Decision Making with Recommendation

Paul Cheung* Yusufcan Masatlioglu†

September 30, 2021

Abstract

Recommendations play an undeniable role in decision-making. While the use of recommendations has become widespread across different realms of economic activity, there is no decision-theoretic work in this area. This paper aims to fill that gap by providing a new theoretical foundation for applied and empirical work on recommendations. We consider both deterministic and stochastic choice under recommendations. The stochastic version of the model enables us to study aggregate behavior observed in the real world. We also provide a parametric stochastic model in which the parameters are uniquely identified with minimal data requirements. This gives us a frugal way to make out-of-sample predictions for application purposes. All three characterizations are based on simple and intuitive behavioral postulates that offer straightforward tests for our models.

1 Introduction

Recommendation is one of the key determinants in decision-making. For instance, we constantly rely on recommendation from our friends, consumer reports, mass media when selecting a movie to see; a book to read; a car to buy; a school to send our children.[1] It is not surprising that many internet sites such as Amazon.com, Netflix, YouTube, LinkedIn, Spotify, Tripadvisor, and Facebook, incorporated similar tools in their website to help their customer. "*Frequently bought together*" and "*Best-selling*" of Amazon.com, and "*the Top Picks*" of Netflix are just a few examples of these recommendation tools.[2]

The choices of an individual are directly influenced by the recommendations they receive. The evidence on recommendations altering choice behavior is conclusive and indisputable across

*Department of Economics, University of Maryland. Email: hycheung@umd.edu

†Department of Economics, University of Maryland. Email: yusufcan@umd.edu

[1] The recommender website, such as Angie's List, HomeAdvisor, Houzz, Thumbtack, shows that people are willing to pay for recommendations, which capture word of mouth wisdom. In order words, consumers value this information.

[2] According to the 2013 data released by McKinsey & Company, recommendation systems drive 35% of purchases at Amazon. Similarly, 75% of what people watch on Netflix is initiated by their product recommendations.

a wide spectrum of economic activities (e.g. for labour market, Horton (2017); for hospitality and tourism, Litvin et al. (2008); for music streaming service, Adomavicius et al. (2018) and Li et al. (2007); for e-commerce (of commodities or goods), Goodman et al. (2013), Häubl and Trifts (2000), Rowley (2000), Senecal and Nantel (2004), and Vijayasarathy and Jones (2000)). Despite its practical importance, we still do not understand how recommendation affects choices. In this paper, we aim at filling the gap by studying theoretically the effect of recommendations on choices by using revealed preference techniques. We propose a simple model of recommendation and investigate the empirical implications of *aggregate* behaviour of the model. More importantly, viewing choice behavior as resulting from an unobservable cognitive process, we identify the primitives of the model from observed choices.

We first propose a deterministic model of recommendation. Our deterministic model is intended to be simple so that we can study the aggregate behavior of such consumers later. In this model, each decision maker is identified with a pre-recommendation choice and a preference ordering. The pre-recommendation choice represents the option either previously chosen or the default. In addition to these, it could be interpreted as the best alternative in her awareness set according to her preference before receiving any recommendation. Under these interpretations, it is natural that the pre-recommendation choice is picked if there is no recommendation since DM perceives no change in her environment. Each decision problem represents a different set of recommendations while the set of feasible alternatives are always fixed.[3] If an alternative is recommended, the decision maker picks the better option between the recommended alternatives and her pre-recommendation choice according to her preference. In our model, recommendation cannot hurt since the decision maker can always ignore the recommendation if it is not better than her pre-recommendation choice. This formulation eliminates any strategic consideration due to trust concerns about the source and intention of recommendation.[4] We see this model as a "rational" benchmark model of recommendation.

Our model is designed with two features in mind. First, a recommendation can only improve the decision maker's welfare. Second, we consider a very basic form of recommendation such as *Best Sellers* or *Award Winners*, where a recommendation is a set of alternatives. This type of recommendation is called recommendation signage. Goodman et al. (2013) examined 52 of the top 100 US retailers and found that at least 31 (60%) retailers that use these forms of recommendations. Recommendation signage helps customers by creating a consideration set (Gershoff

[3]As opposed to the standard choice models, there is no choice set variation in this model. What could vary is the set of recommended alternatives.

[4]Our DM does not need to trust the recommendation since she can always ignore the recommendation. Nevertheless, our framework is rich enough to study other channels that recommendation might be operational. The areas of further research should includes developing more elaborate and complex models of recommendation.

et al., 2001; Gupta and Harris, 2010; Sela et al., 2009; West and Broniarczyk, 1998). There are other examples of this form of recommendation, such as word-of-mouth, recommendation on social media, off-site recommendations with email, and uninformative advertising (not mentioning any attribute of the product). All of these enlarge the awareness of customers. 92% of consumers report that a word-of-mouth recommendation is the top reason they buy a product or service (Nielsen 2012). Abernethy and Butler (1992) show that 37.5% of U.S. TV advertising has no product attribute information. Mayzlin and Shin (2011) show uninformative advertising can be seen as an initiation to assess the product through inspection and it could be very effective for high quality products. In the retail banking industry, Honka et al. (2017) showed that advertising's primary role is largely informative as opposed to persuasive. Gupta and Harris (2010) shows indicate that e-WOM (i.e. electronic word-of-mouth) is likely to result in more time considering the recommended product.

Oftentimes recommendation inspires consumers to pick products that they otherwise would not have noticed (Chen et al. (2019) and Kawaguchi et al. (2021)). Hence, in this paper, recommendation's primary role is to inform customer about the existence and availability of the products.[5] We focus on products where consumers can determine whether they like it or not by inspection prior to purchase. Recommendations might not necessarily reflect what the consumer likes.[6] Due to the possibility of misalignment in preferences, the consumer still needs to go through the details of the product before making a decision.

Our first result identifies three simple and intuitive postulates which characterize the deterministic model. First note that the final choice in our model could be outside of the recommended set. This feature makes our model different from standard choice models. Our behavioral postulates depend on whether or not the final choice belongs to the recommended set. Our first postulate states that removing some of the unchosen alternatives from the recommendation set would not influence the final choice as long as the choice belongs to the recommended set. This postulate is similar to IIA in spirit. The next postulate assumes that if the decision maker does not follow a particular recommendation set, then she does not follow any subset of it. The final postulate states that if the decision maker does not follow the recommendation, she always chooses the same alternative. We show that these postulates characterize the deterministic model. Moreover, we can identify the pre-recommendation choice and "unique" preferences.

[5] Recommendation might also affect choices through by changing consumer's preferences. It is very difficult to disentangle whether a recommendation is informative or persuasive with naturally occurring data. A recent study by Kawaguchi et al. (2021) shows that "the recommendations affect total vending machine sales more through the attention channel than through the utility channel."

[6] One prominent reason is the inherent stochasticity of preference, which we will elaborate in the stochastic version of the model. On the other hand, even if one tries to give an algorithmic recommendation based on observed behavior, one often faces trade-off between efficiency of computation and quality of recommendation (Aggarwal, 2016; Hu et al., 2018).

Since, in many examples we have in mind, the data is given in the form of the aggregate behavior (stochastic choice), we extend our intuition of the deterministic model to this environment. The stochastic choice can be interpreted in two ways: It can be interpersonal, where we observed choice made by different individual; or intrapersonal, where the same individual makes choices under potentially different circumstances. Therefore, in section 3, we introduce a parametric model of recommendation where we capture all the information about the model's predictions with a finite set of alternative-specific parameters. The parametric model is based on two simple motives: Firstly, preference can be stochastic, where we observe the outcome of random preference of individual, [7] or we observe the outcome of a group of (heterogeneous) agents. Secondly, pre-recommendation can also be stochastic: an individual may base their pre-recommendation choice on different channels under different occasions (e.g. their friend suggestion, their last bought items or the recommendation they received last time and etc); on the aggregate level, people generally do not share the same opinions on what to choose when they receive no recommendation. To capture these two sources of randomness, we utilize two sets of alternative-specific parameters: $d(x) \geq 0$ and $w(x) > 0$. d represents the likelihood of being the pre-recommendation choice, hence $\sum_{x \in X} d(x) = 1$. w is a crude measure of the utility value. The higher the w is the higher the choice probability of that alternative. We combine the intuition of our deterministic model and the Luce model to describe the choice probabilities. Given a pre-recommendation choice, non-recommended alternatives are chosen with zero probability except for the pre-recommendation choice. The probabilities of choosing a recommended item depends on its own weight proportional to the total weight of recommended alternatives and the pre-recommendation choice. Since being the pre-recommendation choice is also probabilistic, thus, the odds of selecting an unrecommended alternative, $x \notin R$, is the probability of being the pre-recommendation choice times the choice probability when it is the pre-recommendation choice.

Our parametric model of recommendation has three advantages: i) being tractable, hence very useful in applications, ii) possessing strong uniqueness properties for identifying of the underlying parameters, iii) can function well under limited data scenario. To investigate into the model, we first consider the model under full data assumption. We show that this parametric model is characterized by three simple behavioral tests. The first condition says that all recommended alternatives are chosen with positive probability. The second one is Luce's IIA for recommended alternatives: the odds of choosing a recommended item over another recommended item do not depend on the other items in the recommended set. One would suspect that

[7] In the traditional economics wisdom, there could be *unobserved* underlying factors (e.g. weather) changing the preference.

a similar axiom must hold for off-recommended alternatives. It turns out that that property does not hold for non-recommended alternatives. Instead, there is another well-known property called conditional choice. That is, $\rho(a, R)\rho(R \cup a, X)$ is independent of R. In the classical domain, this axiom is equivalent to Luce's IIA. However, in our framework, this axiom does not hold for alternatives in R. On the other hand, this axiom holds for non-recommended alternatives. Then, we work with limited data. We show that testing the model under limited data requires a stronger axiom, which has two key behavioral implications. It says that an non-recommended item would be chosen less when more items are recommended; and the rate it decreases depends on the relative choice frequencies of the item and each of its rival if they are both recommended. Lastly, assuming the model is correct, we show that we can uniquely identify the parameters of the model with only two data points.[8] This is helpful for real-life applications since it offers a frugal way to make out-of-sample prediction for the effect of recommending different sets of items.

Instead of assuming two sources of independent randomness on pre-recommendation choice and preference in R-Luce, one may suspect there could be potential dependency between the two. To answer this question, we propose the R-RUM model, where we assume a joint distribution on preference and pre-recommendation. Here, stochastic choice data is observed by aggregating a group of different individuals in the same environment, provided that each of them may differ on their default options and their preferences. Our model resembles the classical random utility model (RUM), where the heterogeneous utility types are not observable and they need to be derived from the stochastic choice data. In our model, the type space is more complicated: each type is a pair of the default option and a preference $t = (a, \succ)$. Each type corresponds to a particular deterministic model of recommendation. Each distribution of such types gives rise to a stochastic choice data set. In other words, the choice probability of an alternative is the sum of probabilities of choice functions which select that alternative. We denote this model R-RUM.

We identify all the behavioral implications of R-RUM. We show that three simple tests determine whether stochastic choice data can be represented by R-RUM. The empirical content of the standard RUM is investigated by Falmagne (1978), Barbera and Pattanaik (1986) and McFadden and Richter (1990). The non-negativity of the Block-Marshak polynomials (BM) from Block and Marschak (1959) is equivalent to the choice data being generated by RUM. In our environment, the classical BM are defined for the recommended alternatives. We need to introduce an analogue of BM for the alternatives that are not recommended. We show that the non-negativity of BM for off-recommendation and on-recommendation is necessary but

[8] While d's are uniquely identified, the weights w's are unique up to a scale factor.

not sufficient. Since being recommended has positive effect, the BM of on-recommendation is always greater than the corresponding BM of off-recommendation. Finally, since we treat the default option as if it is always recommended in the deterministic model, there is no difference between on-recommendation and off-recommendation when the only recommended alternative is the default option. These three simple tests fully characterize the empirical content of R-RUM.

Our method of proof is constructive: it offers an algorithm to construct a rationalizing distribution of choice types from the observed choice probabilities. The uniqueness properties of R-RUM can be recovered by that algorithm. The uniqueness result of R-RUM is similar to that of RUM, which is a weaker form of ordinal uniqueness. Given two possible representations, certain marginal distributions of the preferences are essentially unique in our model. This is helpful to gauge the effectiveness of a recommendation from a policy maker perspective.

In all of our models, we utilize the idea that modelers can observe consumers' choice data as a function of their recommendation set. The idea of a recommendation set is a distinct notion from that of a choice set as choices can be outside of a recommendation set. Hence our choice function is both conceptually and mathematically different from the classical choice function. This new choice object, to the best of our knowledge, has yet to be explored in the choice theory literature. Our paper, in terms of modelling strategy, naturally separates itself from other theories in the standard domain, and, in terms of interpretation, enables us to take a fresh perspective into choice data with some of our traditional economic intuition. In the following, we provide a brief discussion for related literature. The classical Luce model, while being a cornerstone model of consumer behavior, attracts a wealth of scholarly attention into developing different generalizations (e.g. Ahumada and Ülkü, 2018; Echenique and Saito, 2019; Echenique et al., 2018; Fudenberg et al., 2015; Gul et al., 2014; Kovach and Tserenjigmid, 2019, 2021; Tserenjigmid, 2021). All of these models involve different relaxations of the Luce IIA axiom. In our environment, with an aim to provide a simple, applicable and tractable parametric model, we employ the Luce IIA axiom for the on-recommendation data and the Luce Choice axiom for the off-recommendation data. On the other hand, there are several different strands of research departing from choice-set variation in the standard model. For example, some studies utilize list variation to study choices (e.g. Guney (2014) and Ishii et al. (2021)) and approval rate (Manzini et al. (2021)); Natenzon (2019) studies how "non-choosable" phantom options affect choices. With a similar spirit to our models, these lines of research are also augmenting the standard choice environment to enhance our understanding of human behaviors.

2 Deterministic

As we mentioned in the introduction, our deterministic model is intended to be simple so that we can study the aggregate behavior of such consumers later. However, our framework is flexible enough to study more involved models of recommendation capturing different aspects of the recommendation process.

In this paper, the set of feasible alternatives is fixed, denoted by X (e.g., all documentaries available at Netflix or all 65 Inch Smart TVs sold at Amazon.com).[9] Hence there is no variation in terms of choice set in our model. The decision maker receives a recommendation in the form of a set of alternatives, say R, which is the source of variation in our model. Any subset of X could constitute a decision problem, including the empty set (no recommendation). While a recommendation can influence choices, it does not constrain it. To capture this, we define a choice rule c to be a function of recommendation set, R, but allow for $c(R)$ to be outside of R. Hence, the only restriction we impose is $c(R) \in X$. Notice that our choice function is different from the one used in the choice theory literature where $c(R)$ is always in R. This distinction will be important when we introduce behavioral properties. In addition, we allow that c is observable only for some recommendation sets but not others. This assumption aims to capture some real world environments where the collection of recommended set is just a fraction of the entire product space. Let $\mathcal{D} \subset 2^X$ denote all possible recommendation sets we have the data for, for instance single alternative recommendations. The following definition captures the choice rule under this framework.

Definition 1. A deterministic choice rule c on domain $\mathcal{D}$ is a mapping from $\mathcal{D}$ to X such that $c(R) \in X$ for all $R \in \mathcal{D}$.

In our model, each decision maker is identified with a pre-recommendation choice and a preference ordering $(a, \succ)$, where $\succ$ is a linear order. The pre-recommendation choice might have different interpretations. It could be either previously chosen, or easily identified. The default option itself could be a previously obtained recommendation.[10] In addition to these interpretations, the pre-recommendation choice could be seen as the best alternative in DM's initial awareness set according to her preference before receiving any recommendation. Hence it is natural that the pre-recommendation choice option is picked if there is no recommenda-

[9] We impose this assumption to capture possible data restriction in naturally occurring data. Having said that, this framework is flexible enough to allow set variations by assuming $c(R, S) \in S$ and $R \in X$ where S represents the set of feasible alternatives and R is the recommended set. This formulation allows the recommended alternatives being unavailable in order to capture "out of stock" concept when $R \setminus S \neq \emptyset$.

[10] Defaults help consumers to reduce the need of engaging deliberative decisions. They also eliminates difficult trade-offs (Thaler & Sunstein, 2008). Use of default is also related decision task complexity (Redelmeier & Shafir, 1995), conflict (Tversky & Shafir, 1992), and/or emotionally difficult decisions (Luce, 1998).

tion. We use the terminologies default option and pre-recommendation choice interchangeably throughout the paper. In this model, when the decision maker receives a recommendation, R, the decision maker picks the best option among $R \cup a$ according to her preference. The decision maker will consider all the recommended alternative before making a final decision. In our model, recommendation brings the recommended alternatives into active consideration (Court et al., 2009; Goodman et al., 2013; Gupta and Harris, 2010). However, the decision maker deliberately assesses each recommended alternative before making a decision. Formally,

Definition 2. A deterministic choice rule c has a recommendation representation on $\mathcal{D}$ if there exist a preference $\succ$ and a default option a such that

$$c(R) = \max(R \cup a, \succ) := c_{(a,\succ)}(R)$$

for all $R \in \mathcal{D}$.[11]

In this model, the decision maker is both willing to and capable of considering all the recommended alternatives. This feature of the model is shared by the classical rational choice model where the decision maker maximizes her preferences among all available alternatives. Because of full consideration, we see this model as a "rational" benchmark model of recommendation. In other words, the consideration set is equal to the recommendation set including the pre-recommendation choice. Notice that the "rationality" requirement is not as severe as the classical model due to the fact that we require that only a limited number of alternatives are recommended in our data. Nevertheless, in the supplementary material of the paper, we study a model where the consideration set may be different from the recommended set to study behavioral factors in this environment.

Two special cases of our model are worth mentioning. The first one is when the pre-recommendation choice is the best alternative according to the DM's preference. In that case, $c_{(a,\succ)}(R) = a$ for all $R \in \mathcal{D}$. In other words, choices are not influenced by recommendations. The second one is when the pre-recommendation choice is the worst alternative according to preference. In that case, the decision-maker behaves as if she is a classical preference maximizer: $c_{(a,\succ)}(R) = \max(R, \succ) \in R$ where the recommendation dictates the chosen alternative. In our model, a recommendation is effective if it offers improvement over the pre-recommendation choice. Otherwise, the recommendation does not influence choices.

For the sake of completeness, we provide a characterization for our simple deterministic

[11] One might interpret $R \cup a$ as the consideration set. We would like to highlight that the actual consideration set could be larger than $R \cup a$. For example, the union of R and the lower counter set of a with respect to $\succ$ could be the actual consideration set of the decision maker.

model of recommendation. This characterization will serve as a benchmark for future works to model bounded rationality in this topic. The first axiom is the famous IIA in the choice theory literature.[12] It states that removing some of the unchosen alternatives from the recommendation set will not influence the final choice. Note that this axiom is satisfied by our model. If the chosen alternative is not the default option, it must be better than all the other recommended alternative including the default. Hence, it will stay the best alternative if we remove some of the unchosen alternatives from the recommended set.

Axiom 1 (IIA)**.** If $c(R) \in R' \subset R$ then $c(R) = c(R')$.

The next axiom requires that when a decision maker is presented with a recommendation set, if they do not choose within that recommendation set, then, when the decision maker is presented with any subset of that recommendation set, they will also not choose within that recommendation set. In our model, this means that the decision maker chooses the default option. If the chosen alternative is the default option, then it must be the best alternative in the larger recommendation set. Hence, it is also the best alternative in the smaller recommendation set.

Axiom 2 (weak-IIA)**.** If $c(R) \notin R$ then $c(R') \notin R'$ for all $R' \subset R$.

The last axiom states that if the decision maker does not follow the recommendation, then she always chooses the same alternative.

Axiom 3 (Consistency)**.** If $c(R) \notin R$ and $c(R') \notin R'$ then $c(R) = c(R')$.

We assume that $\mathcal{D}$ includes all recommendation sets with $|R| \leq 3$.[13] Then our deterministic model is characterized by these three simple axioms.

Theorem 1 (Characterization)**.** Let $\mathcal{D}$ includes all recommendation sets with $|R| \leq 3$. Then c satisfies Axioms 1-3 if and only if c has a recommendation representation.

As far as we know, this is the first characterization of decision-making under recommendation. As we stated before, this is a "rational" benchmark model of recommendation. Here, the decision maker's consideration set includes all the recommended alternatives. We hope that this characterization will guide future research on recommendations.

We now discuss identification and uniqueness properties of our model. It is clear that the pre-recommendation choice must be the choice when there is no recommendation $a := c(\emptyset)$. Similarly, if x is chosen when R is recommended, then x is revealed to be preferred to every alternative

[12] This property is also known as Sen's α axiom (Sen, 1971), Postulate 4 of Chernoff (1954), C3 of Arrow (1959), the Heritage property of Aĭzerman and Aleskerov (1995), or the Heredity property of Aleskerov et al. (2007).

[13] The characterization also works for $\mathcal{D} = 2^X$.

in R. If x is different from $c(\emptyset)$, x is also revealed to be preferred to the pre-recommendation choice. The next proposition states all the uniqueness properties of our deterministic model.

Proposition 1 (Uniqueness)**.** If $(a_1, \succ_1)$ and $(a_2, \succ_2)$ represents the same choice rule, then

i) $a_1 = a_2 := a$,

ii) $L_{\succ_1}(a) = L_{\succ_2}(a) := L_a$, and

iii) $x \succ_1 y$ if and only if $x \succ_2 y$ for all $x, y \in X \setminus L_a$.[14]

This proposition states that the pre-recommendation choice is unique, and the preference is uniquely identified as long as it matters for the choice.

3 A Parametric Stochastic Choice

In the last section, we studied the deterministic environment where we observe the choice made by a single (type of) individual. In this section, we focus on stochastic choice. One interpretation of this randomness is that there is a group of individuals whose types are unknown and we can only observe their aggregate behavior (interpersonal). The other interpretation is the choices of a single individual in different situations (intrapersonal). Hence, our model can be interpreted as both intrapersonal and interpersonal stochastic choice. In the classical choice-set variation framework, this format of data has been well studied. In our framework, to incorporate the idea that decision makers can choose outside of the recommendation R, we must also extend the idea of a (stochastic) choice rule just as we did in our deterministic setting.

Definition 3. A choice rule ρ is a mapping from $X \times \mathcal{D}$ to $[0, 1]$ such that $\sum_{x \in X} \rho(x, R) = 1$.[15]

We first study a parametric model of recommendation with stochastic data based on the logit model of Luce (1959). All the information about the model's predictions will be summarized by a finite set of parameters depending only on alternatives. As is typically the case with parametric models, our parametric model offers three advantages over a more general version we study in the next section. First, the model is very tractable, which is a desirable property for applications. Second, this model possesses strong uniqueness properties. Third, it allows for identification and characterization under limited data.

In this model, each alternative x is represented by two parameters: $d(x) \geq 0$ and $w(x) > 0$. d represents the likelihood of being the pre-recommendation choice, hence $\sum_{x \in X} d(x) = 1$. w is a crude measure of the utility value. An alternative with a high w will be chosen more often than an alternative with a low w. We follow the Luce-logit model to describe the choice probabilities

[14] Throughout the paper, we denote the lower contour set of a with respect to preference $\succ$ as $L_\succ(a)$.

[15] This is different from the classical stochastic choice model where $\rho(x, R) = 0$ for $x \notin R$.

given a default. To mimic our deterministic model, for a fixed default, all non-recommended alternatives are chosen with zero probability except the default. The probabilities of choosing a recommended item depends on its own weight proportional to the total weight of recommended alternatives and the default option. Hence, we first define choice probabilities given a fixed default. The choices for a given pre-recommendation choice, a, can be expressed as:

$$W_a(x, A) = \begin{cases} \dfrac{w(x)}{\sum_{z \in A \cup a} w(z)} & \text{if } x \in A \cup a \\ 0 & \text{otherwise} \end{cases}$$

Note that W_a is itself a parametric choice model where $\sum\limits_{x \in X} W_a(x, A) = 1$. According to this formulation, only the recommended alternatives and the pre-recommendation choice are chosen with positive probability. W_a captures the randomness in preferences as in the logit model. Note that the deterministic model of Section 2 is a limit case of this model since deterministic model has no randomness in preferences.[16] Therefore, we expand on this intuition by also assuming that being default is probabilistic. Let $d(x)$ be the probability of x being the default. Then the choice probability of any alternative is defined as a mathematical expectation: the probabilities of a given default option times the conditional choice probability given that default option. Formally, we have the following.

Definition 4. A choice rule ρ has a Luce recommendation representation (R-Luce) if there exists functions $w : X \to \mathbb{R}_{++}$ and $d : X \to \mathbb{R}_{+}$ with $\sum\limits_{x \in X} d(x) = 1$ such that for $x \in X$,

$$\rho(x, R) = \sum_{a \in X} d(a) W_a(x, R)$$

for all $R \in \mathcal{D}$.

R-Luce model has inherently two types of randomness. While d captures the randomness in pre-recommendation choice, w represents the randomness in preferences. Note that the odds of selecting an non-recommended alternative, $x \notin R$, is the probability of being default times the conditional choice probability when it is the default, that is, for any $x \notin R$

$$\rho(x, R) = \frac{d(x)w(x)}{\sum\limits_{y \in R \cup x} w(y)} = d(x)W_x(x, R)$$

[16] To see how W_a reduces (or, approaches) to deterministic case where DM's type is $(a, \succ)$, we first enumerate all alternatives according to $\succ$, $x_1 \succ x_2 \cdots \succ x_n$. We assign $w(x_i) = \varepsilon^i$ for $i > 1$ and $w(x_1) = 1 - \sum_{i=2}^{n} \varepsilon^i$. By taking ε to *zero*, $(a, \succ)$ becomes the limit case of W_a.

When $x \notin R$, $\rho(x, R)$ is always zero in the standard random utility model since R represents the set of feasible alternatives. However, in our model, the set of alternatives are always the same and R represents only the recommended alternatives. The effective weight of x becomes $d(x)w(x)$, which is strictly less than $w(x)$ if $d(x) < 1$. This can be interpreted as the value of non-recommended alternatives is discounted while the recommended ones stay the same.

For recommended alternatives, $x \in R$, our formula is more involved:

$$\rho(x, R) = \sum_{z \in X} \frac{d(z)w(x)}{\sum\limits_{y \in R \cup z} w(y)}$$

We now study the empirical content of this model. Our first property requires that all recommended alternatives be chosen with positive probability. This is due to the fact that we assume $w(x) > 0$. Then the recommended alternatives are always positively chosen in W_a independent of the default. Hence, they must be chosen with strictly positive probability.

Axiom 4 (Positivity)**.** For $x \in R$, $\rho(x, R) > 0$.

The second condition is the same as Luce's IIA for recommended alternatives. Remember Luce's IIA says that the odds of choosing one alternative over another one do not depend on the feasible set. Our next property requires that Luce's IIA holds for all recommended alternatives:

Axiom 5 (Recommended IIA)**.** For $x, y \in R \cap R'$,

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

One would suspect that a similar axiom must hold for off-recommended alternatives. It turns out that this property does not hold for non-recommended alternatives. Instead, another well-known property is satisfied with a caveat. The property is known as Luce's Choice Axiom:

$$\rho(a, R)\rho(R, R') = \rho(a, R')$$

This property says that the probability of choosing a from R' is equal to the conditional probability that a is chosen from R given that the choice from R' belongs to R. Here, we slightly modify Luce's Choice Axiom. The property says that, for $x \notin R$ and $R \cup x \subseteq R'$, the probability of choosing x first when R is recommended and then choosing $R \cup x$ when R' is recommended is independent of R. This property holds for non-recommended alternatives. That is our next axiom.

Axiom 6 (Off-Recommendation Path Independence)**.** For $x \notin R$ and $R \cup x \subseteq R'$,

$$\rho(x, R)\rho(R \cup x, R') \text{ is independent of } R$$

As long as probabilities are strict, Luce's IIA and Luce's Choice Axiom are equivalent in the usual choice domain where $\rho(R, R) = 1$ for all R. Surprisingly, this equivalence does not hold in our domain since $\rho(R, R)$ could be strictly less than 1 for $R \neq X$. This discussion highlights that equivalence of these two properties depends on the domain to which they apply. Here, we show that recommended and non-recommended alternatives obey different rules. These three properties summarize the entire empirical content of our parametric model.

Theorem 2 (Characterization)**.** Assume $\mathcal{D} = 2^X$. Then ρ satisfies Axiom 4-6 if and only if ρ has a R-Luce representation.

We now demonstrate that we can provide a similar characterization of this model with limited data. Suppose $\mathcal{D}$ includes all recommendation set with $|R| \leq 2$. Under limited data, we need to impose a stronger axiom on the off-recommendation data. The reason why we need this axiom is because Axiom 6 is not strong enough in this limited domain. This concern does not exist if we have full data. To see why we need an additional axiom, consider the following example in Table 1 with $X = \{a, b, c, d\}$. Notice that Axiom 4 and 5 are immediately satisfied. One can even check that Axiom 6 is also satisfied.[17] However, none of these axioms govern how $\rho(c, \{a, b\})$ behaves: Axiom 6 puts restriction on $\rho(c, \{a, b\})$ only if we also observe data on some other recommendations that includes all a, b and c.[18] Nonetheless, we might not observe a recommendation including all three alternatives due to the limited data assumption. Hence, to check for the validity of the model under limited data, we need to impose a stronger axioms.

	$\{a,b\}$	$\{a,c\}$	$\{a,d\}$	$\{b,c\}$	$\{b,d\}$	$\{c,d\}$	$\{a\}$	$\{b\}$	$\{c\}$	$\{d\}$	$\emptyset$
a	5/12	5/12	5/12	1/12	1/12	1/12	3/8	1/8	1/8	1/8	1/4
b	5/12	1/12	1/12	5/12	5/12	1/12	1/8	3/8	1/8	1/8	1/4
c	0	5/12	1/12	5/12	1/12	5/12	1/8	1/8	3/8	1/8	1/4
d	2/12	1/12	5/12	1/12	5/12	5/12	1/8	1/8	1/8	3/8	1/4

Table 1. A stochastic choice data satisfying Axiom 4-6 but it does not have R-Luce representation under limited data

The off-recommendation path independence axiom revolves around the fact that the choice probability of an unrecommended alternative under recommendation set R is tightly related to

[17] Since the data are symmetric, we consider a. Note that $\rho(a, \emptyset)\rho(a, \{a, b\}) = \rho(a, \emptyset)\rho(a, \{a, c\}) = \rho(a, \emptyset)\rho(a, \{a, d\}) = \rho(a, \{b\})\rho(\{a, b\}, \{a, b\}) = \rho(a, \{c\})\rho(\{a, c\}, \{a, c\}) = \rho(a, \{d\})\rho(\{a, d\}, \{a, d\}) = \frac{5}{48}$. Hence it satisfies Axiom 6.

[18] On the other hand, one can immediately see that this data does not have a R-Luce representation. Suppose it does, then $\rho(c, \{a, b\}) = 0$ implies $d(c) = 0$. Yet, $\rho(c, \emptyset) = 1/4$ implies $d(c) = 1/4$. This is a contradiction.

its choice probability when recommended with the set R. It turns out that this dependency can be made more explicit. With the help of Axiom 5, we define the following shorthand: $r(z,x) := \frac{\rho(z,A)}{\rho(x,A)}$ for some A including x and z. Each $r(z,x)$ captures exactly the choice ratio of x and z when both of them are recommended.

Axiom 7 (Off-recommendation Independence)**.** For $x \notin R$,

$$\rho(x,R)(1+\sum_{z\in R} r(z,x)) \text{ is independent of } R$$

The intuition behind this axiom is that, there is a fixed amount for how often x can be chosen when x is not recommended. Notice that as recommendation set R grows, the markup $\sum_{z\in R} r(z,x)$ will be greater and in turns $\rho(x,R)$ will be smaller. Therefore, this axiom dictates that x must be chosen less as more and more alternatives are recommended while x is not included in the recommendation. Also, the rate that $\rho(x,R)$ decreases also depends on how "likable" x is when both x and his rival z are both recommended. i.e. it depends on the $r(z,x)$. In particular, it will decreases less if x is chosen much more often than z when both of them are recommended.

It is clear that the above example violates Axiom 7. Consider the choice probability of c under the recommendation set $\{a,b\}$ and $\emptyset$, we get $\rho(c,\emptyset) = \frac{1}{4} \neq 0 = \rho(c,\{a,b\})(1+r(a,c)+r(b,c))$. Moreover, we can also see that Axiom 5 and 7 imply Axiom 6: Since $r(z,x)$ can be represented with any recommendation set as long as they include both z and x, we consider arbitrary R' such that $x \cup R \subseteq R'$. Then, by simplification, we can get $1+\sum_{z\in R} r(z,x) = \rho(R\cup x, R')/\rho(x,R')$, which basically implies Axiom 6.

Theorem 3. Let $\mathcal{D}$ includes all recommendation sets with $|R| \leq 2$. Then, ρ satisfies Axiom 4-5 and 7 if and only if ρ has a R-Luce representation.

Theorem 3 provides a similar characterization with limited data. Theorem 3 eliminates arguably unrealistic data requirements of Theorem 2. Indeed, many models in decision theory require a similar rich data set, typically choices from all decision problems. In Theorem 3, we are able to show that the identification of the parameters of our model can be inferred from recommendation sets up to size 2. We show that we can identify the parameters of the model uniquely. While d's are uniquely identified, the weights are unique up to a scale factor.

Proposition 2 (Uniqueness)**.** If (w_1,d_1) and (w_2,d_2) represent the same choice rule ρ, then

i) $d_1 = d_2$, and

ii) $w_1 = \alpha w_2$ for some $\alpha > 0$.

Identification

One important advantage of our parametric model is that it has strong identification properties, which offers a frugal way to make out-of-sample prediction. Here, we assume the data is very limited: $\mathcal{D} = \{\emptyset, \{a\}\}$ for some $a \in X$. We now illustrate that observations from only two such simple recommendation sets are sufficient for unique identification for our model. We first state our result.

Proposition 3. If $\rho(x, \emptyset) > \rho(x, \{a\})$ and $\rho(x, \emptyset) > 0$ for all $x \in X \setminus a$, the parameters of the model are fully identified.

There are two requirements here. The first requirement has a normative appeal. It says that when a is recommended, x will be chosen less compared to when there is no recommendation. The reason behind this is that there is a non-zero probability that those whose pre-recommendation choice is x switched to choosing a just because they discovered that a is better due to recommendation. On the other hand, we need to impose the assumption that $\rho(x, \emptyset) > 0$ so that we are able to observe $w(x)$ even when x is not recommended.[19] Since the procedure is simple enough, we will demonstrate the identification here in the main text. Firstly, let $d(x) := \rho(x, \emptyset)$ for every $x \in X$. Then, we identify w. Note that w is unique up to a scaling factor, we let $w(a) = 1$. Note that, for $x \in X \setminus a$, we define

$$w(x) := \frac{\rho(x, \{a\})}{\rho(x, \emptyset) - \rho(x, \{a\})}$$

From this definition, one can immediately see the necessity of the condition $\rho(x, \emptyset) > \rho(x, \{a\})$. Note that, by re-arrangement, we have

$$\rho(x, \{a\})(w(a) + w(x)) = d(x)w(x)$$
$$\rho(x, \{a\}) = d(x)\frac{w(x)}{w(x) + w(a)}$$

Hence, this definition fulfills the model.

Discrete Choice

In the following, we consider the model from the discrete choice perspective. In the framework of discrete choice, a decision maker's random utility is defined as

$$U(x) = v(x) + \varepsilon(x)$$

[19]This positivity property, as argued by McFadden, 1973, cannot be refuted based on any finite data set.

where $\varepsilon(x)$ is known as "random utility shock". From the standard RUM perspective, the event that an alternative is chosen in a choice set is equated to the event that an alternative has the highest realized utility in that choice set. However, from the recommendation set perspective, we cannot equate these two events since not only does an alternative need to achieve the highest utility within the recommendation set, but it also needs to 1) be better than the default option, and 2) be inside the recommendation set or be the default option. Here, we assume independence between the distribution of default options and the random utility shock. Formally, we define a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for the random variable $\varepsilon : \Omega \to \mathbb{R}^X$. We define the following event where x achieves the highest utility within a set A,

$$\omega_{x,A} = \{\omega \in \Omega : v(x) + \varepsilon_\omega(x) \geq v(y) + \varepsilon_\omega(y) \text{ for all } y \in A\}$$

Then, we write down the discrete choice model based on the idea of recommendation set.

Definition 5. A choice rule ρ has a R-logit representation if there exists $v : X \to \mathbb{R}$, $d : X \to \mathbb{R}_+$ with $\sum_{x \in X} d(x) = 1$ and $\varepsilon : \Omega \to \mathbb{R}^X$ which follows a Gumbel distribution with noise parameter λ and is i.i.d. across $x \in X$ such that

$$\rho(x, R) = \begin{cases} \sum_{a \in X} d(a)\mathbb{P}(\omega_{x,R\cup a}) & \text{if } x \in R \\ d(x)\mathbb{P}(\omega_{x,R}) & \text{if } x \notin R \end{cases}$$

Notice that although we break the equivalence between ρ and $\mathbb{P}$, it does not prevent us from finding the closed-form solution for $\mathbb{P}$ according to the standard technique. In fact, as one may immediately expect, the closed-form solution for $\mathbb{P}(\omega_{x,A})$ is, for every $x \in A$,

$$\mathbb{P}(\omega_{x,A}) = \frac{e^{v(x)/\lambda}}{\sum_{z \in A} e^{v(z)/\lambda}}$$

Therefore, the relationship between R-Luce and R-logit are analogous to the relationship between Luce and logit in the classical choice-set variation domain.

4 A General Stochastic Choice

In the section, we introduce our most general model in this framework. In the last section, we assume that a parametric model where the randomness of types are described by two functions. In this section, stochastic choice data is observed by aggregating a group of different individuals in the same environment, provided that each type differs on both their pre-recommendation choice and their preferences. In other words, this general model would allow for potential dependence between preference and pre-recommendation choice, whereas R-Luce assumes independence.[20] To define the type space, firstly, let $\mathcal{P}$ be the set of all (linear order) preference and $\mathcal{T}$ be the set of all pairs of $(a, \succ)$ where $\succ \in \mathcal{P}$ and $a \in X$, with typical type denoted by t. In this paper, a random utility function is a probability measure μ on $\mathcal{T}$ such that $\sum_{t\in\mathcal{T}} \mu_t = 1$, where $\mu_t := \mu(\{t\})$. Then, we state the definition of the model.

Definition 6. A choice rule ρ has a *Recommendation* representation under Random Utility (R-RUM) if there exists a random utility function μ on $\mathcal{T}$ such that

$$\rho(x, R) = \mu(\{t|\ c_t(R) = x\})$$

for every $R \in \mathcal{D}$.

Above definition mimics the classical random utility. As in RUM, R-RUM tests the hypothesis of a group of preference maximizing individuals. In RUM, consumers' tastes vary explicitly. In R-RUM, not only their taste differ but also their pre-recommendation choices. Hence, R-RUM enjoys a much richer type space. In other words, R-RUM is a much richer model compared to RUM. We later consider a subclass of R-RUM where each type shares the same pre-recommendation choice, hence types differ only with respect to their preferences.

In RUM framework, Falmagne (1978) answered the question whether individual preference maximization has any implication for aggregate data. For characterization of RUM, Falmagne (1978) utilizes a well-known concept in that literature: the Block-Marschak polynomials, named after Block and Marschak (1959)'s seminal work on the random utility model. It has shown that a stochastic choice data has a RUM representation if and only if its Block-Marschak polynomials are non-negative. The necessity was originally obtained by Block and Marschak (1959). Falmagne (1978) showed that they are also sufficient.

In our framework, we also utilize Block-Marschak (BM) polynomials. Let $q_\rho(a, R)$ be the Block-Marschak polynomials. i.e. for $a \in R$,

$$q_\rho(a, R) := \sum_{B \supseteq R} (-1)^{|B\setminus R|} \rho(a, B)$$

Note that the Block-Marschak polynomials are defined with respect to choice data ρ. Throughout the paper, we mostly skip denoting ρ and write $q(a, R)$ unless specified otherwise. Interestingly, this definition can be applied to off-recommendation data as well and, as we shall see, it

[20] Notice that our model has a close connection with the standard random utility model (RUM) in the classical domain where $\rho(x, A) = 0$ if $x \notin A$.

has a important role in R-RUM. We define for $a \notin R$,

$$y_\rho(a, R) := \sum_{a \notin B \supseteq R} (-1)^{|B \setminus R|} \rho(a, B)$$

Again, we will skip denoting ρ and write $y(a, R)$ unless specified otherwise.

Figure 1 generalizes the classical network representation of partial order sets for our purposes. Each node represents a subset of the set of alternatives. Each black solid line indicates a subset relationship among subsets. The Block-Marschak polynomials can be thought of as the amount of flow on each black line. In the original network of this Hasse diagram, the degree of each node is equal to the number of alternatives, and inflow and outflow of black lines are always equal for each node.

$$\sum_{a \in R} q_\rho(a, R) = \sum_{b \notin R} q_\rho(b, R \cup b)$$

In RUM, each preference ranking corresponds a path starting from the empty set and ending at the grand set. For example, $\emptyset - \{c\} - \{b, c\} - \{a, b, c\}$ would represent $a \succ_1 b \succ_1 c$.[21] We first highlight that the above equality is no longer true in our model. But we will discuss below how to recover a similar equality and provide a similar visual representation for types in R-RUM.

As opposed to RUM, R-RUM has two sets of BM conditions: one for recommended alternatives, q, and one for non-recommended alternatives, y. To represent the new BM conditions, y's, we introduce new flows, which are represented by dashed red lines. These are always the outflows (or "leakages") from the network. We abuse notation and denoted both nodes and the flows with the same notation. $y_\rho(a, \{c\})$ denotes both the phantom node and the flow to the that node. Interestingly, if we also take into account y's, we recover the equality of inflow and outflow of all black and red lines.[22] That is,

$$\sum_{a \in R} q_\rho(a, R) + \sum_{a \notin R} y_\rho(a, R) = \sum_{b \notin R} q_\rho(b, R \cup b)$$

Given this equality, we can represent each type in R-RUM by a path in the new Hasse diagram. Similar to RUM, each type corresponds a path starting from a phantom node and ending at the grand set. For example, $y_\rho(c, \emptyset) - \emptyset - \{c\} - \{b, c\} - \{a, b, c\}$ would represent $a \succ_1 b \succ_1 c$ with c being the pre-recommendation choice, hence the type is $(c, \succ_1)$. Note that here the pre-recommendation choice is the worst alternative according to $\succ_1$. Similarly, $y_\rho(b, \{c\}) - \{c\} - \{b, c\} - \{a, b, c\}$ would represent the type is $(b, \succ_1)$. Each of these two paths

[21] One can refer to Fiorini (2004) for a network analysis of RUM.

[22] This result is stated as Lemma 1, which is a generalization of Falmagne (1978)'s Theorem 3. We believe that this lemma could be of independent interest since it is model-free. We provide the proof for it in the Appendix.

corresponds a unique type. However, the path $y_\rho(a, \{b, c\}) - \{b, c\} - \{a, b, c\}$ corresponds two types $(a, \succ_1)$ and $(a, \succ_2)$ where $a \succ_2 c \succ_2 b$. Notice that these two types cannot be distinguished because they always choose a. This will be a trivial non-uniqueness of R-RUM.

From the diagram, if $y(a, A) = 0$ for every $a, A \neq \emptyset$ and $a \notin A$, our model is behaviorally/mathematically equivalent to the standard RUM.[23] As one shall see in Proposition 4, by assuming the model is correct, this is equivalent to say that every $\mu_{a,\succ} = 0$ for a not being $\succ$-worst, i.e. it must be that every consumer's default option is the worst alternative. Note that a similar case also holds for our deterministic model.

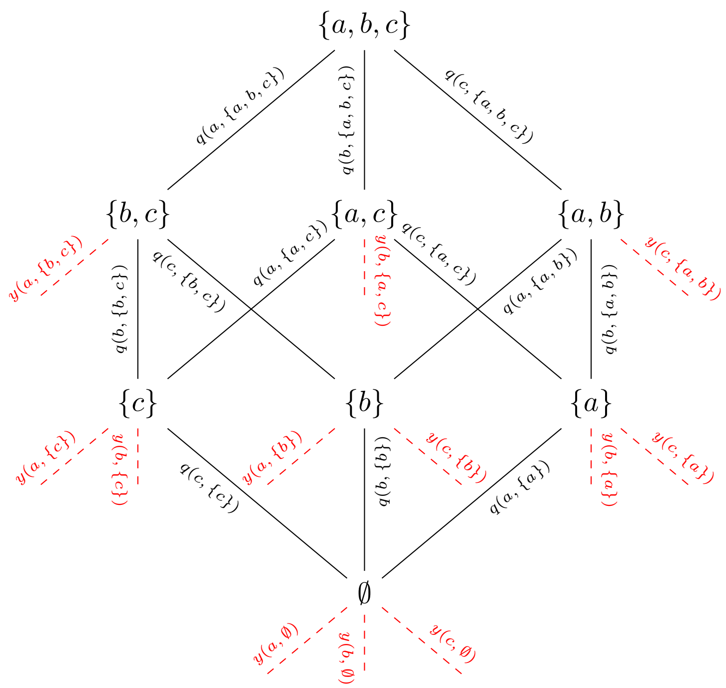

Figure 1. Hasse Diagram of R-RUM for three alternatives

In the following, we introduce the axioms of the model. The first axiom is a non-negative axiom of BM polynomials in the standard domain. This non-negativity is closely related to the non-negativity of BM polynomials in the standard domain. Similar to the idea in RUM, the necessity of the non-negativity of $y(a, R)$ revolves around the fact that $y(a, R)$ itself is a marginal distribution over the type space, in the case where the model is correct. Hence, it must not be negative. One is able to see the actual marginal distribution in Proposition 4.

Axiom 8 (Non-negativity)**.** For $a \notin R$, $y(a, R) \geq 0$.

We so far only apply non-negativity on y's but not q's. By introducing the second axiom, one will see we impose more than just non-negativity on q's. Being also a marginal distribution,

[23] A caveat is that we also need $q(a, a) = y(a, \emptyset)$ as in Axiom 10.

q represents a fraction of the types. More importantly, the difference $q(a,R) - y(a, R \setminus a)$ has a specific economic meaning in our framework: It captures a marginal effect of introducing a to the set $R \setminus a$ for the alternative a. Note that introducing a will only make an impact for a consumer if a is better than their default option and other recommended items. While the difference of $\rho(a, R)$ and $\rho(a, R \setminus a)$ represents the aggregate effect of introducing a to $R \setminus a$, the difference of $q(a, R)$ and $y(a, R \setminus a)$ provides a finer detail: it represents the fraction of those consumers who are marginally better off (i.e. who ranks a just above $R \setminus \{a\}$) and happen to switch their choice to a exactly because a is recommended. Therefore, it must not be negative.

Axiom 9 (Positive marginal recommendation effect)**.** For $a \in R$, $q(a, R) \geq y(a, R \setminus a)$.

The last axiom is an equality axiom, which requires that the weak inequality in Axiom 9 must be equality when R is singleton. Interestingly, we can also understand the intuition of this axiom through the idea of marginal recommendation effect. Notice that the difference between $q(a, a)$ and $y(a, \emptyset)$ represents the consumers who rank a above only "nothing" and happen to switch their choice to a because a is recommended. However, ranking a only above nothing means that a is the least preferred alternative. In such case, those consumers would always prefer to choose their default option instead. Hence, the difference will always be zero.

Axiom 10 (Emptiness)**.** For all a, $q(a, a) = y(a, \emptyset)$.

Since the model is general, we need a rich data set for characterization. For the next theorem, we assume that $\mathcal{D} = 2^X$.[24]

Theorem 4 (Characterization)**.** ρ satisfies Axioms 8-10 if and only if ρ has a R-RUM representation.

The sufficiency proof of the theorem is constructive. We provide an algorithm to compute a full distribution of types in the R-RUM representation. The algorithm is certainly helpful in application if one would like to have an estimate of the type distribution. However, just as RUM in the standard choice domain, there is generally not a unique distribution over types which gives rise to R-RUM data. Yet, there are several uniqueness properties of the model which are useful for policy recommendation. To do this, we first uniquely identify certain marginal distributions from the choice data. Let $L_{\succ}(a)$ represent strict lower contour set of a according to $\succ$. We state the uniqueness result in the following proposition.

Proposition 4 (Uniqueness and identification)**.** If μ^1 and μ^2 represent the same choice rule ρ, then for every $A \subset X$, $b \notin A$ and $i = 1, 2$,

[24] This full data assumption is standard for RUM models.

i) $y_\rho(b,A) = \mu^i\Big(\{(b,\succ)|A = L_\succ(b)\}\Big)$
ii) $q_\rho(b,A\cup b) = \mu^i\Big(\{(a,\succ)|A = L_\succ(b), a\in A\cup b\}\Big)$

Here, property i) says that, $y_\rho(b,A)$ must equal to the probability that b is the default while b is exactly and only better than the set of alternatives A. On the other hand, property ii) says that $q_\rho(b,A)$ must equal to the probability that b is exactly and only better than the set of alternatives A and the default is within $A\cup b$. Note that (i) and (ii) imply that $q_\rho(b,A\cup b) - y_\rho(b,A) = \mu^i\Big(\{(a,\succ)|A = L_\succ(b), a\in A\}\Big)$, where the LHS is exactly the object from our Axiom 9. From here, one can immediately see that it captures the fraction of people who rank b exactly above A while their default is in A. Surely, this fraction of people will switch to b if b is included in the recommendation set. Based on this uniqueness result, we have the following observation.

Remark 1. Suppose ρ is a R-RUM. For every b, $\mu(\{(b,\succ)|\succ\in\mathcal{P}\}) = \rho(b,\emptyset) = \sum\limits_R y(b,R)$.

We first take a look at $\mu(\{(b,\succ)|\succ\in\mathcal{P}\}) = \rho(b,\emptyset)$. Similar to the intuition from the deterministic case, when nothing is recommended, it must be that the DM is only choosing her default option, which captures the fraction of consumer with default b. On the other hand, $\mu(\{(b,\succ)|\succ\in\mathcal{P}\}) = \sum\limits_A y(b,A)$ can be understood through Proposition 4 in terms of the marginal distribution. We take summation over all marginal distributions under the condition that b is the default. Even though the previous two observations immediately imply $\rho(b,\emptyset) = \sum\limits_A y(b,A)$, there is another perspective to look at it. It can be understood through the *implied* recursive structure of the definition of BM polynomials,[25] i.e. for every $b\notin B$

$$y(b,B) = \rho(b,B) - \sum_{A\supset B} y(b,A)$$

One can see immediately it has to hold by putting $B=\emptyset$.

Impact of Recommendation

In the following, we tackle the problem regarding which alternative the policy maker should recommend if he has access to certain estimates of the choice data. In our framework, recommendation always improve welfare since a recommendation does nothing more than make a consumer aware of a product. One immediate implication is that the policy maker should recommend everything. However, this may not be feasible in practice since it may be costly to make a recommendation to a consumer. Also, as the recommendation set gets bigger, the decision maker might suffer from limited consideration issue.[26] Here, we focus our analysis by

[25] One can refer to Barbera and Pattanaik (1986) for some discussions of the recursive nature of BM polynomials.
[26] We explore this issue in the supplementary material.

restricting recommendation to only one alternative.

We first consider the following example in Table 2. Consider two types of individuals with default option c but with preference $a \succ_1 b \succ_1 c$ and $b \succ_2 c \succ_2 a$. The populations as a whole will benefit regardless of whether a or b is recommended. Supposedly, if α is closely to 1, the policy marker might want to recommend a. In this case, those who ultimately stick to c (i.e. people with preference $\succ_2$) are choosing their second-best, and those who choose a due to recommendation is choosing their first-best. On the other hand, if α is close to zero, the policy marker may recommend b and the opposite scenario will apply. Therefore, by making an utilitarian assumption, the policy marker will recommend a if $\alpha \geq 0.5$, and recommend b if $\alpha \leq 0.5$.

	$\succ_1$	$\succ_2$
Default	c	c
	a	b
	b	c
	c	a
μ	α	$1-\alpha$

Table 2. A simple example

Notice that the discussion so far only concern about the *ex post* welfare distribution of the population. However, in the recommendation framework, there is a second dimension where the policy marker might also care about. i.e. the *scope* of impact of a recommendation. Notice that recommending b causes both types to *change* behaviors and enjoy welfare improvement. Therefore, taking these into consideration, the policy maker might still recommend b even if α is greater than 0.5. In fact, he might set a new threshold $\alpha^* > 0.5$ such that he recommends b if and only if $\alpha \leq \alpha^*$. One possible functional assumption to support this α^* would be a simple weighted average as the criterion, where a parameter $\beta \in [0,1]$ will determine the weight put on the scope of impact.

$$\beta * \text{Scope of impact for recommending } a + (1-\beta) * f(\text{Welfare distribution of recommending } a)$$

Notice that although the example above is consider only two preferences ordering under three possible alternatives, one can easily apply the criterion to discuss general case. Depending on the specific functional assumption on f and the choice of β, one can find the corresponding optimal recommendation. However, the key concern here is that the policy maker does not have access to the underlying type distribution, i.e. μ, such that the policy maker can not perform the maximization exercise to find out which recommendation to make. Nevertheless, due to the uniqueness results, it turns out the ingredients for the criterion function can be uniquely identified from choice data ρ. Firstly, to estimate the scope of impact for recommending a, it is basically the fraction of decision makers who switch to choose a if a is recommended compared to the case of no recommendation. Therefore, it is simply $\rho(a,\{a\}) - \rho(a,\emptyset)$. Secondly, for the ex pose welfare distribution, we need to know the fraction of decision makers choosing their nth best after getting recommended a. We can recover it from the following formula. We let $\mathbb{P}(n) \subseteq 2^X$ be the collection of sets containing exactly n elements. Therefore, the fraction who chooses nth best after recommending a can be uniquely identified by

$$\underbrace{\sum_{A\in\mathbb{P}(N-n)} q(a,A) - y(a,A)}_{\text{fraction who switches to } a} + \underbrace{\sum_{A\in\mathbb{P}(N-n)} y(a,A) + \sum_{b\in X\setminus a}\ \sum_{A\in\mathbb{P}(N-n)|a\in A} y(b,A)}_{\text{fraction who does not switch choices}}$$

The above formula can be decomposed into two parts. The first part captures the fraction of DMs who switches to a as their nth best. Notice that this utilizes the marginal recommendation effect as discussed in the characterization, which captures the fraction of subjects who switch to a exactly because a is recommended and a is ranked exactly above the set A. It is straightforward to see that if we sum across all sets that contain $N-n$ elements, we get can the total fractions who switches to a as their nth best alternative. Secondly, the latter part contains two sub-parts. The first sub-part are those whose default option is a as nth best alternative and does not switch their choice; and the second sub-part captures those having another default options and the alternative a happens to be in their lower contour set which contains exactly $N-n$ elements.

R-Luce

It is a well-known fact that Luce is a special case of RUM. In this recommendation environment, one might wonder whether R-Luce belongs to R-RUM. It turns out that it is, which we state in the following.

Theorem 5. Every choice rule ρ that has a R-Luce representation has a R-RUM representation.

As far as we know, there are more than one way to show R-Luce is R-RUM. In the proof, we examine the relationship from the behavioral tests offered by the characterization of R-RUM, where we exploit the fact that Luce is a special case of RUM.

Fixed Default

In some scenario, one might suspect that the default option can be fixed while the individual can have a random preference. To capture this, we only needs to impose the following axiom, which basically says that any alternative outside of the recommendation set, other than the default, must be chosen with zero probability.

Axiom 11. There exists $a \in X$ such that $\rho(X \setminus a, \emptyset) = 0$.

Intuitively, this axiom will serve the purpose of having only one default option. To see this is the case, we utilize Remark 1. Since $\mu(\{(b, \succ) | \succ \in \mathcal{P}\}) = \rho(b, \emptyset) = 0$ for every $b \in X \setminus a$, hence, for any preference $\succ$, the probability measure μ will assign zero probability to $(b, \succ)$ whenever b is not the fixed default a. We put this observation into the following Remark.

Remark 2. ρ satisfies Axiom 8 to 11 if and only if ρ has a R-RUM representation where μ assigns positive probability only on one default.

Fixed Preference

Instead of fixing the default, one might want to investigate the case that the individual has a fixed preference. In this case, we can actually perform revealed preference exercise from the individual choice. We also make the assumption that every default must received positive probability to simplify the exposition of the result. Firstly, we let

$$xPy \text{ if there exists } y \text{ and } R \text{ such that } \rho(y|R) > \rho(y|R \cup x)$$

Notice that under a fixed preference, we are able to reveal the preference in this model. The idea is that, if an alternative is chosen less when an alternative is introduced to the recommended set, it must be because the alternative is worse than the new item. What's more, if every alternative can be the default option, we can fully reveal the preference. Therefore, we make the following axiom.

Axiom 12. P is complete, transitive and asymmetric.

On the other hand, we must also make sure we can shut down all other possible channels where the choice probability reflect the effect coming from other types. To do this, we need to assume that the marginal effect from other types are zero. Thus, we have the following axiom.

Axiom 13. $q(a, A) = 0$ for all a, A except for $A \setminus a = L_P(a)$.

We then state our following characterization results. From the construction, imposing these two axioms, we actually can pin down the weights such that $\mu(\{(a, P)\}) = y(a, A)$ where $A = L_P(a)$.[27]

[27] However, one should note that the model is inherently indistinguishable regarding lower contour set of the default option. Therefore, even though both axioms are imposed, the data may still be generated by two different preferences, where the preference differs only under the lower contour set of the default option. Nonetheless, we do have full control over the upper contour set of the default option.

Remark 3. ρ satisfies Axiom 8 to 10, 12 and 13 if and only if ρ has a R-RUM representation where μ assigns positive probability only on one preference.

5 Conclusion

Recommendation is abundant and prevalent in our lives. In this paper, we consider a rational benchmark model of recommendation: recommendation enlarges the decision maker's consideration set. Based on the recommendation, the decision maker compares it to her default and make the best choice. Supported by our deterministic model of behavior, we introduce stochastic choice models with the idea of aggregate choice data in mind. We propose parametric versions of our general model for tractability and applicability. We show that our models, R-RUM and R-Luce, have a close connection to the classic well-known standard stochastic choice models, RUM and Luce model.

While R-RUM and R-Luce are rational benchmark models of recommendation, just as RUM and Luce are in the standard choice domain, we expect that they can be subject to refinement or generalization according to specific needs under different circumstances. Our framework is rich enough to study other channels that recommendation might be operational. The areas of further research should includes developing more elaborate and complex models of recommendation including strategic recommendations, limited consideration, status quo, behavioral search, satisficing, and temptation.[28] Therefore, we believe that this paper also paves a palpable path for fruitful future research and applications where we can apply the economic wisdom that have accumulated throughout the years for the standard models to this setting.

[28] In the supplement material, we provide several recommendation models under limited considerations.

6 Appendix

In this Appendix, we will provide proofs for the main text.

Proof of Theorem 1

Proof. We first identify the default option a. We set $a := c(\emptyset)$. If $c(R) \notin R$ then by Axiom 3, we have $c(R) = a$. Hence, a is unique. For every distinct $x, y \in R \cup a$, we write

$$xPy \text{ if } x = c(R)$$

First we need to show P is asymmetric: for two distinct x and y, if $(x, y) \in P$ then $(y, x) \notin P$. Assume not. Then there exist R and R' such that $x = c(R)$ and $y = c(R')$ and $\{x, y\} \subset (R \cap R') \cup a$. By Axiom 1 and 2, we must have $x = c(\{x, y\}) = y$, a contradiction.

Note that if the default option is preferred to two distinct alternatives, we cannot reveal the relative ranking of these alternatives. In other words, aPx and aPy then $(x, y) \notin P$ or $(y, x) \notin P$. While P is incomplete for the lower counter set of a, P is complete in the upper counter set of a. To show this, assume xPa and yPa. In other words, there exist R and R' such that $x = c(R)$ and $y = c(R')$. If $x \in R'$ or $y \in R$, we would reveal yPx or xPy, respectively. Assume not. Then consider $\{x, y\}$ as the recommended set. First, a cannot be chosen from $c(\{x, y\})$. Otherwise, $a = c(\{x\})$ by Axiom 2. Then by Axiom 1, x cannot be chosen from $c(R)$. Hence, either xPy or yPx, P is complete for the upper counter set of a.

Claim 1. *If xPy and $x \neq a$ then xPa.*

Proof. xPy implies that there exists R such that $x, y \in R$ and $x = c(R)$. Since $\{x\} \subset R$, Axiom 1 implies $c(\{x\}) = x$ implying xPa. ■

Claim 2. *If $xPyPz$ then xPz.*

Proof. First note that x cannot be a since P is incomplete for the lower counter set of a. If z is equal to a, by Claim 1, we have xPz. If $y = a$, then $c(\{x\})$ is equal to x and $c(\{z\})$ is equal to a. Hence $c(\{x, z\})$ must be x by Axiom 1 and 2. Finally, we assume that x, y, z are distinct from a. Then we consider $c(\{x, y, z\})$. It cannot be a by Axiom 2 and Claim 1. It cannot be z since Axiom 1 and $z = c(\{y, z\})$. Finally, it cannot be y since Axiom 1 and $x = c(\{x, y\})$. Hence, xPz. ■

Take any completion of P, say $\succ$. It is routine to show that $c = c_{(a,\succ)}$. ■

Proof of Proposition 1

Proof. $(a_1, \succ_1)$ and $(a_2, \succ_2)$ represents the same choice rule. For i), note that it is immediate that $a_1 = c(\emptyset) = a_1$. For ii), suppose not, there exists b such that $b \in L_{\succ_1}(a)$ but $b \notin L_{\succ_2}(a)$. Then, we know that $c_{a,\succ_1}(\{b\}) = a \neq b = c_{a,\succ_2}(\{b\})$. Contradiction arises. For iii), suppse not, there exists $x, y \in X \setminus L_a$ such that $x \succ_1 y$ but $y \succ_2 x$. Then, we have $c_{a,\succ_1}(\{x, y\}) = x \neq y = c_{a,\succ_2}(\{x, y\})$. Contradiction arises. ■

Proof of Theorem 2

Proof. We first prove the necessity of the axioms. Suppose the model is correct. The necessity of Axiom 4 is immediate due to assumption of $w(x) > 0$. We then prove the necessity of Axiom 5.

Note that, for $x, y \in B$

$$
\begin{aligned}
&\frac{\rho(x,B)}{\rho(y,B)} \\
=&\frac{\left[\sum_{z\in B}\frac{d(z)}{w(B)}w(x) + \sum_{z\notin B}\frac{d(z)}{w(B\cup z)}w(x)\right]}{\left[\sum_{z\in B}\frac{d(z)}{w(B)}w(y) + \sum_{z\notin B}\frac{d(z)}{w(B\cup z)}w(y)\right]} \\
=&\frac{w(x)\left[\sum_{z\in B}\frac{d(z)}{w(B)} + \sum_{z\notin B}\frac{d(z)}{w(B\cup z)}\right]}{w(y)\left[\sum_{z\in B}\frac{d(z)}{w(B)} + \sum_{z\notin B}\frac{d(z)}{w(B\cup z)}\right]} \\
=&\frac{w(x)}{w(y)}
\end{aligned}
$$

Since B is arbitrary, it immediately implies Axiom 5. We then prove the necessity of Axiom 6. We make the following claim.

Claim 3. *For every $x \in B$ and $x \notin A \subseteq B$, we have $\rho(x,A)\rho(A \cup x, B) = \rho(x, \emptyset)\rho(x,B)$.*

Proof.

$$
\begin{aligned}
&\rho(x,A)\rho(A\cup x,B) - \rho(x,\emptyset)\rho(x,B) \\
=&d(x)\left[\frac{w(x)}{w(A\cup x)}[\rho(x,B)+\rho(A,B)] - \rho(x,B)\right] \\
=&\frac{d(x)}{w(A\cup x)}[w(x)\rho(A,B) - w(A)\rho(x,B)]
\end{aligned}
$$

Note that where

$$
\begin{aligned}
&w(x)\rho(A,B) \\
=&w(x)\left[\sum_{y\in B}\frac{d(y)}{w(B)}w(A) + \sum_{y\notin B}\frac{d(y)}{w(B\cup y)}w(A)\right] \\
=&w(A)\left[\sum_{y\in B}\frac{d(y)}{w(B)}w(x) + \sum_{y\notin B}\frac{d(y)}{w(B\cup y)}w(x)\right] \\
=&w(A)\rho(x,B)
\end{aligned}
$$

Hence, since A is arbitrary, the above claim immediately implies Axiom 6. ■

For sufficiency, we define

$$d(x) := \rho(x,\emptyset) \geq 0 \text{ and } w(x) := \rho(x,X) > 0$$

First, Axiom 6 implies that

$$\rho(x,A)\rho(A\cup x,X) = \rho(x,\emptyset)\rho(x,X)$$

Hence we have representation for off-recommendation, i.e. $x \notin A$:

$$\rho(x, A) = \rho(x, \emptyset)\frac{\rho(x, X)}{\rho(A \cup x, X)} = d(x)\frac{w(x)}{\sum\limits_{z \in A \cup x} w(z)} = d(x)W_x(x, A)$$

For on-recommendation alternative, we first make the following claim.

Claim 4. *Axiom 6 implies that for $\emptyset \neq A \neq X$,*

$$\rho(A, A) - \rho(A, \emptyset) = \rho(A, X)\sum_{y \notin A} \frac{\rho(y, A)}{\rho(y, X)}$$

Proof. To prove this, fix a A, we first consider $x \notin X \setminus A$. By Axiom 6, we have, for every $x \notin X \setminus A$,

$$\begin{aligned}
\rho(x, A)(\rho(x, X) + \rho(A, X)) &= \rho(x, \emptyset)\rho(x, X) \\
\rho(x, A) + \frac{\rho(x, A)}{\rho(x, X)}\rho(A, X) &= \rho(x, \emptyset)
\end{aligned}$$

Summing all $x \notin A$, we have

$$\begin{aligned}
\sum_{x \notin A}\left(\rho(x, A) + \frac{\rho(x, A)}{\rho(x, X)}\rho(A, X)\right) &= \sum_{x \notin A} \rho(x, \emptyset) \\
1 - \rho(A, A) + \rho(A, X)\sum_{x \notin A} \frac{\rho(x, A)}{\rho(x, X)} &= 1 - \rho(A, \emptyset) \\
\rho(A, A) - \rho(A, \emptyset) = \rho(A, X)\sum_{y \notin A} \frac{\rho(y, A)}{\rho(y, X)}
\end{aligned}$$

∎

By Axiom 5, if $x \in A$ then

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

By summing all $y \in A$, we have

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

then

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

Hence, for $x \in A$, by Claim 4 and $\rho(x, A) = \frac{\rho(A,A)\rho(x,X)}{\rho(A,X)}$, we have

$$\begin{aligned}
\rho(x, A) &= \frac{\rho(x, X)}{\rho(A, X)}\left[\rho(A, \emptyset) + \rho(A, X)\sum_{y \notin A} \frac{\rho(y, A)}{\rho(y, X)}\right] \\
&= \frac{\rho(x, X)}{\rho(A, X)}\left[\rho(A, \emptyset) + \rho(A, X)\sum_{y \notin A} \frac{\rho(y, \emptyset)}{\rho(A \cup y, X)}\right]
\end{aligned}$$

Since $\rho(y, A) = \rho(y, \emptyset)\frac{\rho(y,X)}{\rho(A \cup y,X)}$ for $y \notin A$

$$= \rho(x, X)\Big[\frac{\rho(A, \emptyset)}{\rho(A, X)} + \sum_{y \notin A} \frac{\rho(y, \emptyset)}{\rho(A \cup y, X)}\Big]$$

$$= \sum_{y \in A} \frac{d(y)w(x)}{\sum\limits_{z \in A} w(z)} + \sum_{y \notin A} \frac{d(y)w(x)}{\sum\limits_{z \in A \cup y} w(z)} \qquad \text{By construction of } d \text{ and } w$$

$$= \sum_{a \in X} d(a) W_a(x, A)$$

The proof is complete. ■

Proof of Theorem 3

Proof. The proof for necessity of Axiom 4-6 is proven in Theorem 2. We prove necessity of Axiom 7. Suppose the model is correct, let $x \notin A$,

$$\frac{\rho(x, \emptyset)}{\sum_{z \in A \cup x} r(z, x)}$$

$$= \frac{d(x)}{\sum_{z \in A \cup x} \frac{w(z)}{w(x)}} \qquad \text{By the necessity proof of Axiom 5}$$

$$= \frac{d(x)w(x)}{\sum_{z \in A \cup x} w(z)} = \rho(x, A)$$

Since A is arbitrary, it is proven.

We then prove the sufficiency. We first let, for every $x \in X$, $d(x) := w(x, \emptyset)$. We arbitrarily designate $z_0 \in X$ as an "anchored" element. And let $w(z_0) = 1$. Since we have all the binary recommendation sets in our data, we let for every $x \in X$,

$$w(x) = \frac{\rho(x, \{x, z_0\})}{\rho(z_0, \{x, z_0\})}$$

Then, we prove the following claim.

Claim 5. *For any $x, y \in A$ with $|A| \leq k - 1$, we have $\frac{\rho(x,A)}{\rho(y,A)} = \frac{w(x)}{w(y)}$. And similarly, for any $x \in A$ and $B \subseteq A$ we have $\frac{\rho(x,A)}{\rho(B,A)} = \frac{w(x)}{w(B)}$*

Proof. Note that, firstly, for any set $A \supseteq \{x, z_0\}$, we have $w(x) = \frac{\rho(x,A)}{\rho(z_0,A)}$ by Axiom 5. Then, for any $x, y \in A$, we have

$$\frac{\rho(x, A)}{\rho(y, A)} = \frac{\rho(x, A \cup z_0)}{\rho(y, A \cup z_0)} \qquad \text{By Axiom 5}$$

$$= \frac{\rho(x, A \cup z_0)\rho(z_0, A \cup z_0)}{\rho(z_0, A \cup z_0)\rho(y, A \cup z_0)}$$

$$= \frac{w(x)}{w(y)}$$

Hence, the first part is proven. The second part is immediate. ■

We first show that the representation holds for off-recommendation set. Let $x \notin A$, by

Axiom 7, we have

$$\begin{aligned}\rho(x,A) &= \frac{\rho(x,\emptyset)}{1+\sum_{z\in A} r(z,x)}\\ &= \frac{d(x)}{\sum_{z\in A\cup x}\frac{w(z)}{w(x)}} && \text{By Claim 5}\\ &= \frac{d(x)w(x)}{w(A\cup x)}\end{aligned}$$

Hence, the representation holds for off-recommendation alternative.

Then, we show that the representation holds for on-recommendation set (i.e. $x \in A$). The representation for $\rho(x, A)$ is immediately proven if $|A|= 1$. Let $|A|\geq 2$, then for every $x \in A$, we have

$$\begin{aligned}\rho(x,A) + \sum_{y\in A\setminus x}\rho(y,A) + \sum_{y\in X\setminus A}\rho(y,A) &= 1\\ \rho(x,A) + \sum_{y\in A\setminus x}\rho(x,A)\frac{\rho(y,\{x,y\})}{\rho(x,\{x,y\})} &= 1-\sum_{y\in X\setminus A}\rho(y,A) && \text{By Axiom 5}\\ \rho(x,A) + \sum_{y\in A\setminus x}\rho(x,A)\frac{w(y)}{w(x)} &= 1-\sum_{y\in X\setminus A}\rho(y,A) && \text{By Claim 5}\\ \rho(x,A)\frac{w(A)}{w(x)} &= 1-\sum_{y\in X\setminus A}\rho(y,A)\\ \rho(x,A) &= \frac{w(x)}{w(A)}\left[1-\sum_{y\in X\setminus A}\frac{d(y)w(y)}{w(A\cup y)}\right]\\ & && \text{By construction of } \rho(y,A) \text{ for } y\notin A\end{aligned}$$

Then, we prove the following claim.

Claim 6. *For $x \in A$ with $|A|\geq 2$*

$$\frac{w(x)}{w(A)}\left[1-\sum_{y\in X\setminus A}\frac{d(y)w(y)}{w(A\cup y)}\right] = w(x)\left[\frac{d(A)}{w(A)}+\sum_{y\in X\setminus A}\frac{d(y)}{w(A\cup y)}\right]$$

Proof.

$$\begin{aligned}&\frac{w(x)}{w(A)}\left[1-\sum_{y\in X\setminus A}\frac{d(y)w(y)}{w(A\cup y)}\right] - w(x)\left[\frac{d(A)}{w(A)}+\sum_{y\in X\setminus A}\frac{d(y)}{w(A\cup y)}\right]\\ =&w(x)\left[\frac{1-d(A)}{w(A)}-\sum_{y\in X\setminus A}[\frac{d(y)w(y)}{w(A)w(A\cup y)}+\frac{d(y)}{w(A\cup y))}]\right]\\ =&w(x)\left[\sum_{y\in X\setminus A}d(y)[\frac{1}{w(A)}-\frac{w(y)}{w(A)w(A\cup y)}-\frac{1}{w(A\cup y)}]\right]\\ =&0\end{aligned}$$

Hence, this claim is proven. ■

Hence, we have shown that

$$\rho(x, A) = w(x)\left[\frac{d(A)}{w(A)} + \sum_{y \in X \setminus A} \frac{d(y)}{w(A \cup y)}\right]$$

By re-arrangement, one can see that it is the representation for $\rho(x, A)$ where $x \in A$. It is proven.

■

Proof of Proposition 2

Proof. Suppose that (w_1, d_1) and (w_2, d_2) represent the same choice rule. Then, by definition, for every $x \in X$, $d_1(x) = \rho(x, \emptyset) = d_2(x)$. Also, for every $x \in X$, we have

$$\frac{w_1(x)}{\sum_{x \in X} w_1(x)} = \rho(x, X) = \frac{w_2(x)}{\sum_{x \in X} w_2(x)}$$

Hence, $w_1 = \frac{\sum_{x \in X} w_1(x)}{\sum_{x \in X} w_2(x)} w_2$, where $\frac{\sum_{x \in X} w_1(x)}{\sum_{x \in X} w_2(x)} > 0$ by definition. ■

Proof of Theorem 5

Proof. It suffices to show that it satisfies Axiom 8 to 10. In the following, for notational ease, we denote $w(A) := \sum_{x \in A} w(x)$. For Axiom 8, we prove it by using standard results from the relationship between RUM and Luce model. Notice that, for $x \notin R$,

$$\begin{aligned} y(x, R) &= \sum_{x \notin B \supseteq R} (-1)^{|B \setminus R|} \rho(x, B) \\ &= \sum_{x \notin B \supseteq R} (-1)^{|B \setminus R|} d(x) \frac{w(x)}{w(B \cup x)} \\ &= d(x) \sum_{A \supseteq R \cup x} (-1)^{|A \setminus (R \cup x)|} \frac{w(x)}{w(A)} \end{aligned}$$

It is a well-known result that every Luce model has a RUM representation. Hence, the term $\sum\limits_{A \supseteq R \cup x} (-1)^{|A \setminus (R \cup x)|} \frac{w(x)}{w(A)}$ is guaranteed to be non-negative since it is exactly the standard block Marschak polynomials of a Luce model.

For Axiom 9 and 10, we first make the following auxiliary claim.

Claim 7. *For every $A \subset X$ and $z \notin A$,* $\sum\limits_{B \supseteq A} (-1)^{|B \setminus A|} \frac{1}{w(B \cup z)} = 0$

Proof.

$$\begin{aligned} &\sum_{B \supseteq A} (-1)^{|B \setminus A|} \frac{1}{w(B \cup z)} \\ =& \sum_{B \supseteq A} (-1)^{|B \setminus A|} \frac{1}{w(B \cup z)} \\ =& \sum_{z \notin B \supseteq A} (-1)^{|B \setminus A|} \frac{1}{w(B \cup z)} + \sum_{z \in B \supseteq A} (-1)^{|B \setminus A|} \frac{1}{w(B \cup z)} \\ =& \sum_{C \supseteq A \cup z} (-1)^{|C \setminus A \cup z|} \frac{1}{w(C)} + \sum_{B \supseteq A \cup z} (-1)^{|B \setminus (A \cup z)| + 1} \frac{1}{w(B)} \end{aligned}$$

$$=0$$

■

Then, we make the following claim. To prove this, we utilize the following expression of $\rho(x, A)$ for $x \in A$,

$$\rho(x, A) = \rho(x, A \setminus x) + \sum_{z \in A \setminus x} \frac{w(x)}{w(z)} \rho(z, A \setminus z) + \sum_{z \notin A} \frac{w(x)}{w(z)} \rho(z, A)$$

Claim 8. *For every $x \in A$,*

$$q(x, A) = y(x, A \setminus x) + \sum_{z \in A \setminus x} \frac{w(x)}{w(z)} y(z, A \setminus z)$$

Proof.

$$\begin{aligned}
q(x, A) &= \sum_{B \supseteq A} (-1)^{|B \setminus A|} \rho(a, B) && \text{By definition} \\
&= \sum_{B \supseteq A} (-1)^{|B \setminus A|} \left[\rho(x, B \setminus x) + \sum_{z \in B \setminus x} \frac{w(x)}{w(z)} \rho(z, B \setminus z) + \sum_{z \notin B} \frac{w(x)}{w(z)} \rho(z, B) \right] \\
& && \text{by the above expression} \\
&= \sum_{B \supseteq A} (-1)^{|B \setminus A|} \left[\rho(x, B \setminus x) + \sum_{z \in A \setminus x} \frac{w(x)}{w(z)} \rho(z, B \setminus z) + \sum_{z \in B \setminus A} \frac{w(x)}{w(z)} \rho(z, B \setminus z) + \sum_{z \notin B} \frac{w(x)}{w(z)} \rho(z, B) \right] \\
&= y(x, A \setminus x) + \sum_{z \in A \setminus x} \frac{w(x)}{w(z)} y(z, A \setminus z) + \sum_{B \supseteq A} (-1)^{|B \setminus A|} \left[\sum_{z \in B \setminus A} \frac{w(x)}{w(z)} \rho(z, B \setminus z) + \sum_{z \notin B} \frac{w(x)}{w(z)} \rho(z, B) \right] \\
& && \text{By applying the definition of } y \text{ on the first two terms}
\end{aligned}$$

It remains to show that the last sum is zero. Note that

$$\begin{aligned}
& \sum_{B \supseteq A} (-1)^{|B \setminus A|} \left[\sum_{z \in B \setminus A} \frac{w(x)}{w(z)} \rho(z, B \setminus z) + \sum_{z \notin B} \frac{w(x)}{w(z)} \rho(z, B) \right] \\
=& w(x) \sum_{B \supseteq A} (-1)^{|B \setminus A|} \sum_{z \in X \setminus A} \frac{1}{w(z)} \rho(z, B \setminus z) \\
=& w(x) \sum_{B \supseteq A} (-1)^{|B \setminus A|} \sum_{z \in X \setminus A} \frac{1}{w(z)} \frac{w(z) d(z)}{w(B \cup z)} && \text{By the definition of } \rho(z, B) \\
=& w(x) \sum_{z \in X \setminus A} d(z) \sum_{B \supseteq A} (-1)^{|B \setminus A|} \frac{1}{w(B \cup z)} && \text{By switching summation sign} \\
=& 0 && \text{by Claim 7}
\end{aligned}$$

■

Hence, Claim 8 is proven. Hence, to show that Axiom 9 is satisfied, we have

$$q(x, A) - y(x, A \setminus x) = \sum_{z \in A \setminus x} \frac{w(x)}{w(z)} y(z, A \setminus z) \geq 0$$

Also, Axiom 10 is immediately satisfied by putting $A = \emptyset$ into the expression in Claim 8. ■

Proof of Theorem 4

Proof. For the necessity proof, we suppose the data follows the model. We introduce the notation, for $a, b \in X$ and $b \notin A$,

$$M(a,b,A) = \mu\Big(\{(a,\succ) \in \mathcal{T} : A = L_{\succ}(b)\}\Big)$$

where $L_{\succ}(a)$ is the strict lower contour set of a according to $\succ$.

Claim 9. *For $x \notin A$,*
i) $M(x,x,A) = y(x,A)$
ii) $\sum_{a \in A \cup \{x\}} M(a,x,A) = q(x, A \cup \{x\})$

Proof. For i), we prove by strong induction by "stepping down". For $A = X \setminus \{x\}$, we have,

$$\begin{aligned} y(x, X \setminus \{x\}) &= \rho(x, X \setminus \{x\}) && \text{By definition of } y \\ &= \mu\Big(\{(x,\succ) : X \setminus \{x\} = L_{\succ}(x)\}\Big) \\ &= M(x,x,X \setminus \{x\}) \end{aligned}$$

So, i) is true for size of A equals to $|X|-1$. Suppose i) is true for size of $k+1$, $k+2$,...,$|N|-1$. Let $|A|= k$

$$\begin{aligned} y(x,A) &= c(x,A) - \sum_{x \notin B \supset A} y(x,B) \\ &= \sum_{x \notin B \supseteq A} M(x,x,B) - \sum_{x \notin B \supset A} M(x,x,B) && \text{by Definition and induction hypothesis} \\ &= M(x,x,A) \end{aligned}$$

Hence, the proof is complete for i).

For ii), we also prove by strong induction by "stepping down". For $A = X \setminus \{x\}$, we have,

$$\begin{aligned} q(x,X) &= \rho(x,X) && \text{By definition} \\ &= \sum_{a \in A \cup \{x\}} \mu\Big(\{(a,\succ) : X \setminus \{x\} = L_{\succ}(x)\}\Big) \\ &= \sum_{a \in X} M(a,x,X \setminus \{x\}) \end{aligned}$$

So, ii) is true for size of A equals to $|X|-1$. Suppose i) is true for size of $k+1$, $k+2$,...,$N-1$. Let $|A|= k$

$$\begin{aligned} q(x, A \cup \{x\}) &= c(x, A \cup \{x\}) - \sum_{B \supset A \cup \{x\}} q(x,B) \\ &= \sum_{B \supseteq A \cup \{x\}} \sum_{a \in B} M(a,x,B \setminus \{x\}) - \sum_{B \supset A \cup \{x\}} \sum_{a \in B} M(a,x,B \setminus \{x\}) \\ & \qquad \text{by Definition and induction hypothesis} \\ &= \sum_{a \in A \cup \{x\}} M(a,x,A) \end{aligned}$$

Hence, the proof is complete for ii). ■

From this claim, we immediately show Axiom 8 and Axiom 9 since for $x \notin A$, we have

$$\begin{aligned} y(x,A) &= M(x,x,A) \geq 0 \\ q(x, A \cup \{x\}) - y(x,A) &= \sum_{a \in A \cup \{x\}} M(a,x,A) - M(x,x,A) = \sum_{a \in A} M(a,x,A) \geq 0 \end{aligned}$$

Moreover, by putting in $A = \emptyset$ into Claim 1, we have

$$y(x,\emptyset) = M(x,x,\emptyset) = \sum_{a \in \{x\}} M(a,x,\emptyset) = q(x,x)$$

Hence, Axiom 10 is proven. The necessity proof is complete.

For the sufficiency, we need to first prove a lemma for later use.

Lemma 1. For $R \subset X$ and choice rule ρ,

$$\sum_{a \in R} q_\rho(a,R) + \sum_{a \notin R} y_\rho(a,R) = \sum_{b \notin R} q_\rho(b, R \cup b)$$

Proof. We need to show that, for every $R \subset X$,

$$\sum_{a \in R} q(a,R) + \sum_{a \notin R} y(a,R) = \sum_{b \notin R} q(b, R \cup b)$$

We prove by strong induction by "stepping down". For $R = X \setminus \{x\}$, we have

$$\begin{aligned} \text{RHS} &= q(x,X) = \rho(x,X) \\ \text{LHS} &= \sum_{a \in X \setminus \{x\}} q(a, X \setminus \{x\}) + \sum_{a \notin X \setminus \{x\}} y(a, X \setminus \{x\}) \\ &= \sum_{a \in X \setminus \{x\}} q(a, X \setminus \{x\}) + \rho(x, X \setminus \{x\}) \\ &= \sum_{a \in X \setminus \{x\}} \Big[\rho(a, X \setminus \{x\}) - \rho(a,X) \Big] + \rho(x, X \setminus \{x\}) \\ &= \sum_{a \in X} \rho(a, X \setminus \{x\}) - \sum_{a \in X \setminus \{x\}} \rho(a,X) \\ &= 1 - \sum_{a \in X \setminus \{x\}} \rho(a,X) \\ &= \rho(x,X) \end{aligned}$$

Suppose that equality holds for size of $k+1, k+2, \ldots N-1$. Let $|R| = k$,

$$\begin{aligned} &\text{LHS} - \text{RHS} \\ &= \sum_{a \in R} q(a,R) + \sum_{a \notin R} y(a,R) - \sum_{b \notin R} q(b, R \cup b) \\ &= \sum_{a \in R} \left(\rho(a,R) - \sum_{B \supset R} q(a,B) \right) + \sum_{a \notin R} \left(\rho(a,R) - \sum_{a \notin B \supset R} y(a,B) \right) - \sum_{b \notin R} q(b, R \cup b) \end{aligned}$$

$$= \sum_{a \in X} \rho(a, R) - \left[\sum_{a \in R} \sum_{B \supset R} q(a, B) + \sum_{a \notin R} \sum_{a \notin B \supset R} y(a, B) + \sum_{b \notin R} q(b, R \cup b) \right]$$

Since $\sum_{a \in X} \rho(a, R) = 1$, it remains to show that the latter term in the above expression equals 1. We denote $\mathcal{D}_R(i)$ as the collection of superset of R with i element. Hence, we can rewrite

$$\sum_{a \in R} \sum_{B \supset R} q(a, B) + \sum_{a \notin R} \sum_{a \notin B \supset R} y(a, B) + \sum_{b \notin R} q(b, R \cup b)$$

$$= \sum_{i=|R|+1}^{N} \sum_{B \in \mathcal{D}_R(i)} \left[\sum_{a \in R} q(a, B) + \sum_{a \notin B} y(a, B) \right] + \sum_{B \in \mathcal{D}_R(|R|+1)} \sum_{a \notin R} q(a, B) \qquad \text{By rearrangement}$$

$$= \sum_{i=|R|+2}^{N} \sum_{B \in \mathcal{D}_R(i)} \left[\sum_{a \in R} q(a, B) + \sum_{a \notin B} y(a, B) \right] + \sum_{B \in \mathcal{D}_R(|R|+1)} \left[\sum_{a \in B} q(a, B) + \sum_{a \notin B} y(a, B) \right]$$

By taking $i = |R|+1$ from the 1st term and summing it to the second term

$$= \sum_{i=|R|+2}^{N} \sum_{B \in \mathcal{D}_R(i)} \left[\sum_{a \in R} q(a, B) + \sum_{a \notin B} y(a, B) \right] + \sum_{B \in \mathcal{D}_R(|R|+1)} \sum_{a \notin B} q(a, B \cup a)$$

By induction hypothesis

$$= \sum_{i=|R|+2}^{N} \sum_{B \in \mathcal{D}_R(i)} \left[\sum_{a \in R} q(a, B) + \sum_{a \notin B} y(a, B) \right] + \sum_{B \in \mathcal{D}_R(|R|+2)} \sum_{a \notin R} q(a, B) \qquad \text{By rearrangement}$$

=...(repetitively applying induction hypothesis)

$$= \sum_{i=N}^{N} \sum_{B \in \mathcal{D}_R(i)} \left[\sum_{a \in R} q(a, B) + \sum_{a \notin B} y(a, B) \right] + \sum_{B \in \mathcal{D}_R(|N|)} \sum_{a \notin R} q(a, B)$$

$$= \sum_{a \in X} q(a, X)$$

$$= \sum_{a \in X} \rho(a, X) \qquad \text{By definition of } q$$

$$=1$$

Hence, it is proven.

■

Then, we need to introduce new notation. For any $R \subseteq X$, we write Π_R for the set of $|R|!$ permutations on R, with typical element π_R. We write Π for Π_X. Let $\pi_R(i)$ refers to the ith element on the permutation. The type space is now instead specified by $X \times \Pi$, with element (a, π), where $a \in X$ and $\pi \in \Pi$.

For the sufficiency proof, we first construct $F(a, a, A) := y(a, A)$ for $A \subseteq X \setminus \{a\}$. In the following, for every π_R, we denote $a\pi_R$ as the lengthen element of π_R in $\Pi_{R \cup \{a\}}$ where a is inserted at the beginning of the permutation, and similarly, we denote $\pi_R b$ as the lengthen permutation of π_R where b is inserted at the end of the permutation. Anaglously, we denote $a\pi_R b$ where a and b are inserted at the beginning and the end, respectively. Then, we construct, recursively, for $a, b \notin R \cup A$, $R \cap A = \emptyset$ and $b\pi_R a \in \Pi_{R\cup\{a,b\}}$

$$F(a, b\pi_R a, A) = \begin{cases} \dfrac{F(a, \pi_R a, A)(q(b, R \cup A \cup \{a, b\}) - y(b, R \cup A \cup \{a\}))}{\sum_{x\in R\cup A\cup\{a\}} q(x, R \cup A \cup \{a\})} & \text{if denominator is non-zero} \\ 0 & \text{otherwise} \end{cases}$$

Firstly, note that $F \geq 0$ by Axiom 8 and 9.

Claim 10. *For every a, A, $F(a, \pi_R a, A) = \sum_{b\in X\setminus R\cup A\cup\{a\}} F(a, b\pi_R a, A)$*

Proof.

$$\begin{aligned} \sum_{b\in X\setminus A\cup R\cup\{a\}} F(a, b\pi_R a, A) &= \sum_{b\in X\setminus R\cup A\cup\{a\}} \frac{F(a, \pi_R a, A)(q(b, R \cup A \cup \{a, b\}) - y(b, R \cup A \cup \{a\}))}{\sum_{x\in R\cup A\cup\{a\}} q(x, R \cup A \cup \{a\})} \\ &= F(a, \pi_R a, A)\frac{\sum_{b\in X\setminus R\cup A\cup\{a\}}(q(b, R \cup A \cup \{a, b\}) - y(b, R \cup A \cup \{a\}))}{\sum_{x\in R\cup A\cup\{a\}} q(x, R \cup A \cup \{a\})} \\ &= F(a, \pi_R a, A) \qquad \text{by Lemma 1} \end{aligned}$$

■

We show two more properties of F.

Claim 11. *For every non-empty A and $x \notin A$,*
i) $\sum_{a\in A} \sum_{B\subseteq A\setminus\{a\}} \sum_{\pi_B\in\Pi_B} F(a, \pi_B a, A \setminus (B \cup a)) = \sum_{b\in A} q(b, A)$
ii) $\sum_{a\in A} \sum_{B\subseteq A\setminus\{a\}} \sum_{\pi_B\in\Pi_B} F(a, x\pi_B a, A \setminus (B \cup a)) = q(x, A \cup \{x\}) - y(x, A)$

Proof. Note that by expanding the LHS of i) and ii) with the definition of F, one can show that

$$\sum_{a\in A} \sum_{B\subseteq A\setminus\{a\}} \sum_{\pi_B\in\Pi_B} F(a, \pi_B a, A \setminus (B \cup a)) = \sum_{a\in A} F(a, a, A \setminus \{a\}) + \sum_{\substack{C\subseteq A \\ |C|=|A|-1}} \left\{ \sum_{a\in C} \sum_{B\subseteq C\setminus\{a\}} \sum_{\pi_B\in\Pi_B} F(a, c\pi_B a, C \setminus (B \cup a)) : \{c\} = A \setminus C \right\} \quad .....(*)$$

and also

$$\sum_{a\in A} \sum_{B\subseteq A\setminus\{a\}} \sum_{\pi_B\in\Pi_B} F(a, x\pi_B a, A \setminus (B \cup a)) = \frac{q(x, A \cup \{x\}) - y(x, A)}{\sum\limits_{b\in A} q(b, A)} \sum_{a\in A} \sum_{B\subseteq A\setminus\{a\}} \sum_{\pi_B\in\Pi_B} F(a, \pi_B a, A \setminus (B \cup a)) \quad .....(**)$$

Based on this observation, we prove i) and ii) together by induction by the size of A.

For $|A|= 1$, we let $A = \{a\}$. Then, for i), we have

$$\begin{aligned} \sum_{a\in A} \sum_{B\subseteq A\setminus\{a\}} \sum_{\pi_B\in\Pi_B} F(a, \pi_B a, A \setminus (B \cup a)) &= F(a, a, \emptyset) \\ &= y(a, \emptyset) \qquad \text{by construction} \\ &= q(a, a) \qquad \text{by Axiom 10} \end{aligned}$$

For ii), we have

$$\sum_{a\in A}\sum_{B\subseteq A\setminus\{a\}}\sum_{\pi_B\in\Pi_B} F(a, x\pi_B a, A\setminus(B\cup a)) = F(a, xa, \emptyset)$$

$$= \frac{F(a,a,\emptyset)(q(x,\{x,a\})-q(x,\{a\})}{q(a,a)} \quad \text{by construction}$$

$$= q(x,\{x,a\})-q(x,\{a\}) \quad \text{by Axiom 10}$$

Hence, i) and ii) are true for size of A equals 1. Suppose i) and ii) are true for size of $k-1$. Let $|A|=k$. Then, for i), we have, by using $(*)$

$$\sum_{a\in A}\sum_{B\subseteq A\setminus\{a\}}\sum_{\pi_B\in\Pi_B} F(a, \pi_B a, A\setminus(B\cup a))$$

$$=\sum_{a\in A} F(a,a,A\setminus\{a\}) + \sum_{\substack{C\subseteq A\\|C|=|A|-1}} \left\{\sum_{a\in C}\sum_{B\subseteq C\setminus\{a\}}\sum_{\pi_B\in\Pi_B} F(a, c\pi_B a, C\setminus(B\cup a)) : \{c\} = A\setminus C\right\}$$

$$=\sum_{a\in A} y(a,A\setminus\{a\}) + \sum_{\substack{C\subseteq A\\|C|=|A|-1}} \left\{q(c, C\cup\{c\}) - y(c,C) : \{c\} = A\setminus C\right\}$$

by construction and induction hypotheses where $|C|=k-1$

$$=\sum_{a\in A} y(a,A\setminus\{a\}) + \sum_{a\in A}(q(a,A)-y(a,A\setminus\{a\}))$$

$$=\sum_{a\in A} q(a,A)$$

Hence, it is confirmed that $i)$ is true for size of k. Then, for $ii)$, by using $(**)$

$$\sum_{a\in A}\sum_{B\subseteq A\setminus\{a\}}\sum_{\pi_B\in\Pi_B} F(a, x\pi_B a, A\setminus(B\cup a))$$

$$=\frac{q(x,A\cup\{x\})-y(x,A)}{\sum_{b\in A} q(b,A)}\sum_{a\in A}\sum_{B\subseteq A\setminus\{a\}}\sum_{\pi_B\in\Pi_B} F(a, \pi_B a, A\setminus(B\cup a))$$

$$=\frac{q(x,A\cup\{x\})-y(x,A)}{\sum_{b\in A} q(b,A)}\sum_{b\in A} q(b,A) \quad \text{since } i) \text{ is true for size of A to be } k$$

$$=q(x,A\cup\{x\})-y(x,A)$$

Hence, by induction, i) and ii) hold. ■

We then define each individual weight. For $\pi\in\Pi$, we first write $\pi^{t(a)}$ as the "truncated" sequence of π up to a and does not include a. Also, we write $L_\pi(a)$ as the strict lower contour set of a according to π. Hence, we define

$$\hat{\mu}_{a,\pi} := F(a, \pi^{t(a)}a, L_\pi(a))\frac{1}{|L_\pi(a)|!}$$

Firstly, note that $\hat{\mu}_{a,\pi}\geq 0$ due to the fact that $F\geq 0$. On the other hand, notice that the fraction $\frac{1}{|L_\pi(a)|!}$ reflects the fact that the model does not distinguish weights when other alternatives are worse than the default. Here, for succinctness, we assume even weights. Yet, it is non-inconsequential for the construction to work.

In the following, we introduce the notation, for $a, b \in X$ and $b \notin A$,

$$\hat{M}(a,b,A) = \sum\{\hat{\mu}_{a,\pi} : \ A = L_\pi(b)\}$$

Claim 12. *For $x \notin A$,*
i) $\hat{M}(x,x,A) = y(x,A)$
ii) $\sum_{a\in A} \hat{M}(a,x,A) = q(x, A\cup\{x\}) - y(x,A)$

Proof. For i), by using Claim 10, one can show that

$$\hat{M}(x,x,A) = F(x,x,A)$$

Hence, by construction of F, it is proven. For ii), by expanding and using Claim 10, one can see that, for $a \neq x$,

$$\hat{M}(a,x,A) = \sum_{B\subseteq A\setminus\{a\}} \sum_{\pi_B\in\Pi_B} F(a, x\pi_B a, A\setminus(B\cup a))$$

Hence, by putting $\sum_{a\in A}$ on both side and applying Claim 11(i), it is proven. ■

Since the weight are constructed, we have, for $x \in R$,

$$\begin{aligned}
\hat{\rho}(x,R) &= \sum_{A\supseteq R}\sum_{a\in A} \hat{M}(a,x,A\setminus\{x\}) && \text{by Definition}\\
&= \sum_{A\supseteq R}\left[\hat{M}(x,x,A\setminus\{x\}) + \sum_{a\in A\setminus\{x\}} \hat{M}(a,x,A\setminus\{x\})\right]\\
&= \sum_{A\supseteq R}\left[y(x,A\setminus\{x\}) + q(x,A) - y(x,A\setminus\{x\})\right] && \text{By Claim 12}\\
&= \sum_{A\supseteq R} q(x,A)\\
&= \rho(x,R) && \text{By Mobius Inversion}
\end{aligned}$$

for $x \notin R$,

$$\begin{aligned}
\hat{\rho}(x,R) &= \sum_{a\notin B\supseteq R} \hat{M}(x,x,B) && \text{by Definition}\\
&= \sum_{a\notin B\supseteq R} y(x,B) && \text{By Claim 12(i)}\\
&= \rho(x,R) && \text{By Mobius Inversion}
\end{aligned}$$

Hence, the constructed weights explain the data. Since it explains the data, it is immediately that $\sum_{a\in X}\sum_{\pi\in\Pi} \hat{\mu}_{a,\pi} = 1$. The sufficiency proof is complete.

■

Proof of Proposition 4

Proof. It is proven in Claim 9. ■

Fact 1. *Axiom 5 and Axiom 6 are independent.*

Proof. Consider the following two choice rules

$\rho(.)$	a	b
$\{a,b\}$	.5	.5
$\{a\}$	1	0
$\{b\}$	0	1
$\emptyset$	.75	.25

$\rho(.)$	a	b	c
$\{a,b,c\}$	1/3	1/3	1/3
$\{a,b\}$	1/3	5/9	1/9
$\{a,c\}$	1/3	1/9	5/9
$\{b,c\}$	1/9	1/3	5/9
$\{a\}$	2/3	1/6	1/6
$\{b\}$	1/6	2/3	1/6
$\{c\}$	1/6	1/6	2/3
$\emptyset$	1/3	1/3	1/3

In the left choice rule, Axiom 5 is trivially satisfied. Axiom 6 is violated since

$$\rho(a,\emptyset)\rho(\{a\},\{a,b\}) = .375 \neq 0 = \rho(a,\{b\})\rho(\{a,b\},\{a,b\})$$

In the right choice rule, Axiom 5 is violated since $\frac{\rho(a,\{a,b,c\})}{\rho(a,\{a,b\})} = 1 \neq \frac{3}{5} = \frac{\rho(b,\{a,b,c\})}{\rho(b,\{a,b\})}$. Axiom 6 is satisfied, since we have

$$\rho(a,A)\rho(A\cup a,X) = \frac{1}{9} \text{ for every } A \text{ such that } a \notin A$$
$$\rho(b,A)\rho(A\cup b,X) = \frac{1}{9} \text{ for every } A \text{ such that } b \notin A$$
$$\rho(c,A)\rho(A\cup c,X) = \frac{1}{9} \text{ for every } A \text{ such that } c \notin A$$

■

References

Abernethy, A. M., & Butler, D. D. (1992). Advertising information: services versus products. *Journal of Retailing*, *68*(4), 398.

Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2018). Effects of Online Recommendations on Consumers' Willingness to Pay. *Information Systems Research*, *29*(1), 84–102.

Aggarwal, C. C. (2016). Neighborhood-based collaborative filtering. *Recommender systems: The textbook* (pp. 29–70). Springer International Publishing.

Ahumada, A., & Ülkü, L. (2018). Luce rule with limited consideration. *Mathematical Social Sciences*, *93*, 52–56.

Aĭzerman, M. A., & Aleskerov, F. T. o. (1995). *Theory of choice* (Vol. 38). North Holland.

Aleskerov, F., Bouyssou, D., & Monjardet, B. (2007). *Utility maximization, choice and preference* (Vol. 16). Springer Science & Business Media.

Arrow, K. J. (1959). Rational Choice Functions and Orderings. *Economica*, *26*(102), 121.

Barbera, S., & Pattanaik, P. K. (1986). Falmagne and the Rationalizability of Stochastic Choices in Terms of Random Orderings. *Econometrica*, *54*(3), 707.

Block, H. D., & Marschak, J. (1959). *Random orderings and stochastic theories of response* (tech. rep.). Cowles Foundation for Research in Economics, Yale University.

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

Cattaneo, M. D., Ma, X., Masatlioglu, Y., & Suleymanov, E. (2020). A random attention model. *Journal of Political Economy*, *128*(7), 2796–2836.

Chen, Y., Lu, Y., Wang, B., & Pan, Z. (2019). How do product recommendations affect impulse buying? An empirical study on WeChat social commerce. *Information and Management*, *56*(2).

Cherepanov, V., Feddersen, T., & Sandroni, A. (2013). Rationalization. *Theoretical Economics*, *8*(3), 775–800.

Chernoff, H. (1954). Rational Selection of Decision Functions. *Econometrica*, *22*(4), 422.

Court, D., Elzinga, D., Mulder, S., & Vetvik, O. J. (2009). The consumer decision journey. *McKinsey Quarterly.*

Echenique, F., & Saito, K. (2019). General Luce model. *Economic Theory*, *68*(4), 811–826.

Echenique, F., Saito, K., & Tserenjigmid, G. (2018). The perception-adjusted Luce model. *Mathematical Social Sciences*, *93*, 67–76.

Falmagne, J. C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, *18*(1), 52–72.

Fiorini, S. (2004). A short proof of a theorem of Falmagne. *Journal of Mathematical Psychology*, *48*(1), 80–82.

Fudenberg, D., Iijima, R., & Strzalecki, T. (2015). Stochastic Choice and Revealed Perturbed Utility. *Econometrica*, *83*(6), 2371–2409.

Gershoff, A. D., Broniarczyk, S. M., & West, P. M. (2001). Recommendation or Evaluation? Task Sensitivity in Information Source Selection. *Journal of Consumer Research*, *28*(3), 418–438.

Goodman, J. K., Broniarczyk, S. M., Griffin, J. G., & McAlister, L. (2013). Help or hinder? When recommendation signage expands consideration sets and heightens decision difficulty. *Journal of Consumer Psychology*, *23*(2), 165–174.

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

Guney, B. (2014). A theory of iterative choice in lists. *Journal of Mathematical Economics*, *53*, 26–32.

Gupta, P., & Harris, J. (2010). How e-WOM recommendations influence product consideration and quality of choice: A motivation to process information perspective. *Journal of Business Research*, *63*(9-10), 1041–1049.

Häubl, G., & Trifts, V. (2000). Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing Science*, *19*(1), 4–21.

Helmers, C., Krishnan, P., & Patnam, M. (2019). Attention and saliency on the internet: Evidence from an online recommendation system. *Journal of Economic Behavior and Organization*, *161*, 216–242.

Honka, E., Hortaçsu, A., & Vitorino, M. A. (2017). Advertising, consumer awareness, and choice: evidence from the U.S. banking industry. *RAND Journal of Economics*, *48*(3).

Horton, J. J. (2017). The Effects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment. *Journal of Labor Economics*, *35*(2), 345–385.

Hu, J., Liang, J., Kuang, Y., & Honavar, V. (2018). A user similarity-based Top-N recommendation approach for mobile in-application advertising. *Expert Systems with Applications*, *111*, 51–60.

Ishii, Y., Kovach, M., & Ülkü, L. (2021). A model of stochastic choice from lists. *Journal of Mathematical Economics*, 102509.

Kawaguchi, K., Uetake, K., & Watanabe, Y. (2021). Designing Context-Based Marketing: Product Recommendations Under Time Pressure. *Management Science*, *67*(9).

Kovach, M., & Tserenjigmid, G. (2019). Behavioral foundations of nested stochastic choice and nested logit. *Available at SSRN 3437165.*

Kovach, M., & Tserenjigmid, G. (2021). The Focal Luce Model. *American Economic Journal: Microeconomics*, *(forthcoming).*

Li, Q., Myaeng, S. H., & Kim, B. M. (2007). A probabilistic music recommender considering user opinions and audio features. *Information Processing and Management*, *43*(2), 473–487.

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, *29*(3), 458–468.

Lleras, J. S., Masatlioglu, Y., Nakajima, D., & Ozbay, E. Y. (2017). When more is less: Limited consideration. *Journal of Economic Theory*, *170*, 70–85.

Luce, R. D. (1959). *Individual choice behavior.* Wiley, New York.

Manzini, P., & Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, *82*(3), 1153–1176.

Manzini, P., Mariotti, M., & Ulku, L. (2021). Sequential Approval: A Model of "Likes", Paper Downloads and Other Forms of Click Behaviour.

Masatlioglu, Y., Nakajima, D., & Ozbay, E. Y. (2012). Revealed attention. *American Economic Review*, *102*(5), 2183–2205.

Mayzlin, D., & Shin, J. (2011). Uninformative advertising as an invitation to search. *Marketing Science*, *30*(4), 666–685.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 105–142.

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

Natenzon, P. (2019). Random choice and learning. *Journal of Political Economy*, *127*(1), 419–457.

Rowley, J. (2000). Product searching with shopping bots. *Internet Research*, *10*(3), 203–214.

Sela, A., Berger, J., & Liu, W. (2009). Variety, Vice, and Virtue: How Assortment Size Influences Option Choice. *Journal of Consumer Research*, *35*(6), 941–951.

Sen, A. K. (1971). Choice functions and revealed preference. *Review of Economic Studies*, *38*(3), 307–317.

Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, *80*(2), 159–169.

Tserenjigmid, G. (2021). The Order-Dependent Luce Model. *Management Science*.

Vijayasarathy, L. R., & Jones, J. M. (2000). Print and Internet catalog shopping: Assessing attitudes and intentions. *Internet Research*, *10*(3), 191–202.

West, P. M., & Broniarczyk, S. M. (1998). Integrating Multiple Opinions: the Role of Aspiration Level on Consumer Response to Critic Consensus. *Journal of Consumer Research*, *25*(1), 38–51.

Supplementary Materials:
Limited Consideration with Recommendation

1 Introduction

It is well-recognised that recommendation tends to alter behavior in the way that aligns with recommendation (Adomavicius et al. (2018), Goodman et al. (2013), Häubl and Trifts (2000), Horton (2017), Li et al. (2007), Litvin et al. (2008), Rowley (2000), Senecal and Nantel (2004), and Vijayasarathy and Jones (2000)). One obvious implication is that recommended alternative tends to be chosen more frequently when compared to no recommendation. We call this *positive recommendation effect*. In the full consideration perspective, we show that this effect has been embedded in every versions of the recommendation model.[29]

On the other hand, researchers have also found that sometimes recommendation can cause some other alternatives to be chosen more frequently, whereas these alternatives might not even be recommended in the first place (e.g. Helmers et al. (2019) and Kawaguchi et al. (2021)). This effect is known as the *spillover effect*. To have a better understanding of this effect, we need to take a deeper dive into the cognitive process of the decision maker, with the help of the conceptual foundation on consideration set mapping.

Theories on consideration set mapping, which can be traced back to Masatlioglu et al. (2012), make the simple assumption that people cannot pay attention to all products. This idea flourishes and sparks a plethora of literature to study different aspects and properties underneath the consideration process (e.g. Brady and Rehbeck (2016), Cattaneo et al. (2020), Cherepanov et al. (2013), Lleras et al. (2017), and Manzini and Mariotti (2014)). However, one cannot directly import knowledge from the limited consideration literature into the recommendation framework. After-all, the limited consideration literature were not intended to take into consideration the possible external impact of recommendation. In this framework, we are interested in finding out, given a set of recommended alternatives, what alternatives will finally fall into the consideration set of the individual. To achieve this, formally, we consider a correspondence $\Gamma : \mathcal{D} \to X$ as consideration correspondence. Given R, $\Gamma(R)$ is the set of alternatives the decision maker considers.

We do not impose any restriction on the consideration correspondence Γ. Firstly, recommended items might not be considered. i.e. we can have $\Gamma(R) \cap R \neq \Gamma(R)$. Secondly, un-

[29]Notice that Positive Marginal Recommendation Effect axiom implies immediately Positive Recommendation Effect. Since the parametric and deterministic are special cases of the general model, the Positive Recommendation Effect holds everywhere.

recommended items might be considered. i.e. we can have $\Gamma(R) \cap R^c \neq \emptyset$. While the former captures the cases where the decision maker ignores some of the recommended alternative, the latter can cover *attentional* spillover effect where the recommendation attract attention towards to non-recommended alternatives. Hence, in terms of theoretical construct, Γ is more general than the usual consideration set mapping in the limited consideration literature.

2 Deterministic

Note that, however, this representation alone does not help us to make further inference or prediction about decision maker's behavior or consideration. In fact, any choice rule can be rationalized under this framework, where the decision maker only considers the chosen alternative. We put this observation into the following remarks.

Remark 4. Any deterministic choice rule has a recommendation representation under limited consideration.

Definition 7. A deterministic choice rule c has a recommendation representation under limited consideration on $\mathcal{D}$ if there exist a preference $\succ$ and consideration correspondence Γ such that

$$c(R) = \max(\Gamma(R), \succ) := c_{(\Gamma,\succ)}(R)$$

for all $R \in \mathcal{D}$.

Therefore, we need to put meaningful behavioral restriction into the unobserved consideration. In the following, we consider two non-parametric conditions.

Irrelevance of Recommending Considered Alternative (IRC)

Note that in the recommendation domain, it seems reasonable to impose the following condition on consideration mapping. The property dictates that recommending already considered alternative does not affect the consideration set.

$$x \in \Gamma(R) \text{ implies } \Gamma(R \cup x) = \Gamma(R)$$

We call this condition as Irrelevance of Recommending Considered Alternative (IRC). This condition is saying "what can be considered while non-recommended, if recommended, would not changed my consideration". Another way of stating this condition is:

$$\Gamma(R \cup x) \neq \Gamma(R) \text{ implies } x \notin \Gamma(R)$$

In this way, it says, "if my consideration changes with an additional recommended item, this item was not considered before". Note that this condition is vacuous if we assume $\Gamma(R) \subset R$. One of the implications of this property is

$$\text{If } R \subset \Gamma(\emptyset) \text{ then } \Gamma(R) = \Gamma(\emptyset).$$

There are several examples of this consideration set mapping under recommendations.

Example 1 (Thorough consideration and/or ignoring recommendation)**.** These two examples is under the umbrella of "constant consideration set". On one extreme, the decision maker considers everything before making a decision ($\Gamma(R) = X$ for all R); on the other extreme, the decision maker ignore recommendation altogether and focuses attention on one particular set ($\Gamma(R) = A$ for all R). It is clear that in either case an recommending an considered item will not change the decision maker's consideration. ⌟

Example 2 (Consider only the recommended)**.** The decision maker might only consider the recommended item. In this case, recommending an considered item does not even change the recommendation itself ($\Gamma(R) = R \cup \Gamma(\emptyset)$ for all R). ⌟

Example 3 (Spill-Over consideration of similar item)**.** A recommendation might activate consideration of the similar items for the decision maker. For example, Coca-cola and Pepsi might always come together into decision maker's consideration set even if only one of them is recommended. Therefore, if both of them are recommended, it will not change the consideration set. Formally, let $\{C_i\}$ be a partition of X, where each C_i correspondences category i. The decision maker only consider the first category in the absence of any recommendation ($\Gamma(\emptyset) = C_1$). The decision maker considers all the categories as long as some alternative is recommended in that category ($\Gamma(R)$ is the union of all categories intersecting with R and C_1). ⌟

It turns out this condition is fully characterized by one axiom. We call this axiom as Dynamic Consistency. After initial choice, if the decision maker is allowed to make choice again with the same recommendation including her choice, she will make the same choice.

Axiom 14 (Dynamic Consistency)**.** $c(R) = c(R \cup c(R))$.

Again this axiom trivially holds in the standard choice environment where $c(R) \in R$.

Theorem 6. *c satisfies Axiom 14 if and only if c has a recommendation representation under limited consideration with IRC.*

Proof. Necessity is obvious. For sufficiency, let $\Gamma(R)$ be $\{c(R)\}$ for all $R \in \mathcal{D}$. Take any preference $\succ$, then $(\Gamma, \succ)$ immediately explains the data. Then, to check IRC, we let $x \in \Gamma(R)$. Note that by construction, $\Gamma(R) = \{c(R)\}$ and $\Gamma(R \cup c(R)) = \{c(R \cup c(R))\}$. By Axiom 14, we have $c(R \cup c(R)) = c(R)$. Hence, the proof is complete. ■

The proof of the theorem makes it clear that this axiom is clearly the basic assumption in this environment. However, this restriction does not allow us to make any inference about the decision-maker's preferences. To be able to make inference, we will impose a property used in the consideration set literature, called Attention Filter (AF). That is,

$$x \notin \Gamma(R \cup x) \text{ implies } \Gamma(R \cup x) = \Gamma(R)$$

or, equivalently,

$$\Gamma(R \cup x) \neq \Gamma(R) \text{ implies } x \in \Gamma(R \cup x)$$

Notice that this condition capture almost the same intuition from the unawareness argument when availability set varies. It turns out that, if we want to jointly characterize these two conditions for one consideration mapping, we only need one additional axiom from Masatlioglu et al. (2012). Firstly, we let

$$yPx \text{ if there exists } R \text{ such that } y = c(R \cup x) \neq c(R).$$

This axiom is simply that P must be acyclic, which appears in Masatlioglu et al. (2012).

Axiom 15. P has no cycle.

Theorem 7. c satisfies Axiom 14 and 15 if and only if c has a recommendation representation under limited consideration with IRC and AF.

Proof. Necessity is immediate. For sufficiency, we let yPx if there exists R such that $y = c(R \cup x) \neq c(R)$. Let $\succ$ be a completion of the transitive closure of P. Then, we let

$$\Gamma(R) := \{c(R)\} \cup (\{t : c(R) \succ t\} \cap R)$$

$(\Gamma, \succ)$ trivially explains the data. We need to show that Γ satisfies IRC and AF. To show AF, assume $\Gamma(R \cup x) \neq \Gamma(R)$. We now show that x must be in $\Gamma(R \cup x)$. If $c(R \cup x) \neq c(R)$, then we must have $c(R \cup x)Px$ $(c(R \cup x) \succ x)$. Hence, $x \in \Gamma(R \cup x)$. If $c(R \cup x) = c(R)$ then, by the definition of Γ, the only alternative belonging the set $\Gamma(R \cup x) \setminus \Gamma(R)$ must be x. Hence, $x \in \Gamma(R \cup x)$.

To show IRC, assume $x \in \Gamma(R)$. By the definition of Γ, i) $c(R) \succ x$ and $x \in R$, or ii) $c(R) = x$. The later and Axiom 14 imply that $c(R) = c(R \cup x)$. Hence, $\Gamma(R \cup x) = \Gamma(R)$. The former case implies that we cannot have $xPc(R)$. Hence, we must have $c(R \cup x) = c(R)$. Since $x \in R$, $\Gamma(R \cup x) = \Gamma(R)$ by the definition of Γ. This completes the proof. ■

3 Stochastic Choice

Above our limited attention model was deterministic, which potentially limits the applicability of the model when the naturally occurring data is stochastic. In this section, we consider cases where the decision maker pays attention to the different subset of options every time she is confronted with the same recommendations. We call a mapping $\mu : \mathcal{X} \times \mathcal{X} \to [0, 1]$ attention rule if $\sum_{T \in \mathcal{X}} \mu(T|R) = 1$ for all $R \in \mathcal{X}$.

Definition 8. A choice rule ρ has a recommendation representation under limited consideration on $\mathcal{X}$ if there exist a preference $\succ$ and an attention rule μ such that

$$\rho(a|R) = \sum_{T \in \mathcal{X}} \mathbb{1}(a \text{ is } \succ\text{-best in } T) \cdot \mu(T|R)$$

for all $a \in R$ and $R \in \mathcal{X}$.

Positive Recommendation and Attention Diversion

It turns out that to generalize IRC into the stochastic domain we only need the following condition,

$$\text{For any } x \in T,\ \mu(T|R) \leq \mu(T|R \cup x)$$

We call this condition Positive Recommendation in Consideration (PRC). On the other hand, from Cattaneo et al. (2020), we know that AF can be generalized in the following condition:

$$\text{For any } x \notin T,\ \mu(T|R) \geq \mu(T|R \cup x)$$

which we will call Attention Diversion. In the following, we demonstrate the two conditions in Table 3. We label the inequalities driven by Attention Diversion in Red and the inequalities driven by Positive Recommendation in Attention in Blue.

We can see that $\mathbin{\rotatebox[origin=c]{-90}{$\leq$}}$ reflects the idea positive recommendation effect in attention, where every set which contains the newly recommended item will get more attention. This condition permits us to also capture *spillover effect* discussed in the literature. For example, you only consider the drinks on the same row as the recommended item on a vending machine, where items as the

T	$\{a\}$	$\{b\}$	$\{c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$	$\{a,b,c\}$
$\mu(.\|\{a,b,c\})$							
	$\rotatebox{90}{$\geqslant$}$	$\rotatebox{90}{$\geqslant$}$	$\rotatebox{90}{$\leqslant$}$	$\rotatebox{90}{$\geqslant$}$	$\rotatebox{90}{$\leqslant$}$	$\rotatebox{90}{$\leqslant$}$	$\rotatebox{90}{$\leqslant$}$
$\mu(.\|\{a,b\})$							
	$\rotatebox{90}{$\geqslant$}$	$\rotatebox{90}{$\leqslant$}$	$\rotatebox{90}{$\geqslant*	$\rotatebox{90}{$\leqslant$}$	$\rotatebox{90}{$\geqslant*	$\rotatebox{90}{$\leqslant$}$	$\rotatebox{90}{$\leqslant$}$
$\mu(.\|\{a\})$							
	$\rotatebox{90}{$\leqslant$}$	$\rotatebox{90}{$\geqslant*	$\rotatebox{90}{$\geqslant*	$\rotatebox{90}{$\leqslant$}$	$\rotatebox{90}{$\leqslant$}$	$\rotatebox{90}{$\geqslant*	$\rotatebox{90}{$\leqslant$}$
$\mu(.\|\emptyset)$							

Table 3. An illustration of two conditions. Note that $\rotatebox{90}{$\geqslant* is not captured in the standard domain.

recommended item are having spilled over attention. On the other hand, $\rotatebox{90}{$\geqslant$}$ captures an idea of *crowding out effect*: If the set does not contain the new recommended alternative, it is meant to get less attention.

To fully characterize these two conditions, it turns out that we also only need two simple condition. Firstly, the following condition is simple, which says that an item must be weakly chosen more if it is recommended.

Axiom 16 (Positive Recommendation)**.** $\rho(x|R) \leq \rho(x|R \cup x)$.

Note that this axiom alone can fully characterize the model with only the IRC. Since we are also assuming Attention Diversion, we must also need an additional axiom. To achieve this, similar to deterministic case, we let

$$yPx \text{ if there exists } R \text{ such that } \rho(y|R) < \rho(y|R \cup x)$$

Similar to the deterministic case, we only need P to be acyclic, from Cattaneo et al. (2020).

Axiom 17. P has no cycle.

Theorem 8. ρ satisfies Axiom 16 and 17 if and only if ρ has a recommendation representation under limited consideration with Positive Recommendation in Consideration and Attention Diversion.

Proof. Necessity is immediate. For sufficiency, we let yPx if there exists R such that $\rho(y|R) < \rho(y|R \cup x)$. Let $\succ$ be a completion of the transitive closure of P. Define $L_\succ(x)$ be the lower contour set of x according to $\succ$ and we set

$$\mu(T|R) := \begin{cases} \rho(x|R) & \text{if there exists } x \in T \text{ s.t. } T = x \cup (L_\succ(x) \cap R) \\ 0 & \text{otherwise} \end{cases}$$

We first check whether it is well-defined. Suppose it is not, i.e. there exists $x, x' \in T$ such that $x' \neq x$ and $T = x \cup (L_\succ(x) \cap R) = x' \cup (L_\succ(x') \cap R)$. Then we must have $x \in L_\succ(x') \cap R$ and $x' \in L_\succ(x) \cap R$, which imply $x \succ x'$ and $x' \succ x$. It is a contradiction.

$(\mu, \succ)$ clearly explains the data, hence the representation holds. We now show μ satisfies our two properties. We first check Attention Diversion. We let $x \notin T$ and we want to show $\mu(T|R) \geq \mu(T|R \cup x)$. If $\mu(T|R \cup x) = 0$, then the condition holds trivially. Suppose not, i.e., $\mu(T|R \cup x) > 0$. By definition, there exists $y \in T$ such that $T = y \cup (L_{\succ}(y) \cap (R \cup x))$ and $\mu(T|R \cup x) := \rho(y|R \cup x)$. Then, since $x \notin T$, it means that $x \notin L_{\succ}(y)$ and $T = y \cup (L_{\succ}(y) \cap R)$. Hence, by definition, it must be that $\mu(T|R) := \rho(y|R)$. Since $x \succ y$, (y, x) does not belong to P, which implies $\rho(y|R) \geq \rho(y|R \cup x)$. Hence, $\mu(T|R) \geq \mu(T|R \cup x)$.

To verify PRC, let $x \in T$. We want to show $\mu(T|R) \leq \mu(T|R \cup x)$. If $x \in R$, there is nothing to prove, hence assume $x \notin R$. If $\mu(T|R) = 0$, then the condition holds trivially. Suppose $\mu(T|R) > 0$. By definition, there exists $y \in T$ such that $T = y \cup (L_{\succ}(y) \cap R))$ and $\mu(T|R) := \rho(y|R)$. Note that since $x \in T$ but $x \notin R$, we must have $x = y$. Therefore, by definition, it must be that $\mu(T|R \cup x) := \rho(x|R \cup x)$. By Axiom 16, we know that $\rho(x|R \cup x) \geq \rho(x|R)$. The proof is complete. ■