Theoretical Economics 13 (2018), 61–85 1555-7561/20180061

On path independent stochastic choice

David S. Ahn
Department of Economics, University of California, Berkeley

Federico Echenique
Division of the Humanities and Social Sciences, California Institute of Technology

Kota Saito
Division of the Humanities and Social Sciences, California Institute of Technology

We investigate stochastic choice when only the average and not the entire distribution of choices is observable, focusing attention on the popular Luce model. Choice is path independent if it is recursive, in the sense that choosing from a menu can be broken up into choosing from smaller submenus. While an important property, path independence is known to be incompatible with continuous choice. The main result of our paper is that a natural modification of path independence, which we call *partial path independence*, is not only compatible with continuity, but ends up characterizing the ubiquitous Luce (or logit) rule.

Keywords. Luce model, stochastic choice, logit model, path independence.

JEL classification. D01, D11.

1. Introduction

This paper studies stochastic choice in linear environments where only average choices, rather than the entire distributions of choices, are observed. A distribution over choices describes the probability that each alternative is chosen, while average choice refers to the mean of the probability distribution over alternatives. Average statistics are simpler to interpret and report than entire distributions. While average choice data are less rich, for the ubiquitous Luce model of stochastic choice, average statistics lose no power in terms of identification. We provide a novel characterization of the Luce model for the average choice environment via a relaxation of the classic path independence axiom.

Studying average choice has several compelling motivations. Most salient is that the average, or other aggregate statistics of choices, are more parsimonious than a description of the entire distribution. For example, a restaurant that lists n different items

David S. Ahn: dahn@econ.berkeley.edu
Federico Echenique: fede@hss.caltech.edu
Kota Saito: saito@caltech.edu
We want to thank the Editor and two anonymous referees for very helpful comments on an earlier draft. We are also grateful to Michihiko Kandori and Jay Lu for suggestions on some alternative interpretations of our results. Echenique thanks the National Science Foundation for its support through Grants SES 1558757 and CNS-1518941. Saito thanks the National Science Foundation for its support through Grant SES 1558757.

https://doi.org/10.3982/TE2653

on its menu effectively offers 2^n different possible meals that depend on which items are included in the meal. You can order a hamburger with fries, but you can also get it with coleslaw or you could order salmon and a salad instead. A random choice must track how many agents purchase each different package of items: describing the entire distribution over meals requires $2^n - 1$ numbers. Alternatively, tracking an average by recording the proportion of agents who order each item has a dimensionality of n. To report the average, one simply has to record the number of hamburgers, salmons, portions of fries, and coleslaw consumed. The resulting statistic, the *average choice* from the restaurant, is the object of study in our paper.[1] Another example concerns financial decision. Consider the contrast between tracking an individual's monthly investment decisions over a lifetime of employment and tracking her total portfolio at retirement.

Of course, this simplicity comes with possible disadvantages. The average disregards potentially important information regarding the distribution of choices. In the restaurant example, seeing the proportion of diners ordering each item ignores possible correlations across items. However, in this paper we focus on the Luce rule, and show that the implied Luce weights maintain the same canonical level of uniqueness with only average choice data that they enjoy with information about the full distribution. Therefore, our exercise illuminates a novel methodological advantage of the Luce formula, namely that higher-order moments of the choice distribution are gratuitous for identifying its parameters. That is, while the benefits of restricting attention to average choice data are general, these benefits come with no loss of information when specialized to the Luce model. This is a consequence of Luce's sharp functional form: the weights of all objects can be uniquely identified by observing the average choice from doubletons, since the location of the average along the segment joining the two choices uniquely identifies the ratio of their Luce weights.

Beyond identification, we further contribute to the understanding of the Luce model by providing a full characterization for average choice data. Our main condition relaxes the classic path independence condition made famous by Plott (1973). Plott's condition requires that choices from large menus can be recursively computed by considering the subchoices from submenus and then choosing from those subchoices. More precisely, if x is chosen from submenu A and y is chosen from submenu B, then whichever is chosen from the pair $\{x, y\}$ is also chosen from the combined menu $A \cup B$. Path independence allows the decomposition of a large choice problem into smaller subproblems without affecting the final selection.

Our generalization, which we call partial path independence, only requires that if x is chosen from A and y is chosen from B, then the choice from $A \cup B$ is a convex combination—in other words, a lottery—of x and y. The important departure from Plott path independence is that the weights on x and y *can* depend on the sets A and B from which they were chosen. In other words, we allow for some degree of path dependence. To satisfy Plott path independence, the average choice from $A \cup B$ must be equal to the average choice from $A \cup B'$ for any other set B' that also yields y as its average choice. This is inconsistent with the Luce formula because if B' has many more elements, it will carry more weight than B when combined with A.

[1]We thank the co-editor for suggesting this example.

The resulting equivalence of our relaxed form of path independence to the Luce model is technically interesting in view of two earlier theoretical literatures. The first is a set of impossibility theorems for path independence. Although it is a well known necessary condition for rational deterministic choice, Kalai and Megiddo (1980) and Machina and Parks (1981) showed that classic path independence has restrictive implications for average stochastic choice: *path independence cannot be satisfied by any continuous model of stochastic choice*. In contrast, we demonstrate that an intuitive relaxation of path independence is not only compatible with continuous stochastic choice, but ends up characterizing the ubiquitous Luce model.

Average choice and its linear structure also yield additional interpretations for path independence that are absent in the general case. One prominent new interpretation is in terms of prior selection. An agent faces a set of possible priors (probability distributions) over states of the world and must choose one prior. There are many models of multiple priors and many theories of which prior gets selected, but these are usually guided by some consequentialist principle. For example, a prior may be chosen because it is pessimistic. Path independence can be understood as a recursive, or consistent, guide to choosing a prior: choices from larger sets $A \cup B$ are derived from the choices one would make from smaller sets. This imposes a consistency on choices as one goes from smaller to larger sets. The negative results of Kalai–Megiddo and Machina–Parks mean that continuity cannot be reconciled with such consistency or recursivity. In contrast, our result says that if one allows for the selections from $A \cup B$ to depend on the level of ambient uncertainty when a prior from A and a prior from B are selected, as reflected in the priors in the sets A or B, then this impossibility is avoided. In fact, the Luce model emerges as the only prior selection rule that is both continuous and partially path independent.

The second connected literature is the axiomatic theory of the Luce rule. By exploiting the convexity of linear spaces, our characterization departs from the standard understanding of the Luce model, which is well known to be equivalent to independence of irrelevant alternatives (IIA). When the Luce model is generalized or altered to accommodate alternative patterns of choice, IIA is almost always implicated. Our characterization exposes new and interesting features of the Luce model, in particular its exact connection to a classic axiom for deterministic choice. The relaxation of path independence illuminates the importance of convexity in decompositions for the Luce model. Aside from being necessary for the Luce model, IIA and our path independence condition do not share any immediate resemblance. In Section 4.1, we discuss the relation between IIA and our path independence axioms more carefully. Specifically, we explain why our main result cannot be proven by simply showing that our path independence implies IIA.

We hope that our axiomatization allows for subsequent generalizations of the Luce model that relax path independence rather than IIA. As an immediate illustration of the usefulness of using average choice, we sharpen the model and provide the first, to our knowledge, axiomatization of the *linear* Luce model. It is difficult to overstate the importance of the Luce model or its equivalent formulation as the logit model of discrete choice in applications. Logit is the workhorse specification of demand in most structural models of markets and a standard part of graduate econometric training. But in nearly all these applications, the utility function is specified to be linear in attributes. Despite their prevalence, the full empirical content of linear logit models was previously unknown. For example, in his influential paper justifying the logit model for empirical estimation, McFadden (1974) offers linearity as an axiom (McFadden's Axiom 4) rather than deriving linearity as a necessary implication of more basic conditions phrased on choice data. We show that an average choice version of the von Neumann–Morgenstern independence axiom ensures that the Luce utility function is ordinally equivalent to a linear function.

Finally, aside from providing novel theoretical foundations, our associated modification of path independence is practically useful as it is statistically very well behaved, and better suited than IIA for an empirical assessment of Luce's model, benefits that flow naturally from the reduced complexity and dimensionality of average choice as a primitive. In particular, the IIA axiom is more cumbersome and more volatile than partial path independence because testing IIA requires estimating the ratios of relative choice probabilities. These ratios are statistically more complicated objects than the average. With small samples, they are difficult to estimate robustly. Even with large samples, estimates of these ratios can have an arbitrarily large variance. We discuss these issues in detail in Section 4.4, where we demonstrate formally the improved behavior of the average.

The empirical difficulty in testing the IIA axiom is not a new observation. In fact, Luce (1959) already understood the complications of testing IIA when he introduced the condition. Using calculations similar to ours in Section 4.4, Luce concludes that a reliable test of IIA requires large data sets with several thousands of observations. The tests are, therefore, impracticable for most experimental designs.[2] Our axiomatization for average choice is much better behaved because averages can be estimated more efficiently than ratios of probabilities.

We close the Introduction by discussing some related papers. Although the primitives of their model are very different, Billot et al. (2005) use an axiom with a similar syntax to our partial path independence. They take as primitives an arbitrary set C of *cases* and a finite set Ω of states of the world. They characterize a function p mapping sequences in C to beliefs over Ω. Their *concatenation* axiom says that the beliefs induced by the concatenation of two data bases must lie in the line segment connecting the beliefs induced by each data base separately. This is analogous to partial path independence.

The environments and interpretations of the axioms are obviously different. Moreover, there are clear mathematical differences between their concatenation and our partial path independence. Their concatenation axiom has some additional bite in their application to ordered lists because it applies to data bases even when they share common cases. Concatenation is different from set union because it counts multiple copies

[2]Luce suggested using psychophysical experiments to obtain the kinds of sample sizes needed to test IIA.

of the same item and treats a list such as (a, b, b, b) as a separate object rather than the set $\{a, b\}$. This is technically useful because it allows calibration of a case's importance by considering duplicates of that case. This calibration plays a crucial first step in the proof of Billot et al. (2005).[3] This analysis in their proof is not possible in our application because our primitive of sets does not distinguish duplicates.

Several recent papers in economic theory have proposed generalizations of the Luce model, but in the classic environment of full random choice rather than our average choice setting. Gul et al. (2014) recently proposed an extension of the Luce model to address well known difficulties of the Luce model that arise when objects have common attributes. Gul et al. (2014) also provide an alternative axiomatization of the Luce model in rich environments. Dogan and Yildiz (2016) also propose an extension of the Luce model. They offer their main axiom, which they call *stochastic path independence*, as an analog of path independence for standard random choice environments with information about the full distribution, as opposed to the average choice settings studied in this paper and in the original papers by Kalai and Megiddo (1980) and Machina and Parks (1981). While both are inspired by path independence, we see no substantive connection between our condition and theirs; for example, their condition is incompatible with continuity for random choice, whereas we explicitly invoke continuity for average choice as an additional assumption.

In our paper, we focus on average choices that are continuous, but a recent generalization of our result due to Hamze (2017a) shows that the partial path independence axiom characterizes the class of all interior (even possibly discontinuous) Luce rules when the continuity axiom is dropped. Hamze (2017b) also discusses the case without an interiority axiom.

Our results can be viewed as the selection of a representative element from a set of vectors, with the representative being the average choice. An axiomatic literature in social choice studies electing representatives (see Chambers 2008), but to the best of our knowledge there are no results closely connected to ours. Perhaps the closest results are about choosing a representative prior in a Bayesian setting; see the discussion in Section 4.5. The work of Chambers and Hayashi (2010), who show that it is impossible to select a representative that is always consistent with Bayesian updating, is close in spirit to our discussion in Section 4.5.

2. Average choice and the Luce model

2.1 *Notation*

We begin by formally describing our model. Definitions of standard mathematical terms are provided in Appendix A.1.

If $A \subseteq \mathbf{R}^n$, then conv A denotes the the set of all convex combinations of elements in A, termed the *convex hull* of A. Let $\text{conv}^0 A$ denote the relative interior of conv A.

A *preference relation* is a weak order; that is, a complete and transitive binary relation.

[3]See pp. 1131–1132 of Billot et al. (2005).

A function $f: Y \to \mathbf{R}_+$ is a *finite support probability measure* on Y if there is a finite subset $A \subseteq Y$ such that $\sum_{x \in A} f(x) = 1$ and $f(x) = 0$ for $x \notin A$. In this case, we say that the set A is a *support* for f. When Y is a finite set, we use the term *lottery* to refer to a probability measure with finite support. Let $\Delta(Y)$ denote the set of all lotteries, and for $A \subset Y$, let $\Delta(A)$ denote the set of lotteries whose supports are contained in A.

2.2 *Primitives*

Let X be a compact and convex subset of $\mathbf{R}^n$ with $n \geq 2$. Without loss of generality, assume X is of full dimensionality n. An important special case is when X is a set of probability measures over a finite set. For example, if we let P be a finite set of *prizes* and let $X = \Delta(P)$ be the set of all lotteries over P parameterized as a subset of $\mathbf{R}^n$ with $n = |P| - 1$, we get the setting of Gul and Pesendorfer (2006). Or we can let S be a finite set of *states of the world* and let $X = \Delta(S)$ be the set of beliefs over S. The latter is the setting of Section 4.5.

Let $\mathcal{A}$ denote the family of all finite subsets of X, with the interpretation that $A \in \mathcal{A}$ is a menu of available options. An *average choice* is a function

$$\rho^*: \mathcal{A} \to X$$

such that $\rho^*(A) \in \operatorname{conv} A$ for all $A \in \mathcal{A}$. That is, an average choice takes a menu of options to a weighted average of those options, where weighted averages are represented as convex combinations of A.

2.3 *Luce model*

The more standard model of random choice does not reduce distributions to their averages, but records the shape of the entire choice distribution. A *stochastic choice* is a function

$$\rho: \mathcal{A} \to \Delta(X)$$

such that $\rho(A) \in \Delta(A)$.

A stochastic choice $\rho: \mathcal{A} \to \Delta(X)$ is a *continuous Luce rule* if there is a continuous function $u: X \to \mathbf{R}_{++}$ such that

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

Every stochastic choice ρ induces an average choice ρ^* in the obvious manner. The average choice ρ^* is *rationalized* by the stochastic choice ρ if

$$\rho^*(A) = \sum_{x \in A} x\rho(x, A)$$

for all $A \in \mathcal{A}$. An average choice ρ^* is *continuous Luce rationalizable* if there is a continuous Luce rule ρ that rationalizes ρ^*.

2.4 *A characterization of average choices satisfying partial path independence, interiority, and continuity*

Our main axiom modifies the *path independence* axiom proposed by Plott (1973). Plott's axiom imposes recursive structure on decomposed choice, in the sense that choice from $A \cup B$ is obtained by choosing first from A and B separately, and then from a set consisting of the two chosen elements. The idea is that knowing the chosen elements from A and B is sufficient to predict the choice from their union. Plott originally intended his condition for abstract determinate choice. For the same average choice environment that we consider, Kalai and Megiddo (1980) and Machina and Parks (1981) argue that Plott's axiom is too strong. They show that Plott path independence leads to an impossibility result. Our result is that a natural modification of Plott's axiom avoids the negative implications of the original axiom and instead provides an unanticipated characterization of the Luce model. So while Luce models are inconsistent with Plott's version of path independence, they are completely characterized by our modification.

Our notion of path independence generalizes Plott's version of path independence. It asserts that the average choice from $A \cup B$ is a convex combination of the average choice from A and the average choice from B. Formally, our axiom is as follows.

Partial path independence. *If $A \cap B = \varnothing$, then*

$$\rho^*(A \cup B) \in \operatorname{conv}\{\rho^*(A), \rho^*(B)\}.$$

As mentioned, the main relaxation of Plott's path independence is that the weights assigned to the choices $\rho^*(A)$ and $\rho^*(B)$ from the submenus can depend on the particular submenus from which they were chosen. More specifically, if $\rho^*(A) = \rho^*(A')$, Plott path independence implies $\rho^*(A \cup B) = \rho^*(A' \cup B)$. However, in the Luce model, the weight of A or A' versus B depends on the number of elements in A or A' and on their utility values. For example, if u is a constant function, then the weight of $\rho^*(A)$ and $\rho^*(A')$ versus $\rho^*(B)$ in $\rho^*(A \cup B)$ and $\rho^*(A' \cup B)$ is proportional to the cardinalities of A and A'. We formally discuss the relationship between our version of path independence and Plott's original axiom in Section 4.2.

We argue that partial path independence has its own normative appeal, separately from its relationship to Plott's condition. One interpretation is that the choice from $A \cup B$ must be a randomization of $\rho^*(A)$ and $\rho^*(B)$. If we think of the agent's choice as a random variable X and think of a menu A as being the conditioning event $X \in A$, then the expected choice $\rho^*(A)$ can be written as $\mathbf{E}(X|X \in A)$ and path independence is analogous to the tower property of averages with respect to conditional averages:

$$\mathbf{E}(X|X \in A \cup B) = \mathbf{E}(X|X \in A)\mathbf{Pr}(X \in A|X \in A \cup B) + \mathbf{E}(X|X \in B)\mathbf{Pr}(X \in A \cup B).$$

In this interpretation, the weight on $\rho^*(A)$ versus $\rho^*(B)$ is determined exactly by the probability of the optimal element living in A versus living in B. The formula is directly satisfied by the Luce formula, where the probability of choosing an outcome in a menu is computed as the conditional probability of its Luce weight with respect to the summed Luce weights over the menu. Here, our axiom is about conditional expectations rather than conditional probabilities. Notice that other important rules with background randomness, such as the random expected utility representation of Gul and Pesendorfer (2006) or the more general family of random utility models, fail to satisfy partial path independence. This is because the "probability space" parameterized in these representations is not directly defined on the choice objects themselves, but rather on indirect artifacts like utility functions or strict orders.

Our second condition is analogous to the positivity axiom in Luce's characterization of his model. Luce positivity requires that all elements have strictly positive probability, that is, $\rho(x, A) > 0$ whenever $x \in A$. In our model, the analogous requirement is that the average choice is in the strict interior of the convex hull of A since all elements have strictly positive weight in computing that average.

Interiority. *For all $A \in \mathcal{A}$,*

$$\rho^*(A) \in \operatorname{conv}^0 A.$$

Our final axiom imposes a limited version of continuity.

Continuity. *Let $x \notin A$. For any sequence x_n in X, if $x_n \to x$, then*

$$\rho^*\big(A \cup \{x\}\big) = \lim_{n\to\infty} \rho^*\big(A \cup \{x_n\}\big).$$

Continuity is a technical condition, but note that the stronger assumption that ρ^* is a continuous function of A is inconsistent with the Luce model because of the effects of nearby but identical objects. The nature of the discontinuity is closely related to the classic blue bus–red bus objections to IIA. We discuss these issues in Section 4.

Our main finding is that partial path independence, interiority, and continuity characterize the continuous Luce model.

Theorem 1. *An average choice is continuous Luce rationalizable if and only if it satisfies partial path independence, interiority, and continuity.*

We can gain some intuition for Theorem 1 from Figure 1. The central observation is that partial path independence pins down $\rho^*(A)$ uniquely from the value of $\rho^*(A')$ when $A' \subsetneq A$.[4] We construct a rationalizing Luce model from the average choice from sets of small cardinality and then use path independence to show that ρ^* must always coincide with the average choice generated from that Luce model.

Figure 1 shows how $\rho^*(A)$ is uniquely determined. The points x and y are extreme points of $\operatorname{conv}(A)$, where $A = \{x, y, z, w, q, r\}$. Partial path independence forces $\rho^*(A)$

[4] See also the discussion in Section 4.1, where we compare path independence and Luce's IIA, and argue that we cannot use IIA to the same effect as path independence.

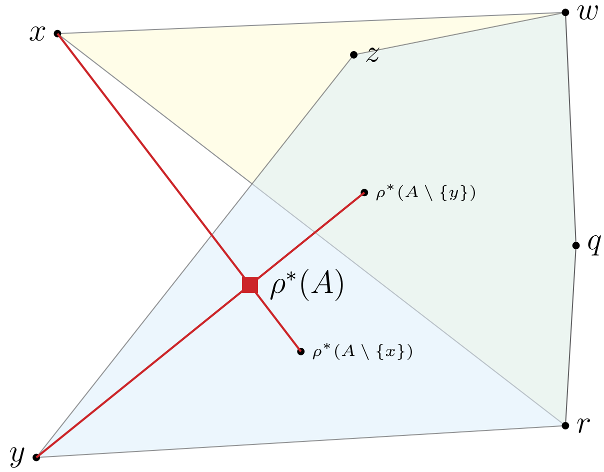

Figure 1. The set $A = \{x, y, z, w, q, r\}$: $\rho^*(A)$ is determined from partial path independence, $\rho^*(A \setminus \{x\})$, and $\rho^*(A \setminus \{y\})$.

simultaneously to lie on the line segment joining x and $\rho^*(A \setminus \{x\})$ and lie on the line segment joining y and $\rho^*(A \setminus \{y\})$. The main step in the proof (Lemma 6) establishes that, except for some "nongeneric" sets A, the intersection of the line segments x—$\rho^*(A \setminus \{x\})$ and y—$\rho^*(A \setminus \{y\})$ is a singleton, that is, $\rho^*(A)$ is uniquely identified from $\rho^*(A \setminus \{x\})$ and $\rho^*(A \setminus \{y\})$.

Figure 1 also illustrates why the line segments x—$\rho^*(A \setminus \{x\})$ and y—$\rho^*(A \setminus \{y\})$ have a singleton intersection. We choose x and y to lie on a proper face of conv A, so there is a hyperplane supporting A at that face. If there were two points in the intersection of x and $\rho^*(A \setminus \{x\})$ and y and $\rho^*(A \setminus \{y\})$, then all four points x, $\rho^*(A \setminus \{x\})$, y, and $\rho^*(A \setminus \{y\})$ would lie in the same line. This implies that $\rho^*(A \setminus \{x\})$ and $\rho^*(A \setminus \{y\})$ lie on the hyperplane as well. But since these two average choices are in the relative interior of the respective sets, it implies that A also lies in the hyperplane. This qualifies when A is not "not generic": this argument works whenever conv A has a nonempty interior.

Finally, as mentioned in the Introduction, the Luce representation is uniquely identified from average choice; in exactly the same manner it is identified up to a scalar multiple when observing the richer domain of full random choice. When ρ^* is Luce rationalizable and the Luce rule is defined by (1), then we say that the average choice function is *continuous Luce rationalizable by* u. The Luce model carries the same level of identification for its parameters with average choice that it enjoys with richer stochastic choices, that is, average choices provide sufficient data to pin down the Luce weights: If an average choice function is continuous Luce rationalizable by $u \in \mathbb{R}^X_{++}$ and by $v \in \mathbb{R}^X_{++}$, then there exists a positive real number α such that $u = \alpha v$.[5]

[5]In particular, this means that if we strengthen the partial path independence axiom to require that if $A \cap B = \varnothing$, then $\rho^*(A \cup B) = \frac{1}{2}\rho^*(A) + \frac{1}{2}\rho^*(B)$, we characterize the simple average. This is a well known result.

3. Linear Luce

We now specialize the representation result. Specifically, we consider continuous Luce models in which

$$u(x) = f(v \cdot x),$$

with $v \in \mathbf{R}^n$ and f a continuous strictly monotonic function. We term this the *linear Luce model*.

The assumption that u is linear "is maintained in the great majority of applications" of the multinomial logit (Train 1986, p. 35). To our knowledge, our results on linear Luce and the strictly affine Luce model are the first axiomatic characterizations of any version of Luce model with linearity in object attributes. Note that the function f can be thought of as measuring how important errors are, if we think of the Luce rule as a random utility model where there are "errors" in the utility function.

A stochastic choice $\rho : \mathcal{A} \to \Delta(X)$ is a *linear Luce rule* if there is $v \in \mathbf{R}^n$ and a monotone increasing and continuous function $f : \mathbf{R} \to \mathbf{R}_{++}$ such that

$$\rho(x, A) = \frac{f(v \cdot x)}{\sum_{y \in A} f(v \cdot y)}.$$

Consider the following axiom, which captures the kind of independence property that is normally associated with von Neumann–Morgenstern expected utility theory.

Independence. *If $\rho^*(\{x, y\}) = (1/2)x + (1/2)y$, then for all $\lambda \in \mathbf{R}$ and all z s.t. $\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z \in X$,*

$$\rho^*\big(\{\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z\}\big) = \lambda \rho^*\big(\{x, y\}\big) + (1-\lambda)z.$$

Note that if ρ^* is continuous Luce rationalizable with utility u, then $u(x) = u(y)$ if and only if $\rho^*(\{x, y\}) = (1/2)x + (1/2)y$. So the meaning of the independence axiom is that $u(x) = u(y)$ if and only if, for all $\lambda \in \mathbf{R}$ and all z s.t. $\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z \in X$, $\rho^*(\{\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z\}) = \lambda \rho^*(\{x, y\}) + (1-\lambda)z$.

The diagram in Figure 2 is a geometric illustration of independence. Note that the average choice $\rho^*(\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z)$ must be translated with λ from $\rho^*(x, y)$. Since we have $u(x) = u(y)$, this forces a rationalizing rule to translate indifference curves in λ in a similar fashion, so the indifference curves passing through the points $\lambda x + (1-\lambda)z$ and $\lambda y + (1-\lambda)z$, and through $\lambda' x + (1-\lambda')z$ and $\lambda' y + (1-\lambda)z'$, must be translates of the indifference curve passing through x and y.

The independence axiom ensures that u satisfies von Neumann–Morgenstern independence. The immediate implication of the axiom (Lemma 8) says that u satisfies a weak version of independence, but this version can be shown to suffice for the result (Lemma 9). The result is as follows.

Theorem 2. *An average choice is continuous linear Luce rationalizable if and only if it satisfies independence, partial path independence, interiority, and continuity.*

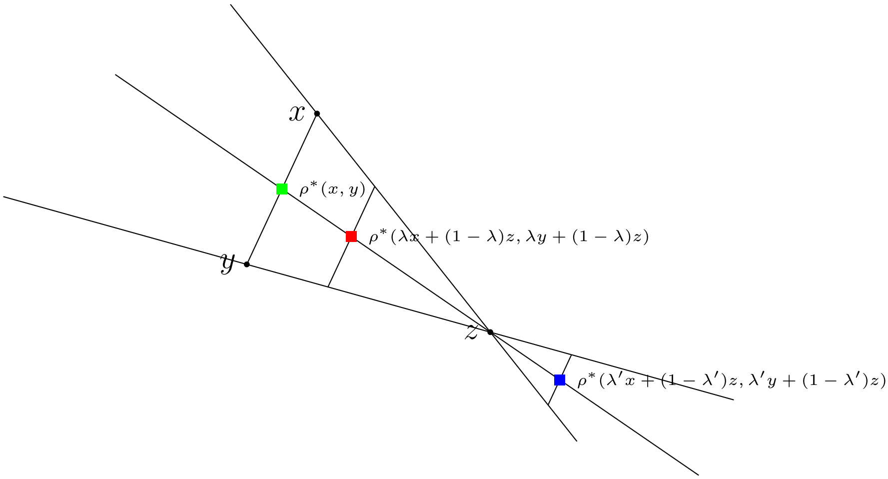

FIGURE 2. Independence.

4. Discussion

We conclude by discussing several features of our model and characterization. We begin by comparing our model and axiom to standard stochastic choice and Luce's IIA axiom, and we explain why the conditions are quite distant. We next compare our axiom to Plott's original path independence conditions, and show that they are disjoint in the sense that no average choice can satisfy both axioms. We next turn to our limited version of continuity and explain why more general continuity of ρ^* is violated by the Luce model. Finally, we compare the small-sample properties of testing our partial path independence condition to testing the classic IIA axiom, and we conclude that average choices are better behaved than ratios of choice probabilities.

4.1 *Partial path independence and IIA*

Taking a stochastic choice ρ as primitive, Luce (1959) famously showed that the Luce model is characterized by the IIA axiom

$$\frac{\rho(x, \{x, y\})}{\rho(y, \{x, y\})} = \frac{\rho(x, A)}{\rho(y, A)}$$

for any $A \ni x, y$.

While they both are invoked in characterizations of the Luce model, IIA and path independence are very different conditions. Consider the violation of path independence illustrated in Figure 3.

The average choice $\rho^*(\{x, y, z\})$ is outside the line segment connecting z with $\rho^*(\{x, y\})$. Instead, $\rho^*(\{x, y, z\})$ is above the line. If we write $\rho^*(\{x, y, z\})$ as a weighted average of the vectors x, y, and z, then this means that the weight on x, relative to the weight placed on y, is clearly higher than the relative weight placed on x in $\rho^*(\{x, y\})$. This implies a violation of Luce's IIA for any stochastic choice ρ that rationalizes ρ^*.

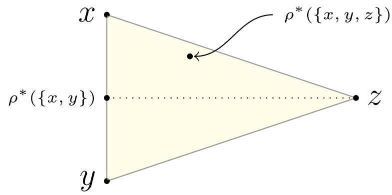

FIGURE 3. Path independence and Luces' IIA.

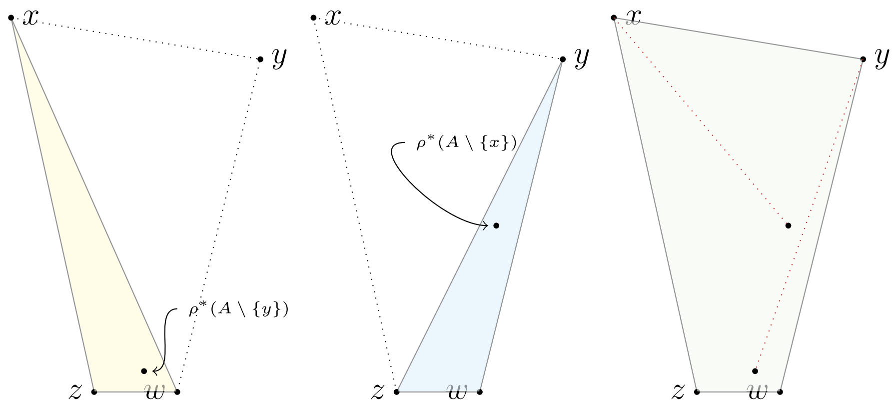

FIGURE 4. Violation of partial path independence: $A = \{x, y, z, w\}$.

So for at least some special cases as illustrated in Figure 3, a violation of path independence for the average choice ρ^* can sometimes imply a violation of IIA for any rationalizing stochastic choice ρ. However, the example is a red herring, because in fact there is no direct general relationship between path independence and IIA, and it is not straightforward to prove Theorem 1 by way of IIA. Consider the violation of path independence in Figure 4.

Figure 4 exhibits a violation of partial path independence. The figure depicts the average choice from some subsets of $A = \{x, y, z, w\}$, and these choices are incompatible with partial path independence. The diagram on the left shows where $\rho^*(\{x, z, w\})$ is in $\text{conv}\{x, z, w\}$, while the diagram in the middle depicts $\rho^*(\{y, z, w\})$ in $\text{conv}\{y, z, w\}$. It should be clear, as illustrated in the diagram on the right of Figure 4, that partial path independence cannot be satisfied because the lines x–$\rho^*(A \setminus \{x\})$ and y–$\rho^*(A \setminus \{y\})$ do not cross. However, in this case there is no required violation of IIA. Partial path independence takes conditions on the averages from smaller sets and imposes restrictions on the averages from larger sets. In fact, our proof proceeds by induction on the cardinality of the menu, and uses these restrictions to argue the inductive step from smaller to larger cardinalities. Alternatively, IIA imposes restriction across ratios, and not on averages. It would be surprising if there was an immediate connection between these different statistics.

Another way to see the difference between the axioms is to consider the gap in the information encoded in stochastic choice and average choice. While every stochastic choice reduces to a unique average choice, the reverse association is less crisp as several stochastic choice functions can rationalize the same average choice. For a menu of linearly independent vectors, the distribution is identified by the average. But when the menu is not an affinely independent set (as for all menus with more than n elements), the background distribution cannot be directly inferred from the average. For this reason, satisfying path independence is not generally informative of the veracity of the IIA assumption.

4.2 *Plott path independence*

Plott (1973) introduced the now-classic version of path independence. In our model, Plott's notion of path independence translates into the following condition.

Plott path independence. *If $A \cap B = \varnothing$, then*

$$\rho^*(A \cup B) = \rho^*\big(\{\rho^*(A), \rho^*(B)\}\big).$$

It is important to see how our path independence is different from Plott's. We require only that $\rho^*(A \cup B)$ be a strict convex combination of $\rho^*(A)$ and $\rho^*(B)$, but the weights in this convex combination do not need to coincide with the weights in the choice from the set $\{\rho^*(A), \rho^*(B)\}$. So the elements of $A \cup B$ are allowed to influence the average choice $\rho^*(A \cup B)$ by way of affecting the weights placed on $\rho^*(A)$ and $\rho^*(B)$.

Although these two versions of path independence are similar, they are fundamentally different in their implications. We already explained how our version of path independence crucially allows the weight of $\rho^*(A)$ to depend on the size and utility of the set A. A more direct way to see the tension is through the required conditions. As Kalai and Megiddo (1980) and Machina and Parks (1981) already observed, Plott path independence is incompatible with continuity. Plott's original version of path independence and our interiority axiom are also mutually incompatible. This follows as a corollary of Theorem 1 of Kalai and Megiddo (1980) that shows that Plott path independence implies that there exist x, y such that $\rho^*(A) = \rho^*(\{x, y\})$, which is incompatible with interiority.

4.3 *Continuity and Debreu's example*

The continuity axiom we are using is perhaps not the first such axiom one would think of (it certainly was not the first we thought of). A stronger axiom would demand that ρ^* be continuous, but this is incompatible with the Luce rule.

The reason for the failure relates to the following well known "blue bus–red bus" violation of IIA due to Debreu (1960) and Tversky (1972). An agent chooses a mode of transportation. Choosing x means taking a blue bus while choosing y means taking a cab. Alternative z is also a bus, but of a color red rather than blue. Debreu argues that the presence of z should strictly decrease the relative probability of choosing x over y, as the agent would consider whether to take a bus or a taxi and then be indifferent over which bus to take.[6]

[6] Debreu actually used an example of two different recordings of the same Beethoven symphony versus a suite by Debussy.

To see the connection to Debreu's example, suppose that ρ^* is continuous Luce rationalizable, with a continuous function u. Let z_n be a sequence in X converging to $x \in X$. Then

$$\rho^*(\{x, y\}) \neq \frac{2u(x)x + u(y)y}{2u(x) + u(y)} = \lim_{n\to\infty} \rho^*(\{x, y, z_n\}).$$

Thus ρ^* must be discontinuous.

The lack of continuity of ρ^* is similar to the blue bus–red bus example. The Luce model demands that z_n be irrelevant for the relative choice of x over y, even when z_n becomes very similar to x. In that sense, the lack of continuity of ρ^* is related to the blue bus–red bus phenomenon.

4.4 *Sampling error*

Theoretical studies of stochastic choice assume that choice probabilities are observed perfectly. But we intend our axioms to be useful as empirical tests, so that one can empirically decide whether observed data are consistent with the theory.

In an empirical study, however, choice probabilities must be estimated from sample frequencies. In the individual interpretation of the stochastic choice model (see the discussion in the Introduction), one agent makes a choice repeatedly from a set of available alternatives. This allows us to estimate the stochastic choice, but not to perfectly observe it. In the population interpretation (again, see the Introduction), we can observe the choices of a group of agents. The group may be large, and the fraction with which an agent makes a choice may be close to the population fraction of that choice, but there is likely some important randomness due to sampling. Here we want to argue that our axiomatization is better suited to dealing with such errors.

In his discussion of the Luce mode, Luce (1959) makes the point that testing his axiom, IIA, requires a large sample size. We follow up on Luce's remark and compare the efficiency of the IIA test statistic with the test statistic needed to test for partial path independence, the average of the choices.

Fix a set of alternatives A. Suppose that the population choices from A are given by $p \in \Delta(A)$, which comes from some continuous Luce rule $u : X \to \mathbf{R}_{++}$. We do not observe p but instead a sample $X_1, \dots, X_k$ of choices, with $X_i \in A$ for $i = 1, \dots, n$. The X_i are independent and distributed according to p. The probability p is estimated from the empirical distribution

$$p_x^k = \frac{|i : X_i = x|}{k}.$$

We have two alternative routes for testing the Luce model:

(i) Test the model using Luce's original axiomatization, by computing relative probabilities

$$\frac{p_x^k}{p_y^k}$$

for $x, y \in A$.

(ii) Test the model using our axiomatization, by computing average choice

$$\mu^k = \sum_{x \in A} x p_x^k.$$

We argue that option (ii) is better for two reasons. One reason is that, in small samples, it is more reliable to use the average than to use relative probabilities. The other reason is that, even in a large sample, the test statistic in (i) can have a very large variance relative to the test statistic in (ii). Specifically, for any M, there is an instance of the Luce model in which the asymptotic variance of the statistic in (i) is M times the asymptotic variance of the statistic in (ii).

We proceed to formalize the second claim.

Standard calculations yield

$$\sqrt{k}\left(\frac{p_x^k}{p_y^k} - \frac{p_x}{p_y}\right) \xrightarrow{d} N\left(0, \frac{2p_x^2}{p_y^2}\right).$$

Alternatively, $\sqrt{k}(\mu^k - \mu) \xrightarrow{d} N(0, \Sigma)$, where

$$\Sigma = (\sigma_{l,h}) \quad \text{and} \quad |\sigma_{l,h}| \leq \max\{x_l x_k : x \in A\}.$$

It is obvious that the entries of Σ are bounded and the asymptotic variance of p_x^k / p_y^k can be taken to be as large as desired by choosing a Luce model in which $u(x)/u(y)$ is large. So for any M, there is a Luce model for which the asymptotic variance of p_a^k / p_b^k relative to $\max\{\sigma_{l,h}\}$, the largest element of Σ, is greater than M.

4.5 *On Bayesian updating and path independence*

Up to now, we have interpreted our model as a reduction of stochastic choice. Here we consider a different interpretation: The elements of A are interpreted as (posterior) beliefs, and the randomization over the elements of a set A is viewed as a signal structure. In this reinterpretation of our model, partial path independence restricts two-stage signals in what we view as a normatively appealing way.[7]

Consider the following example. There are two states of the world: in state 1, the stock market goes up, and in state 2, it goes down. An agent first receives information from one of two different news sources. The first source is cable TV, denoted by A. The second source is a newspaper, denoted by B. Source A may send a number of possible signals, each one resulting in a different posterior. Think of A then as a finite collection of vectors $x \in \mathbf{R}^2$, with x_s being the probability of state s in signal realization x. Associated to A is a probability distribution $\rho(x, A)$ for $x \in A$. So $\rho(x, A)$ is the probability of observing signal x when receiving news from source A. Similarly, if the agent gets to read newspaper B, then she may acquire one of a finite set of possible posteriors. We

[7]We thank Michihiro Kandori and Jay Lu for comments leading to the discussion in this section. The interpretation discussed here is closer to the spirit of Billot et al. (2005), which we described in the Introduction.

view B as the set of possible posteriors, and associate a probability $\rho(x, B)$ with each $x \in B$.

Now, when the agent learns that she will receive news from source A, she will have beliefs $\rho^*(A)$ about the different states of the world. Bayesian updating demands that $\rho^*(A) = \sum_{x \in A} x\rho(x, A)$; in our language, that $\rho^*(A)$ is the average choice from A. Indeed, Bayesian updating means that for each $x \in A$, $x_s = \Pr(x|s)\Pr(s)/\Pr(x)$, where $\Pr(x|s)$ is the probability of the signal x in state s, and $\Pr(x)$ is the unconditional probability of signal x in signal structure A. Write $\Pr(x) = \rho(x, A)$ and $\Pr(s) = \rho^*(A)_s$. Then by adding over $x \in A$, one obtains $\sum_{x \in A} \rho(x, A)x = \rho^*(A)$. In other words, the average choice from A is the prior belief. The same is true for B and $\rho^*(B)$.

Given the interpretation of A and B as two sources of news, and given the resulting interpretation of $\rho^*(A)$ and $\rho^*(B)$ as the beliefs the agent should have upon learning that she will receive news from A or B, what can we say about $\rho^*(A \cup B)$? Before learning the source of news, our agent will have some beliefs $\rho^*(A \cup B)$ about the state of the world. It stands to reason that this belief should be a convex combination of $\rho^*(A)$ and $\rho^*(B)$. If our agent is Bayesian, she should ascribe some probability to getting news from source A and some probability to getting news from B. Then her beliefs, before she learns whether the news will come from A or B, should be a convex combination of the beliefs she would hold upon learning where the news will come from. Therefore, $\rho^*(A \cup B)$ should be a convex combination of $\rho^*(A)$ and $\rho^*(B)$. But this of course is our partial path independence axiom.

More generally, let S be a finite set, and interpret the elements of S as states of the world. Each $x \in X = \Delta(S)$ is a *belief* over S. We are going to interpret $A \in \mathcal{A}$ and $\rho(\cdot, A)$ as a *signal structure*. Specifically, suppose an agent has a prior belief $z \in \Delta(S)$ over S but can observe a signal that takes as many values as there are elements in some set A. Conveniently, we can identify a signal realization with its induced posterior belief that it gives rise to, so $x \in A$ is the value of the signal for which the updated posterior belief is x. The probability of observing $x \in A$ from the signal structure described by A and ρ is $\rho(x, A)$. This allows us to parameterize signaling structures as subsets of X. Bayesian updating means that

$$z = \sum_{x \in A} \rho(x, A)x = \rho^*(A).$$

In other words, the average choice from A is the prior belief.

Consider now the signal structure given by $\{\rho^*(A), \rho^*(B)\}$, together with some choice probabilities to be specified. We may think of the structure as representing signals whose realization is revealed in two stages. First, we learn whether we will observe a signal value from A or from B. Second, we learn the specific signal value in $A \cup B$. If we are told to expect a signal realization from A, then Bayesian updating dictates the belief $\rho^*(A)$, which is the expected posterior from signal structure A. Similarly, if we are told that the signal will come from B, our belief should be $\rho^*(B)$.

Partial path independence now demands that the prior from $A \cup B$ should be a weighted average of the intermediate beliefs $\rho^*(A)$ and $\rho^*(B)$. This demand is reasonable because the two intermediate signals ("the signal will come from A" and "the signal will come from B") summarize the a priori information contained in the signals in A and in B. In other words, if we are told that the signal will come from A, then the belief will be $\rho^*(A)$, and if we are told it will come from B, then the belief will be $\rho^*(B)$. So a priori our belief should be an expectation that places some weight on $\rho^*(A)$ and some weight on $\rho^*(B)$.

Path independence demands more than this. It says that the weights placed on $\rho^*(A)$ and $\rho^*(B)$ should be the same regardless of the quality of information in A or B. The weights should only depend on what the expected posteriors are under A and B, but not on the identity or number of the ultimate signals (if $\rho^*(A) = \rho^*(A')$ and $\rho^*(B) = \rho^*(B')$, then $\rho^*(A \cup B) = \rho^*(A' \cup B')$). At a prima facie level, this seems like an unreasonable requirement. At a structural level, the impossibility results of Kalai–Megiddo and Machina–Parks imply that it is not possible to satisfy path independence and continuity.

In contrast, Theorem 1 says that partial path independence and continuity pin down a unique model for the probability distribution of signals. The model has the Luce form. It says that there is a function u, which we can think of as measuring the likelihood of each signal, such that the probability of each signal $x \in A$ is a kind of conditional probability: $\rho(x, A) = u(x)/\sum_{y\in A} u(y)$.

Appendix A: Proof of Theorem 1

A.1 *Definitions from convex analysis*

Let $x, y \in \mathbf{R}^n$. The *line* passing through x and y is the set $\{x + \theta(y - x) : \theta \in \mathbf{R}\}$. The *line segment* joining x and y is the set $\{x + \theta(y - x) : \theta \in [0, 1]\}$. A subset of $\mathbf{R}^n$ is *affine* if it contains the line passing through any two of its members. A subset of $\mathbf{R}^n$ is *convex* if it contains the line segment joining any two of its members.

If A is a subset of $\mathbf{R}^n$, the *affine hull* of A is the intersection of all affine subsets of $\mathbf{R}^n$ that contain A. An *affine combination* of members of A is any finite sum $\sum_{i=1}^{l} \lambda_i x_i$ with $\lambda_i \in \mathbf{R}$ and $x_i \in A$, $i = 1, \dots, l$ and $\sum_{i=1}^{l} \lambda_i = 1$. The affine hull of A is equivalently the collection of all affine combinations of its members. If A is a subset of $\mathbf{R}^n$, the *convex hull* of A is the intersection of all convex subsets of $\mathbf{R}^n$ that contain A. A set A is *affinely independent* if none of its members can be written as an affine combination of the rest of the members of A.

A point $x \in A$ is *relative interior for* A if there is a neighborhood N of x in $\mathbf{R}^n$ such that the intersection of N with the affine hull of A is contained in A. The *relative interior* of A is the set of points $x \in A$ that are relative interior for A.

A *polytope* is the convex hull of a finite set of points. The *dimension* of a polytope P is $l - 1$ if l is the largest cardinality of an affinely independent subset of A. A vector $x \in A$ is an *extreme point* of a set A if it cannot be written as the convex combination of the rest of the members of A. A *face* of a convex set A is a convex subset $F \subseteq A$ with the property that if $x, y \in A$ and $(x + y)/2 \in F$, then $x, y \in F$. If F is a face of a polytope P, then F is also a polytope and there is $\alpha \in \mathbf{R}^n$ and $\beta \in \mathbf{R}$ such that $P \subseteq \{x \in \mathbf{R}^n : \alpha \cdot x \leq \beta\}$ and $F = \{x \in P : \alpha \cdot x = \beta\}$. If a polytope has dimension l, then it has faces of dimension $0, 1, \dots, l$ (Corollary 2.4.8 in Schneider 2013).

A.2 *Proof of Theorem 1*

The following axiom seems to be weaker than partial path independence, but it is not.

ONE-POINT PATH INDEPENDENCE. *If x is an extreme point of A, then*

$$\rho^*(A) \in \text{conv}^0\{x, \rho^*(A \setminus \{x\})\}.$$

We use one-point path independence in the proof of sufficiency in Theorem 1. Since partial path independence is satisfied by any continuous Luce rationalizable average choice, one-point path independence and path independence are equivalent (at least under the hypothesis of continuity).

We now proceed with the formal proof.

LEMMA 3. *Interiority and partial path independence imply one-point path independence.*

PROOF. Suppose that x is an extreme point of A. If $\{x\} = A$, then the result holds obviously. Consider the case where there exists $y \in A$ such that $y \neq x$. By partial path independence, we know

$$\rho^*(A) \in \text{conv}\{x, \rho^*(A \setminus \{x\})\}.$$

So to show the lemma, it suffices to show that $\rho^*(A) \notin \{x, \rho^*(A \setminus \{x\})\}$. By way of contradiction, suppose first that $\rho^*(A) = x$. Since there exists $y \in A$ such that $y \neq x$, $\rho^*(A) = x \notin \text{conv}^0 A$. This contradicts interiority.

Suppose, second, that $\rho^*(A) = \rho^*(A \setminus \{x\})$. Since x is an extreme point, $\rho^*(A) = \rho^*(A \setminus \{x\}) \notin \text{conv}^0 A$. This again contradicts interiority. □

The second lemma establishes necessity, and is also needed in the proof of sufficiency.

LEMMA 4. *If ρ^* is continuous Luce rationalizable, then it satisfies continuity, interiority, and partial path independence.*

PROOF. Interiority and continuity of the average choice are direct consequences of the positivity and continuity of u. Partial path independence is also simple: Note that when $\text{conv}\, A \cap \text{conv}\, B = \varnothing$, then A and B are disjoint. So it follows that

$$\begin{aligned}\left(\sum_{x\in A} u(x) + \sum_{y\in B} u(y)\right)\rho^*(A \cup B) &= \sum_{x\in A} u(x)x + \sum_{x\in B} u(x)x \\ &= \rho^*(A)\sum_{x\in A} u(x) + \rho^*(B)\sum_{x\in B} u(x).\end{aligned}$$

Since $A \cap B = \varnothing$,

$$\rho^*(A \cup B) \in \text{conv}\{\rho^*(A), \rho^*(B)\}.$$

□

To show sufficiency, we prove one preliminary lemma.

Lemma 5. *Suppose that* ρ^* *satisfies one-point path independence. For any* A*, then there exists* $\{\lambda_z\}_{z\in A}$ *such that* $\rho^*(A)=\sum_{z\in A}\lambda_z z$ *and* $\lambda_z>0$ *for all* $z\in A$.

Proof. The proof is by induction on $|A|$. Consider any A such that $|A|=2$. Then $A=\{x,y\}$ for some x,y such that $x\neq y$. Since both x and y are extreme points, by one-point path independence, $\rho^*(\{x,y\})=\lambda x+(1-\lambda)y$ for some $\lambda\in(0,1)$, as desired.

In general, suppose that the claim holds for A such that $|A|\leq k$; that is, $\rho^*(A)=\sum_{z\in A}\lambda_z z$ and $\lambda_z>0$ for any point $z\in A$.

Fix A such that $|A|=k+1$ and choose any extreme point $x\in A$. By one-point path independence and the induction hypothesis, there exists $\mu\in(0,1)$ such that $\rho^*(A)=\mu\rho^*(A\setminus x)+(1-\mu)x=\sum_{y\in A\setminus x}\mu\lambda_y y+(1-\mu)x$, where all of the coefficients are positive. □

The proof of the sufficiency of the axioms relies on two key ideas. One is that when A is affinely independent, $\rho^*(A)$ has a unique representation as a convex combination of the elements of A. Affine independence holds for sets A of small cardinality, and it allows us to construct a utility giving a Luce representation for such small sets. The other idea is the following lemma, which is used to show that one-point path independence determines average choice uniquely. We use the lemma to finish our proof by induction on the cardinality of the sets A.

Lemma 6. *Suppose that* ρ^* *satisfies one-point path independence. Let* A *be a finite set with* $|A|\geq 3$*, and let* $x,y\in A$ *with* $x\neq y$*. If* x *and* y *are extreme points of* A *and there is a proper face* F *of* $\operatorname{conv}(A)$ *with* $\dim(F)\geq 1$ *and* $x,y\in F$*, then*

$$\operatorname{conv}^0\big(\{x,\rho^*(A\setminus\{x\})\}\big)\cap\operatorname{conv}^0\big(\{y,\rho^*(A\setminus\{y\})\}\big)$$

is a singleton.

Proof. First, $\rho^*(A)\in\operatorname{conv}^0\{x,\rho^*(A\setminus\{x\})\}$ and $\rho^*(A)\in\operatorname{conv}^0\{y,\rho^*(A\setminus\{y\})\}$, as ρ^* satisfies one-point path independence. So

$$\varnothing\neq\operatorname{conv}^0\{x,\rho^*(A\setminus\{x\})\}\cap\operatorname{conv}^0\{y,\rho^*(A\setminus\{y\})\}.$$

Since $x,y\in F$, there is a vector p and a scalar α with $F\subseteq\{z\in X:p\cdot z=\alpha\}$—one of the hyperplanes supporting $\operatorname{conv}A$—and such that $\operatorname{conv}A\subseteq\{z\in X:p\cdot z\geq\alpha\}$. We prove that if there is

$$z^1,z^2\in\operatorname{conv}^0\{x,\rho^*(A\setminus\{x\})\}\cap\operatorname{conv}^0\{y,\rho^*(A\setminus\{y\})\},$$

with $z^1\neq z^2$, then $A\subseteq F$, contradicting that F is a proper face of $\operatorname{conv}(A)$.

Now $z^1,z^2\in\operatorname{conv}^0\{x,\rho^*(A\setminus\{x\})\}$ implies that there is $\theta_x,\theta_{\rho^*(A\setminus\{x\})}\in\mathbf{R}$ such that $x=z^2+\theta_x(z^1-z^2)$ and $\rho^*(A\setminus\{x\})=z^2+\theta_{\rho^*(A\setminus\{x\})}(z^1-z^2)$. Similarly, we have $y=z^2+\theta_y(z^1-z^2)$ and $\rho^*(A\setminus\{y\})=z^2+\theta_{\rho^*(A\setminus\{y\})}(z^1-z^2)$.

Then as a consequence of $p\cdot x=\alpha=p\cdot y$, we have

$$p\cdot z^2+\theta_x p\cdot(z^1-z^2)=p\cdot z^2+\theta_y p\cdot(z^1-z^2).$$

Since $\theta_x \neq \theta_y$ (as $x \neq y$), we obtain that $p \cdot (z^1 - z^2) = 0$. So

$$\alpha = p \cdot x = p \cdot z^2 + p \cdot \theta_x \big(z^1 - z^2\big) = p \cdot z^2.$$

Then

$$p \cdot \rho^* \big(A \setminus \{x\}\big) = p \cdot \big(z^2 + \theta_{\rho^*(A \setminus \{x\})} \big(z^1 - z^2\big)\big) = p \cdot z^2 = \alpha.$$

But $\rho^*(A \setminus \{x\}) = \sum_{z \in A \setminus \{x\}} \lambda_z z$ for some $\lambda_z > 0$ for all $z \in A \setminus \{x\}$, by Lemma 5. Then $p \cdot z \geq \alpha$ for all $z \in A$ implies that $p \cdot z = \alpha$ for all $z \in A \setminus \{x\}$.

This means that $A \subseteq F$, contradicting that F is a proper face of conv A. □

The proof of sufficiency proceeds by first constructing a stochastic choice ρ, and then arguing that it is a Luce rule that rationalizes ρ^*. In Step 1, we define $\rho(A)$ for A with $|A| = 2$. In Step 2, we define $\rho(A)$ for A with $|A| = 3$. In Steps 3 and 4, we define ρ on all of $\mathcal{A}$ by constructing a continuous Luce rule and using this rule to define ρ. The average choice defined from the Luce rule must be path independent, so Lemma 6 is used to show that the average choice from the constructed Luce rule coincides with ρ^*.

Step 1. Let $x, y \in X$, $x \neq y$. Since $\rho^*(\{x, y\})$ is in the relative interior of $\{x, y\}$ (Lemma 5), there is a unique $\theta \in (0, 1)$ with $\rho^*(\{x, y\}) = \theta x + (1 - \theta) y$. Define $\rho(x, \{x, y\}) = \theta$ and $\rho(y, \{x, y\}) = 1 - \theta$.

Moreover, by continuity of ρ^*, we have

$$\rho\big(x_n, \{x_n, y\}\big) \to \rho\big(x, \{x, y\}\big), \rho\big(y, \{x_n, y\}\big) \to \rho\big(y, \{x, y\}\big) \quad \text{as } x_n \to x. \tag{2}$$

Thus we have defined $\rho(A)$ for A with $|A| = 2$, and established that ρ satisfies (2).

Step 2. Now turn to A with $|A| = 3$. Let $A = \{x, y, z\}$.

Case 1. Consider the case when the vectors x, y, and z are affinely independent. Such collections of three vectors exist because $n \geq 2$. Then there are unique $\rho(x, A)$ $\rho(y, A)$, and $\rho(z, A)$ (nonnegative and adding up to 1) such that $\rho^*(A) = x\rho(x, A) + y\rho(y, A) + z\rho(z, A)$. Since $\rho^*(A)$ is in the relative interior of conv(A) (Lemma 5) in fact, $\rho(x, A), \rho(y, A), \rho(z, A) > 0$.

By one-point path independence, there is $\theta \in (0, 1)$ such that

$$\rho^*(A) = \theta z + (1 - \theta)\rho^*\big(A \setminus \{z\}\big) = \theta z + (1 - \theta)\big[x\rho\big(x, \{x, y\}\big) + y\rho\big(y, \{x, y\}\big)\big].$$

The vectors x, y, and z are affinely independent, so the weights $\rho(x, A)$, $\rho(y, A)$, and $\rho(z, A)$ are unique. This implies that $\rho(x, A) = (1 - \theta)\rho(x, \{x, y\})$ and $\rho(y, A) = (1 - \theta)\rho(y, \{x, y\})$. Hence

$$\frac{\rho(x, A)}{\rho(y, A)} = \frac{\rho\big(x, \{x, y\}\big)}{\rho\big(y, \{x, y\}\big)}. \tag{3}$$

In the same way, we can show $\frac{\rho(z, A)}{\rho(y, A)} = \frac{\rho(z, \{z, y\})}{\rho(y, \{z, y\})}$.

Case 2. Consider now the case when the vectors x, y, and z are not affinely independent. Choose a sequence $\{z_n\}$ such that

(a) x, y and z_n are affinely independent for all n

(b) $z = \lim_{n\to\infty} z_n$

(c) $\rho(\{x, y, z_n\})$ converges.

To see that it is possible to choose such a sequence, note that x, y, and z_n are affinely independent if and only if $x - z_n$ and $y - z_n$ are not collinear. Now $x - z$ and $y - z$ are collinear, so there is $\theta \in \mathbf{R}$ with $x - z = \theta(y - z)$. For each n, the ball with center z and radius $1/n$ has full dimension, so the intersection of this ball with the complement in X of the line passing through x and z (which is also the line passing through y and z) is nonempty. By choosing z_n in this ball, but outside of the line passing through x and z, we obtain a sequence that converges to z. Then we have that

$$\frac{(x - z_n)_i}{(y - z_n)_i} = \frac{\theta(y - z)_i + (z - z_n)_i}{(y - z)_i + (z - z_n)_i} = \frac{\theta + \dfrac{(z - z_n)_i}{(y - z)_i}}{1 + \dfrac{(z - z_n)_i}{(y - z)_i}},$$

a ratio that is not a constant function of i, as z_n is not on the line passing through y and z. Finally, by going to a subsequence if necessary, we can ensure that condition (c) holds because the simplex is compact.[8] Define $\rho(\{x, y, z\})$ to be the limit of $\rho(\{x, y, z_n\})$.

In Case 1, we have shown that (3) hold for sets of three affinely independent vectors. So $\rho(x, \{x, y, z_n\})/\rho(y, \{x, y, z_n\}) = \rho(x, \{x, y\})/\rho(y, \{x, y\})$ for all n. Hence, $\rho(x, A)/\rho(y, A) = \rho(x, \{x, y\})/\rho(y, \{x, y\})$. In particular, this means that $\rho(x, A), \rho(y, A) \in (0, 1)$, as $\rho(x, \{x, y\}), \rho(y, \{x, y\}) \in (0, 1)$.

Again, the fact that (3) holds for sets of three affinely independent vectors implies

$$\frac{\rho(z_n, \{x, y, z_n\})}{\rho(y, \{x, y, z_n\})} = \frac{\rho(z_n, \{z_n, y\})}{\rho(y, \{z_n, y\})}.$$

Remember that, by (c), $\frac{\rho(z_n,\{x,y,z_n\})}{\rho(y,\{x,y,z_n\})} \to \frac{\rho(z,A)}{\rho(y,A)}$, and by (2), ρ and $\frac{\rho(z_n,\{z_n,y\})}{\rho(y,\{z_n,y\})} \to \frac{\rho(z,\{z,y\})}{\rho(y,\{z,y\})}$ as $n \to \infty$. Then

$$\frac{\rho(z, A)}{\rho(y, A)} = \frac{\rho(z, \{z, y\})}{\rho(y, \{z, y\})}.$$

In particular, $\rho(z, A) \in (0, 1)$ because $\rho(z, \{z, y\}) \in (0, 1)$.

Thus we have established that, for any three distinct vectors $x, y, z \in X$ (affinely independent or not), (3) holds with $A = \{x, y, z\}$.

Step 3. Now we turn to the definition of ρ on $\mathcal{A}$. The definition proceeds by induction. We use (3) to define a utility function $u : X \to \mathbf{R}_{++}$ and a Luce rule. Then we show by induction, and using Lemma 6, that ρ rationalizes ρ^*.

Fix $y_0 \in X$. Let $u(y_0) = 1$. For all $x \in X$, let $u(x) = \frac{\rho(x,\{x,y_0\})}{\rho(y_0,\{x,y_0\})}$. Note that u is a continuous function, as $x \mapsto (\rho(x, \{x, y_0\}), \rho(y_0, \{x, y_0\}))$ is continuous, and that $u > 0$. By (3), we obtain

$$\frac{u(x)}{u(y)} = \frac{\rho(x, \{x, y_0\})}{\rho(y_0, \{x, y_0\})}\frac{\rho(y_0, \{y, y_0\})}{\rho(y, \{y, y_0\})} = \frac{\rho(x, \{x, y\})}{\rho(y, \{x, y\})} = \frac{\rho(x, \{x, y, z\})}{\rho(y, \{x, y, z\})}$$

for all $x, y, z \in X$.

[8] For all n, $\rho(\{x, y, z_n\}) \in \{(a, b, c) \in \mathbf{R}^3 | a, b, c \geq 0 \text{ and } a + b + c = 1\}$.

Let $\rho_u(A)$ be the Luce rule defined by u: for all $x \in A \in \mathcal{A}$,

$$\rho_u(x, A) = \frac{u(x)}{\sum_{y \in A} u(y)}.$$

This definition of ρ_u coincides with the definition of ρ we have given for A with $|A| \leq 3$ because

$$\frac{\rho_u(x, \{x, y\})}{\rho_u(y, \{x, y\})} = \frac{u(x)}{u(y)} = \frac{\rho(x, \{x, y\})}{\rho(y, \{x, y\})} \quad \text{and} \quad \frac{\rho_u(x, \{x, y, z\})}{\rho_u(y, \{x, y, z\})} = \frac{u(x)}{u(y)} = \frac{\rho(x, \{x, y, z\})}{\rho(y, \{x, y, z\})}$$

for all $x, y, z \in X$. So we simply write ρ instead of ρ_u.

Note also that, by definition of ρ,

$$\begin{aligned} \rho^*(\{x, y\}) &= x\rho(x, \{x, y\}) + y\rho(y, \{x, y\}) \quad \text{and} \\ \rho^*(\{x, y, z\}) &= x\rho(x, \{x, y, z\}) + y\rho(y, \{x, y, z\}) + z\rho(z, \{x, y, z\}) \end{aligned} \tag{4}$$

for all $x, y, z \in X$. This establishes the desired result for A with $|A| \leq 3$.

Step 4. Let

$$\bar{\rho}(A) = \sum_{x \in A} x\rho_u(x, A).$$

We prove that $\bar{\rho}(A) = \rho^*(A)$ for all $A \in \mathcal{A}$, which finishes the proof of the theorem.

The proof proceeds by induction on the size of A. We have established (see (4)) that $\bar{\rho}(A) = \rho^*(A)$ for $|A| \leq 3$. Suppose then that $\bar{\rho}(A) = \rho^*(A)$ for all $A \in \mathcal{A}$ with $|A| \leq k$. Let $A \in \mathcal{A}$ with $|A| = k + 1$. We prove that $\bar{\rho}(A) = \rho^*(A)$.

Case 1. Suppose first that $\dim(\text{conv}\, A) \geq 2$. Then, by Corollary 2.4.8 in Schneider (2013), there is a proper face F of $\text{conv}(A)$ with dimension ≥ 1. Let $x, y \in A$ be two distinct extreme points of $\text{conv}(A)$, such that $x, y \in F$. Such x and y exist because F has dimension at least 1. Note that $\bar{\rho}(A \setminus \{x\}) = \rho^*(A \setminus \{x\})$, so we have $\bar{\rho}(A \setminus \{y\}) = \rho^*(A \setminus \{y\})$ by the inductive hypothesis. So the intersections

$$\begin{aligned} &\text{conv}^0(\{x, \bar{\rho}(A \setminus \{x\})\}) \cap \text{conv}^0(\{y, \bar{\rho}(A \setminus \{y\})\}) \\ &\quad = \text{conv}^0(\{x, \rho^*(A \setminus \{x\})\}) \cap \text{conv}^0(\{y, \rho^*(A \setminus \{y\})\}) \end{aligned} \tag{5}$$

coincide. Lemma 6 implies that the intersection in (6) is a singleton.

By Lemma 4, $\bar{\rho}$ satisfies partial path independence. Since $\bar{\rho}$ and ρ^* both satisfy partial path independence on $A \in \mathcal{A}$,

$$\bar{\rho}(A) \in \text{conv}^0(\{x, \bar{\rho}(A \setminus \{x\})\}) \cap \text{conv}^0(\{y, \bar{\rho}(A \setminus \{y\})\}) \tag{6}$$

and

$$\rho^*(A) \in \text{conv}^0(\{x, \rho^*(A \setminus \{x\})\}) \cap \text{conv}^0(\{y, \rho^*(A \setminus \{y\})\}). \tag{7}$$

Therefore, by (5), (6), and (7), we have $\bar{\rho}(A) = \rho^*(A)$.

Case 2. Consider the case when $\dim(A)=1$. Let $z\in A$ be an extreme point of conv A. Let $\{z_n\}$ be a sequence with $z_n\to z$, such that $A\setminus\{z\}\cup\{z_n\}$ has dimension ≥ 2. This is possible because X has dimension larger than 3. The we obtain that $\bar{\rho}(A)=\rho^*(A)$ by continuity of ρ^* and $\bar{\rho}$.

Appendix B: Proof of Theorem 2

Lemma 7. *If ρ^* is continuous linear Luce rationalizable, then it satisfies independence.*

Proof. Let ρ^* be continuous linear Luce rationalizable, with $u(x)=f(v\cdot x)$. We check that it satisfies independence. Note first that

$$\begin{aligned}\rho^*\big(\big\{\lambda x+(1-\lambda)z,\lambda y+(1-\lambda)z\big\}\big)&=\mu\big(\lambda x+(1-\lambda)z\big)+(1-\mu)\big(\lambda y+(1-\lambda)z\big)\\&=\lambda\big(\mu x+(1-\mu)y\big)+(1-\lambda)z,\end{aligned}$$

where

$$\mu=\frac{f\big(v\cdot\big(\lambda x+(1-\lambda)z\big)\big)}{f\big(v\cdot\big(\lambda x+(1-\lambda)z\big)\big)+f\big(v\cdot\big(\lambda y+(1-\lambda)z\big)\big)}.$$

Suppose first that $u(x)=u(y)$. Then $v\cdot x=v\cdot y$, as f is monotone increasing. Then $f(v\cdot(\lambda x+(1-\lambda)z))=f(v\cdot(\lambda y+(1-\lambda)z))$. This means that $\mu=1/2$. So $\mu x+(1-\mu)y=\rho^*(\{x,y\})$, because $u(x)=u(y)$ implies that $\rho^*(\{x,y\})=\frac{1}{2}x+\frac{1}{2}y$. □

Let ρ^* be continuous Luce rationalizable with utility function u. Write $x\sim y$ when $u(x)=u(y)$.

Lemma 8. *If ρ^* satisfies independence, then it satisfies the property*

$$x\sim y \textit{ iff } \forall\lambda\in\mathbf{R}\,\forall z\in X\big[\lambda x+(1-\lambda)z,\lambda y+(1-\lambda)z\in X\Rightarrow\lambda x+(1-\lambda)z\sim\lambda y+(1-\lambda)z\big].$$

Proof. Let $x\sim y$, $\lambda\in\mathbf{R}$ and $z\in X$. Let μ be such that

$$\begin{aligned}\rho^*\big(\big\{\lambda x+(1-\lambda)z,\lambda y+(1-\lambda)z\big\}\big)&=\mu\big(\lambda x+(1-\lambda)z\big)+(1-\mu)\big(\lambda y+(1-\lambda)z\big)\\&=\lambda\big(\mu x+(1-\mu)y\big)+(1-\lambda)z.\end{aligned}$$

Then independence implies that

$$\begin{aligned}\lambda\big(\mu x+(1-\mu)y\big)+(1-\lambda)z&=\rho^*\big(\big\{\lambda x+(1-\lambda)z,\lambda y+(1-\lambda)z\big\}\big)\\&=\lambda\rho^*\big(\{x,y\}\big)+(1-\lambda)z,\end{aligned}$$

and thus $\mu x+(1-\mu)y=\rho^*(\{x,y\})$. Since $x\sim y$, we must have $\mu=1/2$. Hence

$$\frac{1}{2}=\frac{u\big(\lambda x+(1-\lambda)z\big)}{u\big(\lambda x+(1-\lambda)z\big)+u\big(\lambda y+(1-\lambda)z\big)}.$$

Therefore, $\lambda x+(1-\lambda)z\sim\lambda y+(1-\lambda)z$.

Conversely, if $\lambda x+(1-\lambda)z \sim \lambda y+(1-\lambda)z$ for all λ and all z, then $x \sim y$ by continuity of u and the fact that $\lambda x+(1-\lambda)z, \lambda y+(1-\lambda)z \in X$. □

The property in Lemma 8 is weaker than the standard von Neumann–Morgenstern independence property (restricted to $\lambda \in (0,1)$). Using Lemma 8, however, we can establish the stronger independence property, as stated in the next lemma. Then the proof of Theorem 2 follows from the expected utility theorem: the preference relation represented by u is a weak order, it satisfies continuity (as u is continuous), and it satisfies independence.

Lemma 9. *If ρ^* is continuous Luce rationalizable and it satisfies independence, then it satisfies the property*

$$u(x) \geq u(y) \text{ iff } \forall \lambda \in [0,1]\ \forall z \in X\, u\big(\lambda x+(1-\lambda)z\big) \geq u\big(\lambda y+(1-\lambda)z\big).$$

Proof. Note that by Lemma 8, independence implies that

$$x \sim y \text{ iff } x+\theta(z-x) \sim y+\theta(z-y)$$

for all scalars θ and for all z such that $x+\theta(z-x), y+\theta(z-y) \in X$.

Suppose toward a contradiction that $u(x) \geq u(y)$ but that $u(\lambda x+(1-\lambda)z) < u(\lambda y+(1-\lambda)z)$. By continuity of u, there is $\lambda^* < \lambda$ such that $u(\lambda^* x+(1-\lambda^*)z) = u(\lambda^* y+(1-\lambda^*)z)$.

Let $\lambda^* x+(1-\lambda^*)z = x'$. Then $x = x' - \frac{1-\lambda^*}{\lambda^*}(z-x')$. Similarly, $y = y' - \frac{1-\lambda^*}{\lambda^*}(z-y')$, where $\lambda^* y+(1-\lambda^*)z = y'$. Then independence and $u(x') = u(y')$ imply that $u(x) = u(y)$. Then $u(\lambda x+(1-\lambda)z) < u(\lambda y+(1-\lambda)z)$ is a violation of independence. □

References

Billot, Antoine, Itzhak Gilboa, Dov Samet, and David Schmeidler (2005), "Probabilities as similarity-weighted frequencies." *Econometrica*, 73, 1125–1136. [64, 65, 75]

Chambers, Christopher P. (2008), "Consistent representative democracy." *Games and Economic Behavior*, 62, 348–363. [65]

Chambers, Christopher P. and Takashi Hayashi (2010), "Bayesian consistent belief selection." *Journal of Economic Theory*, 145, 432–439. [65]

Debreu, Gerard (1960), "Review of R. Duncan Luce, Individual choice behavior: A theoretical analysis." *American Economic Review*, 50, 186–188. [73]

Dogan, Sehart and Kemal Yildiz (2016), "A preference based extension of the Luce rule." Report, Bilkent. [65]

Gul, Faruk, Paulo Natenzon, and Wolfgang Pesendorfer (2014), "Random choice as behavioral optimization." *Econometrica*, 82, 1873–1912. [65]

Gul, Faruk and Wolfgang Pesendorfer (2006), "Random expected utility." *Econometrica*, 74, 121–146. [66, 68]

Hamze, Hamed (2017a), "Combination axiom and applications." Report, Caltech. [65]

Hamze, Hamed (2017b), "On weighted average aggregation." Report, Caltech. [65]

Kalai, Ehud and Nimrod Megiddo (1980), "Path independent choices." *Econometrica*, 48, 781–784. [63, 65, 67, 73]

Luce, R. Ducan (1959), *Individual Choice Behavior a Theoretical Analysis*. John Wiley and Sons, New York. [64, 71, 74]

Machina, Mark J. and Robert P. Parks (1981), "On path independent randomized choice." *Econometrica*, 49, 1345–1347. [63, 65, 67, 73]

McFadden, D. (1974), "Conditional logit analysis of qualitative choice behavior." In *Frontiers in Econometrics* (P. Zarembka, ed.), 105. Academic Press, New York. [64]

Plott, Charles R. (1973), "Path independence, rationality, and social choice." *Econometrica*, 41, 1075–1091. [62, 67, 73]

Schneider, Rolf (2013), *Convex Bodies: The Brunn–Minkowski Theory*. Cambridge University Press, Cambridge. [77, 82]

Train, Kenneth (1986), *Qualitative Choice Analysis*. MIT Press, Cambridge. [70]

Tversky, Amos (1972), "Elimination by aspects: A theory of choice." *Psychological review*, 79, 281. [73]

Co-editor Ran Spiegler handled this manuscript.

Manuscript received 30 September, 2016; final version accepted 5 May, 2017; available online 9 May, 2017.