RANDOM CHOICE AS BEHAVIORAL OPTIMIZATION

FARUK GUL
Princeton University, Princeton, NJ 08544, U.S.A.

PAULO NATENZON
Washington University in St. Louis, Saint Louis, MO 63130, U.S.A.

WOLFGANG PESENDORFER
Princeton University, Princeton, NJ 08544, U.S.A.

RANDOM CHOICE AS BEHAVIORAL OPTIMIZATION

By Faruk Gul, Paulo Natenzon, and Wolfgang Pesendorfer[1]

We develop an extension of Luce's random choice model to study violations of the weak axiom of revealed preference. We introduce the notion of a stochastic preference and show that it implies the Luce model. Then, to address well-known difficulties of the Luce model, we define the *attribute rule* and establish that the existence of a well-defined stochastic preference over attributes characterizes it. We prove that the set of attribute rules and random utility maximizers are essentially the same. Finally, we show that both the Luce and attribute rules have a unique consistent extension to dynamic problems.

Keywords: Luce rule, attribute rule, random utility, dynamic random choice.

1. INTRODUCTION

IN EXPERIMENTAL SETTINGS, subjects routinely violate the weak axiom of revealed preference.[2] Often, these violations occur in a manner inconsistent with any deterministic theory.[3] There are many reasons why a static utility maximization model might be inadequate when analyzing consumption choices across multiple periods: there may be income effects, dynamic effects such as intertemporal complementarities, preference for variety, unobserved changes in consumers' budgets or expectations, and so forth. The theory we present here aims to deal with weak axiom violations that occur even when none of these factors are present. Instead, we explore random choice as a theory of *behavioral optimization*, that is, not as a model of measurement error but as a model of a consumer whose rationality is constrained by behavioral limitations such as limited cognitive abilities or limited attention.

The Luce rule (Luce (1959)) is a well-known behavioral optimization model that retains the simplicity of a deterministic theory. Each option s has a Luce value, v_s, so that the probability of choosing s from a set A containing s is

$$\rho_s(A) := \frac{v_s}{\sum_{t \in A} v_t}.$$

We can interpret the Luce value as a measure of desirability: s is *stochastically preferred* to t if, for any set A that contains neither s nor t, the agent is more

---

[1]This research was supported by National Science Foundation Grant SES-1060073.

[2]See, for example, Sippel (1997) or Manzini, Mariotti, and Mittone (2010).

[3]In one of Sippel's (1997) experiments, 11 of 12 subjects did not make the same choice when the same choice set was presented again. In fact, empirical and experimental studies almost always interpret individual choice behavior as probabilistic. For example, Hey and Orme (1994), a standard reference for the study of individual choice under uncertainty, interpreted subjects' behavior as stochastic.

likely to choose s from $A \cup \{s\}$ than t from $A \cup \{t\}$. Luce values represent this stochastic preference: s is stochastically preferred to t if and only if $v_s \geq v_t$.

These attractive features notwithstanding, the empirical literature on random choice has documented systematic violations of the Luce model.[4] Debreu (1960) anticipated the best known such violation and identified the main shortcoming of Luce's model: consider two items s_1 and s_2 that are very similar (a yellow bus and a red bus) and a third dissimilar option t (a train). Then, it may be that each item is chosen with probability $1/2$ from every two-element subset of $\{s_1, s_2, t\}$, but t is chosen from $\{s_1, s_2, t\}$ more frequently than each of the other two options. It is easy to check that this behavior cannot be generated (nor approximated) by any Luce rule. The problem that Debreu's example identifies is more generally referred to as the "duplicates problem" in the discrete choice estimation literature.

Our model, the *attribute rule*, addresses the shortcomings of the Luce model but retains Luce's idea that choice is governed by desirability values. It does so by reinterpreting the choice objects as *bundles of attributes*. Attributes, or at least their relevance, are subjective; they are properties of the decision maker and not of the choice objects. A main contribution of this paper is to provide a model that derives the collection of relevant attributes from observed random choices.

To see how our model works and how it relates to the Luce model, let Z be the collection of attributes, let X_s be the set of attributes that s has, and let $X(A) = \bigcup_{s \in A} X_s$. Hence, $X(A)$ is the set of attributes represented in A. An *attribute value*, w, maps attributes to positive reals, while an *attribute intensity*, η, maps attribute and option pairs to natural numbers. Let $w(X) = \sum_{x \in X} w_x$ and $\eta^x(A) = \sum_{s \in A} \eta_s^x$. Then, the probability of choosing s from A (containing s) is

$$\rho_s(A) := \sum_{x \in X_s} \frac{w_x}{w(X(A))} \cdot \frac{\eta_s^x}{\eta^x(A)}.$$

Hence, in an attribute rule, the decision maker first chooses a relevant attribute according to a Luce-type formula and then picks one option that has that attribute according to another Luce-type formula. The attribute rule reduces to a Luce rule when no pair of alternatives shares a common attribute.

Let $\rho(A, A \cup C) = \sum_{s \in A} \rho_s(A \cup C)$ be the probability that the agent chooses an alternative in A. In the Luce model, if $\sum_{s \in A} v_s \geq \sum_{t \in B} v_t$, then $\rho(A, A \cup C) \geq \rho(B, B \cup C)$ for all C disjoint from A and B. Thus, the Luce model satisfies the following independence assumption: if $A \cup B$ and $C \cup D$ are disjoint, then

(I) $\rho(A, A \cup C) \geq \rho(B, B \cup C)$ implies $\rho(A, A \cup D) \geq \rho(B, B \cup D)$.

[4]See Luce (1977) for a discussion of a number of these empirical studies.

Our first result, Theorem 1, shows that in a rich setting, with sufficient variety in options and option sets, the Luce rule is the only rule satisfying (I). We can interpret (I) as a context-independence requirement: if the agent is more likely to reject alternatives in C in favor of alternatives in A than he is to reject alternatives in C in favor of alternatives in B, then this ranking of A and B is preserved when C is replaced with D. The attribute rule weakens independence by requiring that the sets C and D above *share no common attributes* with A and B. Theorem 2, our main result, shows that a rich[5] random choice rule is an attribute rule if and only if it satisfies this weak version of independence together with the requirement that duplicates can be removed without affecting the choice probabilities of the remaining alternatives. Thus, allowing for duplicates and attribute overlap is what distinguishes the attribute rule from the Luce rule.

Section 5 applies the attribute rule to dynamic choice problems. In dynamic choice, it is often the case that two distinct choices are consistent with the same outcome. We would expect the overlap in outcomes to affect dynamic behavior just as attribute-overlap affects static choice behavior. By identifying a choice's attributes as the *outcomes consistent with that choice*, the attribute model provides a natural framework for analyzing random choice in dynamic settings. In Theorem 4, we characterize the *recursive attribute rule*, an adaptation of our model to dynamic settings.

1.1. *Related Literature*

Block and Marschak (1960) showed that Luce rules are random utility maximizers. Holman and Marley (see Luce and Suppes (1965)) and McFadden (1978) used the Gumbel distribution to construct a random utility for the Luce model. Their construction facilitated the estimation of Luce values as a function of (observable) background parameters. Falmagne (1978) characterized the set of all random utility maximizers. In Theorem 3 below, we show that the Block and Marschak theorem extends to attribute rules: every attribute rule is a random utility maximizer. The converse is almost true: while there are some random utility maximizers that are not attribute rules, every random utility maximizer can be approximated by an attribute rule.

Fudenberg and Strzalecki (2013) characterized a sequence of Luce rules in a dynamic setting. Their model involved consumption in every period and accommodated time discounting. However, their model did not address the overlap of continuation problems that is the focus of our analysis.

In Section 3 below, we discuss the relation between the attribute rule and two existing approaches that deal with the duplicates problem: Tversky's (1972) elimination by aspects (EBA) model and the (cross-)nested logit models familiar from the discrete choice estimation literature (Ben-Akiva and Lerman (1985), Train (2009)).

---

[5]As before, richness is a requirement on the variety of potential choices and choice sets.

2. REVEALED STOCHASTIC PREFERENCE

Let S be a nonempty set of choice objects. A set, $\mathcal{A}$, of countable subsets of S is a *proper collection* if (i) $\{s\} \in \mathcal{A}$ for all $s \in A$, (ii) $A \subset B \in \mathcal{A}$ implies $A \in \mathcal{A}$, and (iii) $A, B \in \mathcal{A}$ implies $A \cup B \in \mathcal{A}$. One example of a proper collection is the set of all finite subsets of S.

To simplify the statements below, we use the following notational convention:

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

$$As := A \cup \{s\}.$$

Given any proper collection $\mathcal{A}$, let $\mathcal{A}_+ = \mathcal{A} \setminus \{\emptyset\}$. A function $\rho: \mathcal{A} \times \mathcal{A}_+ \to [0, 1]$ is a (random) *choice rule* if, for all $A \in \mathcal{A}_+$, $\rho(\cdot, A)$ is countably additive and

(rcr) $\rho(A, A) = 1.$

The equation (rcr) is the feasibility constraint; ρ must choose among options available in A. Countable additivity is the requirement that $\rho(\cdot, A)$ is a probability. When $\mathcal{A}$ is the collection of all finite subsets of A, countable additivity is just additivity. We write $\rho_s(A)$ rather than $\rho(\{s\}, A)$.

Independence, stated below, requires that the stochastic preference is complete. Formally, set A is *stochastically preferred* to set B if $\rho(A, AC) > \rho(B, BC)$ for all $C \in \mathcal{A}_+$ such that $AB \cap C = \emptyset$; the set A is *stochastically indifferent* to B if $\rho(A, AC) = \rho(B, BC)$ for $C \in \mathcal{A}_+$ such that $AB \cap C = \emptyset$. Thus, A is stochastically preferred to B if options in A are chosen more frequently from AC than options in B from BC.

INDEPENDENCE: $\rho(A, AC) \geq \rho(B, BC)$ *implies* $\rho(A, AD) \geq \rho(B, BD)$ *if* $C, D \in \mathcal{A}_+$ *and* $AB \cap CD = \emptyset$.

If independence holds, stochastic preference is a complete binary relation $\succcurlyeq_\rho$ on $\mathcal{A}$. The decision maker stochastically prefers A to B ($A \succ_\rho B$) if he is more likely to choose from A when faced with AC than he is to choose from B when confronting BC for *any* C that is disjoint from A and B.

Let $v: \mathcal{A} \to \mathbb{R}_{++}$ and $v_s := v(\{s\})$. Such a function v is a *Luce value* if it is countably additive. Hence, $v(\emptyset) = 0$ and, for all $A \in \mathcal{A}_+$,

$$v(A) = \sum_{s \in A} v_s.$$

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

(ℓ) $\rho_s(A) = \dfrac{v_s}{v(A)}$

whenever $s \in A \in \mathcal{A}$. We say that the Luce value v *represents* ρ if equation (ℓ) holds for all such s, A. Clearly, every Luce value represents a unique choice rule.

It is easy to see that Luce rules satisfy independence since

$$A \succcurlyeq_\rho B \quad \text{if and only if} \quad v(A) \geq v(B).$$

Hence, v represents the stochastic preference of the Luce rule.

Theorem 1, below, shows that in a setting with a sufficient variety of options and option sets, that is, in a rich setting, the Luce rule is the only choice rule that satisfies independence and, therefore, the only rule that admits a stochastic preference. Next, we state this richness condition.

RICHNESS: *For* $A \neq \emptyset$, C *and* $\delta \in (0, 1)$, *there is* B *such that* $B \cap C = \emptyset$ *and* $\rho(A, AB) = \delta$.

Richness requires that the probability of choosing something in A can be varied continuously in the interval $(0, 1)$ by pairing A with an appropriate collection of other alternatives (B). Moreover, there is enough variety that, for any given choice set C, we can choose the alternatives in B to be disjoint from C. This last condition is a technical requirement that is satisfied in all of the examples below. Together with independence, richness also ensures that every option is chosen with positive probability from every option set that contains it. Example 1, below, illustrates a rich setting.

EXAMPLE 1: Alternatives are differentiated along two dimensions, for example, speed (x) and comfort (y). Let $F \subset R_+$ be the set of feasible x, y combinations. Let F^* be a finite subset of the efficient frontier of F and assume that F^* contains at least one extreme option, $(0, y^*)$ or $(x^*, 0)$, and at least one nonextreme option (x, y) such that $x, y > 0$. The choice objects are lotteries π on F^* that yield each element $i \in F^*$ with strictly positive probability $\pi_i > 0$. The two dimensions are complements, so that the lottery π has the Luce value $v_\pi = \sum_F \pi_i x_i y_i$. Let $\mathcal{A}$ be the collection of finite subsets of such lotteries. It is easy to verify that this example satisfies richness.

Example 1 illustrates how a sparse setting can be enriched with the help of lotteries. Note that Example 1 works because the set F^* contains at least one alternative (a boundary alternative) which, by itself, would never be chosen. In this case, full support lotteries yield a rich setting.

THEOREM 1: *A rich choice rule satisfies independence if and only if it is a Luce rule*.

It is easy to verify that two Luce values $v, \hat{v}$ represent the same Luce rule if and only if $\hat{v} = \alpha v$ for some $\alpha > 0$. We interpret richness as a technical assumption analogous to small event continuity (P6) in Savage (1954). Both assumptions are idealizations that facilitate the calibration of probabilities. In a sparse setting, there may be choice rules that satisfy independence but are not Luce rules. However, the fact that independence is satisfied for those choice rules can be viewed as an artifact of the sparse setting.

Example 1 provides a rich setting with a continuum of alternatives. Example 2, below, illustrates a rich setting with a countable number of alternatives.

EXAMPLE 2: Let S be the set of all strictly positive rational numbers and let $\mathcal{A}$ be the collection of subsets of S that are summable. That is, $\mathcal{A} = \{A \subset S \mid \sum_{s \in A} s < \infty\}$. It is easy to verify that $\mathcal{A}$ is a proper collection. Let $v(A) = \sum_{s \in A} s$ and let ρ be the Luce rule that v represents. In Appendix, we demonstrate that ρ is rich.

3. ATTRIBUTES

The following is a slight modification of an example proposed by Debreu (1960).

EXAMPLE 3: Let $S = \{s_1, s_2, s_3, t\}$. Assume that s_1, s_2, s_3 are transparently similar options, for example, buses of three different colors, while t is a train. Let $A = \{s_1, s_2, s_3\}$ be the set of buses. If the agent stochastically prefers buses over trains but is unresponsive to color, then we might have

$$\rho_{s_i}(\{s_i, t\}) = 0.6 \quad \text{and} \quad \rho_{s_i}(At) = 0.2$$

for all i. That $\rho_{s_i}(\{s_i, t\}) > \rho_t(\{s_i, t\})$ suggests that every bus is stochastically preferred to the train, whereas $\rho_{s_i}(At) < \rho_t(At)$ suggests that the train is stochastically preferred to every bus. Thus, we have a violation of independence.

Debreu's example suggests that independence is violated because the decision maker has a stochastic preference over *attributes* and options that share attributes compete more for the decision maker's attention than those that do not share attributes.[6] In the above example, options in A all have the same single attribute and, as a result, the decision maker treats them as if they were a single option. The Luce model is inadequate for describing the resulting behavior.

In the following example, alternatives have multiple attributes.

---

[6]Tversky (1972) called this the similarity effect.

EXAMPLE 4: Let $S = \{r, s, t\}$ and assume that the options represent three different airlines that service two destinations. Option r services both destinations, while s services only the first and t only the second destination. The decision maker is stochastically indifferent between the two destinations and does not care about the airline. The following choice rule describes such a situation:

$$\rho_r(\{r, s\}) = 3/4; \quad \rho_r(\{r, t\}) = 3/4; \quad \rho_r(\{r, s, t\}) = 1/2.$$

It is easily verified that the choice rule above is not a Luce rule. (For any Luce rule, the first two choice probabilities imply that $\rho_r(\{r, s, t\}) = 3/5$.) Of the two attributes, r and s share the first while r and t share the second. Therefore, intuitively, the set $\{s, t\}$ is a duplicate of r.

Examples 3 and 4 suggest specific choices of attributes. In Example 3, it seems natural that every collection of buses is treated like a single alternative. In Example 4, the choice of airlines is the result of a more basic choice of destinations. In each case, we could preserve stochastic independence (and hence the Luce model) if we focused on *what is really driving the choice*. The difficulty is that the designation of attributes is rarely clear-cut. Even in the stylized examples above, the choice rule may be responsive to the buses' color or to other characteristics of the airlines. If that is the case, then buses are not exact duplicates and r is not an exact duplicate of $\{s, t\}$. Thus, a satisfactory model must also deal with less clear-cut examples.

A more basic challenge is that the designation of attributes is subjective and must be derived from behavior (i.e., the choice rule). If the decision maker cares about the color of the means of transportation, then the fact that buses share similar physical characteristics is irrelevant and, instead, what matters is the color-attribute. Thus, how the decision maker groups objects into duplicates cannot be decided based on physical characteristics of the objects. Therefore, duplicates and attributes are a property of the choice rule and not of the objects.

In Example 3, the yellow and red buses are duplicates because we can replace a yellow bus with a red bus in any option set without affecting the choice probabilities of remaining alternatives. The next definition extends this notion of a duplicate to option sets: A and B are *duplicates* if replacing A with B has no effect on the probabilities of choosing elements that are not in A or B.

DEFINITION: A, B are *duplicates* if $AB \cap C = \emptyset$ and $s \in C$ implies $\rho_s(AC) = \rho_s(BC)$.

We write $A \sim B$ if A is a duplicate of B. The relation $\sim$ is symmetric and reflexive. Next, we define the notion of overlap of two option sets: when A and B have elements in common, they overlap. Even if A and B have no elements in common, they overlap if there are duplicates of A and B that have

elements in common. In Example 4 above, the two-destination airline r and the one-destination airline s overlap because $B = \{s, t\}$ is a duplicate of r and $B \cap \{s\} \neq \emptyset$.

DEFINITION: $A, B \in \mathcal{A}$ are *non-overlapping* if $A \sim A'$, $B \sim B'$ implies $A' \cap B' = \emptyset$.

We write $A \perp B$ if A and B are non-overlapping. Our first substantive assumption says that duplicates are treated like a single option. Specifically, if A and B' are duplicates, then adding A to a choice set that contains B' does not alter the odds of choosing options that do not overlap with A.

ELIMINATION OF DUPLICATES: $A \sim B' \subset B \perp C$ *and* $s \in C$ *implies* $\rho_s(BC) = \rho_s(ABC)$.

As we illustrated in Example 3 above, duplicates may lead to violations of independence. Recall that independence requires that $\rho(A, AC) \geq \rho(B, BC)$ implies $\rho(A, AD) \geq \rho(B, BD)$ for $C \cap AB = D \cap AB = \emptyset$ and $C, D \in \mathcal{A}_+$. In Example 3, independence fails because D or C overlap with A or B. Weak independence, below, applies only if A and B have no overlap with C and D and allows it to fail otherwise.

WEAK INDEPENDENCE: $\rho(A, AC) \geq \rho(B, BC)$ *implies* $\rho(A, AD) \geq \rho(B, BD)$ *if* $C, D \in \mathcal{A}_+$ *and* $AB \perp CD$.

Theorem 2, below, shows that strongly rich[7] choice rules that satisfy elimination of duplicates and weak independence are *attribute rules*. A collection of attributes, a function that assigns each attribute a value, and a function that assigns to each object and attribute an intensity define an attribute system. Every attribute system represents a unique choice rule. Henceforth,

$$\mathcal{A} = \mathcal{A}_f,$$

where $\mathcal{A}_f$ is the (proper) collection of all finite subsets of S. Therefore, when we wish to be explicit about the domain of a choice rule ρ, we can simply write (ρ, S). Let $|A|$ be the cardinality of the set A.

An *attribute set* Z is an arbitrary index set and each element of Z is an attribute. An *attribute intensity* is a function $\eta : Z \times S \to \mathbb{N} \cup \{0\}$ that measures the degree to which s has attribute x. We write η_s^x rather than $\eta(x, s)$ and let $\eta^x(A) = \sum_{s \in A} \eta_s^x$. We say that option s has attribute x if $\eta_s^x > 0$. We assume that each object has a nonempty and finite set of attributes: for each $s \in S$ there exists $x \in Z$ such that $\eta_s^x > 0$ and $X_s = \{x \in Z \mid \eta_s^x > 0\}$ is finite.

---

[7]For a definition of strong richness, see below.

Let $X(A) := \{x \in Z \mid \eta^x(A) > 0\}$ for all $A \in \mathcal{A}_+$. We call $X(A)$ the set of attributes that are active in A; these are the attributes that at least one member of A possesses.

Throughout this section, we assume that $\{s \in S \mid \eta^x_s > 0\} = \{s \in S \mid \eta^y_s > 0\}$ implies $x = y$. Hence, without risk of confusion, we identify each attribute x with the set $\{s \in S \mid \eta^x_s > 0\}$ and write $s \in x$ to mean s has attribute x.[8]

An *attribute value* is a function $w : Z \to \mathbb{R}_{++}$ that measures the desirability of attribute x. Again, we write w_x rather than $w(x)$ and let $w(X) := \sum_{x \in X} w_x$ for all nonempty finite $X \subset Z$. We call (w, η) an *attribute system*.

We say that η is *simple* if $\eta^x_s = 0$ or 1 for all x, s. An object $s \in S$ is an *archetype* for $x \in Z$ if $\eta^y_s = 1$ if $y = x$ and 0 otherwise. An attribute system is *complete* if every attribute has multiple (i.e., at least two) archetypes. The choice rule ρ is a (complete) attribute rule if there exists a (complete) attribute system (w, η) such that

$$(1) \qquad \rho_s(A) = \sum_{x \in X(A)} \frac{w_x}{w(X(A))} \cdot \frac{\eta^x_s}{\eta^x(A)}.$$

We say that the attribute system (w, η) *represents* ρ if equation (1) holds for all $s \in A \in \mathcal{A}_+$. Clearly, every attribute system (w, η) on $\mathcal{A}$ represents a unique choice rule ρ.

EXAMPLE 5: To define an attribute rule that is consistent with Example 3 above, let $Z = \{x, y\}$, where $x = A$ is the bus-attribute and $y = \{t\}$ is the train-attribute. Let $w_x = 3$, $w_y = 2$ and let η be the simple intensity such that $\eta^x_s = 1$ if and only if $s \in x = A$ and $\eta^y_s = 1$ if and only if $s = t$. Then, $\rho_{s_i}(\{s_i, t\}) = 0.6$ and $\rho_{s_i}(At) = 0.2$, as required.

EXAMPLE 6: For an attribute rule that is consistent with Example 4, let $Z = \{x, y\}$, where $x = \{r, s\}$ and $y = \{r, t\}$. Set $w_x = w_y = 1$ and let η be the simple attribute intensity such that $\eta^x_r = \eta^x_s = 1$, $\eta^y_r = \eta^y_t = 1$. Then, $\rho_r(\{r, s\}) = \rho_r(\{r, t\}) = 3/4$ and $\rho_r(\{r, s, t\}) = 1/2$, as required.

Note that every Luce rule is an attribute rule. To see this, let v be a Luce value. Choose the attribute set $Z = S$ and set $w_s = v_s$ for all $s \in S$. Define the simple attribute intensity η such that $\eta^x_s = 1$ if and only if $x = s$. Then, (1) can be restated as follows:

$$\rho_s(B) = \frac{w_x}{w(X(B))} = \frac{v_s}{v(B)}.$$

---

[8]Theorem 2 shows that this entails no loss of generality in sufficiently rich environments. When S is finite, we permit $\{s \in S \mid \eta^x_s > 0\} = \{s \in S \mid \eta^z_s > 0\}$ for $x \neq z$ and with some abuse of notation, we still write $t \in x$ to mean $t \in \{s \in S \mid \eta^x_s > 0\}$.

Hence, if each object has an exclusive attribute, that is, if each s is an archetype and has no duplicates, the attribute rule reduces to the Luce model.

To establish our main result, we require that there is sufficient variety in the set of options and option sets. The richness assumption of Theorem 2 below strengthens the richness assumption of Theorem 1. First, we define *fine* option sets.

DEFINITION: B is *fine* if, for every partition $A_1 A_2 \cdots A_n = A \sim B$, there is a partition $B_1 B_2 \cdots B_n = B$ with $B_i \sim A_i$ for all i.

In Example 4 above, $\{s, t\}$ is a fine set, whereas its duplicate $\{r\}$ is not fine.[9] Let $\mathcal{M} = \{A \in \mathcal{A} \mid A \text{ is fine}\}$ be the fine option sets. We strengthen the richness assumption of the previous section in two ways. First, we require an analogous richness of non-overlapping option sets. Second, we require a rich collection of fine duplicates.

STRONG RICHNESS: *For $A \neq \emptyset$, C and $\delta \in (0, 1)$, there is* (i) B *such that* $B \perp C$ *and* $\rho(A, AB) = \delta$ *and* (ii) $D \in \mathcal{M}$ *such that* $D \cap C = \emptyset$ *and* $D \sim A$.

Strong richness is satisfied if each attribute has an archetype and each archetype has many duplicates. The following example of a strongly rich random choice rule illustrates such a setting.

EXAMPLE 7: Let $S = \mathbb{N} \times \mathbb{R}_{++}$ and let $Z = \mathbb{R}_{++}$. For $s = (i, r) \in S$, let $w_r = r$ and let the simple attribute intensity η be such that $\eta_{ir'}^r = 1$ if and only if $r' = r$. In this example, a choice object s is a pair $(i, r) \in \mathbb{N} \times \mathbb{R}_{++}$. The first coordinate describes an irrelevant feature analogous to the color of the bus in Debreu's example. Since each $s = (i, r)$ has a single attribute $r \in Z$, every option is an archetype. It is easy to check that (w, η) is a complete attribute system and that the ρ it represents satisfies strong richness.

The choice rule above is an example of a particular type of attribute rule that we call *extended Luce rules*: take any Luce rule ρ^o on some set of alternatives S. Then, let $S_e = I \times S$, where I is any index set. Finally, set $Z = S$, $w_s = v_s$, and $\eta_{is}^s = 1$ and $\eta_{is'}^s = 0$ if $s \neq s'$. Thus, for $i, j \in I$, $s \in S$, (i, s) and (j, s) are duplicates and each $S_i = \{(i, s) \mid s \in S\}$ is a "copy" of the original S; that is, $\rho_{is}(A_i) = \rho_s^o(A)$ for all $s \in A \subset S$, where $A_i = \{(i, s) \mid s \in A\}$. It is easy to see that if ρ^o is a rich Luce rule, then ρ is a strongly rich attribute rule. Hence, while a rich Luce rule is not a strongly rich attribute rule, it has a strongly rich extension.

---

[9]The proof of Theorem 2 reveals that when ρ is an attribute rule, an option set A is fine if and only if it consists of archetypes.

THEOREM 2: *A strongly rich choice rule satisfies weak independence and elimination of duplicates if and only if it is a complete attribute rule*.

Theorem 2 shows that in an environment with a sufficient variety of options and option sets, the only rule that satisfies weak independence and elimination of duplicates is the attribute rule. In a sparse environment, other rules may satisfy those two requirements but those rules cannot be extended without violating one or both of them. The necessity part of Theorem 2 requires completeness—a richness requirement specific to the attribute rule. It ensures that options are non-overlapping if and only if they share no common attribute. Note that any attribute rule can be completed by adding options (at most two archetypes per attribute).

Without strong richness, there can be several representations for the same attribute rule. However, strong richness yields the following uniqueness result:

PROPOSITION 1: *The attribute systems* (w, η) *and* $(\hat{w}, \hat{\eta})$ *represent the same strongly rich, complete attribute rule if and only if* $\hat{w} = \alpha \cdot w$ *and* $\hat{\eta}^x = \beta_x \cdot \eta^x$ *for* $\alpha, \beta_x > 0$.

Note that Proposition 1 applies not just to complete attribute systems but to all attribute systems. It states that the w is unique up to a positive multiplicative constant, while the η is unique up to a positive constant for each x. The uniqueness of η implies that the set of attributes, Z, is uniquely identified.

If η is simple, the choice rule that (w, η) represents satisfies *weak stochastic transitivity*:

$$\rho_r(\{r, s\}) > 1/2 \text{ and } \rho_s(\{s, t\}) > 1/2 \quad \text{implies} \quad \rho_r(\{r, t\}) > 1/2.$$

There is a good deal of evidence suggesting that choice rules may violate this property.[10] The following example illustrates how a non-simple attribute rule can accommodate failures of weak stochastic transitivity.

EXAMPLE 8: Let $A = \{r, s, t\}$ and assume there are three attributes, $Z = \{1, 2, 3\}$. Each attribute value is 1, that is, $w_x = 1$ for all $x \in Z$. Option r has attributes 1, 3, s has attributes 1, 2, and t has attributes 2, 3. In particular, $\eta_r^1 = \eta_s^2 = \eta_t^3 = 4$, $\eta_r^3 = \eta_s^1 = \eta_t^2 = 1$, and $\eta_r^2 = \eta_s^3 = \eta_t^1 = 0$. This attribute system represents the choice rule ρ such that

$$\rho_r(\{r, s\}) = \rho_s(\{s, t\}) = \rho_t(\{r, t\}) = 3/5.$$

In this example, binary comparisons are "won" by the alternative that has a higher value of the common attribute. For example, r is chosen over s with

---

[10]For a detailed review of this evidence, see Rieskamp, Busemeyer, and Mellers (2006).

probability 3/5 because $\eta_r^1 > \eta_s^1 > 0$. This "matchup-effect" creates a violation of stochastic transitivity: r matches up well against s, s matches up well against t, and t matches up well against r.

4. RELATED MODELS OF RANDOM CHOICE

4.1. *Random Utility Maximization*

The theoretical literature on random choice has focused largely on interpreting random choice as *random utility maximization*. In this section, we briefly discuss this literature and relate the attribute model to random utility maximization. Often, the random utility literature assumes finitely many alternatives. Thus, to relate the attribute model to that literature, it is convenient to consider a finite setting.

Let $S = \{1, \ldots, n\}$ be the set of alternatives and let $\mathcal{A}_+$ be the set of all nonempty subsets of S. In this case, a choice rule ρ can be identified with a vector $q \in \mathbb{R}_+^{n(2^n-1)}$, where $q_{iA} = \rho(\{i\}, A)$ for all $A \in \mathcal{A}_+$. Such a vector q satisfies

$$(2) \qquad q_{iA} \leq 1,$$
$$q_{iA} > 0 \quad \text{implies} \quad i \in A,$$
$$\sum_{i \in S} q_{iA} = 1.$$

Let Q be the set of all $q \in \mathbb{R}_+^{n(2^n-1)}$ that satisfy the conditions in (2). Let Q_ℓ be the subset of Q corresponding to Luce rules and Q_a be the subset corresponding to attribute rules. One other class of extensively studied choice rules are *random utility maximizers*. Most econometric models of discrete choice, such as logit, probit, nested logit, etc., are examples of random utility maximizers.

Let U be the set of all bijections from S to S. For any $i \in A \in \mathcal{A}_+$, let

$$[iA] = \{u \in U \mid u_i \geq u_j \ \forall j \in A\}.$$

A function $\pi: U \to [0, 1]$ is a *random utility* if $\sum_{u \in U} \pi(u) = 1$. We identify each such function with an element in $\mathbb{R}_+^{|U|}$. Let $\Pi = \{\pi \in \mathbb{R}_+^{|U|} \mid \sum_{u \in U} \pi(u) = 1\}$ be the set of all random utilities. Hence, Π is the $|U| - 1 = n! - 1$-dimensional unit simplex. Let Q_r be the set of choice rules that maximize a random utility.

DEFINITION: The choice rule q maximizes the random utility π if $q_{iA} = \sum_{u \in [iA]} \pi_u$ for all i, A.

Let Q_r denote the set of random utility maximizers, that is, the set of choice rules that maximize some random utility. Falmagne (1978) provided necessary and sufficient conditions for a choice rule to be an element of Q_r. Block and

Marschak (1960) showed that $Q_\ell \subset Q_r$. Holman and Marley (see Luce and Suppes (1965)) and McFadden (1978) show how to find a random utility π for any Luce rule. Theorem 3, below, shows that every attribute rule is a random utility maximizer, that is, $Q_a \subset Q_r$. Hence, Theorem 3 extends Block and Marschak's result to the attribute rule. For any subset $X \subset \mathbb{R}^k$, let cl X denote the closure of X and let conv X denote its convex hull.

THEOREM 3: $Q_\ell \subset Q_a \subset Q_r = \operatorname{cl} Q_a = \operatorname{cl} \operatorname{conv} Q_\ell$.

Given the fact that Q_r is closed, convex, and contains Q_ℓ (i.e., the Block–Marschak theorem), the argument establishing that Q_r is equal to the closed convex hull of Q_ℓ is not difficult. However, unlike the closure operator, taking convex hulls is not innocuous; as can be seen from Debreu's example, the behavior associated with a mixture of two Luce rules can be very different from the behavior associated with any single Luce rule.

Showing that the closure of Q_a contains the convex hull of Q_ℓ is straightforward. From this and the fact that the closed convex hull of Q_ℓ is equal to Q_r, it follows that the closure of Q_a contains Q_r. The most challenging step in the proof of the theorem above is establishing that every attribute rule is a random utility maximizer.

Theorem 2 shows that once the main hypothesis of the Luce model (i.e., consistent revealed stochastic preference) is modified to deal with the duplicates problem by restricting independence to non-overlapping option sets and eliminating duplicates, we end up with the attribute rules. The random utility model, on the other hand, interprets the choice behavior as the result of probabilistic choice of utility functions. In general, this interpretation does not necessitate a well-defined revealed stochastic preference (i.e., any form of independence). Nevertheless, Theorem 3 establishes that the two approaches yield essentially the same result ($Q_r = \operatorname{cl} Q_a$).

4.2. *Elimination by Aspects*

Tversky (1972) introduced the elimination-by-aspects (EBA) rule that shares certain features with the attribute rule. A choice rule q is an elimination-by-aspects (EBA) rule if there is a scale $u : \mathcal{A}_+ \to \mathbb{R}_+$ such that, for all $i \in A \in \mathcal{A}_+$,

$$q_{iA} = \frac{\displaystyle\sum_{B \in \mathcal{A}} u(B) q_{iA \cap B}}{\displaystyle\sum_{B : B \cap A \neq \emptyset} u(B)}.$$

We can interpret this rule as a modification of the attribute rule. First, the decision maker chooses an attribute and discards all choices that do not have it.

To choose among the remaining alternatives, the decision maker selects a second attribute and again discards all options that do not have it. This process is repeated until a single alternative is left.

Some attribute rules cannot be approximated by any EBA rule. Specifically, Example 4 above (three airlines with two destinations) is not close to any EBA rule. Tversky (1972) showed that every EBA rule is a random utility maximizer. It then follows from Theorem 3 that every EBA can be approximated by some attribute rule and hence, that attribute rules are more permissive than EBA rules.

4.3. *The Nested Logit*

The econometric discrete choice literature uses the associated random utility model to analyze estimation techniques for the Luce model.[11] Generalizations, such as the cross-nested logit model, allow for correlations in the underlying distribution of utilities to address phenomena related to the duplicates problem (see, e.g., Ben-Akiva and Lerman (1985), Wen and Koppelman (2001), and Train (2009)).

In a cross-nested logit, the modeler specifies a collection of (possibly overlapping) subsets that shape the correlation structure of the utility distribution. Like our attributes, these subsets represent shared features. There are two key differences between our approach and cross-nested logit models. First, our model does not assume a set of attributes, but identifies them from observed choice frequencies. Second, the cross-nested logit model has no analogue of our key parameters, attribute values, and intensities. Attribute values and intensities in our model are *context-independent*, that is, the same values apply for every choice problem. It is the context-independence of those parameters that allows us to interpret the attribute rule as a generalization of the context-independence that characterizes Luce's original model. For a *fixed* choice problem, a two-stage cross-nested logit can be written as a two-level Luce rule. However, if the decision problem changes, the parameters of the corresponding two-level Luce rule change as well. Thus, the parameters of the cross-nested logit cannot be interpreted as attribute values and intensities.

5. DYNAMIC CHOICE

In this section, we extend our model to dynamic choice, that is, to a domain in which options are nodes in decision trees. Decision trees reflect the timing and order of decisions, that is, physical descriptions of the choice environment that are, in principle, observable. As we illustrate in the following example, the

---

[11]The Luce model corresponds to a setting where utilities are independent and distributed according to a Gumbel distribution. See Holman and Marley (see Luce and Suppes (1965)) and McFadden (1978).

duplicates problem is especially relevant in dynamic settings. Even if distinct alternatives are non-overlapping, distinct terminal nodes may yield the same physical outcome and, therefore, decision trees may create overlap even if individual alternatives are non-overlapping.

More specifically, suppose s_1, s_2, s_3, and s_4 are four non-overlapping alternatives. Decision problem 1 is a static problem, as in the previous section: $D^1 = \{s_1, s_2, s_3, s_4\}$. Decision problem 2 is a two-stage problem with two actions in stage 1; action 1 leads to the terminal node s_1; action 2 leads to the (second-stage) decision problem $\{s_2, s_3, s_4\}$:

$$D^2 = \big\{s_1, \{s_2, s_3, s_4\}\big\}.$$

Decision problem 3 differs from 2 by the addition of another stage-1 action that excludes s_1 and s_3:

$$D^3 = \big\{s_1, \{s_2, s_4\}, \{s_2, s_3, s_4\}\big\}.$$

Clearly, the actions $\{s_2, s_4\}$ and $\{s_2, s_3, s_4\}$ overlap in the sense defined above. After all, anything that can be achieved by choosing $\{s_2, s_4\}$ can also be achieved by choosing $\{s_2, s_3, s_4\}$. The objective of this section is to characterize dynamic choice behavior that results from applying our notions of duplicates and overlap to dynamic settings. Specifically, we will introduce a new axiom, *consistency*, that requires s_1 to be chosen with the same probability from all three decision problems described above. Theorem 4, below, characterizes attribute rules that satisfy consistency.

For the nonempty set K, let $F(K)$ be the set of all nonempty finite subsets of K. Define, inductively, $\mathcal{T}_1 := F(S)$ and $\mathcal{T}_{n+1} := F(\mathcal{T}_n \cup S)$ for all $n \geq 1$. The set $\mathcal{T}_n$ is the set of all decision problems with no more than n decision stages. Let $\mathcal{T} := \bigcup_{n \geq 1} \mathcal{T}_n$; the set of *decision nodes* is $\Omega = \mathcal{T} \cup S$ and $\mathcal{D} = F(\Omega)$ is the set of all decision problems. Since all option sets are finite, the following fact is straightforward:

FACT: $\mathcal{D} = \mathcal{T}$.

We write D, D' for elements of $\mathcal{D}$ and s, t, s', t' for elements of S. The function $\phi : (\mathcal{D} \cup \{\emptyset\}) \times \mathcal{D} \to [0, 1]$ is a *dynamic choice rule* if, for all $D \in \mathcal{D}$, $\phi(\cdot, D)$ is additive and $\phi(D, D) = 1$. For a dynamic choice rule, the choice objects are decision nodes $\omega \in \Omega$ which can be final outcomes ($\omega \in S$) or (continuation) decision problems ($\omega \in \mathcal{D}$). Hence Ω is analogous to S in the static case.

The definitions of *duplicates, non-overlapping*, and *fine* can be applied to a dynamic choice rule ϕ without modification. The same is true for the axioms *Elimination of Duplicates*, *Weak Independence*, *Strong Richness*, and the notion of an attribute rule. We add one new axiom for dynamic choice rules:

CONSISTENCY: $\{D\}$ *is a duplicate of* D.

Consistency states that delaying the choice from D has no effect on the choice probabilities of alternatives outside D. Hence, $D^1 = \{s_1, \{s_2, s_3\}\}$ leads to the same probability of choosing s_1 as $D^0 = \{s_1, s_2, s_3\}$. In D^0, the choice is between s_1, s_2, and s_3 in the first (and final) stage. In D^1, the stage-1 choice is between s_1 and not-s_1 and the stage-2 choice is between s_2 and s_3.

An attribute system (w, η) where $w: Z \to \mathbb{R}_{++}$ and $\eta: Z \times \Omega \to \mathbb{R}_{++}$ is *recursive* if, for $\omega = D \in \mathcal{D}$,

$$(r) \qquad \eta_\omega^x = \sum_{\omega' \in D} \eta_{\omega'}^x.$$

Hence, for a recursive attribute rule, the attribute intensity of the sub-problem D is simply the sum of the attribute intensities of its decision nodes.

As in Section 3, $X_\omega = \{x \in Z \mid \eta_\omega^x > 0\}$ and $X(D) = \bigcup_{\omega \in D} X_\omega$. Then, ϕ is a *recursive attribute rule* if there is a recursive attribute system (w, η) such that

$$(3) \qquad \phi_\omega(D) = \sum_{x \in X(D)} \frac{w_x}{w(X(D))} \cdot \frac{\eta_\omega^x}{\eta^x(D)}$$

for all $D \in \mathcal{D}$ and $\omega \in D$. As in Section 3, an option $\omega \in \Omega$ is an *archetype* for $x \in Z$ if $\eta_\omega^x = 1$ and $\eta_\omega^y = 0$ for all $y \neq x$. An attribute system is *complete* if every attribute has multiple (i.e., at least two) archetypes. The following theorem characterizes recursive attribute rules.

THEOREM 4: *A strongly rich dynamic choice rule satisfies weak independence, elimination of duplicates, and consistency if and only if it is a complete recursive attribute rule.*

By restricting a dynamic choice rule ϕ to the objects S, we get another choice rule (ϕ, S). We call this choice rule the static rule induced by ϕ; conversely, we call (ϕ, Ω) an *extension* of (ϕ, S). As we demonstrate in Proposition 2, below, the attribute rule can be used to extend Luce rules (or attribute rules) to dynamic settings. Moreover, if the static setting is rich (or strongly rich in the case of an attribute rule), then this extension is unique.

PROPOSITION 2: *The extension of any rich Luce rule or strongly rich, complete attribute rule to a recursive attribute rule is unique, complete, and satisfies strong richness.*

Proposition 2 shows that dynamic extensions of rich Luce rules are *strongly rich* attribute rules. Notice that in this setting, outcomes ($s \in S$) are non-overlapping; the extension to dynamic decision problems creates duplicates and overlap and, as Proposition 2 shows, this is enough to guarantee strong richness.

So far, our dynamic model analyzes decision making in the *first stage* of a multistage choice problem. To extend the model to subsequent stages requires an assumption as to how randomness evolves over time. In the following, we analyze the behavior of a *consequentialist* (see Machina (1989)) agent, that is, an agent whose choice probabilities are independent of the choice history.

For any decision problem $D \in \mathcal{D}$, let $\sigma(D) \subset S$ be the terminal nodes of D. The vector $(\omega_0, \ldots, \omega_n)$ is a *path* of D if $\omega_0 \in \sigma(D)$, $\omega_i \in \omega_{i+1}$ for all $i < n$ and $\omega_n = D$. The (consequentialist) probability of path $h = (\omega_0, \ldots, \omega_n)$ of D is

(c) $$p_h(D) = \phi_{\omega_0}(\omega_1) \cdot \phi_{\omega_1}(\omega_2) \cdots \phi_{\omega_{n-1}}(\omega_n).$$

We use the term "consequentialist" because, in the formula (c) above, the conditional probability of choosing ω_i at node ω_{i+1} of D is the same as the probability of choosing ω_i in the decision problem ω_{i+1}. Thus, dynamic choice probabilities are history-independent. The *outcome probability* of $s \in S$ is the sum of the probabilities of all paths in D with $\omega_0 = s$. We write $p_s(D)$ for the outcome probability of s in problem D. A decision problem $D \in \mathcal{D}$ is *simple* if all of its terminal nodes are distinct; let $\mathcal{D}^s \subset \mathcal{D}$ be the set of simple decision problems. In that case, every $s \in \sigma(D)$ has a unique associated path. We call the consequentialist probability of this path, the outcome probability of s in D. We say that a dynamic choice rule is *invariant* if the outcome probability of s in D is the same as in D' whenever D, D' are two simple decision problems with the same outcomes. That is, $p_s(D) = \phi_s(\sigma(D))$ for all $D \in \mathcal{D}^s$. Proposition 3, below, shows that every extension of a rich Luce rule is invariant. Moreover, extensions of rich Luce rules are the only recursive attribute rules that satisfy this property.

PROPOSITION 3: *A strongly rich, complete, recursive attribute rule is invariant if and only if it is the extension of a rich Luce rule.*

Invariance breaks down once we consider decision problems with multiple paths that lead to the same option or to overlapping options. Hence, the presentation of the choice problem matters. In general, there are two conflicting effects: adding more copies of s to subtrees increases opportunities to choose s but also increases the inclination to delay choosing s at earlier nodes. The first effect increases the outcome probability of s while the second effect decreases it.

More specifically, adding another s to any node of a simple decision problem that already has s as a possible outcome increases the outcome probability of s. However, adding one more s to an arbitrary decision problem may decrease the outcome probability of s. For example, comparing $D = \{s, t, \{s, t, \{s, t\}\}\}$ to $D' = \{s, t, \{s, t, \{t\}\}\}$, we note that $p_s(D) = \phi_s(\{s, t\}) < p_s(D')$. Hence, removing the final opportunity to choose s from D increases the overall probability of choosing s.

While there is some theoretical work on dynamic random choice, empirical and experimental work on the topic is limited. Hence, there are few documented dynamic random choice regularities, and no well-documented dynamic random choice paradoxes. When more evidence becomes available, our consequentialism assumption may prove to be restrictive. In that case, it may be useful to consider versions of the recursive attribute model that permit history-dependence. There is, in principle, no difficulty with constructing such models nor with extending our notions of richness, duplicates, independence, and consistency to them.

6. CONCLUSIONS

When agents follow the Luce rule, choice frequencies reflect stochastic preference: alternative s is stochastically preferred to alternative t if, in any decision problem, s is more likely to be chosen than t. We show that in a sufficiently rich setting, Luce rules are the only stochastic choice rules that admit a context-independent stochastic preference. While context-independence is an attractive feature, it is too restrictive in many settings. For example, in dynamic choice problems, we would expect that adding a future opportunity to choose alternative s disproportionately affects the probability that s is chosen in the current stage of the problem. Thus, future opportunities to choose the same alternative create a wedge between stochastic preference and choice frequencies and context-independence must be relaxed. We introduced the attribute model to account for this effect. In our dynamic extension of the Luce model, choices are themselves (continuation) decision problems and the attributes of a choice are simply the collection of possible outcomes associated with the continuation problem.

The technical innovation of the paper is to use richness assumptions to facilitate an exact calibration of the model's parameters. In a strongly rich setting, attribute rule parameters are uniquely identified up to a positive scalar. In other words, we can infer the set of relevant attributes, their values, and attribute intensities uniquely from the stochastic choice rule. While strongly rich environments have desirable theoretical properties, they represent an idealized setting that may not fit a particular application. For example, strong richness requires that there are multiple archetypes (single-attribute choices) for each attribute. In some settings, such as the dynamic extensions of rich Luce rules analyzed in Section 5, this assumption is automatically satisfied. In other settings, there may be no single-attribute alternatives. The attribute rule can be applied in non-rich settings as well, but the parameters of the model will not be uniquely identified. This is analogous to other choice-theoretic models, such as subjective expected utility theory, where probabilities are uniquely pinned down if there is a finely divisible state space but will not be uniquely identified in a discrete setting.

Both the Luce rule and the attribute rule are random utility maximizers and, therefore, inherit the following monotonicity property: when an alternative is

added to the choice set, the choice probability of the original members of the choice set cannot increase. This property is in conflict with some of the evidence from the marketing literature, most notably the attraction effect (Huber, Payne, and Puto (1982)) and the related compromise effect (Simonson (1989)).

There are at least two possible avenues to extend the attribute rule beyond the scope of random utility. First, recall that the attribute values in an attribute rule must be positive, that is, each additional attribute must increase the frequency that an object is chosen. If we relax this requirement and allow negative attributes, that is, attributes that reduce an object's choice probability, then the resulting random choice rules are no longer random utility maximizers. Specifically, this generalization allows for non-monotonicities as required by the attraction effect. The attraction effect may come about when the decoy option shares an unattractive attribute (a high price) with an existing option. The added alternative dilutes the effect of the negative attribute, thereby creating an attraction effect.

Second, recall that in our analysis of dynamic choice, we interpreted the tree as a description of the timing and order of decisions, that is, a physical description of the choice environment that may, in principle, be observable. Alternatively, we may also interpret decision trees as *frames*, that is, descriptions of how the decision maker perceives the choice problem. Presumably, such perceptions are subjective and can only be identified through their implications on observed behavior.

We noted above that adding more opportunities to choose s may decrease the outcome probability of s. Hence, by identifying a suitable frame and focusing on final choice probabilities, the recursive attribute model can be used to analyze some violations of monotonicity. An analysis of these and other generalizations of the attribute rule is left for future research.

APPENDIX

A.1. *Example 2*

We must prove that the ρ defined in Example 2 is rich. Let $S = \{s_1, s_2, \ldots\}$ be an enumeration of S and consider any $A, C \in \mathcal{A}_+$, and $\delta \in (0, 1)$. Let $\delta_1 = \sum_{s \in A} s$, $\delta_2 = \frac{(1-\delta)\delta_1}{\delta}$ and set $B_0 = \emptyset$. Define B_j for $j = 1, 2, \ldots$ as follows: $B_{j+1} = B_j \cup \{s_{j+1}\}$ if $s_{j+1} \notin AC$ and $\sum_{s \in B_j} s + s_{j+1} \leq \delta_2$; otherwise $B_{j+1} = B_j$. (We set $\sum_{s \in B_0} s = 0$.) Let $B = \bigcup_{j \geq 1} B_j$ and note that $\sum_{s \in B} s = \delta_2$. Hence, $B \in \mathcal{A}$, $B \cap C = \emptyset$ and $\rho(A, AB) = \frac{\delta_1}{\delta_1 + \delta_2} = \delta$ as desired.

A.2. *Proof of Theorem 1*

Verifying that every Luce rule satisfies independence is straightforward. Hence, we will only prove that a rich choice rule that satisfies independence

is a Luce rule. We assume richness (R) and independence (I) throughout the following lemmas.

Define a binary relation $\succcurlyeq_\rho$ on $\mathcal{A}_+$ as follows: $A \succcurlyeq_\rho B$ if and only if $\rho(A, AC) \geq \rho(B, BC)$ for all $C \in \mathcal{A}_+$ such that $AB \cap C = \emptyset$. Let $\sim_\rho$ be the symmetric and $\succ_\rho$ be the strict part of $\succcurlyeq_\rho$.

LEMMA A.1: $\succcurlyeq_\rho$ *is complete and transitive*.

PROOF: Clearly, ρ satisfies independence only if $\succcurlyeq_\rho$ is complete. Next, assume that $A \succcurlyeq_\rho B$ and $B \succcurlyeq_\rho C$. By richness, there exists a $D \in \mathcal{A}$ such that $D \cap ABC = \emptyset$ and $\rho(C, CD) < 1$. Hence, $D \neq \emptyset$. Note that $\rho(A, AD) \geq \rho(B, BD) \geq \rho(C, CD)$; thus independence implies $A \succcurlyeq_\rho C$ as desired. Q.E.D.

DEFINITION: The sequence $A_1, \ldots, A_n \in \mathcal{A}$ is a *test sequence* if the elements are pairwise disjoint and $\rho(A_i, A_i A_{i+1}) = 1/2$ for all $i = 1, \ldots, n-1$.

LEMMA A.2: *For any test sequence* $A_1, \ldots, A_n \in \mathcal{A}_+$, $\rho(A_i, A_i A_j) = 1/2$ *for all* $i \neq j$.

PROOF: If the result is true for $n = 3$, then it is true for all n. So assume $n = 3$ and suppose $\rho(A_1, A_1 A_3) > 1/2$. Independence implies that $A_1 \succ_\rho A_2$. Since $\rho(A_1, A_1 A_2) = 1/2 = \rho(A_3, A_3 A_2)$, independence also implies $A_1 \sim_\rho A_3$. Then, by Lemma A.1, we have $A_3 \succ_\rho A_2$. But $\rho(A_3, A_1 A_3) < 1/2 = \rho(A_2, A_1 A_2)$, contradicting $A_3 \succ_\rho A_2$. A similar argument reveals the impossibility of $\rho(A_1, A_1 A_3) < 1/2$. Hence, $\rho(A_1, A_1 A_3) = 1/2$ as desired. Q.E.D.

LEMMA A.3: *If* $A_1, \ldots, A_n$ *is a test sequence and* $A \in \mathcal{A}_+$ *with* $A \cap A_1 A_2 \cdots A_n = \emptyset$, *then* $\rho(A, AA_i) = \rho(A, AA_1)$ *for all* $i = 1, \ldots, n$.

PROOF: If necessary, use richness to extend the test sequence so that $n \geq 3$. Then, Lemma A.2 implies $A_i \sim_\rho A_j$ for all i, j and hence $\rho(A, AA_i) = \rho(A, AA_1)$ for all i. Q.E.D.

LEMMA A.4: *For all* $A, B \in \mathcal{A}_+$ *with* $A \cap B = \emptyset$, $A \succcurlyeq_\rho B$ *if and only if* $\rho(A, AB) \geq 1/2$.

PROOF: By richness, we can choose $D \in \mathcal{A}_+$ such that $D \cap AB = \emptyset$ and $\rho(B, BD) = 1/2$. Let $B_1 = B$ and $B_2 = D$ and note that B_1, B_2 is a test sequence. Then, by Lemma A.3, $\rho(A, AB) = \rho(A, AD)$ and therefore $\rho(A, AB) \geq 1/2$ if and only if $\rho(A, AD) \geq \rho(B, BD)$, that is, $\rho(A, AB) \geq 1/2$ if and only if $A \succcurlyeq_\rho B$. Q.E.D.

LEMMA A.5: *If* C_1, C_2, C_3, C_4 *is a test sequence, then* $\rho(C_i, C_1 C_2 C_3 C_4) = 1/4$ *for all* $i = 1, 2, 3, 4$.

PROOF: Let $C = C_1 C_2 C_3 C_4$, and without loss of generality, assume $\rho(C_i, C) \geq \rho(C_j, C)$ whenever $i \leq j$. Hence, by Lemma A.4,

(A.1) $C_1 C_2 \succcurlyeq_\rho C_3 C_4$ and $C_1 C_3 \succcurlyeq_\rho C_2 C_4$.

By richness, there exists C_5 such that C_1, C_2, C_3, C_4, C_5 is a test sequence. By Lemmas A.2 and A.4, $C_i \sim_\rho C_j$ for all i, j and hence, by Lemma A.1, we have

$$\begin{aligned}
\text{(A.2)} \quad \rho(C_1 C_2, C_1 C_2 C_3) &= \rho(C_1 C_2, C_1 C_2 C_5) \\
&\geq \rho(C_3 C_4, C_3 C_4 C_5) \\
&= \rho(C_3 C_4, C_2 C_3 C_4).
\end{aligned}$$

And by the same argument,

(A.3) $\rho(C_1 C_3, C_1 C_2 C_3) \geq \rho(C_2 C_4, C_2 C_3 C_4)$.

But we also have

$$\begin{aligned}
\text{(A.4)} \quad 2 &= 2\big[\rho(C_1, C_1 C_2 C_3) + \rho(C_2, C_1 C_2 C_3) + \rho(C_3, C_1 C_2 C_3)\big] \\
&= \rho(C_1 C_2, C_1 C_2 C_3) + \rho(C_1 C_3, C_1 C_2 C_3) + \rho(C_2 C_3, C_1 C_2 C_3) \\
&\geq \rho(C_3 C_4, C_2 C_3 C_4) + \rho(C_2 C_4, C_2 C_3 C_4) + \rho(C_2 C_3, C_2 C_3 C_4) \\
&= 2\big[\rho(C_2, C_2 C_3 C_4) + \rho(C_3, C_2 C_3 C_4) + \rho(C_4, C_2 C_3 C_4)\big] \\
&= 2.
\end{aligned}$$

Equation (A.4) implies that the inequalities in (A.2) and (A.3) must in fact be equalities. Hence $\rho(C_1 C_2, C_1 C_2 C_5) = \rho(C_1 C_2, C_1 C_2 C_3) = \rho(C_3 C_4, C_2 C_3 C_4) = \rho(C_3 C_4, C_3 C_4 C_5)$ and therefore, by independence $C_1 C_2 \sim_\rho C_3 C_4$. By Lemma A.4, we have $\rho(C_1 C_2, C) = \rho(C_3 C_4, C) = 1/2$. Finally since $\rho(C_i, C) \geq \rho(C_j, C)$ for $i \leq j$, we must have $\rho(C_i, C) = 1/4$ for $i = 1, 2, 3, 4$.   Q.E.D.

LEMMA A.6: *If $A_1, \ldots, A_n$ is a test sequence, then $A_i A_j \sim_\rho A_k A_\ell$ for all $i \neq j$ and $k \neq \ell$.*

PROOF: If i, j, k, ℓ are all distinct, then Lemma A.5 implies $\rho(A_i A_j, A_i A_j A_k A_\ell) = 1/2$ and Lemma A.4 implies $A_i A_j \sim_\rho A_k A_\ell$. If $\{i, j, k, \ell\}$ has three distinct elements, assume, without loss of generality, that $j = \ell$. Let $B_1 = A_i, B_2 = A_j, B_3 = A_k$, and note that B_1, B_2, B_3 is a test sequence. By richness, we can choose B_4, B_5 such that B_1, B_2, B_3, B_4, B_5 is a test sequence. By Lemmas A.4 and A.5, $B_1 B_2 \sim_\rho B_4 B_5$ and $B_2 B_3 \sim_\rho B_4 B_5$. Then, Lemma A.1 implies $B_1 B_2 \sim_\rho B_2 B_3$, that is, $A_i A_j \sim_\rho A_k A_\ell$. Finally, if $\{i, j, k, \ell\}$ has two distinct elements, then $A_i A_j = A_k A_\ell$, and by Lemma A.1, we have $A_i A_j \sim_\rho A_k A_\ell$.
          Q.E.D.

LEMMA A.7: *If $A_{i_1}, \ldots, A_{i_{2^n}}$ are distinct elements and $A_{j_1}, \ldots, A_{j_{2^n}}$ are distinct elements of the test sequence $A_1, \ldots, A_m$, then $\bigcup_{k=1}^{2^n} A_{i_k} \sim_\rho \bigcup_{k=1}^{2^n} A_{j_k}$.*

PROOF: The proof is by induction. When $n = 1$, the statement is true by Lemma A.6. Next, assume it is true for n and let $A_{i_1}, \ldots, A_{i_{2^{n+1}}}$ and $A_{j_1}, \ldots, A_{j_{2^{n+1}}}$ be two collections of distinct elements of the same test sequence. Use richness to extend the test sequence $A_1, \ldots, A_m$ with an additional 2^{n+1} elements, namely $A_{m+1}, \ldots, A_{m+2^{n+1}}$. Let $B_k = A_{i_{2k-1}} A_{i_{2k}}$, let $C_k = A_{j_{2k-1}} A_{j_{2k}}$, and let $D_k = A_{m+2k-1} A_{m+2k}$ for $k = 1, \ldots, 2^n$. Lemma A.6 implies that $B_k \sim_\rho B_\ell$ for all k, ℓ. Hence $B_1, \ldots, B_{2^n}$ is a test sequence. By the same argument, $C_1, \ldots, C_{2^n}$ is a test sequence and $D_1, \ldots, D_{2^n}$ is a test sequence. By construction, B_k and D_ℓ are disjoint for every k and ℓ. Moreover, Lemma A.6 implies that $B_k \sim_\rho D_\ell$ for all k, ℓ. Hence we can relabel $B_1, \ldots, B_{2^n}$ and $D_1, \ldots, D_{2^n}$ so that they become distinct elements of the same test sequence. By the inductive hypothesis, $\bigcup_{k=1}^{2^n} B_k \sim_\rho \bigcup_{k=1}^{2^n} D_k$. By an entirely analogous argument, $\bigcup_{k=1}^{2^n} C_k \sim_\rho \bigcup_{k=1}^{2^n} D_k$. Finally, by Lemma A.1, we obtain $\bigcup_{k=1}^{2^{n+1}} A_{i_k} = \bigcup_{k=1}^{2^n} B_k \sim_\rho \bigcup_{k=1}^{2^n} C_k = \bigcup_{k=1}^{2^{n+1}} A_{j_k}$ as desired.   Q.E.D.

LEMMA A.8: *If $A_1, \ldots, A_{2^n+1}$ is a test sequence, then $\rho(A_j, A_1 A_2 \cdots A_{2^n+1}) = 1/(2^n + 1)$ for $j = 1, 2, \ldots, 2^n + 1$.*

PROOF: By richness, we can find A_{2^n+2} such that $A_1, \ldots, A_{2^n+1}, A_{2^n+2}$ is a test sequence. Then for any $j > 1$, Lemmas A.1, A.4, and A.7 imply

$$\rho(A_1, A_1 \cdots A_{2^n+1}) = \rho(A_{2^n+2}, A_2 \cdots A_{2^n+2})$$
$$= \rho(A_{2^n+2}, A_1 \cdots A_{j-1} A_{j+1} \cdots A_{2^n+2})$$
$$= \rho(A_j, A_1 \cdots A_{2^n+1}).$$

Then, the feasibility constraint and the additivity of ρ yield the desired result.   Q.E.D.

LEMMA A.9: *If $A_{i_1}, \ldots, A_{i_n}$ are distinct elements and $A_{j_1}, \ldots, A_{j_n}$ are distinct elements of the test sequence $A_1, \ldots, A_m$, then $\bigcup_{k=1}^{n} A_{i_k} \sim_\rho \bigcup_{k=1}^{n} A_{j_k}$.*

PROOF: Choose an integer k such that $2^k > m \geq n$. By R, we can find $2^k + 1 - n$ distinct elements $A_{m+1}, \ldots, A_{2^k+1}, \ldots, A_{2^k+1+m-n}$ such that $A_1, \ldots, A_{2^k+1+m-n}$ is a test sequence. Let $B = A_{m+1} \cdots A_{2^k+1+m-n}$. Then Lemma A.8 implies

$$\rho(A_{i_1} \cdots A_{i_n}, A_{i_1} \cdots A_{i_n} B) = n/(2^k + 1)$$
$$= \rho(A_{j_1} \cdots A_{j_n}, A_{j_1} \cdots A_{j_n} B).$$

Then, I yields the desired result.   Q.E.D.

LEMMA A.10: *If $A_1, \ldots, A_n$ is a test sequence, then $\rho(A_j, A_1 A_2 \cdots A_n) = 1/n$ for all j.*

PROOF: By R, we can find A_{n+1} such that $A_1, \ldots, A_{n+1}$ is a test sequence. Then for any $j > 1$, Lemmas A.1, A.4, and A.9 imply

$$\rho(A_1, A_1 \cdots A_n) = \rho(A_{n+1}, A_2 \cdots A_{n+1})$$
$$= \rho(A_{n+1}, A_1 \cdots A_{j-1} A_{j+1} \cdots A_{n+1})$$
$$= \rho(A_j, A_1 \cdots A_n).$$

Then, the feasibility constraint and the additivity of ρ yield the desired result.   *Q.E.D.*

LEMMA A.11: *If $A_{i_1}, \ldots, A_{i_k}$ are k distinct elements of the test sequence $A_1, \ldots, A_n$ and $A = \bigcup_{j=1}^{k} A_{i_j}$, $B = \bigcup_{j=1}^{n} A_j$, then $\rho(A, B) = \frac{k}{n}$.*

PROOF: By Lemma A.10, we have $\rho(A_i, B) = \rho(A_j, B) = \frac{1}{n}$ for all i, j. Then, the additivity of ρ yields the desired result.   *Q.E.D.*

LEMMA A.12: *If $A, B \in \mathcal{A}_+$ and $A \cap B = \emptyset$, then $0 < \rho(A, AB) < 1$.*

PROOF: Suppose $\rho(A, AB) = 1$; then R implies that there exists C such that $C \cap AB = \emptyset$ and $\rho(C, AC) > 1/2$. Hence, by Lemma A.4, $C \succ_\rho A$ and therefore $\rho(C, CB) > \rho(A, AB) = 1$, a contradiction. By symmetry, we cannot have $\rho(A, AB) = 0$ either.   *Q.E.D.*

Assume ρ satisfies R and I. Then, choose any $A_o \in \mathcal{A}_+$ and define, $\bar{v}(A_o) = 1$. Then, set $\bar{v}(\emptyset) = 0$ and, for all $B \in \mathcal{A}_+$ such that $B \cap A_o = \emptyset$, let

$$\bar{v}(B) = \frac{\rho(B, BA_o)}{1 - \rho(B, BA_o)}.$$

Finally, for any $B \in \mathcal{A}_+$ such that $A_o \cap B \neq \emptyset$, find $A \in \mathcal{A}$ such that $A \cap BA_o = \emptyset$ and $\rho(A, AB) = 1/2$ and let $\bar{v}(B) = \bar{v}(A)$.

LEMMA A.13: *The function $\bar{v}$ is well-defined and satisfies the following:*
(i) $\bar{v}: \mathcal{A} \to \mathbb{R}_+$ and $v(A) = 0$ if and only if $A = \emptyset$. (ii) $\bar{v}(A) \geq \bar{v}(B)$ if and only if $A \succcurlyeq_\rho B$.

PROOF: To prove that $\bar{v}$ is well-defined, we first note that by Lemma A.12, $\bar{v}(A) < \infty$ for all A disjoint from A_o. Next, suppose A_1, A_2 are such that $A_o B \cap A_1 = A_o B \cap A_2 = \emptyset$ and $\rho(A_1, A_1 B) = \rho(A_2, A_2 B)$. Then $A_1 \sim_\rho A_2$ and hence, $\rho(A_1, A_1 A_0) = \rho(A_2, A_2 A_0)$ and therefore $\bar{v}(A_1) = \bar{v}(A_2)$, proving that $\bar{v}$ is well-defined.

By Lemma A.12, $\bar{v}$ satisfies (i). To prove (ii), choose C such that $C \cap BAA_o = \emptyset$ and $\rho(C, CA_o) = 1/2$. Then, by Lemma A.4, $A_o \sim_\rho C$. For any $D \in \mathcal{A}_+$ with

$D \cap C = \emptyset$, if $D \cap A_o = \emptyset$, then $A_o \sim_\rho C$ implies $\rho(D, DC) = \rho(D, DA_o) = \frac{\bar{v}(D)}{1+\bar{v}(D)}$. If $D \cap A_o \neq \emptyset$, then since $\bar{v}$ is well-defined, we have $\rho(D, DC) = \frac{\bar{v}(D)}{1+\bar{v}(D)}$. Hence, $A \succcurlyeq_\rho B$ if and only if $\rho(A, AC) \geq \rho(B, BC)$ if and only if $\bar{v}(A) \geq \bar{v}(B)$. $\qquad$ Q.E.D.

Let nC denote the union of some n-element test sequence C_i such that each element $C_i \sim_\rho C$ for some C. Then, by Lemma A.11, $\rho(nC, AnC)$ is the same for all such sequences provided $nC \cap A = \emptyset$. Hence, from now on, we will let nC denote the union of any n-element test sequence with each element satisfying $C_i \sim_\rho C$.

LEMMA A.14: *If $nC \sim_\rho A_o$, then $mC \sim_\rho B$ if and only if $\bar{v}(B) = \frac{m}{n}$.*

PROOF: By R it suffices to show the statement holds for A_0, B, nC, mC pairwise disjoint. Assume $nC \sim_\rho A_o$ and hence $\rho(B, BA_o) = \rho(B, BnC)$. Then, Lemma A.11 yields $\rho(mC, (n+m)C) = \frac{m}{n+m}$. By definition, $\rho(mC, (n+m)C) = \rho(B, BnC)$ if and only if $B \sim_\rho mC$. Hence, $\rho(B, BA_o) = \frac{m}{n+m}$ if and only if $mC \sim_\rho B$ and therefore $\bar{v}(B) = \frac{m}{n}$ if and only if $mC \sim_\rho B$. $\quad$ Q.E.D.

LEMMA A.15: $\rho(A, AB) = \frac{\bar{v}(A)}{\bar{v}(A)+\bar{v}(B)}$ *for all $A, B \in \mathcal{A}_+$ such that $A \cap B = \emptyset$.*

PROOF: First assume that $\bar{v}(A), \bar{v}(B)$ are rational numbers. Then, there exist positive integers k, m, n such that $\bar{v}(A) = \frac{k}{n}$ and $\bar{v}(B) = \frac{m}{n}$. Choose C such that $nC \sim_\rho A_o$, that is, C such that $C \cap A_o = \emptyset$ and $\rho(C, CA_o) = \frac{1}{n+1}$. Note that by Lemma A.14, $kC \sim_\rho A$ and $mC \sim_\rho B$ and hence $\rho(kC, (k+m)C) = \rho(A, AmC) = \rho(A, AB)$. But Lemma A.11 implies $\rho(kC, (k+m)C) = \frac{k}{k+m}$, which yields the desired result.

If either $\bar{v}(A)$ or $\bar{v}(B)$ is not a rational number, then for any $\varepsilon > 0$, choose rational numbers, r_1, r_2 such that $r_1 < \bar{v}(A)$, $r_2 > \bar{v}(B)$, and $\frac{r_1}{r_1+r_2} > \frac{\bar{v}(A)}{\bar{v}(A)+\bar{v}(B)} - \varepsilon$. Then, choose C, D such that A, B, C, D are all pairwise disjoint and $\bar{v}(C) = r_1$ and $\bar{v}(D) = r_2$. By the preceding argument, $\rho(C, CD) = \frac{r_1}{r_1+r_2}$, and by Lemma A.13(ii), $\rho(A, AB) \geq \rho(A, AD) \geq \rho(C, CD)$. Hence, $\rho(A, AB) \geq \frac{\bar{v}(A)}{\bar{v}(A)+\bar{v}(B)} - \varepsilon$ for every $\varepsilon > 0$, that is, $\rho(A, AB) \geq \frac{\bar{v}(A)}{\bar{v}(A)+\bar{v}(B)}$. A symmetric argument ensures that $\rho(A, AB) \leq \frac{\bar{v}(A)}{\bar{v}(A)+\bar{v}(B)}$ and hence the desired conclusion. $\qquad$ Q.E.D.

To complete the proof of the theorem, let $v_s = \bar{v}(\{s\})$.

A.3. *Proofs of Theorem 2 and Proposition 1*

We assume strong richness (SR), weak independence (WI), and elimination of duplicates (E) throughout the lemmas below. (Lemmas A.16–A.24 use only SR. The remainder of the proof also uses WI and E.)

LEMMA A.16: *If $A \cap C = B \cap D = \emptyset$, $A \sim B$, and $C \sim D$, then $AC \sim BD$.*

PROOF: Assume $A \cap C = B \cap D = \emptyset = ABCD \cap E$ and $C \sim D$. Let $s \in E$ and $A \sim B$. Then, choose $B^* \sim B$ such that $B^* \cap ABCDE = \emptyset$ and $C^* \sim C$ such that $C^* \cap ABB^*CDE = \emptyset$. By SR, this can be done. Then, $\rho(s, ACE) = \rho(s, AC^*E) = \rho(s, BC^*E) = \rho(s, B^*C^*E) = \rho(s, B^*CE) = \rho(s, B^*DE) = \rho(s, BDE)$ as desired.   *Q.E.D.*

LEMMA A.17: *$\sim$ is an equivalence relation.*

PROOF: By construction, $\sim$ is reflexive and symmetric. To prove it is transitive, assume $A \sim B \sim C$ and let $s \in D$ for D such that $AC \cap D = \emptyset$. By SR, we can choose $E \sim \{s\}$ such that $E \cap ABCD = \emptyset$ and we can choose $D' \sim (D \setminus \{s\})$ such that $D' \cap ABCDE = \emptyset$. Then $\rho(t, AD's) = \rho(t, AD'E)$ for all $t \in AD'$, hence we have $\rho(s, AD's) = \rho(E, AD'E)$. Similarly, we have $\rho(s, CD's) = \rho(E, CD'E)$. Therefore, $\rho(s, AD) = \rho(s, AD's) = \rho(E, AD'E) = \rho(E, BD'E) = \rho(E, CD'E) = \rho(s, CD's) = \rho(s, CD)$ as desired.   *Q.E.D.*

LEMMA A.18: *$A \sim \emptyset$ implies $A = \emptyset$.*

PROOF: If $A \neq \emptyset$, then, by SR, there is B such that $B \cap A = \emptyset$ and $\rho(A, AB) > 0$. Hence, $\rho(B, AB) < 1 = \rho(B, B)$ and, therefore, $\rho(s, B) \neq \rho(s, AB)$ for some $s \in B$, proving that A is not a duplicate of $\emptyset$.   *Q.E.D.*

LEMMA A.19: *If $A \sim B \in \mathcal{M}$ and $A \neq \emptyset$, then there exists an onto mapping $f : B \to A$ such that $s \sim f^{-1}(s)$ for all $s \in A$.*

PROOF: Consider an enumeration $s_1, s_2, \ldots, s_{|A|}$ of the elements of A. Since B is fine, there is a partition $B_1 B_2 \cdots B_{|A|} = B$ such that $B_i \sim \{s_i\}$ for each i. To obtain the desired f, for each i and each $t \in B_i$, let $f(t) = s_i$.   *Q.E.D.*

LEMMA A.20: *$A \in \mathcal{M}$ if and only if $|A| \geq |B|$ for all $B \sim A$.*

PROOF: If $|A| = 0$, the result follows from Lemma A.18. Otherwise, suppose $A \in \mathcal{M}$ and $\{s_1, s_2, \ldots, s_{|B|}\} = B \sim A$. Since A is fine, we can find a partition $A_1 A_2 \cdots A_{|B|} = A$ such that $A_i \sim \{s_i\}$ for all i. By Lemma A.18, $A_i \neq \emptyset$ for all i. Hence $|A| = \sum_i |A_i| \geq |B|$. Conversely, suppose that $|A| \geq |B|$ for all $B \sim A$. By SR, we can find $C \in \mathcal{M}$ with $C \sim A$. By the first part of the proof, $|C| \geq |A|$, and by hypothesis, $|C| \leq |A|$. Since C is fine, there is a bijection $f : C \to A$ such that $f(s) \sim s$ for all $s \in C$. Together with Lemma A.16 and Lemma A.17, this implies $A \in \mathcal{M}$.   *Q.E.D.*

LEMMA A.21: *If $A \subset B \in \mathcal{M}$, then $A \in \mathcal{M}$.*

PROOF: Suppose $A \notin \mathcal{M}$ and $A \subset B$. By Lemma A.20, there is some $C \sim A$ with $|C| > |A|$. By SR, we can choose $D \in \mathcal{M}$ such that $D \cap C = \emptyset$ and $D \sim (B \setminus A)$. By Lemma A.20, $|D| \geq |(B \setminus A)|$. Hence $|CD| = |C| + |D| > |A| + |B \setminus A| = |B|$. By Lemma A.16, $CD \sim B$, thus Lemma A.20 implies $B \notin \mathcal{M}$.
                                                                                Q.E.D.

Let $T = \{s \in S \mid \{s\} \in \mathcal{M}\}$ and let $\mathcal{B}_0$ be the set of all finite subsets of T. Let $\theta : T \to T$ be a selection from the equivalence classes of $(\sim, B)$, that is, θ is any function such that (i) $\theta(s) \sim s$ for all $s \in T$ and (ii) $s \sim t$ implies $\theta(s) = \theta(t)$. Finally, let $T_1 = \{\theta(s) \mid s \in T\}$ and let $\mathcal{B}_1$ be the set of all finite subsets of T_1.

LEMMA A.22: $\mathcal{B}_0 = \mathcal{M}$.

PROOF: $\mathcal{M} \subset \mathcal{B}_0$ follows from Lemma A.21. To show $\mathcal{B}_0 \subset \mathcal{M}$, let $A = \{s_1, s_2, \ldots, s_{|A|}\} \in \mathcal{B}_0$. By SR, there is $D \in \mathcal{M}$ with $D \sim A$. Lemma A.20 implies $|D| \geq |A|$. Since D is fine, there is a partition $D_1 D_2 \cdots D_{|A|} = D$ with $D_i \sim \{s_i\}$ for each i. Since each $\{s_i\}$ is fine, Lemma A.20 implies $|D_i| \leq 1$ for all i. Hence $|D| = |A|$, and Lemma A.20 implies $A \in \mathcal{M}$.   Q.E.D.

LEMMA A.23: $A \subset B \perp C$ implies $A \perp C$.

PROOF: Assume $A' \sim A$ and $C' \sim C$ and choose $D \sim B \setminus A$ such that $D \cap A' = \emptyset$. Then, by Lemma A.16, $D \cup A' \sim (B \setminus A) \cup A = B$. Since $B \perp C$, we have $(D \cup A') \cap C' = \emptyset$ and hence $A' \cap C' = \emptyset$, proving that $A \perp C$.   Q.E.D.

LEMMA A.24: (i) If $s, t \in T_1$ and $s \neq t$, then $\{s\} \perp \{t\}$. (ii) For $A, B \in \mathcal{B}_0$, $A \perp B$ or there is $s \in A, t \in B$ such that $s \sim t$. (iii) For $A, B \in \mathcal{B}_1$, $A \cap B = \emptyset$ if and only if $A \perp B$.

PROOF: (i) Suppose $s, t \in T_1$, $s \sim A \in \mathcal{A}$, $t \sim B \in \mathcal{A}$ and $A \cap B \neq \emptyset$. By Lemma A.22, $s, t \in \mathcal{M}$. Hence, by Lemmas A.18 and A.20, $|A| = |B| = 1$ and, therefore, $A = B$ and, therefore, $s \sim t$ by Lemma A.17. Then, by the definition of θ, $s = \theta(s) = \theta(t) = t$.

(ii) Assume $A, B \in \mathcal{B}_0$, $A' \sim B$, $B' \sim B$, and $s^* \in A' \cap B'$. By Lemmas A.19 and A.22, there are functions f, g mapping A, B onto A', B' such that $f^{-1}(s) \sim s$ and $g^{-1}(t) \sim t$ for all $s \in A'$ and $t \in B'$. It follows from Lemma A.17 that $f^{-1}(s^*) \sim g^{-1}(s^*)$. By Lemma A.21, $f^{-1}(s^*), g^{-1}(s^*) \in \mathcal{M}$, and hence, applying Lemma A.19 again yields an onto function $h : f^{-1}(s^*) \to g^{-1}(s^*)$ such that $h^{-1}(s) \sim s$ for all $s \in g^{-1}(s^*)$. By Lemma A.20, h must be a bijection. Hence, there are $s \in A$ and $t \in B$ such that $t \sim s$.

(iii) Assume $A, B \in \mathcal{B}_1$. That $A \perp B$ implies $A \cap B = \emptyset$ is obvious. To prove the converse, assume that $A \perp B$ does not hold. Then, by part (ii) of this lemma, there are $s \in A$ and $t \in A$ such that $t \sim s$. Then, $t = s$ by part (i) of this lemma and hence $A \cap B \neq \emptyset$.   Q.E.D.

LEMMA A.25: $\rho(s, AB) = \rho(s, AB_0)$ if $s \in A$, $B_0 \in \mathcal{B}_0$, $s \perp B_0$, and $B = \{\theta(t) \mid t \in B_0\}$.

PROOF: Let $B_0 = B^1 B^2 \cdots B^k$, where $s \sim t$ if and only if $i = j$ for all $s \in B^i$ and $t \in B^j$. Hence, $B^1, \ldots, B^k$ is the partition of B_0 that the equivalence relation $\sim$ yields. First, we note that $i \neq j$ implies $B^i \perp B^j$. To see this, note that if B^i and B^j overlap, then by Lemma A.24(ii), there are $s \in B^i$ and $t \in B^j$ such that $s \sim t$ and hence $i = j$.

Let $n(B_0) = \sum_{i=1}^{k} |B^i| - k$. The proof is by induction on $n(B_0)$. If $n(B_0) = 0$, then each B^i is a singleton and hence $B \sim B_0$ by Lemma A.16 and the result follows. Suppose the result holds whenever $n(\hat{B}_0) = n$ and let $n(B_0) = n + 1 \geq 1$. Hence, there is some i such that $|B^i| > 1$. Choose $t, t' \in B^i$ such that $t \neq t'$ and let $\hat{B}_0 = B_0 \setminus \{t\}$. By Lemma A.23, $B^i \perp \{s\}$ and, therefore, E implies $\rho(s, AB_0) = \rho(s, A\hat{B}_0)$. By the inductive hypothesis, $\rho(s, A\hat{B}_0) = \rho(s, AB)$ and hence $\rho(s, AB_0) = \rho(s, AB)$ as desired.   Q.E.D.

For $B \in \mathcal{B}_1$ such that $B \neq \emptyset$, define the choice rule ρ^1 such that

$$\rho^1(s, B) = \rho(s, B).$$

LEMMA A.26: ρ^1 satisfies R, I and therefore is a Luce rule.

PROOF: Assume $A, C \in \mathcal{B}_1$, $A \neq \emptyset$, and $\delta \in (0, 1)$. By SR, there is $\hat{B} \in \mathcal{A}$ such that $\rho(A, A\hat{B}) = \delta$ and $\hat{B} \perp AC$. Again by SR, we can choose $B_0 \in \mathcal{M}$ such that $B_0 \sim \hat{B}$ and $B_0 \cap A = \emptyset$. Hence, $\delta = \rho(A, A\hat{B}) = \rho(A, AB_0)$. Lemma A.22 implies $B_0 \in \mathcal{B}_0$. Lemma A.17 and $B_0 \sim \hat{B} \perp AC$ imply $B_0 \perp AC$. Then, Lemma A.23 implies $B_0 \perp C$ and $B_0 \perp A$. Let $B = \{\theta(s) \mid s \in B_0\}$. Lemma A.25 yields $\delta = \rho(A, AB_0) = \rho(A, AB)$. Lemma A.23 implies that, for all $s \in B_0$, we have $s \perp C$ and, therefore, $\theta(s) \cap C = \emptyset$. Hence, $B \cap C = \emptyset$ and ρ^1 satisfies R. Lemma A.24(iii) and WI imply that ρ^1 satisfies I and hence, by Theorem 1, it is a Luce rule.   Q.E.D.

For $s \in T_1$, define

$$x_s = \{t \in S \mid \exists B \in \mathcal{B}_0 \text{ such that } s \in B \text{ and } B \sim t\}$$

and define $Z = \{x_s \mid s \in T_1\}$. Let v be the Luce value that represents ρ^1 and define

$$w(x_s) = v_s.$$

For $s \in S$, choose any $B \in \mathcal{B}_0$ such that $s \sim B$ and define

$$\eta_{x_t}(s) = \left| \left\{ t' \in B \mid t = \theta(t') \right\} \right|.$$

LEMMA A.27: (i) $w : Z \to \mathbb{R}_{++}$ *is well-defined*. (ii) $\eta : Z \times S \to \mathbb{N} \cup \{0\}$ *is well-defined*.

PROOF: (i) We must show that $x_s = x_t$ implies $s = t$. Since $t \in x_t$, $x_t = x_s$ implies $t \in x_s$ and hence $t \sim B$ for some B such that $s \in B$. By Lemma A.22, $\{t\} \in \mathcal{M}$, which, by Lemma A.20, implies $|B| = 1$ and hence $t \sim s$. It follows that $t = \theta(t) = \theta(s) = s$, as desired.

(ii) We must show that $s \sim B \in \mathcal{B}_0$ and $s \sim B' \in \mathcal{B}_0$ implies $|\{t' \in B \mid t = \theta(t')\}| = |\{t' \in B' \mid t = \theta(t')\}|$. Note that, by Lemma A.17, $s \sim B \in \mathcal{B}_0$ and $s \sim B' \in \mathcal{B}_0$ implies $B \sim B'$. Hence, by Lemmas A.19 and A.20, there is a bijection $f : B' \to B$ such that $t' \sim f(t')$ for all $t \in B$. Thus, $t' \in B'$ such that $t = \theta(t')$ implies $\theta(f(t')) = \theta(t')$ for $f(t') \in B$, proving that $|\{t' \in B \mid t = \theta(t')\}| = |\{t' \in B' \mid t = \theta(t')\}|$. $Q.E.D.$

LEMMA A.28: $\eta_{x_t}(s) > 0$ *if and only if* $s \in x_t$.

PROOF: Note that $\eta_{x_s}(s) = 1$ for all $s \in T_1$. Suppose $s \in x_t$. Then, there exists $B \in \mathcal{B}_0$ such that $t \in B$ and $B \sim s$. Since $\theta(t) = t$, it follows that $\eta_{x_t}(s) > 0$. Conversely, if $\eta_{x_t}(s) > 0$, then there exists $B \in \mathcal{B}_0$ such that $s \sim B$ and $\theta(t') = t$ for some $t' \in B$. If $t \in B$, then $s \in x_t$ as desired. If not, let $B' = (B \setminus \{t'\}) \cup \{t\}$ and note that $B' \sim B$ by Lemmas A.16 and A.17 and hence, $s \in x_t$. $Q.E.D.$

LEMMA A.29: Z *is an attribute set*.

PROOF: By Lemma A.28,

$$X_t := \big\{ x \in Z \mid \eta_x(t) > 0 \big\} = \{ x \in Z \mid t \in x \}.$$

For all $t \in S$, there exists $B \in \mathcal{B}_0$ such that $t \sim B$ (by SR and Lemma A.22). Then, pick any $t' \in B$ and let $s = \theta(t') \in T_1$. If $s \in B$, then $t \in x_s$. Otherwise, let $B' = \{s\} \cup (B \setminus \{t'\})$. By Lemmas A.16 and A.17, $B' \sim B \sim t$ and hence $t \sim B'$ and again, $t \in x_s$. Therefore, $X_t \neq \emptyset$.

To prove that X_t is finite, we will show that $t \sim B \in \mathcal{B}_0$ implies $|X_t| \leq |B|$. Suppose $t \sim B \in \mathcal{B}_0$ and $|X_t| > |B|$. Let $B^* = \{\theta(s) \mid s \in B\}$. Then $|B^*| \leq |B| < |X_t|$. So there must exist at least one $s' \in T_1$ such that $x_{s'} \in X_t$ and $s' \notin B^*$. Since $x_{s'} \in X_t$, there exists some $A \in \mathcal{B}_0$ such that $s' \in A$ and $A \sim t$. Lemma A.17 implies $A \sim B$. Lemmas A.19 and A.20 imply there exists some $s'' \in B$ such that $s'' \sim s'$. But then $s' = \theta(s') = \theta(s'')$, which would imply $s' \in B^*$, a contradiction. $Q.E.D.$

LEMMA A.30: $s \sim t$, $s, t \in B_0 \in \mathcal{B}_0$ *implies* $\rho(s, B_0) = \rho(t, B_0)$.

PROOF: Choose $D \sim t \sim s$ such that $D \cap B_0 = \emptyset$ and let $C = B_0 \setminus \{s, t\}$. Then, $\rho(t', CDs) = \rho(t', CDt)$ for all $t' \in CD$. Hence $\rho(s, CDs) = \rho(t, CDt)$. Therefore, $\rho(s, B_0) = \rho(s, Cst) = \rho(s, CDs) = \rho(t, CDt) = \rho(t, Cst) = \rho(t, B_0)$ as desired. $Q.E.D.$

Lemmas A.27–A.29 imply that (w, η) is an attribute system. Let $\hat{\rho}$ be the choice rule that (w, η) represents. For $s \in T_1$, $X_s = \{x_s\}$ and by SR, $\{t \mid \theta(t) = s\}$ must be infinite. Hence, there are infinitely many archetypes for every $x_s \in Z$ and therefore $\hat{\rho}$ is complete. The following lemma completes the proof of the "only if" part of Theorem 2.

LEMMA A.31: $\rho = \hat{\rho}$.

PROOF: First, we show that $\rho_{s^*}(B_0) = \hat{\rho}_{s^*}(B_0)$ whenever $s^* \in B_0 \in \mathcal{B}_0$. As in Lemma A.25, let $B_0 = B^1 B^2 \cdots B^k$, where, for all $s \in B^i$ and $t \in B^j$, $s \sim t$ if and only if $i = j$. Hence, $B^1, \ldots, B^k$ is the partition on B_0 that $\sim$ yields. In the proof of Lemma A.25, we showed that $i \neq j$ implies $B^i \perp B^j$. Assume without loss of generality that $s^* \in B^1$. If $k = 1$, then $\rho_s(B_0) = \rho_t(B_0)$ for all $s, t \in B_0$ by Lemma A.30 and hence $\rho_{s^*}(B_0) = \frac{1}{|B_0|} = \hat{\rho}_{s^*}(B_0)$ as desired. If $k > 1$, let $B = \{\theta(s) \mid s \in B_0\}$ and let $\hat{s} = \theta(s^*)$. By definition,

$$\rho_s(B) = \frac{v_s}{\sum_{t \in B} v_t} = \frac{w_{x_s}}{\sum_{t \in B} w_{x_t}} = \hat{\rho}_s(B)$$

for all $s \in B$. Let $\hat{B} = \hat{s} B^2 \cdots B^k$ and $B^* = s^* B^2 \cdots B^k$. Since $B^1 \perp B^2 \cdots B^k$, E implies

$$\rho_{s^*}(B^*) = \rho_{\hat{s}}(\hat{B}) = \rho_{\hat{s}}(B) = \frac{w_{x_{\hat{s}}}}{\sum_{t \in B} w_{x_t}} = \hat{\rho}_s(B) = \hat{\rho}_{\hat{s}}(\hat{B}) = \hat{\rho}_{s^*}(B^*)$$

and $\rho_s(B^*) = \rho_s(B_0)$ for all $s \in B^2 \cdots B^k$, that is, $\sum_{s \in B^1} \rho_s(B_0) = \hat{\rho}_{s^*}(B^*)$. But then Lemma A.30 implies

$$\rho_{s^*}(B_0) = \frac{1}{|B^1|} \hat{\rho}_{s^*}(B^*) = \hat{\rho}_{s^*}(B_0)$$

as desired.

To complete the proof of the lemma, consider an arbitrary $A = \{s_1, \ldots, s_n\}$ and define C_i, A_i for $i = 0, \ldots, n$ inductively as follows: $C_0 = \emptyset$, $A_0 = A$. For $i > 0$, choose $C_i \in \mathcal{B}_0$ such that $C_i \sim s_i$ and $C_i \cap AA_{i-1} = \emptyset$ and let $A_i = C_i(A_{i-1} \setminus \{s_i\})$. Then, since $A_n \in \mathcal{B}_0$,

$$\rho_{s_1}(A) = \rho(C_1, A_1) = \cdots = \rho(C_1, A_n) = \hat{\rho}(C_1, A_n).$$

Finally, for each i, since $s_i \sim C_i \in \mathcal{B}_0$, we have $X_{s_i} = X(C_i)$ and therefore $X(A) = X(A_n)$, which together with equation (1) and the definition of η implies

$$\hat{\rho}(C_1, A_n) = \hat{\rho}_{s_1}(A)$$

as desired.   Q.E.D.

To prove the "if" part of Theorem 2, let ρ be a complete attribute rule that (w, η) represents and assume ρ satisfies SR. Let Z be the attribute set.

LEMMA A.32: *Z is infinite and, for every $x \in Z$, there is an infinite number of archetypes.*

PROOF: If Z is finite, then ρ has a denumerable image, contradicting part (i) of SR. Suppose some $x \in Z$ has a finite number of archetypes. Since Z is infinite, we can find some $y \in Z$ with $y \neq x$. Completeness guarantees that we can find two archetypes $s \neq s'$ for x, and an archetype t for y. Consider any $D \in \mathcal{A}_+$ which does not contain t nor contains any archetypes for x. If $\rho(s', ss't) = \rho(s', Ds't)$, then equation (1) implies $\rho(t, ss't) > \rho(t, Ds't)$. Hence, for any such D, we have $D \approx s$, contradicting part (ii) of SR.   *Q.E.D.*

LEMMA A.33: *$A \sim B$ if and only if $\eta^x(A) = \eta^x(B)$ for all $x \in Z$.*

PROOF: Sufficiency follows from equation (1). Next, suppose $\eta^x(A) > \eta^x(B)$ for some $x \in Z$. Lemma A.32 ensures that we can find $y \notin X(A) \cup X(B)$, an archetype t' for y such that $t' \notin AB$, and an archetype s' for x such that $s' \notin ABt'$. By equation (1), if $\rho(t', At') = \rho(t', Bt')$, then $\rho(s', As't') < \rho(s', Bs't')$, hence $A \approx B$, proving necessity.   *Q.E.D.*

LEMMA A.34: *$A \perp B$ implies $\eta^x(s) \cdot \eta^x(t) = 0$ for all $s \in A$, $t \in B$, and $x \in Z$.*

PROOF: Suppose $\eta^x(s) \cdot \eta^x(t) > 0$ for $s \in A$, $t \in B$, and $x \in Z$. By Lemmas A.32 and A.33, there are $A' \sim A$ and $B' \sim B$ such that every $s' \in A' \cup B'$ is an archetype. By Lemma A.33, there is $s' \in A'$ and $t' \in B'$, both archetypes for x. Let $B'' = (B' \setminus \{t'\}) \cup \{s'\}$. Now $B'' \cap A' \neq \emptyset$, $A' \sim A$, and by Lemma A.33, $B'' \sim B$, hence $A \not\perp B$.   *Q.E.D.*

E and WI follow easily from Lemma A.33, Lemma A.34, and equation (1). This completes the proof of the "if" part of Theorem 2.

A.3.1. *Proof of Proposition 1*

Verifying that $w = \alpha \cdot \hat{w}$ for $\alpha > 0$ and $\eta^x = r_x \cdot \hat{\eta}^x$ for $r_x > 0$ rational implies (w, η) and $(\hat{w}, \hat{\eta})$ represent the same choice rule is straightforward and omitted.

For the converse, let (w, η) with attribute set Z be any complete attribute system that represents ρ. It suffices to show that if $(\hat{w}, \hat{\eta})$ with attribute set $\hat{Z}$ is any other attribute system that represents ρ, then there exist $\alpha > 0$ and natural numbers $m_x > 0$ such that $\hat{w} = \alpha \cdot w$ and $\hat{\eta}^x = m_x \cdot \eta^x$. In particular, it must be $\hat{Z} = Z$. Let $X(A) = \{x \in Z \mid \eta^x(A) > 0\}$ and $\hat{X}(A) = \{x \in \hat{Z} \mid \hat{\eta}^x(A) > 0\}$.

CLAIM 1: *$A \perp B$ implies $\hat{X}(A) \cap \hat{X}(B) = \emptyset$.*

PROOF: Choose $C \sim B$ such that $C \cap AB = \emptyset$. The representation and E ensure that $\hat{X}(B) \subset \hat{X}(C) \setminus \hat{X}(A)$. Hence, $\hat{X}(A) \cap \hat{X}(B) = \emptyset$. *Q.E.D.*

CLAIM 2: $C \sim B$ *implies* $\hat{X}(C) = \hat{X}(B)$.

PROOF: Choose $A \neq \emptyset$ such that $A \perp B$. Repeating the argument in the proof of Claim 1 and Claim 1 itself ensure $\hat{X}(B) \subset \hat{X}(C) \setminus \hat{X}(A) = \hat{X}(C)$, and by symmetry, $\hat{X}(C) = \hat{X}(B)$. *Q.E.D.*

CLAIM 3: $C \sim B$ *if and only if* $\hat{\eta}^x(C) = \hat{\eta}^x(B)$ *for all* $x \in \hat{Z}$.

PROOF: By SR, there exists an infinite sequence B_i of pairwise disjoint sets such that $B_i \cap CB = \emptyset$ and $B_i \sim B$ for all i. By Claim 2, $\hat{X}(C) = \hat{X}(B) = \hat{X}(B_i)$ for all i. There exists a subsequence B_{i_j} of B_i such that $\hat{\eta}^x(B_{i_j})$ converges (possibly to infinity) for all $x \in \hat{X}(B)$. Without loss of generality, assume that the subsequence is the sequence itself.

By Claim 2 above, $\lim \hat{\eta}^x(B_i)$ is not zero for all $x \in X(C) = X(B)$. Let k_x be this limit and let $B^n = \bigcup_{i=2}^{n} B_i$. For now, assume $k_x < \infty$ for all $x \in X(C)$. Then, the discreteness of the range of $\hat{\eta}$ ensures that there exists n such that $\hat{\eta}^x(B_i) = k_x$ for all $i \geq n$. Again without loss of generality, assume this $n = 1$ so that $B_i \sim C$ for all i. Hence, we have

$$\rho(C, CB^n) = \rho(B, BB^n) = 1/n.$$

Since $(\hat{w}, \hat{\eta})$ represents ρ, the above equation yields

$$\sum_{x \in X(C)} \frac{\hat{w}_x}{\hat{w}(X(C))} \cdot \frac{\hat{\eta}^x(B)}{\hat{\eta}^x(B) + (n-1)k_x} = 1/n.$$

Some straightforward manipulations of the equation above yield

$$\sum_{x \in X(C)} \hat{w}_x \frac{\hat{\eta}^x(B) - k_x}{\hat{\eta}^x(B)/(n-1) + k_x} = 0.$$

Let $r_x = \hat{\eta}^x(B)/k_x$, then (i) let $n = 2$ and (ii) take the limit as n goes to infinity and divide by 2. The two cases, (i) and (ii), yield

$$\sum_{x \in X(C)} \hat{w}_x \frac{r_x - 1}{r_x + 1} = 0,$$

$$\sum_{x \in X(C)} \hat{w}_x \frac{r_x - 1}{2} = 0.$$

Comparing the two equations above, we note that whenever $r_x - 1 > 0$, the denominator of the terms in the top equation is larger, and whenever $r_x - 1 < 0$,

the denominator of the corresponding terms in the bottom equation is larger. Therefore, unless $r_x = 1$ for all x, the left-hand side of the bottom equation will be larger. Hence, we have $\hat{\eta}^x(B) = k_x$ for all x. By symmetry, the same holds of C.

To conclude the proof, we will show that $k_x = \infty$ is not possible. Suppose $k_x = \infty$ for some x. Let $Y = \{y \in X(C) \mid k_y < \infty\}$. Recall that the discreteness of the range of $\hat{\eta}$ ensures that there exists n such that $\hat{\eta}^y(B_i) = k_y$ for all $i \geq n$. If $Y = \emptyset$, let $n = 1$. Then, since $k_x = \infty$ for all $x \in X(C) \setminus Y$, there exists $m \geq n$ such that $\hat{\eta}^x(B_m) > \hat{\eta}^x(B_n)$ for all such x. Hence, B_m has at least as much as B_n of every attribute and strictly more of some attribute, contradicting the fact that B_m and B_n are duplicates.   *Q.E.D.*

CLAIM 4: *$A \perp B$ if and only if $\hat{\eta}^x(A) \cdot \hat{\eta}^x(B) = 0$ for all $x \in \hat{Z}$.*

PROOF: Suppose $\hat{\eta}^x(A) \cdot \hat{\eta}^x(B) = 0$ for all $x \in \hat{Z}$ and $A' \sim A$, $B' \sim B$. By Claim 3, $\hat{\eta}^x(A') = \hat{\eta}^x(A)$ and $\hat{\eta}^x(B') = \hat{\eta}^x(B)$ for all $x \in \hat{Z}$. Therefore, $A' \cap B' = \emptyset$ and hence $A \perp B$. The converse follows from Claim 1.   *Q.E.D.*

CLAIM 5: *$s \in T$ implies $\hat{X}_s$ is a singleton.*

Recall that T is the set of all s such that $\{s\}$ is fine. Lemma A.22 in the proof of Theorem 2 establishes that A is fine if and only if it is a finite, nonempty subset of T.

PROOF: Since (w, η) is complete, Claim 3 and Lemma A.32 imply that $s \in T$ if and only if s is an archetype for (w, η). Hence, the completeness of (w, η) implies $Z = \{x \mid \{x\} = X_s \text{ for some } s \in T\}$. For all $s \in T$, let x_s denote the unique attribute for (w, η) that contains s.

By Claims 3 and 4,

$$(A.5) \qquad x_s = \bigcup_{y \in \hat{X}_s} y$$

for all $s \in T$. Suppose $y, z \in \hat{X}_s$ and $y \neq z$. Then, without loss of generality, there exists $\emptyset \neq A \subset y \setminus z$. By Claim 3, there exists no B such that $s \in B \sim A$ and then, by Claim 4, $A \not\perp \{s\}$. But, then Claim 3 applied to the representation (w, η) ensures $\eta^{x_s}(A) > 0$. But, then Lemma A.32 ensures there exists $B \sim A$ such that $s \in B$, a contradiction. Hence, $y, z \in \hat{X}_s$ implies $y = z$.   *Q.E.D.*

CLAIM 6: *$Z = \hat{Z}$.*

PROOF: Since (w, η) is complete, equation (A.5) and Claim 5 ensure that $Z \subset \hat{Z}$. For the converse, take $A \subset x \in \hat{Z}$. By SR, there is a fine B such that

$A \sim B$ and hence by Claim 3, $s \in x$ for some $s \in T$. Hence, $\hat{X}_s = \{x\}$, and by equation (A.5), $x = x_s \in Z$. $Q.E.D.$

Choose any $s \in T$ and let $\alpha = \hat{w}_{x_s}/w_{x_s}$. Then, for any $y \in Z$, $y \neq x_s$, pick t such that $x_t = y$. Then,

$$\frac{w_{x_s}}{w_{x_s} + w_{x_t}} = \rho_s(\{s, t\}) = \frac{\hat{w}_{x_s}}{\hat{w}_{x_s} + \hat{w}_{x_t}} = \frac{\alpha \cdot w_{x_s}}{\alpha \cdot w_{x_s} + \hat{w}_{x_t}}.$$

Hence $\hat{w}_y = \hat{w}_{x_t} = \alpha \cdot w_{x_t} = \alpha \cdot w_y$ as desired.

Next, for all $x \in Z$, let $m_x = \hat{\eta}_s^x$ for some $s \in T$ such that $x_s = x$. If $t \in T$ and $x_s = x_t = x$, then $s \sim t$ and therefore $\rho_s(\{s, t\}) = 1/2 = \frac{\hat{\eta}_s^x}{\hat{\eta}_s^x + \hat{\eta}_t^x}$ and hence $\hat{\eta}_t^x = \hat{\eta}_s^x$, so m_x is well-defined.

Take any $x \in Z$ and $t \in S$ and choose $s \in T$ such that $x = x_s$. Then,

$$\frac{w_x}{w(\{s, t\})} \cdot \frac{1}{1 + \eta_t^x} = \rho_s(\{s, t\}) = \frac{\hat{w}_x}{\hat{w}(\{s, t\})} \cdot \frac{m_x}{m_x + \hat{\eta}_t^x}$$
$$= \frac{w_x}{w(\{s, t\})} \cdot \frac{m_x}{m_x + \hat{\eta}_t^x}.$$

Hence, $\hat{\eta}_t^x = m_x \cdot \eta_t^x$ as desired. This completes the proof of the proposition.

A.4. *Proof of Theorem 3*

In Section 3, we have established that $Q_\ell \subset Q_a$. Next, we will prove that $Q_r = \mathrm{cl} \, \mathrm{conv} \, Q_\ell$. Fact 1, below, requires no proof.

FACT 1: *The sets Q and Π are compact and convex.*

The next fact follows immediately from Fact 1 and the definition of "q maximizes π."

FACT 2: *If q_i maximizes π_i for $i = 1, 2$ and $\alpha \in [0, 1]$, then $\alpha q_1 + (1 - \alpha)q_2$ maximizes $\alpha \pi_1 + (1 - \alpha)\pi_2$.*

FACT 3: *The set Q_r is compact and convex.*

PROOF: That Q_r is convex follows from Facts 1 and 2 above. Next, we will prove that Q_r is compact. Falmagne[12] (1978) showed that $q \in Q_r$ if and only if

$$(\text{A.6}) \qquad 0 \leq \sum_{\{B | A \subset B\}} (-1)^{|B \setminus A|} q_{iB}$$

---

[12]Block and Marschak (1960) introduced the inequalities (A.6) and identified them as necessary conditions for $q \in Q$ to be an element of Q_r.

for all $i \in A$ and $A \in \mathcal{A}_+$. Let Q_0 be the subset of $\mathbb{R}_+^{n(2^n-1)}$ that satisfies the inequalities above. Clearly, Q_0 is closed, and by Falmange's theorem, $Q_r = Q \cap Q_0$. Since Q_r is the intersection of a closed and (by Fact 1) a compact set, it, too, is compact.   Q.E.D.

Block and Marschak (1960) were the first to prove the following well-known result:

FACT 4: $Q_\ell \subset Q_r$.

Facts 3 and 4 imply $\mathrm{cl\,conv}\,Q_\ell \subset Q_r$. The fact below establishes the reverse inclusion and yields $\mathrm{cl\,conv}\,Q_\ell = Q_r$.

FACT 5: *For every* $\varepsilon' > 0$ *and* $q \in Q_r$, *there exists* $\hat{q} \in \mathrm{conv}\,Q_\ell$ *such that* $|q - \hat{q}| < \varepsilon'$.

PROOF: Assume $0 < \varepsilon < 1$ and, for $u \in U$, define the Luce value $v^{\varepsilon u}$ such that

$$v_i^{\varepsilon u} = \varepsilon^{n-u_i}.$$

Let δ^u be the degenerate random utility that assigns probability 1 to u, let q^{δ^u} be the choice rule that maximizes δ^u, and let $q^{v^{\varepsilon u}}$ be the Luce rule that $v^{\varepsilon u}$ represents. It is easy to see that $q^{v^{\varepsilon u}} \to q^{\delta^u}$ as $\varepsilon \to 0$. It follows that $q := \sum_{u \in U} \pi_u q^{v^{\varepsilon u}} \in \mathrm{conv}\,Q_\ell$ converges to $q^* := \sum_{u \in U} \pi_u q^{\delta^u}$ as $\varepsilon \to 0$. Note that $\pi = \sum_{u \in U} \pi_u \delta^u$ and hence, by Fact 2 (and a simple inductive argument), $q^* = \sum_{u \in U} \pi_u q^{\delta^u}$ maximizes π. Thus, for every $q^* \in Q_r$, we can find $q \in \mathrm{conv}\,Q_\ell$ arbitrarily close to q^*.   Q.E.D.

LEMMA A.35: $Q_a \subset Q_r$.

PROOF: Let $q \in Q_a$ and let (w, η) be an attribute system that represents q. Let $\bar{v}_x = w_x$ for all $x \in Z$. Interpret the function $\bar{v}$ as a Luce value on Z and let $\bar{\rho}^{\bar{v}}$ be the choice rule that $\bar{v}$ represents, that is, for $\emptyset \neq \bar{A} \subset Z$,

$$\bar{\rho}_x^{\bar{v}}(\bar{A}) = \begin{cases} \bar{v}_x \Big/ \Big(\sum_{y \in \bar{A}} \bar{v}_y \Big), & \text{if } x \in \bar{A}, \\ 0, & \text{if } x \notin \bar{A}. \end{cases}$$

Let $K = |Z|$, let $\bar{U}$ be the set of all bijections from Z to $\{1, \ldots, K\}$, and let $\bar{\Pi}$ be the set of all probability distributions on $\bar{U}$. For $x \in Z$, $\emptyset \neq \bar{A} \subset Z$, define

$$[x\bar{A}] = \{ \bar{u} \in \bar{U} \mid \bar{u}(x) \geq \bar{u}(y) \text{ for all } y \in \bar{A} \}.$$

For $\bar{\pi} \in \bar{\Pi}$, define the choice rule $\bar{\rho}^{\bar{\pi}}$ such that, for all $x \in \bar{A} \subset Z$,

$$\bar{\rho}_x^{\bar{\pi}}(\bar{A}) = \sum_{\bar{u} \in [x\bar{A}]} \bar{\pi}(\bar{u}).$$

Applying Fact 4 to this new setting yields $\bar{\mu} \in \bar{\Pi}$ such that

$$(\text{A.7}) \qquad \bar{\rho}_x^{\bar{v}}(\bar{A}) = \sum_{\bar{u} \in [x\bar{A}]} \bar{\mu}(\bar{u}) = \bar{\rho}_x^{\bar{\mu}}(\bar{A})$$

for all $x \in \bar{A} \subset Z$. For any $\emptyset \neq A \subset S$ and $\bar{u} \in \bar{U}$, let

$$x_{A\bar{u}} = \arg\max_{x \in X(A)} \bar{u}(x)$$

and let $A_{\bar{u}} = \{i \in A \mid i \in x_{A\bar{u}}\}$. Then, let

$$(\text{A.8}) \qquad \rho_i^{\bar{u}}(A) = \begin{cases} \eta_i^{x_{A\bar{u}}} \Big/ \Big(\sum_{j \in A_{\bar{u}}} \eta_j^{x_{A\bar{u}}}\Big), & \text{if } i \in A_{\bar{u}}, \\ 0, & \text{otherwise.} \end{cases}$$

For any $\bar{\pi} \in \bar{\Pi}$, let $\rho^{\bar{\pi}} = \sum_{\bar{u} \in \bar{U}} \bar{\pi}(\bar{u})\rho^{\bar{u}}$. We will prove that $q \in Q_r$ by showing (1) $\rho^{\bar{u}} \in Q_r$ for all $\bar{u} \in \bar{U}$ and (2) $q = \rho^{\bar{\mu}}$. Note that (1) and (2) together establish that q is a convex combination of choice rules that are in Q_r, which, together with Fact 3 above, yields $q \in Q_r$.

Recall that U is the set of all bijections from $S = \{1, \ldots, n\}$ to S; Π is the set of all probabilities on U, and ρ^{π} is the choice rule that maximizes $\pi \in \Pi$.

CLAIM: *For every $\bar{u} \in \bar{U}$, there is $\pi \in \Pi$ such that $\rho^{\bar{u}} = \rho^{\pi}$.*

PROOF: Fix $\bar{u} \in \bar{U}$ and enumerate the attribute set $Z = \{x_1, \ldots, x_K\}$ such that $\bar{u}(x_k) = k$. Let $B^K = \{i \in S : \eta_i^{x_K} > 0\}$ be the set of objects that have attribute x_K, the highest ranked attribute according to $\bar{u}$. For each $k = 1, \ldots, K - 1$ let $B^k = \{i \in S : \eta_i^{x_k} > 0 \text{ and } \eta_i^{x_\ell} = 0 \ \forall \ell > k\}$, the set of objects that have attribute x_k and none of the attributes ranked higher than x_k. For each k such that $B^k \neq \emptyset$ let ρ^k be the Luce rule on B^k given by

$$\rho_i^k(A) = \frac{\eta_i^{x_k}}{\eta^{x_k}(A)}$$

for each $i \in A \subset B^k$, and $\rho_i^k(A) = 0$ for $i \in B^k \setminus A$. By Fact 4, each ρ^k maximizes a random utility π^k on B^k. If the utility function $u \in U$ is such that: (i) $u(i) > u(j)$ for each $i \in B^m$ and $j \in B^\ell$ with $m > \ell$; and (ii) $u(i) > u(j)$ if and only

if $u^k(i) > u^k(j)$ for $i, j \in B^k$, then let $\pi(u) = \prod_{k:B^k \neq \emptyset} \pi^k(u^k)$. Otherwise, let $\pi(u) = 0$. Note that π is a random utility on U and, for each $i \in A_{\bar{u}}$,

$$\rho_i^{\pi}(A) = \sum_{u \in [iA]} \pi(u) = \sum_{u^k \in [iA_{\bar{u}}]} \pi^k(u^k) = \rho_i^k(A_{\bar{u}})$$

for k such that $x^k = x_{A\bar{u}}$. Hence

$$\rho_i^{\pi}(A) = \frac{\eta_i^{x_k}}{\eta^{x_k}(A)} = \frac{\eta_i^{x_{A\bar{u}}}}{\eta^{x_{A\bar{u}}}(A)} = \rho_i^{\bar{u}}(A)$$

and, for $i \notin A_{\bar{u}}$, $\rho_i^{\pi}(A) = 0 = \rho_i^{\bar{u}}(A)$ as desired. $\hspace{2em}$ Q.E.D.

The claim above implies (1). To prove (2), let $[xX(A)] = \{\bar{u} \in \bar{U} \mid x_{A\bar{u}} = x\}$, $X_i = \{x \mid i \in x\}$, and $[X_iX(A)] = \{\bar{u} \in \bar{U} \mid x_{A\bar{u}} \in X_i\}$. Then, (A.7)–(A.8) imply

$$\rho_i^{\bar{\mu}}(A) = \sum_{\bar{u} \in \bar{U}} \bar{\mu}(\bar{u})\rho_i^{\bar{u}}(A) = \sum_{\bar{u} \in [X_iX(A)]} \bar{\mu}(\bar{u})\rho_i^{\bar{u}}(A)$$

$$= \sum_{\bar{u} \in [X_iX(A)]} \bar{\mu}(\bar{u})\frac{\eta_i^{x_{A\bar{u}}}}{\sum_{j \in A_{\bar{u}}} \eta_j^{x_{A\bar{u}}}} = \sum_{x \in X_i} \sum_{\bar{u} \in [xX(A)]} \bar{\mu}(\bar{u})\frac{\eta_i^x}{\sum_{j \in A \cap x} \eta_i^x}$$

$$= \sum_{x \in X_i} \frac{\eta_i^x}{\sum_{j \in A \cap x} \eta_j^x} \sum_{\bar{u} \in [xX(A)]} \bar{\mu}(\bar{u}) = \sum_{x \in X_i} \frac{\eta_i^x}{\sum_{j \in A \cap x} \eta_j^x} \bar{\rho}^{\bar{u}}(x, X(A))$$

$$= \sum_{x \in X(A)} \frac{\eta_i^x}{\sum_{j \in A \cap x} \eta_j^x} \frac{w_x}{\sum_{y \in X(A)} w_y} = q_{iA}.$$

$\hspace{30em}$ Q.E.D.

We conclude the proof of the theorem by showing that $\operatorname{conv} Q_\ell \subset \operatorname{cl} Q_a$. Since we have already established $\operatorname{cl} \operatorname{conv} Q_\ell = Q_r$, this will imply $Q_r \subset \operatorname{cl} Q_a$. Then, Lemma A.35 and Fact 3 yield $\operatorname{cl} Q_a = Q_r$ and conclude the proof.

Let q^j be a Luce rule for $j = 1, \ldots, m$ and let $q = \sum_{j=1}^m \alpha^j q^j$ for $\alpha_j \geq 0$ such that $\sum_{j=1}^m \alpha^j = 1$. Hence, for each j, there exists a Luce value v^j that represents q^j. We can choose rational-valued $\hat{v}^j$'s close to the corresponding v^j's and hence ensure that $\hat{q} = \sum_{j=1}^m \alpha^j \rho^{\hat{v}^j}$ is close to q. Then, we can multiply each $\hat{v}^j$ with a sufficiently large integer M so that Mv_i^j is an integer for all i, j. Let $Z = \{1, \ldots, m\}$, $w_j = \alpha^j$ and $\eta_i^j = Mv_i^j$ and note that (w, η) represents $\hat{q}$.

A.5. *Proof of Theorem 4 and Propositions 2–3*

Throughout the proofs in this section, we distinguish between the duplicates relation for a choice rule on S and a choice rule on Ω by letting $\sim_S$ denote the

former and letting $\sim_\Omega$ denote the latter. Similarly, let $Z_S \subset 2^S \setminus \{\emptyset\}$ denote the attribute set of a choice rule on S and $Z_\Omega \subset 2^\Omega \setminus \{\emptyset\}$ denote the attribute set for a choice rule on Ω.

THE RECURSIVE CONSTRUCTION: *All of the properties listed below are easy to verify*: *For any attribute system (w, η) on S, define its recursive extension $(\hat{w}, \hat{\eta})$ to Ω as follows*: *Let $x_y = \{\omega \in \Omega \mid \sigma(\omega) \cap y \neq \emptyset\}$. Hence, the node ω has attribute x_y if and only if one of its terminal nodes has attribute y. Set $Z_\Omega = \{x_y \mid y \in Z_S\}$. The function $y \to x_y$ is a bijection from Z_S to Z_Ω and $x \to y_x$, where $y_x = \{s \in S \mid s \in \sigma(t)$ for some $t \in x\} = S \cap x$ is its inverse. Let $\hat{w}_{x_y} = w_y$ and define $\hat{\eta}_s^{x_y}$ recursively for all $s \in \Omega$: if $s \in S$, then $\hat{\eta}_s^{x_y} = \eta_s^y$. For $t \in t_n$, $\hat{\eta}_t^{x_y} = \sum_{s \in t} \hat{\eta}_s^{x_y}$. Clearly, every attribute system on Z_S has a unique recursive extension to Z_Ω.*

Conversely, for every recursive $(\hat{w}, \hat{\eta})$ on Ω, there exists a unique (w, η) on S such that $(\hat{w}, \hat{\eta})$ is the recursive extension of (w, η). Moreover, if $(\hat{w}, \hat{\eta})$ is a recursive extension of (w, η) and represents (ϕ, Ω) an extension of (ϕ, S), then (w, η) represents (ϕ, S).

A.5.1. *Proof of Theorem 4*

By Theorem 2, there exists an attribute system (w, η) that represents (ϕ, Ω). To prove the "only if" part of the theorem, we will show that η is recursive. First, note that Claim 3 in the proof of Proposition 1 ensures that $t \sim_\Omega \hat{t}$ if and only if $\eta_t^x = \eta_{\hat{t}}^x$ for all $x \in Z$. Then, the proof of recursiveness follows from a straightforward inductive argument.

To prove that an SR dynamic rule that can be represented by a complete and recursive attribute system is consistent and satisfies E and WI, note that the latter two properties follow from Theorem 2. To prove consistency, note that the recursivity of η immediately implies that $\eta_D^x = \sum_{\omega \in D} \eta_\omega^x$. Hence, $\eta^x(\{D\}) := \eta_D^x = \sum_{\omega \in D} \eta_\omega^x = \eta^x(D)$. Then, appealing to Claim 3 in the proof of Proposition 1 again yields $\{D\} \sim_\Omega D$.

A.5.2. *Proof of Proposition 2*

To show the existence of an SR dynamic extension of any (1) SR attribute rule (ϕ, S) or (2) rich Luce rule (ϕ, S), let $(\hat{w}, \hat{\eta})$ be the recursive extension of (w, η), where in case (1), (w, η) is any complete attribute system that represents (ϕ, S), and in case (2), let $Z_S = \{\{s\} \mid s \in S\}$, $\eta_s^{\{s'\}} = 1$ if $s = s'$ and 0 otherwise and $w_{\{s\}} = v_s$ for some Luce value v that represents (ϕ, S). Then, let (ϕ, Ω) be the dynamic choice rule that $(\hat{w}, \hat{\eta})$ represents. The preceding observations ensure that (ϕ, Ω) is an extension of (ϕ, S). By Theorem 4, it is enough to show that (ϕ, Ω) is SR and $(\hat{w}, \hat{\eta})$ is complete.

Recall that there is a bijection $y \to x_y$ between Z_S and Z_y. Moreover, s is an archetype for y implies s is an archetype for x_y and so are $\{s\}$, $\{\{s\}\}$, etc. Hence, $(\hat{w}, \hat{\eta})$ every attribute has infinitely many archetypes. This proves completeness. Since (w, η) represents (ϕ, Ω), $\hat{\eta}_A^x = \hat{\eta}_B^x$ for all x implies $A \sim_\Omega B$.

With infinitely many archetypes, it is easy to check from the representation that the converse is also true: $\hat{\eta}_A^x = \hat{\eta}_B^x$ for all x if and only if $A \sim_\Omega B$. Therefore, B is a fine duplicate of A if and only if $\hat{\eta}_A^x = \hat{\eta}_B^x$ for x and B consists of archetypes. But since there are infinitely many archetypes, it follows that there are infinitely many fine duplicates, proving part (ii) of SR.

Suppose $A' \sim_\Omega A$ and $B' \sim_\Omega B$ and $A' \cap B' \neq \emptyset$. Then, we must have $\hat{\eta}_A^x \cdot \hat{\eta}_B^x > 0$ for some x. It follows that if $\hat{\eta}_A^x \cdot \hat{\eta}_B^x = 0$ for all x, then A and B are non-overlapping. In case (1) above, by Claim 4 of the proof of Proposition 1, (ϕ, S) is SR implies that $\eta_A^x \cdot \eta_B^x = 0$ if and only if A, B are (ϕ, S)-non-overlapping. Then, since (ϕ, S) is SR, for any A, $\delta \in (0, 1)$, there is B such that $\eta_{\sigma(A)}^x \cdot \eta_B^x = 0$, $\rho(\sigma(A), AB) = \delta$. Hence, $\hat{\eta}_A^x \cdot \hat{\eta}_B^x = 0$ and therefore $\rho(A, AB) = \delta$. Thus, (ϕ, Ω) also satisfies part (i) of SR.

In case 2, since (ϕ, S) satisfies R, for any $t \in \mathcal{T}$, $\delta \in (0, 1)$, there is $B \subset S$ such that $B \cap \sigma(t) = \emptyset$ and $\rho(\sigma(t), \sigma(t)B) = \delta$. Hence, $\frac{v(\sigma(t))}{v(\sigma(t))+v(B)} = \delta$. By the recursive construction, the fact that $B \cap \sigma(t) = \emptyset$ implies $\eta_{\sigma(t)}^x \cdot \eta_B^x = 0$ for all x and therefore $\hat{\eta}_t^x \cdot \hat{\eta}_B^x = 0$ for all x. Therefore, t is non-overlapping with B and hence $\rho_t(tB) = \frac{w(X_t)}{w(X_t)+w(X(B))} = \frac{v(\sigma(t))}{v(\sigma(t))+v(B)} = \delta$, proving that (ϕ, Ω) satisfies (i) of SR in case 2 as well.

For uniqueness, let (ϕ^1, Ω) and (ϕ^2, Ω) be two SR recursive attribute rules that are extension of the same (ϕ, S). By Theorem 4, both have complete recursive attribute systems, $(\hat{w}^1, \hat{\eta}^1)$, $(\hat{w}^2, \hat{\eta}^2)$ respectively, that represent them.

For $i = 1, 2$, let (w^i, η^i) be the unique attribute system on S that has the property that $(\hat{w}^i, \hat{\eta}^i)$ is its recursive extension. Since $(\hat{w}^i, \hat{\eta}^i)$ is complete, so is (w^i, η^i), and since both $(\hat{w}^i, \hat{\eta}^i)$'s are complete representations of the same SR attribute rule, by Proposition 1, $\hat{\eta}^1 = \hat{\eta}^2$ and there exists $\alpha > 0$ such that $w^1 = \alpha w^2$. But since $(\hat{w}^i, \hat{\eta}^i)$ is a recursive extension of (w^i, η^i), we have $\hat{w}^1 = \alpha w^2$ and $\hat{\eta}^1 = \hat{\eta}^2$. Then, Proposition 1 implies $(\phi^1, \Omega) = (\phi^2, \Omega)$.

If (ϕ, S) is a rich Luce rule, then the fact that $(\hat{w}^i, \hat{\eta}^i)$ is a recursive extension of (w^i, η^i) for every $x \in Z_\Omega$ implies that there must be an (ϕ^i, Ω)-archetype $s \in S$. But then, invoking recursiveness again, we note that $Z_\Omega^1 = Z_\Omega^2 = \{[s] \mid s \in S\}$ where $[s] := \{\omega \mid s \in \sigma(\omega)\}$ and $\hat{\eta}^1 = \hat{\eta}^2$. Let $v_s^i = w_{[s]}^i$ for all s and note that since both (w^i, η^i)'s are attribute systems for (ϕ, S), both v^i's are Luce values for (ϕ, S). Then, by the uniqueness of Luce values, $v^1 = \alpha v^2$ for some $\alpha > 0$ and hence $w^1 = \alpha w^2$ and finally, $\hat{w}^1 = \alpha \hat{w}^2$. By Proposition 1, $(\phi^1, \Omega) = (\phi^2, \Omega)$.

A.5.3. *Proof of Proposition 3*

Let (ϕ, Ω) be an invariant, strongly rich, complete, recursive attribute rule. We claim that $s, t \in S$ must be (ϕ, Ω)-non-overlapping. Assume the contrary and let (w, η) be a complete recursive attribute system that represents (ϕ, Ω). By strong richness, we can find $\emptyset \neq C \subset \Omega$ such that C and $\{s, t\}$ are non-overlapping. Then, Claim 1 in the proof of Proposition 1 ensures that $X(\{s\}) \cap X(C) = X(\{t\}) \cap X(C) = \emptyset$, where $X(D) = \{z \mid \eta_\omega^z > 0 \text{ for some } \omega \in D\}$. Let

$X = X(\{s\}) \cap X(\{t\})$, $a = w(X(\{s\}) \setminus X)$, $a' = w(X(\{t\}) \setminus X)$, $b = w(X)$ and $c = w(X(C))$. Also let $b_s = \sum_{x \in X} w(x) \cdot \frac{\eta_s^x}{\eta_s^x + \eta_t^x}$ and note that $b_s < b$.

Let $D = \{s, t, C\}$ and let $D' = \{t, \{s, C\}\}$. The representation ensures $p_s(D) = \phi_s(D) = \frac{a+b_s}{a+a'+b+c}$ while $p_s(D') = (1 - \rho_t(D')) \cdot \phi_s(\{s, C\}) = \frac{a+b_s+c}{a+a'+b+c} \cdot \frac{a+b}{a+b+c}$. Then, since $b_s < b$, verifying that $p_s(D) \neq p_s(D')$ is straightforward. This contradicts the fact that ϕ is invariant.

Since s, t are non-overlapping for all $s, t \in S$, (ϕ, S) satisfies independence. Since (ϕ, Ω) is strongly rich, recursive and $\{s\} \perp \{t\}$ whenever $s, t \in S$ and $s \neq t$, (ϕ, S) must be rich. Therefore, Theorem 1 ensures that (ϕ, S) is a Luce rule.

Next, assume (ϕ, S) is a rich Luce rule and $D \in \mathcal{D}$ is a regular decision problem. Then, for any $s \in S$ and path $(\omega_0, \ldots, \omega_n)$ of D ending in s, that is, path such that $\omega_0 = s$ and $\omega_n = D$, the recursive property yields

$$p_s(D) = \phi\big(s, \sigma(\omega_1)\big) \cdot \phi\big(\sigma(\omega_1), \sigma(\omega_2)\big) \cdots \phi\big(\sigma(\omega_{n-1}), \sigma(D)\big)$$

$$= \frac{v_s}{v(\sigma(\omega_1))} \cdot \frac{v(\sigma(\omega_1))}{v(\sigma(\omega_2))} \cdots \frac{v(\sigma(\omega_{n-1}))}{v(\sigma(D))}$$

$$= \frac{v_s}{v(\sigma(D))}$$

$$= \phi_s\big(\sigma(D)\big),$$

proving that (ϕ, Ω) is invariant.

REFERENCES

BEN-AKIVA, M., AND S. R. LERMAN (1985): *Discrete Choice Analysis: Theory and Applications to Travel Demand*. Cambridge, MA: MIT Press. [1875,1886]

BLOCK, H. D., AND J. MARSCHAK (1960): "Random Orderings and Stochastic Theories of Responses," in *Contributions to Probability and Statistics*, ed. by I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, and H. Mann. Stanford, CA: Stanford University Press, 97–132. [1875,1885,1905,1906]

DEBREU, G. (1960): "Review of 'Individual Choice Behavior: A Theoretical Analysis' by R. D. Luce," *American Economic Review*, 50, 186–188. [1874,1878]

FALMAGNE, J. C. (1978): "A Representation Theorem for Finite Random Scale Systems," *Journal of Mathematical Psychology*, 18 (1), 52–72. [1875,1884,1905]

FUDENBERG, D., AND T. STRZALECKI (2013): "Recursive Stochastic Choice," Working Paper, Department of Economics, Harvard University. [1875]

HEY, J., AND C. ORME (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62 (6), 1291–1326. [1873]

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, 90–98. [1891]

LUCE, R. D. (1959): *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley. [1873]

——— (1977): "The Choice Axiom After Twenty Years," *Journal of Mathematical Psychology*, 15, 215–233. [1874]

LUCE, R. D., AND P. SUPPES (1965): "Preference, Utility, and Subjective Probability," in *Handbook of Mathematical Psychology*, Vol. 3, ed. by R. D. Luce, R. R. Bush, and E. H. Galanter. New York: Wiley, Chapter 19, 249–410. [1875,1885,1886]

MANZINI, P., M. MARIOTTI, AND L. MITTONE (2010): "Choosing Monetary Sequences: Theory and Experimental Evidence," *Theory and Decision*, 69, 327–354. [1873]

MACHINA, M. J. (1989): "Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty," *Journal of Economic Literature*, 27 (4), 1622–1668. [1889]

MCFADDEN, D. (1978): "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, ed. by S. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull. Amsterdam: North-Holland, 75–96. [1875,1885,1886]

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. [1883]

SAVAGE, L. J. (1954): *The Foundations of Statistics*. New York: Wiley. [1878]

SIMONSON, I. (1989): "Choice Based on Reasons: The Case of Attraction and Compromise Effects," *Journal of Consumer Research*, 16, 158–174. [1891]

SIPPEL, R. (1997): "An Experiment on the Pure Theory of Consumer's Behaviour," *The Economic Journal*, 107, 1431–1444. [1873]

TRAIN, K. (2009): *Discrete Choice Methods With Simulation* (Second Ed.). Cambridge: Cambridge University Press. [1875,1886]

TVERSKY, A. (1972): "Elimination by Aspects: A Theory of Choice," *Psychological Review*, 79 (4), 281–299. [1875,1878,1885,1886]

WEN, C.-H., AND F. S. KOPPELMAN (2001): "The Generalized Nested Logit Model," *Transportation Research Part B: Methodological*, 35 (7), 627–641. [1886]

Dept. of Economics, Princeton University, Princeton, NJ 08544, U.S.A.; fgul@ princeton.edu,

Dept. of Economics, Washington University in Saint Louis, Saint Louis, MO 63130, U.S.A.; pnatenzon@wustl.edu,

and

Dept. of Economics, Princeton University, Princeton, NJ 08544, U.S.A.; pesendor@princeton.edu.