \bar{y}|x\bar{y}) = \beta(\bar{y}, \bar{x}|\bar{x}\bar{y})\beta(x, \bar{y}|x\bar{y})$ for any $x \in X_1$ and $x \neq \bar{x}$.

Lemma 8. *If $\tilde{x} \in P_i$ and $\tilde{y} \in P_j$, $i \neq j$, then $\beta(\tilde{x}, \tilde{y}|\tilde{x}\tilde{y}) = \frac{\gamma(\tilde{x})}{\gamma(\tilde{y})}$.*

Proof. By symmetry, it suffices to consider the following cases.

Case i) If $\tilde{x} \notin P_1$ and $\tilde{y} \notin P_1$, then

$$\begin{aligned}
\beta(\tilde{x}, \tilde{y}|\tilde{x}\tilde{y}) &= \beta(\tilde{x}, \tilde{y}|\bar{x}\tilde{x}\tilde{y}) \\
&= \beta(\tilde{x}, \bar{x}|\bar{x}\tilde{x}\tilde{y})\beta(\bar{x}, \tilde{y}|\bar{x}\tilde{x}\tilde{y}) \\
&= \beta(\tilde{x}, \bar{x}|\tilde{x}\bar{x})\beta(\bar{x}, \tilde{y}|\bar{x}\tilde{y}) \\
&= \frac{\gamma(\tilde{x})}{\gamma(\tilde{y})}.
\end{aligned}$$

Case ii) If $\tilde{x} = \bar{x} \in P_1$ and $\tilde{y} \notin P_1$, by construction, $\beta(\bar{x}, \tilde{y}|\bar{x}\tilde{y}) = \frac{1}{\beta(\tilde{y}, \bar{x}|\bar{x}\tilde{y})} = \frac{\gamma(\bar{x})}{\gamma(\tilde{y})}$.

Case iii) If $\tilde{x} \in P_1$, $\tilde{x} \neq \bar{x}$ and $\tilde{y} \notin P_2$, then

$$\beta(\tilde{x}, \tilde{y}|\tilde{x}\tilde{y}) = \beta(\tilde{x}, \tilde{y}|\tilde{x}\bar{y}\tilde{y})$$

$$
\begin{aligned}
&= \beta(\tilde{x}, \bar{y}|\tilde{x}\tilde{y}\bar{y})\beta(\bar{y}, \tilde{y}|\tilde{x}\tilde{y}\bar{y}) \\
&= \beta(\tilde{x}, \bar{y}|\tilde{x}\bar{y})\beta(\bar{y}, \tilde{y}|\tilde{y}\bar{y}) \\
&= \frac{\gamma(\tilde{x})}{\gamma(\bar{y})}\frac{\gamma(\bar{y})}{\gamma(\tilde{y})} \\
&= \frac{\gamma(\tilde{x})}{\gamma(\tilde{y})}.
\end{aligned}
$$

Case iv) If $\tilde{x} \in P_1, \tilde{x} \neq \bar{x}$ and $\tilde{y} = \bar{y} \in P_2$, then by construction, $\beta(\tilde{x}, \bar{y}) = \frac{\beta(\tilde{x},\bar{y})\gamma(\bar{y})}{\gamma(\bar{y})} = \frac{\gamma(\tilde{x})}{\gamma(\bar{y})}$.

Case v) If $\tilde{x} \in X_1, \tilde{x} \neq \bar{x}$ and $\tilde{y} \in X_2, \tilde{y} \neq \bar{y}$. Repeatedly applying Axiom WA, we have

$$
\begin{aligned}
\beta(\tilde{x}\bar{x}, \tilde{y}\bar{y}|\tilde{x}\bar{x}\tilde{y}\bar{y}) &= \beta(\tilde{x}, \tilde{y}\bar{y}|\tilde{x}\tilde{y}\bar{y}) + \beta(\bar{x}, \tilde{y}\bar{y}|\bar{x}\tilde{y}\bar{y}) \\
&= \frac{1}{\beta(\tilde{y}\bar{y}, \tilde{x}|\tilde{x}\tilde{y}\bar{y})} + \frac{1}{\beta(\tilde{y}\bar{y}, \bar{x}|\bar{x}\tilde{y}\bar{y})} \\
&= \frac{1}{\beta(\tilde{y}, \tilde{x}|\tilde{x}\tilde{y}) + \beta(\bar{y}, \tilde{x}|\tilde{x}\bar{y})} + \frac{1}{\beta(\tilde{y}, \bar{x}|\bar{x}\tilde{y}) + \beta(\bar{y}, \bar{x}|\bar{x}\bar{y})}.
\end{aligned}
$$

Similarly,

$$
\beta(\tilde{y}\bar{y}, \tilde{x}\bar{x}|\tilde{x}\bar{x}\tilde{y}\bar{y}) = \frac{1}{\beta(\tilde{x}, \tilde{y}|\tilde{x}\tilde{y}) + \beta(\bar{x}, \tilde{y}|\bar{x}\tilde{y})} + \frac{1}{\beta(\tilde{x}, \bar{y}|\tilde{x}\bar{y}) + \beta(\bar{x}, \bar{y}|\bar{x}\bar{y})}.
$$

Since

$$
\beta(\tilde{x}\bar{x}, \tilde{y}\bar{y}|\tilde{x}\bar{x}\tilde{y}\bar{y})\beta(\tilde{y}\bar{y}, \tilde{x}\bar{x}|\tilde{x}\bar{x}\tilde{y}\bar{y}) = 1,
$$

by Lemma 2, we have

$$
\beta(\tilde{y}, \tilde{x}|\tilde{x}\tilde{y})\beta(\bar{y}, \bar{x}|\bar{x}\bar{y}) = \beta(\bar{y}, \tilde{x}|\tilde{x}\bar{y})\beta(\tilde{y}, \bar{x}|\bar{x}\tilde{y}).
$$

Therefore,

$$
\begin{aligned}
\beta(\tilde{x}, \tilde{y}|\tilde{x}\tilde{y}) &= \beta(\tilde{x}, \bar{y}|\tilde{x}\bar{y})\beta(\bar{x}, \tilde{y}|\bar{x}\tilde{y})\beta(\bar{y}, \bar{x}|\bar{x}\bar{y}) \\
&= \frac{\gamma(\tilde{x})}{\gamma(\bar{y})}\frac{\gamma(\bar{x})}{\gamma(\tilde{y})}\frac{\gamma(\bar{y})}{\gamma(\bar{x})} \\
&= \frac{\gamma(\tilde{x})}{\gamma(\tilde{y})}.
\end{aligned}
$$

□

Lemma 9. *If $A \subseteq P_i, B \subseteq P_j, i \neq j$, then*

$$
\beta(A, B|AB) = \frac{\gamma(A)}{\gamma(B)}.
$$

Proof. Repeatedly applying Axiom Weak ARP*, we have

$$\begin{aligned}
\beta(A, B|AB) &= \sum_{x\in A} \beta(x, B|xB) \\
&= \sum_{x\in A} \frac{1}{\beta(B, x|xB)} \\
&= \sum_{x\in A} \frac{1}{\sum_{y\in B} \beta(y, x|xy)} \\
&= \sum_{x\in A} \frac{1}{\sum_{y\in B} \frac{\gamma(y)}{\gamma(x)}} \\
&= \frac{\gamma(A)}{\gamma(B)}.
\end{aligned}$$

□

Step iii) Construction of the Luce Value v: For each block P_i, fix an arbitrary alternative $\hat{x}_i \in X_i$, and let $v(\hat{x}_i) = 1$. Define $v(x_i) = \beta(x_i, \hat{x}_i|x_i\hat{x}_i)$ for any $x_i \in X_i$ and $x_i \neq \hat{x}_i$.

Lemma 10. *If* $x_i, x'_i \in P_i$ *and* $x_i \neq x'_i$*, then* $\beta(x_i, x'_i|x_ix'_i) = \frac{v(x_i)}{v(x'_i)}$.

Proof. By construction, $v(x_i) = \beta(x_i, \hat{x}_i|x_i\hat{x}_i)$ and $v(x'_i) = \beta(x'_i, \hat{x}_i|x'_i\hat{x}_i)$. Therefore,

$$\begin{aligned}
\beta(x_i, x'_i|x_ix'_i) &= \beta(x_i, x'_i|x_ix'_i\hat{x}_i) \\
&= \beta(x_i, \hat{x}_i|x_ix'_i\hat{x}_i)\beta(\hat{x}_i, x'_i|x_ix'_i\hat{x}_i) \\
&= \beta(x_i, \hat{x}_i|x_i\hat{x}_i)\beta(\hat{x}_i, x'_i|\hat{x}_ix'_i) \\
&= \frac{v(x_i)}{v(x'_i)}.
\end{aligned}$$

□

Step iv) Verification: We verify that the choice rule is an associationistic Luce rule.

$$\begin{aligned}
\rho(x|A) &= \rho(A_1|A)\beta(x, A_1|A) \\
&= \rho(A_1|A)\beta(x, A_1|A_1) \\
&= \rho(A_1|A)\rho(x|A_1) \\
&= \frac{\rho(A_1|A)}{\sum_{i=1}^n \rho(A_i|A)}\rho(x|A_1)
\end{aligned}$$

$$= \frac{1}{\sum_{i=1}^{n} \beta(A_i, A_1|A)} \rho(x|A_1)$$
$$= \frac{1}{\sum_{i=2}^{n} \beta(A_i, A_1|A_1A_i) + 1} \rho(x|A_1). \tag{6}$$

where the second equality follows from $x \in A_1 \in \mathcal{L}$, and the last equality follows from independence of irrelevant blocks.

Following from (6) and Lemma 9,

$$\rho(x|A) = \frac{1}{\sum_{i=2}^{n} \frac{\gamma(A_i)}{\gamma(A_1)} + 1} \rho(x|A_1)$$
$$= \frac{\gamma(A_1)}{\gamma(A)} \rho(x|A_1). \tag{7}$$

By Lemma 10,

$$\rho(x|A_1) = \frac{1}{\beta(A_1, x|A_1)}$$
$$= \frac{1}{\sum_{y \in A_1} \beta(y, x|A_1)}$$
$$= \frac{1}{\sum_{y \in A_1} \beta(y, x|xy)}$$
$$= \frac{1}{\sum_{y \in A_1} \frac{v(y)}{v(x)}}$$
$$= \frac{v(x)}{v(A_1)}.$$

Continuing from (7),

$$\rho(x|A) = \frac{\gamma(A_1)}{\gamma(A)} \frac{v(x)}{v(A_1)}.$$

This concludes the proof for this direction.

Next we show that if a choice rule is an associationistic Luce rule, it satisfies Axiom T and Axiom Weak ARP*. If a choice rule ρ is an associationistic Luce rule, there exists $((P_i)_{i=1}^{n}, \gamma, v)$ such that

$$\rho(x|A) = \frac{\gamma(A \cap P_x)}{\gamma(A)} \frac{v(x)}{v(A \cap P_x)} \tag{a}$$

whenever $x \in A \in \mathcal{A}_+$.

Axiom Weak ARP*. If $A \cap B = \emptyset$ and $AB \perp C$, where $C \in \mathcal{L}$,

$$\beta(AB, C|ABC) = \frac{\frac{\gamma(AB)}{\gamma(AB)+\gamma(C)}}{\frac{\gamma(C)}{\gamma(AB)+\gamma(C)}}$$

$$= \frac{\frac{\gamma(A)}{\gamma(A)+\gamma(C)}}{\frac{\gamma(C)}{\gamma(A)+\gamma(C)}} + \frac{\frac{\gamma(B)}{\gamma(B)+\gamma(C)}}{\frac{\gamma(C)}{\gamma(B)+\gamma(C)}}$$

$$= \beta(A, C|AC) + \beta(B, C|BC).$$

Axiom T. In what follows, we show that the choice rule ρ satisfies Axiom T. Consider the following coarsening of the partition $(P_i)_{i=1}^n$. For each block P_i, if $|P_i| > 1$, P_i remains a block of the new partition. Let B_0 denote the union of the blocks whose cardinality is one; that is, $B_0 = \cup_{j:|P_j|=1} P_j$. After such coarsening, we relabel the blocks of the new partition as $B_0, B_1, B_2, ..., B_m$.

Consider the new partition $(B_i)_{i=0}^m$. Now for any block $B_i, i \geq 1$, if for all $x, y \in B_i$,

$$\frac{v(x)}{v(y)} = \frac{\gamma(x)}{\gamma(y)}, \tag{8}$$

we union this block with B_0. Denote by C_0 the union of all such blocks together with B_0. After such coarsening, we relabel the blocks of new partition as $C_0, C_1, C_2, ..., C_k$. For any $x \in C_0$, we define $\gamma'(x) = v'(x) = \gamma(x)$. Otherwise, define $\gamma'(x) = \gamma(x)$ and $v'(x) = v(x)$.

We show that $((C_i)_{i=0}^k, \gamma', v')$ also represents the choice rule ρ. Take any arbitrary x and A containing x, we need to show that

$$\rho(x|A) = \frac{\gamma'(A \cap C_x)}{\gamma'(A)} \frac{v'(x)}{v'(A \cap C_x)}. \tag{9}$$

If $x \in C_i$ for some $i \geq 1$, it is easy to see that equality (9) holds. If $x \in C_0$, then

$$\begin{aligned}
\frac{\gamma'(A \cap C_x)}{\gamma'(A)} \frac{v'(x)}{v'(A \cap C_x)} &= \frac{\gamma(A \cap C_x)}{\gamma(A)} \frac{\gamma(x)}{\gamma(A \cap C_x)} \\
&= \frac{\gamma(A \cap P_x)}{\gamma(A)} \frac{\gamma(x)}{\gamma(A \cap P_x)} \\
&= \frac{\gamma(A \cap P_x)}{\gamma(A)} \frac{v(x)}{v(A \cap P_x)} \\
&= \rho(x|A),
\end{aligned}$$

where the first equality follows from definition of γ' and v', and the third equality follows from (8). Therefore, $((C_i)_{i=0}^k, \gamma', v')$ also represents the choice data.

Note that if $x, y \in C_i$, then $x \sim y$. Therefore, to show transitivity, it suffices to show that for any $x \in C_i$, whenever $x \sim y$ for some y, $y \in C_i$. Suppose not, there exists some $y \sim x$, and yet $y \in C_j$ for some $j \neq i$. Note that either x or y belongs to C_k for some $k \geq 1$.

Without loss of generality, we assume $y \in C_k$ for some $k \geq 1$. Our construction ensures that there exists $z \in C_k$ such that $z \neq y$, and

$$\frac{v'(y)}{v'(z)} \neq \frac{\gamma'(y)}{\gamma'(z)}. \tag{10}$$

Therefore,

$$\begin{aligned}
\beta(x, y|xyz) &= \frac{\frac{\gamma'(x)}{\gamma'(x)+\gamma'(y)+\gamma'(z)} \frac{v'(x)}{v'(x)}}{\frac{\gamma'(y)+\gamma'(z)}{\gamma'(x)+\gamma'(y)+\gamma'(z)} \frac{v'(y)}{v'(y)+v'(z)}} \\
&= \frac{\gamma'(x)}{\gamma'(y)+\gamma'(z)} \frac{v'(y)+v'(z)}{v'(y)} \\
&\neq \frac{\gamma'(x)}{\gamma'(y)+\gamma'(z)} \frac{\gamma'(y)+\gamma'(z)}{\gamma'(y)} \\
&= \frac{\gamma'(x)}{\gamma'(y)} \\
&= \beta(x, y|xy),
\end{aligned}$$

where the inequality follows from (10). But since $y \sim x$, $\beta(x, y|xy) = \beta(x, y|xyz)$. We have a contradiction. This concludes the proof that Axiom T is satisfied.

A.4 Proof of Theorem 3

It follows from the proof of Theorem 2 (the necessity part) that we can construct $((C_t)_{t=0}^k, \gamma', v')$ that represents ρ such that $(C_t)_{t=0}^k$ is a coarsening of $(P_i)_{i=0}^n$. In particular, C_0 contains some blocks of $(P_i)_{i=0}^n$, and C_t is a block of the partition $(P_i)_{i=0}^n$ for $t = 1, 2, \ldots, k$. Furthermore, the salience value γ' and the Luce value v' satisfy that i) $\frac{\gamma'_x}{v'_x} = 1$ for all $x \in C_0$; and ii) $|C_t| \geq 2$, and there exist $x, y \in C_t$ such that $\frac{\gamma'_x}{v'_x} \neq \frac{\gamma'_y}{v'_y}$ for $t = 1, 2, \ldots, k$.

It follows that for any $x \in X \backslash C_0$, there exists y in the same block as x with $\frac{\gamma'_x}{v'_x} \neq \frac{\gamma'_y}{v'_y}$. For any z not in the same block as x, it follows from the proof of Theorem 2 (the necessity part) that $\beta(x, z|xyz) \neq \beta(x, z|xz)$. By the definition of S_ρ, $x \notin S_\rho$. By construction, $C_0 \subset S_\rho$. Therefore, $C_0 = S_\rho$.

Suppose that $((P'_j)_{j=1}^m, \gamma', v')$ represents ρ. It follows that $(P'_j)_{j=1}^m$ must be a refinement of $(C_t)_{t=0}^k$. We want to show that C_t is a block of $(P'_j)_{j=1}^m$ for $t \in \{1, 2, ..., k\}$. Suppose to the contrary, there exists $\hat{t} \geq 1$ such that $C_{\hat{t}} = \cup_{s=1}^{\hat{s}} P'_{j_s}$ with $\hat{s} \geq 2$. For any $s \in \{1, 2, ..., \hat{s}\}$, $|P'_{j_s}| \geq 2$, and there exists $x, y \in P'_{j_s}$ such that $\frac{\gamma'_x}{v'_x} \neq \frac{\gamma'_y}{v'_y}$. However, it follows from the proof of Theorem 2 (the necessity part) that for all $s_1, s_2 \in \{1, 2, ..., \hat{s}\}$, $s_1 \neq s_2$, there exist $x, y \in P'_{j_{s_1}}$

and $z \in P'_{j_{s_2}}$ such that $\beta(x, z|xyz) \neq \beta(x, z|xz)$. This implies that x and z are not associated with each other. However, both x and z are contained in $C_{\hat{t}}$. We arrive at a contradiction. Therefore, C_t is a block of $(P'_j)_{j=1}^m$ for $t \in \{1, 2, ..., k\}$. This concludes the proof, since C_t is also a block of $(P_i)_{i=0}^n$ for $t \in \{1, 2, ..., k\}$, and $\cup_{t=1}^k C_t = X \backslash S_\rho$.

A.5 Missing Proof for Section 5

Proof of Proposition 1. *ii)* $\Rightarrow$ *i)* is straightforward. For *i)* $\Rightarrow$ *ii)*, it suffices to consider the following cases.

Case i) Suppose that x and y are in the same block, z is in a different block. In this case, we have a contradiction, since

$$\begin{aligned}\rho(x|xyz) &= \frac{\gamma_x + \gamma_y}{\gamma_x + \gamma_y + \gamma_z} \frac{v_x}{v_x + v_y} \\ &< \frac{v_x}{v_x + v_y} \\ &= \rho(x|xy).\end{aligned}$$

Case ii) Suppose that x and z are in the same block, y is in a different block. In this case,

$$\begin{aligned}&\rho(x|xyz) > \rho(x|xy) \\ &\quad\Rightarrow \frac{\gamma_x + \gamma_z}{\gamma_x + \gamma_y + \gamma_z} \frac{v_x}{v_x + v_z} > \frac{\gamma_x}{\gamma_x + \gamma_y} \\ &\quad\Rightarrow \frac{v_x}{v_z} > \frac{\gamma_x}{\gamma_z} \frac{\gamma_x + \gamma_y + \gamma_z}{\gamma_y}.\end{aligned}$$

Case iii) Suppose that y and z are in the same block, x is in a different block. In this case, we have a contradiction, since

$$\begin{aligned}\rho(x|xyz) &= \frac{\gamma_x}{\gamma_x + \gamma_y + \gamma_z} \\ &< \frac{\gamma_x}{\gamma_x + \gamma_y} \\ &= \rho(x|xy).\end{aligned}$$

Proof of Proposition 2. Suppose that the decision maker follows the associationistic Luce rule, $\rho(x|xyz) > \rho(x|xy)$, and $\rho(z|yzw) > \rho(z|yz)$. It follows from Proposition 1 that x, z and w are in the same block, y is in a different block, and

$$\frac{v_x}{\gamma_x} > \frac{v_z}{\gamma_z} \frac{\gamma_x + \gamma_y + \gamma_z}{\gamma_y},$$

$$\frac{v_z}{\gamma_z} > \frac{v_w}{\gamma_w}\frac{\gamma_y + \gamma_z + \gamma_w}{\gamma_y}.$$

Therefore,

$$\begin{aligned}\frac{v_x}{\gamma_x} &> \frac{v_w}{\gamma_w}\frac{\gamma_x + \gamma_y + \gamma_z}{\gamma_y}\frac{\gamma_y + \gamma_z + \gamma_w}{\gamma_y} \\ &> \frac{v_w}{\gamma_w}\frac{\gamma_x + \gamma_y + \gamma_w}{\gamma_y}.\end{aligned}$$

By Proposition 1, we have $\rho(x|xyw) > \rho(x|xy)$.

References

Block, H. D., and J. Marschak (1960): "Random Orderings and Stochastic Theories of Responses," in *Contributions to Probability and Statistics*, pp. 97–132. Stanford, CA: Stanford University Press.

Brady, R. L., and J. Rehbeck (2016): "Menu-Dependent Stochastic Feasibility," *Econometrica*, 84(3), 1203–1223.

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(3), 225–243.

Echenique, F., K. Saito, and G. Tserenjigmid (2014): "The Perception-Adjusted Luce Model," mimeo. California Institute of Technology.

Gul, F., P. Natenzon, and W. Pesendorfer (2014): "Random Choice as Behavioral Optimization," *Econometrica*, 82(5), 1873–1912.

Huber, J., J. W. Payne, and C. Puto (1982): "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis," *Journal of Consumer Research*, 9(1), 90–98.

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

Luce, R. D. (1959): *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Mandelbaum, E. (2016): "Associationist Theories of Thought," in *The Stanford Encyclopedia of Philosophy*, ed. by E. N. Zalta. Summer 2016 edn.

Manzini, P., and M. Mariotti (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82(3), 1153–1176.

Manzini, P., M. Mariotti, and L. Mittone (2010): "Choosing Monetary Sequences: Theory and Experimental Evidence," *Theory and Decision*, 69(3), 327–354.

McFadden, D. (2001): "Economic Choices," *American Economic Review*, 91(3), 351–378.

Rieskamp, J., J. R. Busemeyer, and B. A. Mellers (2006): "Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice," *Journal of Economic Literature*, 44(3), 631–661.

Simonson, I., and A. Tversky (1992): "Choice in Context: Tradeoff Contrast and Extremeness Aversion," *Journal of Marketing Research*, 29(3), 281–295.

Sippel, R. (1997): "An Experiment on the Pure Theory of Consumer's Behaviour," *Economic Journal*, 107(444), 1431–1444.

Tversky, A. (1969): "Intransitivity of Preferences," *Psychological Review*, 76(1), 31–48.

——— (1972): "Elimination by Aspects: A Theory of Choice," *Psychological Review*, 79(4), 281–299.

Yildiz, K. (2016): "List-Rationalizable Choice," *Theoretical Economics*, 11(2), 587–599.