Random Choice as Behavioral Optimization†

Faruk Gul
Paulo Natenzon
and
Wolfgang Pesendorfer‡

November 2013

Abstract

We develop a new model of random choice 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. Our new model, the attribute rule, addresses well-known difficulties of the Luce model that arise when objects have common attributes. We show that the existence of a well-defined stochastic preference over attributes characterizes our model. We show that the set of attribute rules and random utility maximizers is essentially the same. Finally, we show that Luce and attribute models have a unique consistent extension to dynamic problems.

† This research was supported by a grant from the National Science Foundation (SES-1060073).

‡ Faruk Gul, Department of Economics, Princeton University, fgul@princeton.edu; Paulo Natenzon, Department of Economics, Washington University in St. Louis, pnatenzon@wustl.edu; Wolfgang Pesendorfer, Department of Economics, Princeton University; pesendor@princeton.edu.
1. Introduction

In empirical and experimental studies, individual choice behavior is almost always interpreted probabilistically. This interpretation is obviously appropriate when the data represents the choices of a group of observationally indistinguishable decision makers. But even when analyzing a single individual, deterministic theories may be inadequate.

For example, in experimental settings, subjects routinely violate the weak axiom of revealed preference (WARP) and, therefore, cannot be thought of as maximizing a deterministic utility function. However, deterministic theories that permit violations of WARP are unlikely to find much empirical support; while it is reasonable to expect that a particular decision maker may choose 1 from the set \{1, 2\} on one occasion and choose 2 from \{1, 2, 3\} on another occasion, it is unlikely that this decision maker would choose 1 from \{1, 2\} and 2 from \{1, 2, 3\} on every occasion. To put it differently, a deterministic theory may be too crude for measuring an individual’s tendency to choose 1 over 2.

There are many reasons why a simple deterministic choice model of the sort studied in graduate demand theory courses might be too crude 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 violations of WARP that are observed even when none of these factors are present. Hence, our goal is to develop random choice as a theory of behavioral optimization, that is, not as a model of measurement error but as a behavioral model of the consumer.

Thurstone (1927) was the first person to observe that individual behavior inconsistent with any deterministic explanation might still be amenable to analysis with a stochastic model of behavior. To facilitate this analysis, he proposed what later became known as the probit model. Subsequent research related to Thurstone’s work can be described as the random utility approach to random choice.

In this formulation, the decision maker is defined by a fixed probability distribution over utility functions (with no ties) and each time she faces a choice problem, she selects a utility function from that distribution and chooses the element from the option set that maximizes it. Thus, the Thurstone approach offers a convenient device for analyzing
behavioral optimization. An individual behaves “as if” she is a collection of agents, each of whom satisfies WARP, but at any given time a random agent from the collection is asked to make the choice. The only parameter of the model is a distribution of utility functions, or alternatively, the joint distribution of the utilities of all objects. Though the random utility model if flexible, it can be difficult to identify a priori plausible restrictions on the joint distribution of utility profiles\(^1\) or even more difficult to translate restrictions on the random choice rule to restrictions on the distribution of random utilities.

Perhaps motivated by this fact, Luce (1959) introduces what was later shown to be a more restrictive but more “intuitive” approach with far fewer parameters. In Luce’s formulation, there exists a strictly positive number, \(v_s\), for every option, \(s\), so that the probability of choosing \(s\) from \(A\) is

\[
\rho_s(A) := \frac{v_s}{\sum_{t \in A} v_t}
\]

whenever \(s \in A\). We refer to the numbers \(v_s\) as Luce values and to the corresponding random choice rule as a Luce rule. We can extend \(v\) to an additive (or countably additive) function on appropriate collection of sets of choice objects; that is, let \(v(A) = \sum_{s \in A} v_s\) and rewrite (l) as follows:

\[
\rho_s(A) := \frac{v_s}{v(A)}
\]

It is possible to interpret an option’s Luce value as a measure of its desirability; an option \(s\) is stochastically preferred to \(t\) if \(v_s \geq v_t\) in the sense that \(\rho_s(\{s\} \cup A) \geq \rho_t(\{t\} \cup A)\) whenever \(s, t \notin A\). Hence, each Luce rule contains its own fully rational benchmark. Such a benchmark facilitates the measurement of the decision maker’s behavioral limitations.

These attractive features notwithstanding, the empirical literature on random choice has documented systematic violations of the Luce model. Debreu (1960) anticipates the best known such violation and identifies 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

\(^1\) Natzenon (2013) offers a dynamic extension of the probit model to facilitate such restrictions.
of the other two options. The problem that Debreu’s example identifies is more generally referred to as the “duplicates problem” in the discrete choice estimation literature.

We develop a new model of random choice that addresses this and other shortcomings of the Luce model without completely abandoning Luce’s intuitive approach. That is, we study a class of choice rules that can be described with desirability values but also permit duplicates. Our new model, the attribute rule is essentially as general as the random utility model, deals with the duplicates problem and preserves some of the convenience of the Luce approach by reinterpreting the choice objects as bundles of attributes. To see how our model works and how it relates to the Luce model, let $Z$ be the collection of attributes and let $X_s$ be the set of attributes that $s$ has and finally, 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, $\eta$, maps attribute and option pairs to natural numbers. We let $w(X) = \sum_{x \in X} w_x$ and $\eta^x(A) = \sum_{s \in A} \eta^x_s$. We say that the random choice rule $\rho$ is an attribute rule if there is an attribute system $(w, \eta)$ such that

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

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. Call two options non-overlapping if they have no common relevant attributes. The Luce rule is the special case of the attribute rule in which all pairs of distinct options are non-overlapping.

Note that attributes, or at least their relevance, are subjective; they are properties of the decision maker and not of the objects. We infer that options $s_1$ and $s_2$ have common attributes from the fact that the decision maker treats them like a single option (and not because they happen to both be buses.) The main contribution of this paper is to provide a model that derives a collection of attributes from observed random choices.

Debreu’s example suggests two (relevant) attributes: the bus-attribute, $x$, and the train-attribute $y$. The buses, $s_1, s_2$, have all attributes in common and hence are duplicates. We identify $s_1, s_2$ as duplicates by noting that the decision maker chooses $t \in A$ from
\{s_1\} \cup A with the same probability that she chooses the same \( t \) from \{s_2\} \cup A for every
\( A \) such that \( A \cap \{s_1, s_2\} = \emptyset \). Note that the set \( \{s_1, t\} \) contains \( s_1 \), a duplicate of \( s_2 \), and
an alternative, \( t \), that does not overlap with \( s_1 \). The first substantive axiom of our model
is the assertion that adding a duplicate does not affect the probability of choosing a non-
overlapping option. We call this property elimination of duplicates. In Debreu’s example,
elimination of duplicates means that the probability of choosing the train is unchanged if
the red bus is added to an option set that already contains a yellow bus.

The attribute rule can also address versions of the duplicates problem that are less
stark. For example, it may be that adding more buses slightly reduces the probability of
choosing a train instead of leaving it unchanged. An attribute rule in which buses share a
common attribute but are not exact duplicates—for example, one in which color is relevant
but less important—would yield such a choice rule.

Let \( \rho(A, B) = \sum_{s \in A} \rho_s(B) \). Our only other substantive assumption, independence,
is a property that attribute rules share with Luce rules: for every \( C, D \) that do not overlap
with \( A \) and \( B \),

\[ \rho(A, A \cup C) \geq \rho(B, B \cup C) \implies \rho(A, A \cup D) \geq \rho(B, B \cup D). \]

The Luce rule entails the implicit additional hypothesis that two disjoint sets of options
are always non-overlapping.

Our first result, Theorem 1, shows that in a rich setting, with sufficient variety in
options and option sets, independence characterizes the Luce rule. That is, in a rich
environment, a random choice rule satisfies independence if and only if it is a Luce rule.
In sparse settings, other rules may satisfy independence but such rules cannot be extended
to environments with a large variety of options and option sets in a way that preserves
independence.

Call an option an archetype if it has a unique attribute. We say that an attribute
rule is complete if each attribute has at least two archetypes. Our main result, Theorem
2, shows that if the random choice rule is rich\(^2\) then it satisfies elimination of duplicates
and weak independence, if and only if it is a complete attribute rule.

\(^2\) As before, richness is a requirement on the variety of potential choices and choice sets.
Theorems 1 and 2 reveal that the possibility of overlap together with the elimination of duplicates is what distinguishes the two and enables the latter to resolve the duplicates problem.

Most econometric models of discrete choice such as the logit, the probit, the nested logit, etc., are special examples of random utility models. Formally, a random utility is a probability distribution over utility functions and a random choice rule is a random utility maximizer if there is a random utility such that the probability of choosing \( s \) from \( A \) is equal to the probability of drawing a utility function that attains its maximum in \( A \) at \( s \).

Theorem 3 relates Luce rules and attribute rules to random utility maximizers. We show that every attribute rule is a random utility maximizer and that the set of random utility maximizers is the closure of the set of attribute rules. Hence, every random utility maximizer can be approximated by an attribute rule.

Finally, in section 5 we study dynamic choice problems. The duplicates problem is especially relevant in dynamic settings because a single choice alternative may be available in multiple continuation problems; hence continuation problems overlap even if choice objects are non-overlapping. Formally, the set dynamic choice problems are a subset of static choice problems; in particular, these are static choice problems in which the choice objects are structured. In Theorem 4, by incorporating one additional assumption, consistency, we take advantage of this structure and show that a strongly rich attribute rule is recursive if and only if it satisfies consistency. We show that every rich Luce rule has a unique consistent extension to dynamic choice problems and this extension is a strongly rich attribute rule. In a dynamic choice problem, the agent must make more than one choice. For the case where choice probabilities are history independent, we analyze the effect of the dynamic structure on outcome probabilities in the context of a simple class of decision problems.

1.1 Related Literature

Block and Marschak (1960) show that Luce rules are random utility maximizers. Holman and Marley (see Luce and Suppes, 1965) and McFadden (1978) use the Gumbel distribution to construct a random utility for the Luce model. Their construction facilitates the estimation of Luce values as a function of (observable) background parameters. Falmagne (1978) characterizes 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.

Tversky (1972) introduces the elimination by aspects (EBA) model to address shortcomings of the Luce model. The EBA model assigns aspects of differing values to alternatives and postulates the following stochastic algorithm: with a probability proportional to its value, the decision maker chooses an attribute (or aspect) and eliminates all options that do not have it. Next, the decision maker chooses a second attribute and again eliminates all options that do not have it. This process is repeated until a single alternative is left. In each round, the decision maker considers only those attributes that at least one of the remaining options possesses. In contrast, the attribute rule corresponds to the following stochastic algorithm: with a probability proportional to its value, the decision maker picks a single attribute and chooses an option that has this attribute with probability proportional to the intensity with which it has this attribute. Thus, an attribute rule chooses an attribute among those represented in the choice set according to a Luce-type process and then chooses among all options that share the attribute according to a different Luce-type process while EBA uses other attributes to break ties. Both attribute rules and EBAs are special cases of random utility maximizers. Hence, Theorem 3 below ensures that all EBAs can be approximated by some attribute rule. The converse, however, is not true. There are attribute rules that cannot be approximated by an EBA.

The models discussed above belong to a large class that we call behavioral optimization models. Such models weaken standard economic rationality requirements (i.e., WARP). Strotz’ (1955) model of consistent planning and Simon’s (1978) model of satisficing behavior are two well-known behavioral optimization models. More recent examples include the Pollak (1968) and Peleg and Yaari (1973) models of consistent planning, Köszegi and Rabin’s (2006) model of status quo dependent choice, Manzini and Mariotti’s (2007) model of sequentially rationalizable choice and Salant’s (2009) model of bounded rationality.
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 $\rho : \mathcal{A} \times \mathcal{A}_+ \rightarrow [0, 1]$ is a (random) choice rule if for all $A \in \mathcal{A}_+$, $\rho(\cdot, A)$ is countably additive and

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

The equation $(rcr)$ is the feasibility constraint; $\rho$ must choose among options available in $A$. Countable additivity is the requirement that $\rho(\cdot, A)$ is a probability. In particular, when $\mathcal{A}$ is the collection of all 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 well-defined. Formally, a fixed the set $a$ is stochastically preferred to $b$ 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 requires this binary relation to be complete.

**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, there is a binary relation $\succeq_\rho$ on $\mathcal{A} \times \mathcal{A}$ that represents stochastic preference. 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 \( v : \mathcal{A} \to \mathbb{R}_{++} \) is \( v_s = v(\{s\}) \). Such a \( v \) 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, \( \rho \), a *Luce rule* if there exists a Luce value \( v \) such that

\[
\rho_s(A) = \frac{v_s}{v(A)} \tag{1}
\]

whenever \( s \in A \in \mathcal{A} \). We say that the Luce value \( v \) *represents* \( \rho \) if equation (1) 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

\[
A \succeq_\rho B \text{ if and only if } v(A) \geq v(B)
\]

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, the Luce rule is the only choice rule that satisfies independence and, therefore, the only rule that admits a stochastic preference. Next, we state the richness requirement.

**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 ensure that every option is chosen with positive probability in some degenerate choice set. Together with independence, richness ensures that every option is chosen with positive probability from every option set that contains it. The following example is a rich choice rule.

**Example 1:** Let \( S = (0, 1) \), let \( \mathcal{A} = \mathcal{A}_f \), (i.e., all finite subsets of \( S \)) and let \( v(A) = \sum_{s \in A} s \) for all \( A \in \mathcal{A}_+ \). It is straightforward to verify that the Luce rule that \( v \) represents is rich.

**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 rules \( v, \hat{v} \) represent the same Luce value if and only if \( \hat{v} = \alpha v \) for some \( \alpha > 0 \). We interpret richness as a technical assumption analogous to Savage’s small event continuity (P6). Both assumptions are idealizations that facilitate
the calibration of probabilities. In a sparse setting, there may be choice rules that satisfy independence and do not correspond to a Luce rule. 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 $A$ that are summable. That is, $\mathcal{A} = \{ A \subset S \mid \Sigma_{s \in A} s < \infty \}$. It is easy to verify that $\mathcal{A}$ is a proper collection. Let $v(A) = \Sigma_{s \in A} s$ and let $\rho$ be the Luce rule that $v$ represents. In Appendix A, we demonstrate that $\rho$ 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

$$\rho_{s_i}(\{s_i, t\}) = .6; \rho_{s_i}(At) = .2$$

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 than options that don’t share attributes for the decision maker’s attention.\(^3\) In the above example, alternatives in the set $A$ share the same 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.

---

\(^3\) Tversky calls 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. This situation might be described by the following choice rule:

\[
\rho_r(\{r, s\}) = \frac{3}{4}; \rho_r(\{r, t\}) = \frac{3}{4}; \rho_r(\{r, s, t\}) = \frac{1}{2}
\]

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 \( \rho_r(\{r, s, t\}) = \frac{3}{5} \). In this example, the set \( \{s, t\} \) is a duplicate of \( r \). Notice that this example requires two attributes; \( r \) and \( s \) share the first, while \( r \) and \( t \) share the second.

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 bus’ color or to the name of the airline. 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 $s_1$ and the one-destination airline $s_2$ overlap because $B = \{s_2, s_3\}$ is a duplicate of $s_1$ and $B \cap \{s_2\} \neq \emptyset$.

**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 alternatives from $A$ to a choice set that contains $B'$ does not alter the odds of choosing options that have no 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, requires independence to hold 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\(^4\) 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

\(^4\) For a definition of strong richness, see below.
an intensity defines 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 \( \rho \), we can simply write \((\rho, S)\). Let \( |A| \) be the cardinality of the set \( A \).

An attributes set is a nonempty set \( Z \) and an attribute intensity is a function \( \eta : Z \times S \to \mathbb{N} \cup \{0\} \). An elements \( x \in Z \) is naturally identified with a nonempty subset of \( S \); those objects that have the attribute \( x \). To avoid confusion between these subsets of \( S \) and elements of \( \mathcal{A}_+ \), we will use \( x, y, z \) to denote the former. We write \( \eta^x_s \) rather than \( \eta(x, s) \) and let \( \eta^x(A) = \sum_{s \in A} \eta^x_s \). We say that option \( s \) has attribute \( x \) if \( \eta^x_s > 0 \). Hence, \( \eta^x > 0 \) if and only if \( s \in x \). We assume that each object has a nonempty and finite set of attributes; that is, for all \( s \in S \), there exists \( x \in Z \) such that \( \eta_s(x) > 0 \) and \( X_s = \{ x \mid \eta^x_s > 0 \} \) is finite.

We let \( \eta^x(A) := \sum_{s \in A} \eta^x_s \) and \( X(A) := \{ x \in Z \mid \eta^x(A) > 0 \} \) for all nonempty \( A \in \mathcal{A} \). We call \( X(A) \) the set of attributes that are active in \( A \); that is, these are the attributes that at least one member of \( A \) has.

An attribute value is a function \( w : Z \to \mathbb{R}^+ \). 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 \((w, \eta)\) an attribute system.

We say that \( \eta \) 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 \( \eta^x_s = 1 \) 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 \( \rho \) is a (complete) attribute rule if there exists a (complete) attribute system \((w, \eta)\) such that

\[ \rho_s(A) = \sum_{x \in X(A)} \frac{w_x}{w(X(A))} \cdot \frac{\eta^x_s}{\eta^x(A)} \]  

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

**Example 3:** To define an attribute rule that is consistent with example 3 above, let \( Z = \{1, 2\} \) where 1 is the bus-attribute and 2 is the train-attribute. Let \( w_1 = 3, w_2 = 2 \)
and let \( \eta \) be the simple intensity such that \( \eta^1_s = 1 \) if and only if \( s \in A \) and \( \eta^2_s = 1 \) if and only if \( s = t \). Then, \( \rho_s(\{s_i, t\}) = .6 \) and \( \rho_s(At) = .2 \), as required.

**Example 4:** For an attribute rule that is consistent with example 4, let \( Z = \{1, 2\} \) where 1 and 2 are the two destinations. Set \( w_1 = w_2 = 1 \) and let \( \eta \) be the simple attribute intensity such that \( \eta^1_r = \eta^1_s = 1 \), \( \eta^2_r = \eta^2_t = 1 \) and \( \eta^x_s = 0 \) in all other cases. Then, \( \rho_r(\{r, s\}) = 3/4 \) 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 \( \eta \) such that \( \eta^x_s = 1 \) if and only if \( x = s \). Then, (1) can be restated as follows:

\[
\rho_s(B) = \frac{w_x}{w(X(B))} = \frac{v_s}{v(B)}
\]

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 \( A_1A_2\cdots A_n = A \sim B \) there is a partition \( B_1B_2\cdots B_n = B \) 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.\(^{5}\) Let \( \mathcal{M} = \{A \in \mathcal{A} | A \) 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 \( D \sim A \).

The following example satisfies all assumptions of Theorem 2.

---

\(^{5}\) The proof of Theorem 2 reveals that when \( \rho \) is an attribute rule, an option set \( a \) is fine if and only if it consists of archetypes.
Example 5: Let $S = \mathbb{N} \times \mathbb{R}_{++}$ and let $Z = \mathbb{R}_{++}$. For $s = (i, r) \in A$, let $w_{ir} = r$ and let the simple attribute value $\eta$ be such that $\eta_{ir'} = 1$ 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 $(w, \eta)$ is a complete attribute value and that the $\rho$ it represents satisfies strong richness.

The choice rule above is an example of a particular type of attribute rule that we call \textit{extended Luce rules}: take any Luce rule $\rho$ on some set of alternatives $S$. Then, let $S_e = I \times S$ where $I$ is any infinite set. Then, set $Z = S$, $w_s = v_s$ and $\eta_{is}^{s} = 1$ and $\eta_{is'}^{s} = 0$ if $s \neq s'$. Thus, $S_e$ has $|I|$ many duplicates of each $(i, s)$ while each $S_i = \{(i, s) \mid s \in S\}$ is a “copy” of the original $S$; that is, if $\rho^0$ is the original Luce rule that $\rho_i(A_i) = \rho^0_s(A)$ for all $s \in A \subset S$ where $A_i = \{(i, s) \mid s \in A\}$. It is easy to see that if $\rho^0$ is a rich Luce rule, then $\rho$ 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.

\textbf{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 weak independence and elimination of duplicates but those rules cannot be extended without violating one or both of these properties. Note that completeness is itself a richness requirement; one that is specific to attribute rules. Any attribute rule can be completed by adding more options (at most two archetypes per attribute). Completeness also ensures that two options are duplicates only if they have all attributes in common. The “if” part of Theorem 2 shows that if a rich attribute rule is complete (or once it is completed), it satisfies weak independence and elimination of duplicates. Without strong richness, there can be many representations for the same attribute rule. However, strong richness yields a strong uniqueness result:

\textbf{Proposition 1:} The attribute systems $(w, \eta)$ and $(\hat{w}, \hat{\eta})$ represent the same strongly rich choice rule if and only if $\hat{w} = \alpha \cdot w$ and $\hat{\eta}_x = \beta_x \cdot \eta_x$ for $\alpha, \beta_x > 0$. 

14
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 $\eta$ is unique up to a positive constant for each $x$. An immediate consequence of Proposition 1 is that for complete attribute systems, $w$ is unique up to a positive multiplicative constant and $\eta$ is unique.

If $\eta$ is simple, the choice rule that $(w, \eta)$ represents satisfies weak stochastic transitivity:

$$\rho_r(\{r, s\}) > \frac{1}{2} \text{ and } \rho_r(\{s, t\}) > \frac{1}{2} \implies \rho_r(\{r, t\}) > \frac{1}{2}$$

There is a good deal of evidence suggesting that choice rules may violate this property.\textsuperscript{6} The following example illustrates how a non-simple attribute rule can accommodate failures of weak stochastic transitivity.

**Example 6:** Let $A = \{r, s, t\}$ and assume there are three attributes, $Z = \{1, 2, 3\}$. Each attribute value is 1, i.e., $w_x = 1$ for all $x \in Z$. Option $r$ has attributes 1, 2, $s$ has attributes 2, 3 and $t$ has attributes 1, 2. In particular, $\eta^1_r = \eta^2_s = \eta^3_t = 4$, $\eta^3_r = \eta^1_s = \eta^2_t = 1$ and $\eta^2_r = \eta^3_s = \eta^1_t = 0$. This attribute rule represents the choice rule $\rho$ such that

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

In this example, binary comparisons are “won” by the alternative with the higher attribute value on the common attribute. For example, $r$ is chosen over $s$ with probability $3/5$ because $\eta^1_r > \eta^1_s > 0$. 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 $r$.

\textsuperscript{6} For a detailed review of this evidence, see Rieskamp, Busemeyer and Mellers (2006).
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 \textit{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 $A_+$ be the set of all non-empty subsets of $S$. In this case, a choice rule $\rho$ can be identified with a vector $q \in \mathbb{R}_+^{2^n - 1}$ where $q_{iA} = \rho(\{i\}, A)$ for all $A \in A_+$. Such a vector $q$ satisfies

\begin{align*}
q_{iA} &\leq 1 \\
q_{iA} &> 0 \text{ implies } i \in A \\
\sum_{i \in S} q_{iA} &= 1
\end{align*}  \tag{2}

Let $Q$ be the set of all $q \in \mathbb{R}_+^{2^n - 1}$ that satisfy equation (2). Let $Q_l$ be the subset of $Q$ corresponding to Luce rules and $Q_a$ be the subset corresponding to attribute rules.

One class of extensively studied choice rules are \textit{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 $i \in A \in A_+$, let

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

A function $\pi : U \to [0, 1]$ is a \textit{random utility} if $\sum_{u \in U} \pi(u) = 1$. We identify each such function with an element in $\mathbb{R}_+^{\lvert U \rvert}$. Let $\Pi = \{\pi \in \mathbb{R}_+^{\lvert U \rvert} \mid \sum_{u \in U} \pi_u = 1\}$ be the set of all random utilities. Hence, $\Pi$ is the $\lvert U \rvert - 1 = n! - 1$-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 $\pi$ if $q_{iA} = \sum_{u \in [iA]} \pi_u$ for all $i, A$.  

16
Let $Q_r$ denote the set of random utility maximizers, that is, the set of choice rules that maximize some random utility. Falmagne (1978) provides necessary and sufficient conditions for a choice rule to be an element of $Q_r$. Block and Marschak (1960) show that $Q_l \subset Q_r$. Holman and Marley (Luce and Suppes, 1965) and McFadden (1978) show how to find a random utility $\pi$ for any Luce rule. Theorem 3, below, shows that every attribute rule is a random utility maximizer, i.e., $Q_a \subset Q_r$. Hence, Theorem 3 extends Block and Marschak’s result to the attribute rule.

For any subset $X \subset \mathbb{R}^k$, let $\text{cl}X$ denote the closure of $X$ and let $\text{conv}X$ denote its convex hull.

**Theorem 3:** $Q_l \subset Q_a \subset Q_r = \text{cl} Q_a = \text{cl conv} Q_l$.

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 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 $Q_l$ 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 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 ($Q_r = \text{cl} Q_a$).

4.2 Elimination by Aspects

Tversky (1972) introduces 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 \( u : A_+ \to \mathbb{R}_+ \) such that for all \( i \in A \in A_+ \),

\[
q_{iA} = \frac{\sum_{B \in A} u(B) q_{iA \cap B}}{\sum_{B: B \cap A \neq \emptyset} u(B)}
\]

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.

There are \( q \in Q_a \) that cannot be approximated by any EBA rule. Specifically, example 3 above (3 airlines with 2 destinations) is not close to any EBA rule. Tversky (1972) shows 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 formula for the attribute rule is similar to the two-level nested logit formula of discrete estimation theory. In a two-level nested logit, the probability of choice (or outcome) \( i = A = \{1, \ldots, n\} \) depends on some structural parameters, observable variables and non-observables. Each nest is a subset of \( A \) and the collection of nests forms a partition.

To relate the attribute rule to the nested logit, consider the following example: \( A = \{1, \ldots, 5\} \), \( X = \{\{1,2,3\}, \{4,5\}\} \). Then, the nested logit probability of choosing \( i \) (from \( A \)) is

\[
P_i = \frac{e^{V_i/\gamma} (\sum_{j \in X_i} e^{V_j/\gamma})^{\gamma - 1}}{(\sum_{j \in X_i} e^{V_j/\gamma})^\gamma + (\sum_{z \in Z \setminus X_i} e^{V_i/\gamma})^\gamma}
\]

The \( V_j \)'s are the observable parameters. They, together with the unobservables, the structural parameter \( \gamma \) and other structural parameters yield a random utility model that implies the formula above. The modeler specifies the particular nesting and estimates the structural parameters from observed choice and the realized \( V_j \)'s.

If we interpret \( Z \) to be the set of attributes (alternatives 1, 2, 3 share one attribute and alternatives 4, 5 the other) then the nested logit resembles an attribute rule. Despite their apparent similarity, the two models are different and address different problems. The nested logit is geared towards estimation; it provides a framework for combining a priori
structural assumptions and relevant observables to estimate the remaining parameters and enabling the modeler to predict future choice from these observables. It provides a theory of choice from $A$ but no theory of choice from subsets of $A$.

In our model, the nests are not given and there is no analogue of the $V_j$’s. Instead, the observable variable is the choice set. The proof of Theorem 2 provides a stylized statement of how the attributes (nests) can be determined from observed choices given sufficiently rich sets of options and option sets.

5. Dynamic Choice

In this section, we extend our model to dynamic choice; that is, to a domain where 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 to the modeler. As we illustrate in the following example, the duplicates problem is especially relevant in dynamic settings. Even if choice alternatives are non-overlapping, the same alternative may be available in multiple continuation problems and, therefore, decision trees may create overlap even if individual alternatives are non-overlapping.

Suppose the decision maker must choose one of four non-overlapping alternatives $s_1, s_2, s_3$ and $s_4$. Decision problem #1 is a static problem, as analyzed in the previous section. As before, we represent this problem as the set $\{s_1, s_2, s_3, s_4\}$. Decision problem #2 is a two stage problem; in stage one, the agent takes one of two actions; the first action leads to the terminal node $s_1$; the second action leads to the (second-stage) decision problem $\{s_2, s_3, s_4\}$. We represent problem #2 as

$$\{s_1, \{s_2, s_3, s_4\}\}$$

Decision problem #3 is again a two stage problem but differs from #2 by the addition of an action that excludes $s_1$ and $s_3$; we represent decision problem #3 as

$$\{s_1, \{s_2, s_4\}, \{s_2, s_3, s_4\}\}$$

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; it requires that \(s_1\) is 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 non-empty finite subsets of \(K\). Define \(S_0 = S\), and, inductively, define for all \(n \geq 1\)

\[
\mathcal{T}_n := F(S_{n-1})
\]

\[
S_n := \mathcal{T}_n \cup S
\]

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 \(\phi : \mathcal{D} \times \mathcal{D} \to [0, 1]\) 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 set of choice objects are the decision nodes \(\Omega\) (i.e., analogous to \(S\) in the static case). We write \(\omega, \omega'\) for generic elements of \(\Omega\); note that a choice object \(\omega \in D\) can be a final outcome \(\omega = s \in S\) or another decision problem \(\omega = D' \in \mathcal{D}\).

The definitions of duplicates, non-overlapping and fine can be applied to a dynamic choice rule \(\phi\) 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 making the alternatives in the set \(D\) available for consideration at a later stage has no effect on choosing other alternatives. 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 \((w, \eta)\) where \(w : Z \to \mathbb{R}_{++}\) and \(\eta : \Omega \times Z \to \mathbb{R}_{++}\) is recursive if, for \(\omega = D \in \mathcal{D}\),

\[
\eta^x_\omega = \sum_{\omega' \in D} \eta^x_{\omega'}
\]

Hence, if \((w, \eta)\) is recursive, then the attribute intensity of the sub-problem \(D\) is simply the sum of the attribute intensities of its decision nodes. Thus, as suggested by consistency, the fact that the alternatives in \(D\) may be considered later has no effect on the attribute intensity of the available alternatives.

As in section 3, \(X_\omega = \{x \in Z | \eta^x_\omega > 0\}\) and \(X(D) = \bigcup_{\omega \in D} X_\omega\). Then, \(\phi\) is a recursive attribute rule if there is a recursive attribute system \((w, \eta)\) such that

\[
\phi_\omega(D) = \sum_{x \in X(D)} \frac{w_x}{w(X(D))} \cdot \frac{\eta^x_\omega}{\eta^x(D)}
\]

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 \(\eta^x_\omega = 1\) and \(\eta^y_\omega = 0\) 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 \(\phi\) to the objects \(S\) we get another choice rule \((\phi, S)\). We call this choice rule the static rule induced by \(\phi\); conversely, we call \((\phi, \Omega)\) an extension of \((\phi, S)\). As we demonstrate in Proposition 2, below, the attribute rule can be used to extend Luce rules or attribute rule 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 (strongly rich attribute rule) to a recursive attribute rule is unique and satisfies strong richness.

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

To this point, our dynamic model analyzes decision making in the first stage of a multi-stage 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 (1982)) agent, that is, an agent whose choice probabilities are independent of the choice history.

For any decision problem $D \in D$ 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 $(\omega_0, \ldots, \omega_n)$ of $D$ is

$$\phi_{\omega_0}(\omega_1) \cdot \phi_{\omega_1}(\omega_2) \cdots \phi_{\omega_{n-1}}(\omega_n) \quad (c)$$

We use the term “consequentialist” because, in the formula (c) above, the conditional probability of choosing $\omega_i$ at node $\omega_{i+1}$ of $D$ is the same as choosing $\omega_i$ in the decision problem $\omega_{i+1}$. 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 $D \in D$ is simple if all of its terminal nodes are distinct and write $D^s \subset D$ for the simple decision problem. In that case, every possible outcome in $D$ has a unique associated path. We say that a dynamic choice rule is $s$-invariant if, for every simple decision problem the outcome probabilities depend only on the terminal nodes, that is, $p_s(D) = \phi_s(\sigma(D))$ for all $D \in D^s$. Proposition 3, below, shows that every extension of a Luce rule is $s$-invariant. Moreover, extensions of Luce rules are the only recursive attribute rules that satisfy this property.

**Proposition 3:** A strongly rich recursive attribute rule is $s$-invariant if and only if it is the extension of a rich Luce rule.

Invariance breaks down once we consider decision problems where multiple paths lead to the same outcome. To illustrate the effect of this multiplicity, we focus on a simple special case with two non-overlapping alternatives $a$ and $b$. Let $D^1 = \{a, b\}$ and recursively define $D^{n+1} = \{a, b, D^n\}$. Hence, $D^n$ is an $n$–stage decision problem where in all but the
final stage, the agent must decide between choosing $a$ or $b$ or delaying the decision to the next stage. In the final stage, the decision is between $a$ and $b$. Choice is governed by a recursive attribute rule $\phi$ and the probability of a path is given by formula (c). Let $
abla := \phi_a(\{a, b\})$. Then, $\phi_a(D^n) = \alpha/n$ and

$$p_a(D^n) = \alpha$$

for all $n \geq 1$. Thus, the decision problem $D^n$ leads to the same probability of outcome $a$ as the static problem $\{a, b\}$.

Next, we modify $D^n$ by removing one of its $a$–terminal nodes. The decision problem $D^n_k$ is identical to $D^n$ except that in stage $k$ the alternative $a$ is not available. Thus, stage $k$ of $D^n_k$ is the decision problem $\{b, D^{n-k}\}$. The figure below depicts the decision problems $D^n_k$ for $n = 3$ and $k = 1, 2, 3$.

A straightforward calculation shows that

$$p_a(D^3_1) = \alpha - \frac{\alpha(1 - \alpha)}{3}$$

$$p_a(D^3_2) = \alpha - \frac{\alpha(1 - \alpha)(2 - \alpha)}{12}$$

$$p_a(D^3_3) = \alpha + \frac{\alpha(1 - \alpha)}{6}$$

The decision problem $D^n_k$ differs from $D^n$ in that it has one fewer terminal node with outcome $a$. This change has two effects; the first ‘direct’ effect occurs in period $k$. Since $a$ is not available in that period, the decision maker is forced to delay her choice if $a$ is what she would choose at that point. This reduces the probability of choosing $a$ since, at the
next decision stage, $a$ may no longer be her choice. The second, ‘indirect,’ effect occurs in decision stages prior to $k$. Future opportunities to choose $a$ intensify the attractiveness to delay a choice and, since $a$ is available in fewer (future) decision nodes, the decision maker is less inclined to delay a choice of $a$ than a choice of $b$. This effect increases the probability of outcome $a$. If $k = n$ this effect always dominates whereas if $k = 1$ this effect is absent. Proposition 4 summarizes these observations:

**Proposition 4:** $p_a(D^n_j)$ is strictly increasing in $j$ with $p_a(D^n_2) < \alpha < p_a(D^n_n)$ for $n \geq 3$. 
6. Appendix

6.1 Example 2

We must prove that the $\rho$ defined in Example 2 is rich. Let $S = \{s_1, s_2, \ldots\}$ be an enumeration of $S$ and consider any $A, C \in A_+$, 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, $B \in A$, $B \cap C = \emptyset$ and $\rho(A, AB) = \frac{\delta_1}{\delta_1 + \delta_2} = \delta$ as desired.

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

Lemma A1: $\succeq_{\rho}$ is complete and transitive.

Proof: Clearly, $\rho$ satisfies independence only if $\succeq_{\rho}$ is complete. Next, assume that $A \succeq_{\rho} B$ and $B \succeq_{\rho} C$. By richness, there exists a $D \in A$ 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 $A \succeq_{\rho} C$ as desired.

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

Lemma A2: For any test sequence $A_1, \ldots, A_n \in A_+$, $\rho(A_i, A_i A_j) = \frac{1}{2}$ 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 $\rho(A_1, A_1 A_3) > \frac{1}{2}$. Independence implies that $A_1 \succ_{\rho} A_2$. Since $\rho(A_1, A_1 A_2) = \frac{1}{2} = \rho(A_3, A_3 A_2)$, independence also implies $A_1 \sim_{\rho} A_3$. Then, by Lemma A1 we have $A_3 \succ_{\rho} A_2$. But $\rho(A_3, A_1 A_3) < \frac{1}{2} = \rho(A_2, A_1 A_2)$ contradicting $A_3 \succ_{\rho} A_2$. A similar
argument reveals the impossibility of $\rho(A_1, A_1 A_3) < \frac{1}{2}$. Hence, $\rho(A_1, A_1 A_3) = \frac{1}{2}$ as desired.

\[\text{Lemma A3: If } A_1, \ldots, A_n \text{ is a test sequence and } A \in \mathcal{A}_+ \text{ with } A \cap A_1 A_2 \cdots A_n = \emptyset, \text{ then } \rho(A, AA_i) = \rho(A, AA_1) \text{ for all } i = 1, \ldots, n.\]

\text{Proof: If necessary, use richness to extend the test sequence so that } n \geq 3. \text{ Then, Lemma A2 implies } A_i \sim_\rho A_j \text{ for all } i, j \text{ and hence } \rho(A, AA_i) = \rho(A, AA_1) \text{ for all } i. \]

\[\text{Lemma A4: For all } A, B \in \mathcal{A}_+ \text{ with } A \cap B = \emptyset, A \succeq_\rho B \text{ if and only if } \rho(A, AB) \geq \frac{1}{2}.\]

\text{Proof: By richness, we can choose } D \in \mathcal{A}_+ \text{ such that } D \cap AB = \emptyset \text{ and } \rho(B, BD) = \frac{1}{2}. \text{ Let } B_1 = B \text{ and } B_2 = D \text{ and note that } B_1, B_2 \text{ is a test sequence. Then, by Lemma A3, } \rho(A, AB) = \rho(A, AD) \text{ and therefore } \rho(A, AB) \geq \frac{1}{2} \text{ if and only if } \rho(A, AD) \geq \rho(B, BD); \text{ that is, } \rho(A, AB) \geq \frac{1}{2} \text{ if and only if } A \succeq_\rho B.\]

\[\text{Lemma A5: If } C_1, C_2, C_3, C_4 \text{ is a test sequence, then } \rho(C_i, C_1 C_2 C_3 C_4) = \frac{1}{4} \text{ for all } i = 1, 2, 3, 4.\]

\text{Proof: Let } C = C_1 C_2 C_3 C_4 \text{ and without loss of generality assume } \rho(C_i, C) \geq \rho(C_j, C) \text{ whenever } i \leq j. \text{ Hence, by Lemma A4,}

\[C_1 C_2 \succeq_\rho C_3 C_4 \text{ and } C_1 C_3 \succeq_\rho C_2 C_4. \quad (A1)\]

\text{By richness, there exists } C_5 \text{ such that } C_1, C_2, C_3, C_4, C_5 \text{ is a test sequence. By Lemma A3, } C_2 \sim_\rho C_3 \sim_\rho C_5 \text{ and hence, by Lemma (A1) we have}

\[\rho(C_1 C_2, C_1 C_2 C_3) = \rho(C_1 C_2, C_1 C_2 C_5) \]
\[\geq \rho(C_3 C_4, C_3 C_4 C_5) \quad (A2)\]
\[= \rho(C_3 C_4, C_2 C_3 C_4).\]

\text{And by the same argument,}

\[\rho(C_1 C_3, C_1 C_2 C_3) \geq \rho(C_2 C_4, C_2 C_3 C_4). \quad (A3)\]
But we also have
\[
2 = 2[\rho(C_1, C_1C_2C_3) + \rho(C_2, C_1C_2C_3) + \rho(C_3, C_1C_2C_3)]
\]
\[
= \rho(C_1C_2, C_1C_2C_3) + \rho(C_1C_3, C_1C_2C_3) + \rho(C_2C_3, C_1C_2C_3)
\]
\[
\geq \rho(C_3C_4, C_2C_3C_4) + \rho(C_2C_4, C_2C_3C_4) + \rho(C_2C_3, C_2C_3C_4)
\]  \hspace{1cm} \text{(A4)}
\]
\[
= 2[\rho(C_2, C_2C_3C_4) + \rho(C_3, C_2C_3C_4) + \rho(C_4, C_2C_3C_4)]
\]
\[
= 2.
\]
Equation (A4) implies that the inequalities in (A2) and (A3) must in fact be equalities. Hence \(\rho(C_1C_2, C_1C_2C_5) = \rho(C_1C_2, C_1C_2C_3) = \rho(C_3C_4, C_2C_3C_4) = \rho(C_3C_4, C_3C_4C_5)\) and independence imply \(C_1C_2 \sim_\rho C_3C_4\). By Lemma A4, we have \(\rho(C_1C_2, C) = \rho(C_3C_4, C) = \frac{1}{2}\). Finally since \(\rho(C_i, C) \geq \rho(C_j, C)\) for \(i \leq j\), we must have \(\rho(C_i, C) = \frac{1}{4}\) for \(i = 1, 2, 3, 4\).

\text{Lemma A6:}\quad \text{If } A_1, \ldots, A_n \text{ is a test sequence, then } A_iA_j \sim_\rho A_kA_\ell \text{ for all } i \neq j \text{ and } k \neq \ell.

\text{Proof:}\quad \text{If } i, j, k, \ell \text{ are all distinct, then Lemma A5 implies } \rho(A_iA_j, A_iA_jA_kA_\ell) = 1/2 \text{ and Lemma A4 implies } A_iA_j \sim_\rho A_kA_\ell. \text{ If } \{i, j, k, \ell\} \text{ has three distinct elements assume, without loss of generality, that } j = \ell. \text{ Let } B_1 = A_i, B_2 = A_j, B_3 = a_k \text{ and note that } B_1, B_2, B_3 \text{ is a test sequence. By richness, we can choose } B_4, B_5 \text{ such that } B_1B_2, B_3, B_4, B_5 \text{ is a test sequence. By Lemmas A4 and A5, } B_1B_2 \sim_\rho B_3B_4 \text{ and } B_2B_3 \sim_\rho B_4B_5. \text{ Then Lemma A1 implies } B_1b_B \sim_\rho B_2B_3, \text{ that is, } A_iA_j \sim_\rho A_kA_\ell. \text{ Finally, if } \{i, j, k, \ell\} \text{ has two distinct elements, then } A_iA_j = A_kA_\ell \text{ and by Lemma A1 we have } A_iA_j \sim_\rho A_kA_\ell.

\text{Lemma A7:}\quad \text{If } A_{i_1}, \ldots, A_{i_{2^n}} \text{ are distinct elements and } A_{j_1}, \ldots, A_{j_{2^n}} \text{ are distinct elements of the test sequence } A_1, \ldots, A_m, \text{ then } \bigcup_{k=1}^{2^n} A_{i_k} \sim_\rho \bigcup_{k=1}^{2^n} A_{j_k}.

\text{Proof:}\quad \text{The proof is by induction: When } n = 1, \text{ the statement is true by Lemma A6. Next, assume it is true for } n \text{ and let } A_{i_1}, \ldots, A_{i_{2^{n+1}}} \text{ and } A_{j_1}, \ldots, A_{j_{2^{n+1}}} \text{ be two collections of distinct elements of the same test sequence. Use richness to extend the test sequence } A_1, \ldots, A_m \text{ with an additional } 2^{n+1} \text{ elements, namely } A_{m+1}, \ldots, A_{m+2^{n+1}}. \text{ Let } B_k = A_{i_{2^{k-1}}}, A_{i_{2^k}}, \text{ let } C_k = A_{j_{2^{k-1}}}A_{j_{2^k}} \text{ and let } D_k = A_{m+2^{k-1}}A_{m+2^k} \text{ for } k = 1, \ldots, 2^n. \text{ Lemma}
A6 implies that $B_k \sim \rho B_\ell$ for all $k, \ell$. 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 $D_\ell$ are disjoint for every $k$ and $\ell$. Moreover, Lemma A6 implies that $B_k \sim \rho D_\ell$ for all $k, \ell$. 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 A1 we obtain $\bigcup_{k=1}^{2^n+1} A_i = \bigcup_{k=1}^{2^n} B_k \sim \rho \bigcup_{k=1}^{2^n} C_k = \bigcup_{k=1}^{2^n+1} A_j$ as desired. \[ \square \]

**Lemma A8:** If $A_1, \ldots, A_{2^n+1}$ is a test sequence, then $\rho(A_j, A_1 A_2 \cdots A_{2^n+1}) = \frac{1}{2^{n+1}}$ 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 A1 and A7 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 $\rho$ yield the desired result. \[ \square \]

**Lemma A9:** 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 A8 implies

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

and I then yields the desired result. \[ \square \]

**Lemma A10:** If $A_1, \ldots, A_n$ is a test sequence, then $\rho(A_j, A_1 A_2 \cdots A_n) = \frac{1}{n}$ for all $j$. 28
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 A1 and A9 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 $\rho$ yield the desired result. \hfill \square

Lemma A11: 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 A10, we have $\rho(A_i, B) = \rho(A_j, B) = \frac{1}{n}$ for all $i, j$. Then, the additivity of $\rho$ yields the desired result. \hfill \square

Lemma A12: If $A, B \in A_+$ 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 $\rho(C, AC) > \frac{1}{2}$. Hence, by Lemma A4, $C \succ \rho A$ and therefore $\rho(C, CB) > \rho(A, AB) = 1$, a contradiction. By symmetry, we cannot have $\rho(A, AB) = 0$ either. \hfill \square

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

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

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

Lemma A13: The function $\bar{v}$ is well-defined and satisfies the following

(i) $\bar{v} : A \to \mathbb{R}_+$ and $\bar{v}(A) = 0$ if and only if $A = \emptyset$.

(ii) $\bar{v}(A) \geq \bar{v}(B)$ if and only if $A \succeq \rho B$.

Proof: To prove that $\bar{v}$ is well-defined, we first note that by Lemma A12, $\bar{v}(A) < \infty$ 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 $A_1 \succeq \rho A_2$ 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 A12, \( \bar{v} \) satisfies (i). To prove, (ii), choose \( C \) such that \( C \cap BAA_o = \emptyset \) and \( \rho(C, CA_o) = \frac{1}{2} \). Then, by Lemma A4, \( A_o \sim_{\rho} C \). For any \( D \in A_+ \) 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, \( A \geq_{\rho} B \) if and only if \( \rho(A, AC) \geq \rho(B, BC) \) if and only if \( \bar{v}(A) \geq \bar{v}(B) \). \( \square \)

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 A11, \( \rho(nC, anC) \) 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 element test sequence with each element satisfying \( C_i \sim_{\rho} C \).

**Lemma A14:** If \( nC \sim_{\rho} A_o \), then \( mC \sim_{\rho} B \) if and only if \( \bar{v}(B) = \frac{m}{n} \).

**Proof:** Assume \( nC \sim_{\rho} A_o \) and hence \( \rho(B, BA_o) = \rho(B, BnC) \). Then, Lemma A11 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 \( mC \sim_{\rho} B \). \( \square \)

**Lemma A15:** \( \rho(A, AB) = \frac{\bar{v}(A)}{\bar{v}(A) + \bar{v}(B)} \) for all \( A, B \in A_+ \) such that \( A \cap B = \emptyset \).

**Proof:** First assume that \( \bar{v}(A), \bar{v}(B) \) are rational numbers. Then, there exists 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 A14, \( kC \sim_{\rho} A \) and \( mC \sim_{\rho} B \) and hence \( \rho(kC, (k + m)C) = \rho(a, AmC) = \rho(A, AB) \). But Lemma A11 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 \( \epsilon > 0 \), choose rational numbers, \( r_1, r_2 \) such that \( r_1 < \bar{v}(A), r_2 > \bar{v}(B) \) and \( \frac{r_1}{r_1 + r_2} > \frac{\bar{v}(A)}{\bar{v}(A) + \bar{v}(B)} - \epsilon \). 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 \( \rho(C, C < D) = \frac{r_1}{r_1 + r_2} \) and by Lemma A13(ii), \( \rho(A, AB) \geq \rho(A, AD) \geq \rho(C, CD) \). Hence, \( \rho(A, AB) \geq \frac{\bar{v}(A)}{\bar{v}(A) + \bar{v}(B)} - \epsilon \) for every \( \epsilon > 0 \); 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. \( \square \)

To complete the proof the theorem, let \( v_s = \bar{v}({s}) \).
7. 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 A16–A24 use only SR. The remainder of the proof also uses WI and E).

Lemma A16: 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. \hfill \Box

Lemma A17: $\sim$ 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, $\rho(s, AD) = \rho(s, AD's) = \rho(E, AD'E) = \rho(E, BD'E) = \rho(E, CD'E) = \rho(s, D's) = \rho(s, CD)$ as desired. \hfill \Box

Lemma A18: $A \sim \emptyset$ 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 $\emptyset$. \hfill \Box

Lemma A19: 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 $B_1B_2 \cdots B_{|A|} = B$ 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 $f(t) = s_i$. \hfill \Box
Lemma A20: \( A \in \mathcal{M} \) if and only if \(|A| \geq |B|\) for all \( B \sim A \).

Proof: 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 A18, \( A_i \neq \emptyset \) 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 A16 and Lemma A17 this implies \( A \in \mathcal{M} \). \( \square \)

Lemma A21: If \( A \subset B \in \mathcal{M} \), then \( A \in \mathcal{M} \).

Proof: Suppose \( A \notin \mathcal{M} \) and \( A \subset B \). By Lemma A20 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 A20, \(|D| \geq |(B \setminus A)|\). Hence \(|CD| = |C| + |D| > |A| + |B \setminus A| = |B|\). By Lemma A16, \( CD \sim B \), thus Lemma A20 implies \( B \notin \mathcal{M} \). \( \square \)

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, \( \theta \) 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 A22: \( \mathcal{B}_0 = \mathcal{B}_1 \).

Proof: \( \mathcal{M} \subset \mathcal{B}_0 \) follows from Lemma A21. 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 A20 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 A20 implies \(|D_i| \leq 1\) for all \( i \). Hence \(|D| = |A|\), and Lemma A20 implies \( A \in \mathcal{M} \). \( \square \)

Lemma A23: \( A \subset B \perp C \) 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 A16, \( D \cup A' \sim (B \setminus A) \cup A = B \). Since \( B \perp C \), we have \((D \cup A') \cap C' = \emptyset \) and hence \( A' \cap C' = \emptyset \) proving that \( A \perp C \). \( \square \)
Lemma A24:  
(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, \( s, t \in T_1 \), \( s \sim A \in A \) and \( t \sim B \in \mathcal{A} \) such that \( A \cap B = \emptyset \). By Lemma A22, \( s, t \in \mathcal{M} \). Hence, by Lemma A20, \( |A| = |B| = 1 \) and therefore \( A = B \) and therefore \( s \sim t \) by Lemma A17. Then, by the definition of \( \theta \), \( s = \theta(s) = \theta(t) = t \).

(ii) Assume \( A, B \in \mathcal{B}_0 \), \( A' \sim B \), \( B' \sim B \) and \( s^* \in A' \cap B' \). By Lemma A19, 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 A17 that \( f^{-1}(s^*) \sim g^{-1}(s^*) \). By Lemma A21, \( f^{-1}(s^*), g^{-1}(s^*) \in \mathcal{M} \) and hence applying Lemma A19 again yields an onto function \( h : f^{-1}(s^*) \rightarrow g^{-1}(s^*) \) such that \( h^{-1}(s) \sim s \) for all \( s \in g^{-1}(s^*) \). By Lemma A20, \( h \) 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 \( A \cap B = \emptyset \).

Lemma A25:  \( \rho(s, AB) = \rho(s, AB_0) \) 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 \( B_0 = B_1^1 B_2^2 \cdots B_k^k \) where \( s \sim t \) if and only if \( i = j \) for all \( s \in B_i \) and \( t \in B_j \). Hence, \( B_1^1, \ldots, B_k^k \) 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 A24(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 A16 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 A23, \( B^i \perp \{s\} \) 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.

For \( B \in \mathcal{B}_1 \) such that \( B \neq \emptyset \) define the choice rule \( \rho^1 \) such that

\[
\rho^1(s, B) = \rho(s, B)
\]
Lemma A26: $\rho^1$ satisfies $R$, $I$ and therefore is a Luce Rule.

Proof: Assume $A, C \in B_1$, $A \neq \emptyset$ and $\delta \in (0,1)$. By SR, there is $\hat{B} \in A$ such that $\rho(A, A\hat{B}) = \delta$ and $\hat{B} \perp AC$. Again by SR, we can choose $B_0 \in M$ 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 A22 implies $B_0 \in B_0$. Lemma A17 and $B_0 \sim \hat{B} \perp AC$ imply $B_0 \perp AC$. Then, Lemma A23 implies $B_0 \perp C$ and $B_0 \perp A$. Let $B = \{\theta(s) \mid s \in B_0\}$. Lemma A25 yields $\delta = \rho(A, AB_0) = \rho(A, AB)$. Lemma A23 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 $\rho^1$ satisfies R. Lemma A24(iii) and WI imply that $\rho^1$ satisfies I and hence, by Theorem 1, it is a Luce rule.

For $s \in S_1$, define
\[ x_s = \{t \in S \mid \exists B \in B_0 \text{ such that } s \in B \text{ and } B \sim t\} \]
and define $Z = \{x_s \mid s \in S_1\}$. Let $v$ be the Luce value that represents $\rho^1$ and define
\[ w(x_s) = v_s \]
For $s \in S$ choose any $B \in B_0$ such that $s \sim B$ and define
\[ \eta_{x_t}(s) = |\{t' \in B \mid t = \theta(t')\}| \]

Lemma A27: (i) $w : Z \rightarrow \mathbb{R}_{++}$ is well defined. (ii) $\eta : Z \times S \rightarrow \mathbb{N} \cup \{0\}$ 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 A22, $\{t\} \in M$ which, by Lemma A20, implies $|b| = 1$ and hence $t \sim s$. It follows that $t = \theta(t) = \theta(s) = s$, as desired.

(ii) We must show that $s \sim B \in B_0$ and $s \sim B' \in B_0$, implies $|\{t' \in B \mid t = \theta(t')\}| = |\{t' \in B' \mid t = \theta(t')\}|$. Note that, by Lemma A17, $s \sim B \in B_0$ and $s \sim B' \in B_0$ implies $B \sim B'$. Hence, by Lemmas A19 and A20, there is a bijection $f : B' \rightarrow B$ 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 $|\{t' \in B \mid t = \theta(t')\}| = |\{t' \in B' \mid t = \theta(t')\}|$. 

Lemma A28: $\eta_{x_t}(s) > 0$ if and only if $s \in x_t$. 34
Hence, there are infinitely many archetypes for every \( s \in S_1 \). Suppose \( s \in x_t \). Then, there exists \( B \in B_0 \) such that \( t \in B \) and \( B \sim s \). Since \( \theta(t) = t \), it follows that \( \eta_x(t) > 0 \). Conversely, if \( \eta_x(t) > 0 \), then there exists \( B \in B_0 \) 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 \( B' = (b\{t'} \cup \{t} \) and note that \( B' \sim B \) by Lemmas A16 and A17 and hence, \( s \in x_t \).

**Lemma A29:** \( Z \) is an attribute set.

**Proof:**

By Lemma A28,

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

For all \( t \in S \), there exists \( b \in B_0 \) such that \( t \sim b \) (by SR and Lemma A22). 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 \( B' = \{s\} \cup (B\{t'} \). By Lemmas A16 and A17, \( B' \sim B \sim t \) 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 \( t \sim B \in B_0 \) implies \( |X_t| \leq |B| \). Suppose \( t \sim B \in B_0 \) 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 \( s' \in S_1 \) such that \( c_{s'} \in X_t \) and \( s' \not\in B^* \). Since \( c_{s'} \in X_t \), there exists some \( A \in B_0 \) such that \( s' \in A \) and \( A \sim t \). Lemma A17 implies \( A \sim B \). Lemmas A19 and A20 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.

**Lemma A30:** \( s \sim t, s, t \in B_0 \in B_0 \) 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 \( C = B_0 \{s, t\} \). 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.

Lemmas A27–A29 imply that \((w, \eta)\) is an attribute system. Let \( \hat{\rho} \) be the choice rule that \((w, \eta)\) 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 A31: \( \rho = \hat{\rho} \)

**Proof:** First, we show that \( \rho_{s^*}(B_0) = \hat{\rho}_{s^*}(B_0) \) whenever \( s^* \in B_0 \in B_0 \). As in Lemma A25, 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 on Lemma A25, 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 A30 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 A30 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 \( C_i \in B_0 \) such that \( C_i \sim s_i \) and \( C_i \cap A A_{i-1} = \emptyset \) and let \( A_i = C_i(A_{i-1} \backslash \{ s_i \}) \). Then, since \( A_n \in B_0 \),

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

Finally, for each \( i \), since \( s_i \sim C_i \in B_0 \) we have \( X_{s_i} = X(c_i) \) and therefore \( X(A) = X(A_n) \), which together with equation (5) and the definition of \( \eta \) implies

\[
\rho^*(C_1, A_n) = \hat{\rho}_{s_1}(A)
\]

as desired. \( \square \)

To prove the “if” part of Theorem 2, let \( \rho \) be a complete attribute rule that \( (w, \eta) \) represents and assume \( \rho \) satisfies SR. Let \( Z \) be the attribute set.
Lemma A32: \( Z \) is infinite and for every \( x \in Z \), there is an infinite number of archetypes.

Proof: If \( Z \) is finite then \( \rho \) 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 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 (5) implies \( \rho(t, ss't) > \rho(t, Ds't) \). Hence for any such \( D \) we have \( D \not\sim s \), contradicting part (ii) of SR. \( \square \)

Lemma A33: \( A \sim B \) if and only if \( \eta_x(A) = \eta_x(B) \) for all \( x \in Z \).

Proof: Sufficiency follows from equation (5). Next, suppose \( \eta_x(A) > \eta_x(B) \) for some \( x \in Z \). Lemma A32 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 (5), if \( \rho(t', At') = \rho(t', Bt') \), then \( \rho(s', As't') < \rho(s', Bs't') \), hence \( A \not\sim B \) proving necessity. \( \square \)

Lemma A34: \( A \perp B \) 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 A32 and Lemma A33, there are \( A' \sim A \) and \( B' \sim B \) such that every \( s' \in A' \cup B' \) is an archetype. By Lemma A33, 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 A33 \( B'' \sim B \), hence \( A \not\perp B \). \( \square \)

E and WI follow easily from Lemma A33, Lemma A34 and equation (3). This completes the proof of the “if” part of Theorem 2. \( \square \)

7.1 Proof of Proposition 1

Throughout the proof, let \((w, \eta)\) and \((\hat{w}, \hat{\eta})\) be two attribute systems. Let \( Z \) be the domain of \( w \) and \( X(A) = \{x \in Z \mid \eta^x(A) > 0\} \). Let \( \hat{Z}, \hat{X} \) be the corresponding objects for \((\hat{w}, \hat{\eta})\). 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 \((w, \eta)\) and \((\hat{w}, \hat{\eta})\) represent the same choice rule is straightforward and omitted.

For the converse, let \((w, \eta)\) with attribute set \( Z \) be any complete attribute system that represents \( \rho \). We will show that if \((\hat{w}, \hat{\eta})\) is any other attribute system that represents \( \rho \), then there exists \( \alpha > 0 \) and natural numbers \( m_x > 0 \) such that \( \hat{w} = \alpha \cdot w \) and \( \hat{\eta}^x = m_x \cdot \eta^x \).
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, \( \hat{X}(A) \cap \hat{X}(B) = \emptyset. \) \( \square \)

Claim 2: \( C \sim B \implies \hat{X}(C) = \hat{X}(B). \)

Proof: Choose \( A \) such that \( A \perp B. \) Repeating the argument in the proof of Claim 1 and Claim 1 itself ensure \( \hat{X}(B) \subset \hat{X}(C) \setminus \hat{X}(A). \) Hence, \( \hat{X}(A) \cap \hat{X}(B) = \emptyset. \) \( \square \)

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 \( B \) 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

\[
\rho(C, CB^i) = \rho(B, BB^n) = 1/n
\]

Since \( (\hat{w}, \hat{\eta}) \) represents \( \rho, \) 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:

\[
\sum_{x \in X(C)} \hat{w}(x) \frac{r_x - 1}{r_x + 1} = 0
\]

\[
\sum_{x \in X(C)} \hat{w}(x) \frac{r_x - 1}{2} = 0
\]

38
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 $Y = \{y \in X(C) | k_y < \infty\}$. 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.

\[\square\]

**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 $A \perp B$.

For the converse, assume $\hat{\eta}^x(A) \cdot \hat{\eta}^x(B) \neq 0$ for some $x \in \hat{Z}$ and $A \perp B$. Choose $C$ such that $C \sim B$ and $C \cap AB = \emptyset$. Then, by Claim 3, $\hat{\eta}^x(AC) > \hat{\eta}^x(A)$ and therefore, $\rho(A, AB) < \rho(A, ABC)$ contradicting E.

**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 A22 in the proof of Theorem 2 establishes that $A$ is fine if and only if it is a nonempty subset of $T$.

**Proof:** Since $(w, \eta)$ is complete, Claim 3 and Lemma A32 imply that $s \in T$ if and only if $s$ is an archetype for $(w, \eta)$. Hence, the completeness of $(w, \eta)$ implies $Z = \{x | \{x\} = X_s$ for some $s \in T\}$. For all $s \in T$, let $x_s$ denote the unique element attribute for $(w, \eta)$ that contains $s$.

By Claims 3 and 4,

$$x_s = \bigcup_{y \in \hat{X}_s} y \quad (A5)$$
for all \( s \in T \). Suppose \( y, z \in \mathring{X}_s \) and \( y \neq z \). Then, without loss of generality, there exists \( A \in y \setminus x \). By Claim 3, there exists no \( B \) such that \( s \in B \sim A \) and then, by Claim 4, \( A \notin \{ s \} \). But, then Claim 3 applied to the representation \((w, \eta)\) ensures \( \eta^x_s(A) > 0 \). But, then Lemma A32 ensures there exists \( B \sim A \) such that \( s \in B \), a contradiction. Hence, \( y, z \in \mathring{X}_s \) implies \( y = z \) and then the display equation above ensures \( \{ x \} = X_s = \mathring{X}_s \) and therefore \( \mathring{X}_s \) is a singleton.

**Claim 6:** \( Z = \mathring{Z} \)

**Proof:** Since \((w, \eta)\) is complete, equation (A5) and Claim 5 ensure that \( Z \subset \mathring{Z} \). For the converse, take \( A \in x \in \mathring{Z} \). 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, \( \mathring{X}_s = \{ x \} \) and by construction \( x = x_s \in Z \).

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 \( m_x = \eta^x_s \) 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 \( \rho_s(\{ s, t \}) = 1/2 = \frac{\hat{\eta}^x_s}{\eta^x_s + \hat{\eta}^x_t} \) and hence \( \eta^x_t = \eta^x_s \), so \( m_x \) is well-defined.

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

\[
\frac{w_x}{w(\{ s, t \})} \cdot \frac{1}{1 + \eta^x_t} = \rho_s(\{ s, t \}) = \frac{\hat{w}_x}{\hat{w}(\{ s, t \})} \cdot \frac{m_x}{\hat{m}_x + \hat{\eta}^x_t} = \frac{w_x}{w(\{ s, t \})} \cdot \frac{m_x}{m_x + \hat{\eta}^x_t}
\]

Hence, \( \hat{\eta}^x_t = m_x \cdot \eta^x_s \) as desired. This completes the proof of the proposition. \( \Box \)
8. Proof of Theorem 3

In section 3, we have established that $Q_l \subset Q_a$. Next, we will prove that $Q_r = \text{cl conv } Q_l$. Fact 1, below, requires no proof.

**Fact 1:** The sets $Q$ and $\Pi$ are compact and convex.

The next fact follows immediately from Fact 1 and the definition of “$q$ maximizes $\pi$.”

**Fact 2:** If $q_i$ maximizes $\pi_i$ 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 (1978) showed that $q \in Q_r$ if and only if

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

(A6)

for all $i \in A$ and $A \in \mathcal{A}_+$. Let $Q_0$ be the subset of $\mathbb{R}_+^{2^n-1}$ 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) compact set, it too is compact. 

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

**Fact 4:** $Q_l \subset Q_r$.

Facts 3 and 4 imply $\text{cl conv } Q_l \subset Q_r$. The fact below establishes the reverse inclusion and yields $\text{cl conv } Q_l = Q_r$.

**Fact 5:** For every $\epsilon' > 0$ and $q \in Q_r$, there exists $\hat{q} \in \text{conv } Q_l$ such that $|q - \hat{q}| < \epsilon'$.

**Proof:** Assume $0 < \epsilon < 1$ and, for $u \in U$, define the Luce value $v^\epsilon u$ such that

$$v_i^\epsilon u = \epsilon^{n-u_i}$$

Let $\delta^u$ be the degenerate random utility that assigns probability 1 to $u$, let $q^{\delta^u}$ be the choice rule that maximizes $\delta^u$ and let $q^{v^\epsilon u}$ be the Luce rule that $v^\epsilon u$ represents. It is easy to see that $q^{v^\epsilon u} \to q^{\delta^u}$ as $\epsilon \to 0$. It follows that $q := \sum_{u \in U} \pi_u q^{v^\epsilon u} \in \text{conv } Q_l$ converges

---

7 Block and Marschak (1960) introduced the inequalities (A6) and identified them as necessary conditions for $q \in Q$ to be an element of $Q_r$. 

---
to \( q^* := \sum_{u \in U} \pi_u q^u \) as \( \epsilon \to 0 \). Note that \( \pi = \sum_{u \in U} \pi_u \delta^u \) and hence, by Fact 2 (and a simple inductive argument), \( q^* = \sum_{u \in U} \pi_u q^u \) maximizes \( \pi \). Thus, for every \( q^* \in Q_r \) we can find \( q \) \( \in \text{conv} \ Q_l \) arbitrarily close to \( q^* \). 

Lemma A35: \( Q_a \subset Q_r \).

**Proof:** Let \( q \in Q_a \) and let \( (w, \eta) \) be the 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 \( \bar{A} \subset Z \),

\[
\bar{\rho}^\bar{v}_x(\bar{A}) = \begin{cases} 
\bar{v}_x / \left( \sum_{y \in \bar{A}} \bar{v}_y \right) & \text{if } x \in \bar{A} \\
0 & \text{if } x \notin \bar{A}
\end{cases}
\]

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

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

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

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

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

\[
\bar{\rho}^{\bar{\mu}}_x(\bar{A}) = \sum_{\bar{u} \in [x \bar{A}]} \bar{\mu}(\bar{u}) = \bar{\rho}^{\bar{\pi}}_x(\bar{A}) \tag{A7}
\]

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

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

and let \( A_{\bar{u}} = \{ i \in A | i \in x_{A \bar{u}} \} \). Then, let

\[
\bar{\rho}^{\bar{\mu}}_i(A) = \begin{cases} 
\frac{\eta_i^{x_{A \bar{u}}}}{\sum_{j \in A_{\bar{u}}} \eta_j^{x_{A \bar{u}}}} & \text{if } i \in A_{\bar{u}} \\
0 & \text{otherwise}
\end{cases}
\]

\( \tag{A8} \)

42
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) $q = \rho^{\mu}$. 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 $S$; $\Pi$ is the set of all probabilities on $U$ and $\rho_{\pi}$ 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 of Claim: Fix $\bar{u}$, let $x_k$ be such that $u(x_k) = k$ and define $B_k \subset A$ as follows: $B_k = A^{x_k}$, i.e., $B_K$ is the set of choice objects that have the highest ranked (according to $\bar{u}$) attribute. For $k = 1, \ldots, K - 1$

$$B_k = A^{x_k} \setminus \bigcup_{j=k+1}^{K} B_k$$

For $k < K$, $B_k$ are the objects that have the $k$-th highest ranked attribute but none of the higher ones. Let $k_{\bar{u}} = \max\{k | \bigcup_{j=k}^{K} B_j = A\}$. Let $J^{\bar{u}} \subset \{k_{\bar{u}}, \ldots, K\}$ be such that $k \in J^{\bar{u}}$ if and only if $B_k$ is non-empty and $k \geq k_{\bar{u}}$.

We claim there exists $\pi \in \Pi$ such that (i) $\pi(u) > 0$ implies $u_i > u_j$ whenever $i \in B_l, j \in B_m$ for $l > m$ and (ii) for all $k \in J^{\bar{u}}$ and for every $B'_k \subset B_k$ with $i \in B'_k$,

$$\sum_{u \in |B'_k|} \pi(u) = \frac{\eta^x_i}{\sum_{j \in B_k} \eta^x_j} \quad (A9)$$

for $x$ such that $\bar{u}(x) = k$.

To prove the existence of $\pi$ satisfying (i) and (ii), let $U_{\bar{u}}$ be the set of all $u \in U$ that satisfies condition (i). Let $k \in J^{\bar{u}}$ and let $b_k = \{i_1, \ldots, i_{m_k}\}$ ordered so that $u_{i_1} > u_{i_2} > \ldots > u_{i_{m_k}}$. For $r = 1, \ldots, m_k$ let $S^{u_k}(r) = \sum_{j=r}^{m_k} \eta^x_{i_j}$. For $u \in U_{\bar{u}}$, let

$$\pi(u) = \prod_{k \in J^{\bar{u}}} \prod_{r=1}^{m_k} \frac{\eta^x_{i_r}}{S^{u_k}(r)}$$

Verifying that $\pi$ satisfies equation (A9) is straightforward. Conditions (i) and (ii) ensure that $\rho^{\bar{u}} = \rho_{\pi}$ since $a_{\bar{u}} \subset b_k$ for some $k \in J^{\bar{u}}$. $\square$
The claim above implies (1). To prove (2), let \([xX(A)] = \{ \bar{u} \in \bar{U} \mid x_{A\bar{u}} = x \}\) and \([X_iX(A)] = \{ \bar{u} \in \bar{U} \mid x_{a\bar{u}} \in X_i \}\). Then, \((A7)-(A9)\) imply,

\[
\rho_i^{\bar{u}}(A) = \sum_{\bar{u} \in \bar{U}} \bar{\mu}(\bar{u}) \rho_i^{\bar{u}}(A) = \sum_{\bar{u} \in [X_iX(A)]} \bar{\mu}(\bar{u}) \rho_i^{\bar{u}}(A) = \sum_{\bar{u} \in [X_iX(A)]} \bar{\mu}(\bar{u}) \sum_{j \in A\bar{u}} \eta_{ji}^{x_{A\bar{u}}}
\]

\[
= \sum_{x \in X_i} \sum_{\bar{u} \in [xX(A)]} \bar{\mu}(\bar{u}) \sum_{j \in a_x} \eta_{ji}^x = \sum_{x \in X_i} \sum_{j \in A \cap x} \eta_{ji}^x \sum_{\bar{u} \in [xX(A)]} \bar{\mu}(\bar{u})
\]

\[
= \sum_{x \in X_i} \frac{\eta_{ji}^x}{\sum_{j \in A \cap x} \eta_{ji}^x} \rho_i^{\bar{u}}(x, X(A))
\]

\[
= \sum_{x \in X_i} \frac{\eta_{ji}^x}{\sum_{j \in A \cap x} \eta_{ji}^x} \sum_{y \in X(A)} w_x \frac{w_y}{y} = q_iA.
\]

We conclude the proof of the Theorem by showing that \(\text{conv} Q_l \subset \text{cl} Q_a\). Since we have already established \(\text{cl} \text{conv} Q_l = Q_r\), this will imply \(Q_r \subset \text{cl} Q_a\). Then, Lemma A35 and Fact 3 yield \(\text{cl} Q_a = Q_r\) 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 \(\hat{v}^j\)s and hence ensure that \(\hat{q} = \sum_{j=1}^m \alpha_j \hat{v}^j\) 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 \(\eta_{ji}^j = Mv_i^j\) and note that \((w, \eta)\) represents \(\hat{q}\).

9. Proof of Theorem 4 and Propositions 2-4

Through out the proofs in this section, we distinguish between the duplicates relation for a choice rule on \(S\) and a choice rule on \(\Omega\) by letting \(\sim_S\) denote the former and let \(\sim_\Omega\) 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 \(\Omega\).

The Recursive Construction

All of the properties listed below are easy to verify: For any attribute system \((w, \eta)\) on \(S\) define its the recursive extension \((\hat{w}, \hat{\eta})\) to \(\Omega\) as follows: Let \(x_y = \{ \omega \in \Omega \mid \sigma(\omega) \cap y \neq \emptyset \}\).
Hence, the node $\omega$ 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 $y \rightarrow x_y$ is a bijection from $Z_S$ to $Z_{\Omega}$ and $x \rightarrow y$ where $y_x = \{s \in S \mid s \in \sigma(t) \text{ for some } t \in x\} = S \cap x$ is its inverse. Let $\hat{w}_{x_y} = w_y$ and define $\hat{\eta}_{x_y}$ recursively for all $s \in \Omega$: if $s \in S_0$, then $\hat{\eta}_{x_y} = \eta_y$. For $t \in t_n$, $\hat{\eta}_{t_y} = \sum_{s \in t} \hat{\eta}_{s_y}$. Clearly, every attribute system on $Z_S$ has a unique recursive extension to $Z_{\Omega}$.

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

### 9.1 Proof of Theorem 4

By Theorem 2, there exists an attribute system $(w, \eta)$ that represents $(\phi, \Omega)$. To prove the only if part of the Theorem, we will show that $\eta$ is recursive. First, note that claim 3 in the proof of Proposition 1 ensures that $t \sim_{\Omega} \hat{t}$ if and only if $\eta_{x} = \hat{\eta}_{x}$ 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 $\eta$ immediately implies that $\hat{\eta}_{D} = \sum_{\omega \in D} \eta_{\omega}$. Hence, $\eta^x(D) := \hat{\eta}_{D} = \sum_{\omega \in D} \eta_{\omega} = \eta^x(D)$. Then, appealing to Claim 3 in the proof of Proposition 1 again yields $\{D\} \sim_{\Omega} D$. 

### 9.2 Proof of Proposition 2

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

Recall that there is a bijection between $y \rightarrow x_y$ 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 $(w, \eta)$
represents \((\phi, \Omega), \hat{\eta}^x_A = \hat{\eta}^x_B\) 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; that is, \(\hat{\eta}^x_A = \hat{\eta}^x_B\) for all \(x\) if and only if \(A \sim_{\Omega} B\). Therefore, \(B\) is a fine duplicate of \(A\) if and only if \(\hat{\eta}^x_A = \hat{\eta}^x_B\) 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 \(B \cap B' \neq \emptyset\). Then, we must have \(\hat{\eta}^x_A \cdot \hat{\eta}^x_B > 0\) for some \(x\). It follows that if \(\hat{\eta}^x_A \cdot \hat{\eta}^x_B = 0\) for all \(x\), \(A\) and \(B\) are nonoverlapping. Also, in case (1) above, by Claim 4 of the proof of Proposition 1, \((\phi, S)\) is SR implies that \(\eta^x_A \cdot \eta^x_B = 0\) if and only if \(A, B\) are \((\phi, S)\)-nonoverlapping. Then, since \((\phi, S)\) is SR, for any \(A, \delta \in (0, 1)\), there is \(B\) such that \(\eta^x_{\sigma(A)} \cdot \eta^x_B = 0\), \(\rho(\sigma(A), AB) = \delta\). Hence, \(\hat{\eta}^x_A \cdot \hat{\eta}^x_B = 0\) and therefore \(\rho(A, AB) = \delta\); that is, \((\phi, \Omega)\) also satisfies part (i) of SR.

In case 2, since \((\phi, S)\) satisfies \(R\), for any \(t \in T, \delta \in (0, 1)\), there is \(B \subset S\) such that \(B \cap \sigma(t) = \emptyset\) and \(\rho(\sigma(t), \sigma(t)B) = \delta\); that is, \(\frac{\upsilon(\sigma(t))}{\upsilon(\sigma(t)) + \upsilon(B)} = \delta\). By the recursive construction, the fact that \(B \cap \sigma(t) = \emptyset\) implies \(\eta^x_{\sigma(t)} \cdot \eta^x_B = 0\) for all \(x\) and therefore \(\hat{\eta}^x_t \cdot \hat{\eta}^x_B = 0\) for all \(x\). Therefore, \(t\) is nonoverlapping with \(B\) and hence \(\rho_t(tB) = \frac{\upsilon(X_t)}{\upsilon(X_t) + \upsilon(X(B))} = \frac{\upsilon(\sigma(t))}{\upsilon(\sigma(t)) + \upsilon(B)} = \delta\), proving that \((\phi, \Omega)\) satisfies (i) of SR in case 2 as well.

For uniqueness, let \((\phi^1, \Omega), (\phi^2, \Omega)\) be two SR recursive attribute rules that are extension of the same \((\phi, S)\). 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.

If \((\phi, S)\) is an SR attribute rule, then for \(i = 1, 2\), let \((w^i, \eta^i)\) be the unique attribute system on \(S\) that has with the property that \((\hat{w}^i, \hat{\eta}^i)\) is its recursive extension. Since \((\hat{w}^i, \hat{\eta}^i)\) is complete, so is \((w^i, \eta^i)\) 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 extensions of \((w^i, \eta^i)\), we have \(\hat{w}^1 = \alpha \hat{w}^2\) and \(\hat{\eta}^1 = \hat{\eta}^2\). Then, Proposition 1 implies \((\phi^1, \Omega) = (\phi^2, \Omega)\).

If \((\phi, S)\) is a rich Luce rule, then the fact that \((\hat{w}^i, \hat{\eta}^i)\) is a recursive extension of \((w^i, \eta^i)\) for every \(x \in Z\) there must be an \((\phi^i, \Omega)\)-archetype \(s \in S\). But then, invoking recursiveness again, we note that \(Z^1 = Z^2 = \{s\} | s \in S\) and \(\hat{\eta}^1 = \hat{\eta}^2\). Let \(v^i_s = w_{\{s\}}\) for all \(s\) and note that since both \((w^i, \eta^i)\)'s are attribute systems for \((\phi, S)\), both \(v^i\)'s are Luce values for \((\phi, S)\). 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, \((\phi^1, \Omega) = (\phi^2, \Omega)\). \qed
9.3 Proof of Proposition 3

Let \((\phi, \Omega)\) be a strongly rich recursive attribute rule. We claim that \(s\) and \(s'\) must be \((\phi, \Omega)\)-nonoverlapping for every \(s, s' \in S\). Assume the contrary. Since \((\phi, \Omega)\) is SR, we can find \(\emptyset \neq C \subset S\) such that \(C\) is nonoverlapping with \(\{s, s'\}\). Let \(B = \{s, s'\} \cup C\) and \(A = \{s\} \cup C\). Since \(s, s'\) are overlapping and \(C\) does not overlap with \(\{s, s'\}\), we have

\[
\frac{\rho_s(B)}{\rho_s(B) + \rho(C; B)} < \frac{\rho_s(A)}{\rho_s(A) + \rho(C; A)}
\]

proving that \((\phi, S)\) cannot be a Luce rule.

Next, assume \((\phi, S)\) 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

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

proving that \((\phi, \Omega)\) is reducible.

9.4 Proof of Proposition 4

We first show that \(p_a(D^n_2) > p_a(D^n_1)\) for all \(n \geq 2\). When \(n = 2\) we have

\[
p_a(D^2_2) = \phi_a(\{a, b, \{b\}\}) = \alpha > \left[ \alpha + \frac{(1 - \alpha)}{n} \right] \cdot \alpha = \phi^{b_1}(\{b, D^1\}) \cdot \phi_a(D^1) = p_a(D^2_1)
\]

and when \(n \geq 3\) a straightforward calculation yields

\[
p_a(D^n_2) - p_a(D^n_1) = \frac{\alpha(1 - \alpha)[\alpha(n - 2) + (n - 1)^2]}{n(n - 1)^2} > 0.
\]

Next, note that for \(n > k\) we have

\[
\phi_a(\{a, b, D^n_{k+1}\}) = \frac{\alpha}{n} = \phi_a(\{a, b, D^n_k\})
\]

and

\[
\phi_b(\{a, b, D^n_{k+1}\}) = \frac{\alpha}{n + 1} = \phi_b(\{a, b, D^n_k\})
\]
hence choice probabilities are the same for decision problems \( D^n_k \) and \( D^n_{k+1} \) from stage 1 to stage \( k - 1 \). Together with the observation above, this implies that \( p_a(D^n_k) \) is strictly increasing in \( k \).

To show that \( p_a(D^n_2) < \alpha \) when \( n \geq 3 \), note that

\[
\alpha - p_a(D^n_2) = \alpha - \phi_a(D^n_2) - \phi_{D^n_1}(D^n_2) \cdot \phi_{D^n_2}(D^n_1) \cdot p_a(D^{n-2}) \\
= \alpha - \frac{\alpha}{n-1} - \left[ 1 - \frac{\alpha}{n-1} - \frac{(1 - \alpha)}{n} \right] \left[ \alpha + \frac{(1 - \alpha)}{n-1} \right] \alpha \\
= \frac{\alpha(1 - \alpha)[2 - \alpha + n(n - 3)]}{n(n - 1)^2} \\
> 0
\]

Finally, to show that \( p_a(D^n_n) > \alpha \), note that \( p_a(D^n_2) = \alpha = p_a(D^n_2) \) and, for all \( n \geq 3 \), \( \phi_a(D^n_n) > \phi_a(D^n) \), \( \phi_b(D^n_n) = \phi_b(D^n) \). Since \( p_a(D^n) = \alpha \), the result follows. \( \square \)
References


