Decision Making with Recommendation

Paul H.Y. Cheung* Yusufcan Masatlioglu†

April 5, 2023

Abstract

Recommendations play an undeniable role in decision-making. The empirical literature argues that recommendation can influence demand through two distinct channels: i) by enlarging awareness (attention channel), or ii) by altering preferences (utility channel). In this paper, we develop a framework to study these two channels. We illustrate when and how one can distinguish through which channel the recommendations affect choices. We offer both deterministic and probabilistic models. While deterministic models aim to identify the basic observable behavioral differences between these two channels, our probabilistic models are suitable for econometric estimation, which is crucial for studying aggregate behavior used in empirical work. Our parametric models offer unique identification under minimal data requirements. This enables us to make out-of-sample predictions for counterfactual analysis for policy design purposes. In addition, we offer simple and intuitive behavioral postulates characterizing each model so that one can test our models.

Keywords: Recommendation, Revealed Preference, Attention

1 Introduction

Recommendation is one of the key determinants in decision-making nowadays. For instance, we constantly rely on recommendations from our friends, consumer reports, and mass media when selecting a movie to see; a book to read; a car to buy; or a school to send our children.[1] As digital platforms and their offerings continue to grow, consumers face an unprecedented number of online product and service options. As a result, recommendations have become indispensable for online shopping. Many internet sites such as Amazon, Netflix, Spotify, Tripadvisor, and Facebook incorporate recommendation tools to help customers with the burden of choice. "*Amazon's Choice*" of Amazon, "*Superhost*" of Airbnb, and "*Top 10*" of Netflix are just a few examples of these recommendation tools.

*Naveen Jindal School of Management, University of Texas at Dallas. Email: paul.cheung@utdallas.edu

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

[1] Some online service websites, such as Angi, HomeAdvisor, Houzz, Thumbtack recommend services for consumers' specific projects, which shows that people are willing to pay for recommendations.

Individual choices are directly influenced by what is recommended to them.[2] The evidence on recommendations influencing choice behavior is conclusive across a wide spectrum of economic activities (e.g., for the labor 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, there is no theoretical framework to understand and study how recommendations affect choices. In this paper, we are interested in how recommendations affect the choice process. We use revealed preference techniques to shed light on the underlying choice process due to recommendation. We aim to investigate the empirical implications of *aggregate* behavior of the models. More importantly, viewing choice behavior as a result of an unobservable cognitive process, we identify the primitives of the model from observed choices.

We first propose a new framework where recommendation takes the simplest form: each decision problem represents a different set of recommendations. Here, the analyst does not observe any qualitative information about recommended products. There are two main reasons why simplifying the amount of knowledge about the nature of recommendations. First, in real life, recommendation signs such as "Best Seller," "Award Winner," and "Editor's Pick" are commonplace. Secondly, the analyst might not be aware of the way in which the decision-maker perceives the qualitative recommendation information. We assume that across decision problems, different products may be recommended, but the source and nature of recommendations would remain the same. As a result, we will be able to determine the decision-maker's subjective evaluation of the recommended products. Further, we can compare the impact of recommendations from different sources, which can be interpreted as a measure of agents' trust in different sources of recommendations.

Several studies in this literature have suggested that increased sales of substitutes can be attributed either to preference or attention. While Chen et al. (2019), Goodman et al. (2013), Gupta and Harris (2010), and Mayzlin and Shin (2011) claim recommendations' primary role is to inform customers about the existence and the availability of the products, Adomavicius et al. (2013, 2018), Bairathi et al. (2022), Cosley et al. (2003), and Kawaguchi et al. (2021) state that recommendations might also affect choices by influencing consumers' preferences. One of the puzzles in this literature is that it is unclear how to distinguish between these two channels. In this paper, we ask whether it is possible to determine whether a recommendation operates through the attention channel or the preference channel from observed choices. Distinguishing these two channels is essential for policymakers/firms designing effective recommendations, especially when contextual factors may influence the effects of recommendations on preference and attention in a distinct way.

Section 2 introduces our new framework. Attention and preferences play a distinct role in influencing choices in these examples. In the attention channel, consumers are unaware of some

[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.

of the feasible products, and recommending a particular product makes the customer aware of the product. In the preference channel, recommendations affect the perceived valuation of a product even though she is aware of the product. As an illustration of the richness of our framework, we provide several examples of recommendation models that capture different aspects of the recommendation process.

To shed light on the puzzle discussed above, Section 3 focuses on the baseline models introduced in Section 2. The reasons are two folded. First, each consumer's behavior is intended to be simple enough to study the aggregate behavior of heterogeneous consumers in a tractable way. Second, we depart from the standard model minimally with our baseline models. As a result of its simplicity, the classical choice model is one of the most widely applied. Tractability and simplicity are two important factors for application and estimation purposes. For this reason, we want to stay as close as possible to the rational framework to assess how much explanatory power each channel can provide.

Variations in the set of feasible alternatives are the source of identification results in many papers using the revealed preference technique. Another novel feature of our model is that we only vary the set of recommendations while the set of feasible alternatives is always fixed.[3] While our main aim is to study aggregate data, we start our analysis with the deterministic choice data to build a better understanding. As we said, our deterministic model is intended to be simple so that we can study the aggregate behavior of such consumers in a tractable way. We then study the behavioral implications of our models with the probabilistic choice data. In either environment, we propose two models in which recommendation operates either through attention or preferences. The model capturing the attention channel is called *Recommendation via Attention (RA)* and the model operating through the preference channel is called *Recommendation via Preference (RP)*. We then provide behavioral characterizations of these models to determine whether they are behaviorally distinct from each other.

Our baseline model for the preference channel assumes that the recommendations only affect choices by increasing the valuation of the product. Being recommended only increases the utility of the recommended alternative (positive recommendation).[4] In the deterministic RP model, a decision maker (henceforth, DM) is aware of all of the products, but the utility of an alternative may be affected by recommendation. As in classical choice theory, the DM picks the alternative with the highest utility.

In the baseline model for the attention channel, we consider a DM who only pays attention to a subset of the available alternatives. We assume that the recommendations enlarge the awareness set of the DM. Formally, each DM can be identified with a pre-recommendation choice and a fixed preference ordering. The pre-recommendation choice represents the most preferred alternative in the DM's awareness set according to her preferences before receiving any recommendation. It is natural that the pre-recommendation choice is picked if there is no rec-

[3] We could allow both types of variations: recommendation and feasible set. Such a theory demands much richer data than the standard choice. Our framework is flexible enough to use such data if it is available.

[4] This formulation eliminates negative recommendation where the utility of an alternative goes down. Nevertheless, our framework is rich enough to study negative recommendations as well.

ommendation since there are no changes in the DM's decision environment. If some alternatives are recommended, the DM picks the preferred option between the recommended alternatives and her pre-recommendation choice according to her preferences. Therefore, recommendations cannot hurt the DM since she can always ignore the recommendation if it is not preferred to her pre-recommendation choice.

Our first result identifies the simple postulate underneath these two models, which is the Independence of Irrelevant Recommended Alternatives (IIRA). IIRA states that removing an alternative from the recommendation set does not affect the choice if the removed item was not initially chosen. This postulate is normatively appealing since it mimics the Independence of Irrelevant Alternative (IIA), which is the cornerstone of classical choice theory.[5] One can show that any RP model must satisfy IIRA. Moreover, any choice behavior compatible with IIRA can be represented by the RP model. Hence, IIRA is the only behavioral implication of the RP model.

While IIRA is a necessary condition for the RA model, it is not sufficient. The discrepancy in the behavioral pattern between the RA and RP models can be captured in another postulate - the Sandwich Property. If two nested recommendations result in the same choice, then DM must choose the same option when she receives any recommendation between these recommendations. We show that IIRA and the Sandwich Property fully characterize the RA model. These results imply that when the Sandwich Property fails, the attention channel cannot explain choices. It is surprising since in the standard domain, once choices satisfy IIA (the counterpart of IIRA), we cannot distinguish whether choices can be attributed to attention or preference.

In many examples, the data is given as the aggregate behavior, which can be captured by probabilistic choice. Therefore, we extend our intuition of the deterministic model to this environment. The probabilistic choice can be interpreted in two ways: i) different individuals make choices (interpersonal), or ii) the same individual makes choices under potentially different circumstances (intrapersonal). As in the deterministic environment, we ask whether two channels - attention or preference - can be distinguished in the probabilistic environment. In Section 4.1, we introduce two parametric models of recommendations capturing the two channels based on intuition from the deterministic models. Each model has a finite set of alternative-specific parameters. In both models, we incorporate the randomness in preferences through the idea of the Luce model (also known as Multinomial Logit).[6] Therefore, we call them RA-Luce and RP-Luce models.

The RP-Luce model captures the preference channel by boosting utility when some alternatives are recommended. Firstly, when items are not recommended, there is a crude measure of the baseline utility value represented by u. When items are recommended, the baseline utility is weakly increasing, which is captured by an alternative-specific recommendation value $v(x)$, such that an item now has the utility of $u(x) + v(x)$ instead of $u(x)$. In the deterministic model,

[5]IIA states that choices should not depend on the presence of unchosen alternatives.

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

the decision maker compares every product before making a decision. The RP-Luce maintains this feature of the model such that the choice probability of an alternative will be the share of the utility weight it has relative to the total utility weight in the choice set, where some utility weights are increased due to recommendation. Formally, the choice probability of x given recommendation R under the preference channel is given by

$$\rho^{RP}(x,R) = \begin{cases} \dfrac{u(x)+v(x)}{\sum\limits_{y\in R}[u(y)+v(y)] + \sum\limits_{z\notin R} u(z)} & \text{if } x \in R \\ \dfrac{u(x)}{\sum\limits_{y\in R}[u(y)+v(y)] + \sum\limits_{z\notin R} u(z)} & \text{otherwise} \end{cases}$$

We show that two simple behavioral postulates can fully characterize this model. The first postulate, called Strong Luce-IIA, mimics the famous Luce-IIA in our framework. Indeed, Strong Luce-IIA implies that the choice ratio between recommended items stays constant (Recommended Luce-IIA). This postulate also implies that the choice ratio between non-recommended items also stays constant. The second postulate is R-Regularity, which says that a non-recommended alternative must be chosen more often when the recommendation set gets smaller. Our characterization results do not require the knowledge of full data. We show that if the choice data includes all recommendation sets with size two or less, the characterization holds. In addition, we can uniquely pin down recommendation boosts and the utility weights up to a scaling factor.

The RA-Luce model assumes that alternatives have a fixed utility value, denoted by u. Besides the randomness in preference, RA-Luce also captures the randomness in pre-recommendation choice, which is probabilistic on the aggregate level as people do not necessarily make the same choices when they do not receive any recommendation. To capture this randomness, we include an additional parameter, $d(x) \geq 0$, which represents the likelihood of x being the pre-recommendation choice. Therefore, $\sum_{x\in X} d(x) = 1$. Given a pre-recommendation choice, not recommended alternatives are chosen with zero probability except for the pre-recommendation choice. The probability 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, the odds of selecting a not recommended alternative, $x \notin R$, is the probability of being the pre-recommendation choice times the choice probability when it is the pre-recommendation choice. Formally, the choice probability of x given recommendation R under the attention channel is given by

$$\rho^{RA}(x,R) = \begin{cases} \sum\limits_{z\in X} d(z) \dfrac{u(x)}{\sum\limits_{y\in R\cup z} u(y)} & \text{if } x \in R \\ d(x) \dfrac{u(x)}{\sum\limits_{y\in R\cup x} u(y)} & \text{otherwise} \end{cases}$$

We show that two axioms can fully characterize this model. The first axiom is the Recommended Luce-IIA, in which the choice ratio between two recommended items stays constant across recommendation sets. Due to the characterizations in the deterministic model, one might suspect that a similar Luce-IIA property applies to not recommended items. It turns out that this property does not hold for non-recommended alternatives. Instead, not recommended items satisfy another well-known property called conditional choice, which states that $\rho(a, R)\rho(R \cup a, X)$ is independent of R. In the classical domain, this axiom is equivalent to Luce IIA. However, in our framework, this axiom does not hold for recommended alternatives; it only holds for not recommended alternatives.[7] We also demonstrate how one is able to identify the parameters of the model from data. In particular, we show that we can uniquely identify the parameters of the RA-Luce model with only two data points.[8] This is helpful for real-life applications since it offers a frugal way to make an out-of-sample prediction for the effect of recommending different sets of items.

Given two characterization theorems, we identify when one can distinguish between the attention channel and the preference channel. In the deterministic framework, any behavior explained by the attention channel can be accommodated by the preference channel. This no longer holds in the probabilistic world. Our characterization results show that RA-Luce and RP-Luce are independent. In other words, there are choice behaviors that are only explained by RA-Luce but not RP-Luce, and vice versa. Therefore, these two channels can be distinguished by observing richer data and parametric models.

While parametric models are analytically and computationally tractable but do not account for the possibility that tastes differ randomly between individuals. We also investigate the behavioral implications of non-parametric versions of our baseline models of recommendation. As in the standard Random Utility Model (RUM), each type is assumed to make deterministic choices with the corresponding model. We call these models as RP-RUM and RA-RUM capturing both attention and preference channels.

We identify all the behavioral implications of RA-RUM and RP-RUM. The empirical content of the standard RUM is investigated by Falmagne (1978), Barbera and Pattanaik (1986) and McFadden and Richter (1990) who show that the non-negativity of the Block-Marshak polynomials (BM) is equivalent to the choice data being generated by RUM. In our environment, the classical BM can be defined for both recommended and non-recommended alternatives. Our first axiom is the non-negativity of these BM polynomials. This assumption fully characterizes RP-RUM. For RA-RUM, however, the non-negativity of BMs is necessary but not sufficient. Since being recommended has a greater marginal effect in the RA-Luce model, the BM of recommended alternative is always greater than the corresponding BM of non-recommendation, which we call Positive Marginal Recommendation. It turns out these two properties fully characterize RA-RUM.

[7] A caveat is that we need a stronger axiom when there is limited data, which will be elaborated in the corresponding section.

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

Literature Review

In all of our models, we utilize the idea that modelers can observe consumers' choice data as a function of their recommendation set. Since choices are allowed to be outside of a recommendation set, our choice function is conceptually and mathematically different from the classical choice function. Therefore, our paper naturally separates itself from other theories in the standard domain in terms of modeling strategy. Moreover, it enables us to take a fresh perspective on choice data with some of our traditional economic intuition.

We would like to mention a recent paper by Ke et al. (2021). While the focus of their paper is distinct from ours, their primitive enjoys a similar feature to our choice function. They propose a belief-updating model where a DM receives information from an unknown source. Their model differs from other updating rules where the posterior always belongs to the information set. Similar to our primitive, they allow posterior being outside of the information set.

The classical Luce model attracts a variety of scholarly attention to developing different generalizations of it (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, 2022a, 2022b; Tserenjigmid, 2021). All of these models involve different relaxations of the Luce IIA axiom. In our RA-Luce and RP-Luce models, we employ the Luce IIA axiom for the recommended alternatives to provide a simple applicable, and tractable parametric model. 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 rates (Manzini et al. (2021)). Natenzon (2019) and Guney et al. (2018) study how non-choosable phantom options affect choices. In a similar spirit to our model, these lines of research are also augmenting the standard choice environment to enhance our understanding of human behavior.

2 Framework

Let X be a set of finite alternatives (e.g., all documentaries available at Netflix or all 65 Inch Smart TVs sold at Amazon). We define a recommendation set, denoted by $R \subset X$, as the set of recommended products for the decision-maker. This is the simplest form of recommendation where the recommendation set does not include any qualitative information about recommended products. There are two main reasons for this assumption. First, non-price recommendation signs, such as "Best Seller," "Award Winner," and "Editor's pick" are commonplace in real life. Second, the outside observer might not know how the decision-maker perceives different qualitative recommendation information. For example, the decision-maker does not distinguish between 4-star rated and 5-star rated products as long as they are both recommended. Moreover, the source of the recommendation can also affect the influence of the recommendation. While we vary the set of recommended alternatives, we assume that the source and the nature of the recommendation are the same. This will allow us to identify the decision-maker's subjective evaluations of the recommended products for a fixed source. In addition, we will be able to compare the effect of recommendations across different sources.

In this framework, a decision problem describes not only which alternatives are available, denoted by S, but also which alternatives are recommended, denoted by R. Therefore, (R, S) constitutes a decision problem where $R, S \subset X$ and $S \neq \emptyset$. We define a general choice function c where $c(R, S) \in S$ and $R \in X$ where S represents the set of feasible alternatives and R is the recommended set. There are two types of variations in the general model: recommendation and feasible set. In other words, this framework is more "data hungry" than the standard choice since the analyst must observe choices from all decision problems.

In this paper, to eliminate this data requirement, we assume that the set of available alternatives is fixed. Hence there is no variation in terms of the availability of alternatives in our model. After this restriction, any subset of X could constitute a decision problem, including the empty set which corresponds to not receiving any recommendations. While recommendations can influence choices, it does not constrain them. To capture this, we define a choice rule c as a function of the 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 differs 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. Additionally, we allow c to be observable only for some recommendation sets but not others. This assumption aims to capture some real-world environments in which the collection of the recommended sets 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 instance, $\mathcal{D}$ can include all recommendation sets with the size of 1. The following definition captures the choice rule under this framework. Depending on the results we provide in this paper, we highlight the necessary requirements in the domain.

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

Notice that if the DM always maximizes according to true underlying preference with full consideration, recommendations should not affect choices even if the DM receives different recommendations. In other words, $c(R)$ is a constant function. However, as discussed in the introduction, there is plentiful evidence that recommendation indeed affects choice. In particular, there are two different channels through which recommendations can affect choice behavior: attention and utility. Here, we provide several models of recommendation. These examples illustrate that our framework is flexible enough to study models of recommendation capturing different aspects of the recommendation process.

2.1 Examples

Our examples are categorized into two classes: preference channel and attention channel. The reasons are two folded. Firstly, oftentimes attention and preference can be intertwined in explaining choice data. For example, in the attention literature, if choices satisfy IIA, there is no room to distinguish whether DM is choosing her most preferred alternative or only pays attention to the chosen alternative. We would like to know whether we can distinguish between choice driven by attention and choice driven by the utility from choice data involving recommendations. Secondly, given that the two different channels can be relatively more prominent than the other in different circumstances, studying the different models can also help us identify better descriptive models for explaining behavior in different contexts. All of these require us to first have a deeper understanding of the behavioral implications of each of the models.

We first discuss models operating through the preference channel. The first one is the baseline model of the preference channel since, as in the classical choice theory, each alternative has a subjective value, and the decision-maker selects the alternative with the highest value. The subjective value is independent of the other available alternatives and recommendations. This assumption is relaxed in the context-dependent model where the recommended product's value depends on the recommended alternatives. Finally, the last example allows the valuation of the non-recommended product also be affected by the recommended alternatives.

Preference Channel

- Baseline Model: The recommendations only affect choices by increasing the valuation of the product. Hence, being recommended only increases the relative ranking of the recommended alternative in the decision maker's preferences (positive recommendation). As in the standard theory, there is "no context effect" in the sense that the valuation of the product only depends on whether it is recommended or not. Specifically, it does not depend on the other recommended alternatives. Assume that the decision maker has in mind two functions, u and v, from X to $\mathbb{R}$. When the product x is recommended ($x \in R$), she assigns the utility of $u(x) + v(x)$ to x; otherwise, $u(x)$ is the value of x. She chooses the alternative maximizing utility. The positive recommendation is modeled by assuming $v(x)$ to be non-negative. Formally, the maximization problem can be written as follow:

$$\max_{x \in X} U_R(x) \quad \text{where} \quad U_R(y) = \begin{cases} u(y) + v(y) & \text{if } y \in R \\ u(y) & \text{if } y \notin R \end{cases} \tag{RP}$$

  We denote choices as $c_{(u,v)}(R) = \text{argmax}_{x \in X} \; U_R(x)$.[9] We refer to this model as the *model of recommendation via preference* (henceforth, RP) when $v(x) \geq 0$. We say a choice rule c has an RP representation on $\mathcal{D}$ if there exists (u, v) with $v \geq 0$ such that $c = c_{(u,v)}$.

- Context-Dependent Model: This model extends the baseline model by allowing the valuation of the recommended product to depend on the other recommended alternatives. For example, given R, the decision-maker has a total utility stock $V(R)$, which is proportionally distributed among recommended alternatives. She assigns the utility of $u(x) + \frac{v(x)}{\sum_{y \in R} v(y)} V(R)$ if $x \in R$; otherwise assigns the utility of $u(x)$. Note that the valuations

[9]Since our domain is finite and choices are unique, we assume U_R is a one-to-one function.

of non-recommended alternatives are not changed. Again, the highest utility alternative will be chosen. If $V(R)$ is fixed, the effect of the recommendation decreases with the size of the recommendation set. Note that if $V(R)$ equals to $\sum_{y \in R} v(y)$, this model becomes equivalent to the baseline model. More generally, the objective function can be written as follows:

$$U_R(x) = \begin{cases} u(x) + v(x, R) & \text{if } x \in R \\ u(x) & \text{if } x \notin R \end{cases}$$

where v is a function from $X \times 2^X$ to $\mathbb{R}$.

- Category-Dependent Model: As opposed to the first two models, here we consider a model where the valuation of the non-recommended product can change. For example, recommending only Tesla might increase the valuation of the Porsche Taycan. This model captures the spillover effect of recommendations due to valuation. Formally, the decision-maker groups alternatives into distinct categories $\{C_1, \ldots, C_n\}$ that partition the set of all alternatives X. Given R, she assigns utility of $u(x) + v(x)$ to x if $x \in C_i$ and $C_i \cap R \neq \emptyset$; otherwise assigns utility of $u(x)$. More generally, the objective function can be written as follows:

$$U_R(x) = \begin{cases} u(x) + v(x) & \text{if } x \in R \\ u(x) + w(x) & \text{if } x \in \bigcup\limits_{C_i \cap R \neq \emptyset} C_i \setminus R \\ u(x) & \text{otherwise} \end{cases}$$

where w is a function from X to $\mathbb{R}$. In the above example, $v = w$.

Similar to the preference channel, we provide three classes of models operating through the attention channel. The decision-maker has a well-defined utility function, u, in all these models. In other words, the subjective value is independent of the other available alternatives and recommendations. The recommendation only affects the consideration set. The first one is called the baseline model of the attention channel since, as in the classical choice theory, the decision-maker is both willing to and capable of considering all the recommended alternatives. This assumption is relaxed in the limited attention model, where the decision maker might not consider all recommended products. Finally, the last example allows for enlarging the awareness of the decision-maker. Recommending certain products triggers the consideration of non-recommended products.

Attention Channel

- Baseline Model: The recommendations only affect choices by enlarging the awareness of the decision-maker. The decision-maker has a stable preference ranking. Without any recommendation, the decision-maker chooses one alternative, denoted by a, which might not be the best option according to her preference due to limited awareness. Given R, her consideration set, denoted by $\Gamma(R) \subseteq X$, becomes $R \cup a$. She then picks the best option among $R \cup a$ according to her preferences. As in the classical rational choice model, the decision-maker can and will consider all the recommended alternatives. Formally, the maximization problem can be written as follow:

$$\max_{x \in \Gamma(R)} u(x) \quad \text{where } \Gamma(R) = R \cup a \tag{RA}$$

We denote the maximizer of above maximization problem as $c_{(a,u)}$ given (a, u). We refer to this model as the *model of recommendation via attention* (henceforth, RA). We also say a choice rule c has an RA representation on $\mathcal{D}$ if there exists a pair (a, u) such that $c = c_{(a,u)}$.

- Limited Attention Model: This model allows the decision-maker exhibits limited attention. The decision-maker has, again, a stable preference ranking, and a is the choice in the absence of a recommendation. Given R, her consideration set, $\Gamma(R)$, is a subset of R including a. She then picks the best option in her consideration set. Here, the decision-maker might not consider all the recommended alternatives due to her cognitive limitations. Hence, the only requirement is that $\Gamma(R)$ is a subset of $R \cup a$ and $a \in \Gamma(R)$.

- Expanded Awareness Model: A recommendation might activate consideration of similar items for the decision maker. For example, Toyota and Honda might always come together into the decision maker's consideration set, even if only one is recommended. Formally, given R, her consideration set, $\Gamma(R)$, is a superset of $R \cup a$. In particular, $\Gamma(R) \setminus R$ are the items that are "similar" to some items in R. Hence, the only requirement is that $\Gamma(R)$ is a superset of $R \cup a$.

These examples highlight the richness of our framework. The simplest model in each channel is the baseline one. Depending on the context, one modeling option might be more appropriate than the others.

3 Deterministic Choice: Baseline Models

As the first step in understanding the implication of recommendation, we focus on the baseline models in this paper. The reasons are two folded. First, each consumer's behavior is intended to be simple enough to study the aggregate behavior of heterogeneous consumers in a tractable way. Second, our baseline models offer a minimal departure from the standard model. The classical choice model is the most widely applied model due to its simplicity, even though sometimes explanatory power can be compensated. Therefore, we would like to stay close to the rational framework as much as possible to understand how much explanatory power we can gain by taking a minimal departure for each channel.

3.1 Behavioral Characterization

We would like to know when we can distinguish RP and RA models from the observed choices. This requires us to look into the behavioral implications of the two models and see if we can set them apart from the observed choices.

The first behavioral postulate is the key axiom in this environment, which states that removing some of the unchosen alternatives from the recommendation set does not influence the final choice. Note that it shares a similar flavor as the famous IIA axiom (*i.e.* Independence of Irrelevant Alternative) in the choice theory literature.[10] We call this the Independence of Irrelevant Recommended Alternatives, abbreviated by IIRA.

IIRA. *If* $c(R) \notin R \setminus R'$ *and* $R' \subseteq R$*, then* $c(R) = c(R')$*.*

Since this axiom looks similar to the standard IIA, one might think they are equivalent. While the standard IIA applies to classical choice functions in which the winner always belongs to the choice set, our postulate applies to new choice objects in which the winner can be a recommended or non-recommended item. This seemingly small distinction has important implications. Indeed, if we include "$c(R)$ belongs R'" in the premise of IIRA, then IIRA becomes the standard IIA, which is stated below as IIRA(1). However, IIRA(1) is strictly weaker than IIRA.

IIRA(1). *If* $c(R) \in R'$ *and* $R' \subseteq R$*, then* $c(R) = c(R')$*.*

In other words, seemingly similar axioms have distinct implications in our framework. Hence the existing results in the literature on the standard framework might not be valid in this framework. Since IIRA(1) is weaker than IIRA, we would like to state the counterpart of IIRA(1). To this, we include "$c(R)$ is not in R" into the premise and call it IIRA(2). It states that the winner will persist outside the recommendation set when fewer alternatives exist.

IIRA(2). *If* $c(R) \notin R$ *and* $R' \subseteq R$*, then* $c(R) = c(R')$*.*

It is easy to see that IIRA(1) and IIRA(2) are equivalent to IIRA. Notice that both RA and RP models satisfy IIRA. To see this, first, consider IIRA(1), so $c(R) \in R'$. Then, the RA model says that $c(R)$ must be better than all the other recommended alternatives in R and the default option. Hence, it stays the best alternative if we remove some of the unchosen alternatives from the recommended set. On the other hand, the RP model says that $c(R)$ is evaluated according to the recommended version, and it must be better than all the other recommended and non-recommended alternatives. Therefore, the evaluated ranking of $c(R)$ stays the best alternative even if some alternatives are downgraded.

Consider the second postulate, IIRA(2), so $c(R) \notin R$. Then, the RA model requires that the chosen alternative must be the default option, which has to be preferred to all recommended alternatives in R. Hence, it must also be better than all alternatives in R' since R' is a smaller set. On the other hand, the RP model says that the non-recommended item $c(R)$ is already

[10] 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).

better than the recommended item in R and other non-recommended items. Hence, the non-evaluated version of $c(R)$ is still superior to other alternatives when some other alternatives are not evaluated anymore.

We now consider a new property in this framework– the Sandwich Property. As the name suggests, if two nested recommendations result in the same choice, then the decision maker chooses the same option when she receives any recommendation between these recommendations.

Sandwich Property. *If $R' \subseteq R \subseteq R''$ and $c(R'') = c(R')$, then $c(R) = c(R')$.*

Note that, in the standard choice domain where $c(R) \in R$, the sandwich property is implied by IIA.[11] Since $c(R'') \in R \subset R''$, IIA would imply $c(R) = c(R'')$. However, in our framework, the sandwich property has distinct implications from IIA, due exactly to the fact that choices can be outside of the recommendation set. RA model satisfies the sandwich property. To see this, if $c(R') = c(R'')$, then the chosen item must be preferred to every recommended item in R'' and the default option. Hence, it must still be better than any recommended item for any in-between recommendation set. Last but not least, it turns out that Sandwich Property is the only behavior choice pattern where RA and RP differ.

Theorem 1 (Characterization)**.** Let $\mathcal{D}$ includes all recommendation sets with $|R| \leq 3$. Then,
1) c has an RP representation if and only if c satisfies IIRA;
2) c has an RA representation if and only if c satisfies IIRA and Sandwich Property.

As far as we know, this is the first characterization of decision-making under recommendation. As we stated before, these two models capture two distinct channels of recommendation. The theorem shows that these models are nested. It is a surprising result that the preference channel fully encapsulates the explanatory power of the attention channel while retaining room for us to distinguish the RP model from the RA model. In the standard environment, if the choice behavior satisfies the classical IIA, choices can be attributed both to limited attention or to preferences. Here, it turns out that we can actually distinguish them. Therefore, our intuition based on the standard framework may mislead us here.

Theorem 1 also informs us how to distinguish between these two channels. Given these results, if the Sandwich property fails, we can confidently reject the attention channel. The simplest example of a failing Sandwich property is given as follows.

Example 1 (Violating Sandwich Property)**.** Let $c(\emptyset) = x$, $c(\{y\}) = y$ and $c(\{x, y\}) = x$. This example violates Sandwich Property. Also, one can check that the RA model cannot explain it. Suppose it does. Note that since $c(\emptyset) = x$, it reveals that RA-DM's default option is x. Then, $c(\{y\}) = y$ reveals that y is better than x. Therefore, we must have $y = c(\{x, y\})$, a contradiction. On the other hand, RP-DM can exhibit this behavior if $u(x)+v(x) > u(y)+v(y) > u(x) > u(y)$. ⌟

[11] A version of the sandwich property appears in Masatlioglu and Nakajima (2013), which is stronger than Weak-WARP of Manzini and Mariotti (2007).

We now discuss the identification and uniqueness properties of our model. Note that our domain is finite. Hence, we can only attain ordinal uniqueness. For the RA model, it is easy to see that the pre-recommendation choice must be the choice when there is no recommendation $a := c(\emptyset)$. Moreover, if x is chosen when R is recommended, then x is revealed to be preferred to every alternative in R. If x is different from $c(\emptyset)$, x is also revealed to be preferred to the pre-recommendation choice. Lastly, although the lower contour set of a is the same, the preference relation among those alternatives cannot be pinned down. The next proposition states all the uniqueness properties of the RA model.

Proposition 1 (Uniqueness)**.** Let (a_1, u_1) and (a_2, u_2) be two RA representations of the same choice data. Then, for $x, y \in X$

i) $a_1 = a_2$,

ii) $u_1(a_1) > u_1(x)$ if and only if $u_2(a_2) > u_2(x)$, and

iii) $u_1(x) > u_1(y) > u_1(a_1)$ if and only if $u_2(x) > u_2(y) > u_2(a_2)$

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

Secondly, one can show a similar result for the RP model. Assume (u_1, v_1) and (u_2, v_2) represent the same choice rule. Notice that when there is no recommendation, alternatives are evaluated according to u_i. Therefore, the best alternative in X according to u_1 and u_2 must be the same. Also, if x is different from $c(\emptyset)$, and x is chosen when it is recommended, it also revealed to be preferred to the best alternative in X. Lastly, the lower contour set of the best alternative in X must be the same.

Proposition 2 (Uniqueness)**.** Let (u_1, v_1) and (u_2, v_2) be two RP representations of the same choice data. Then, for $x, y \in X$

i) $\text{argmax}_{x \in X} u_1(x) = \text{argmax}_{x \in X} u_2(x) := a$,

ii) $u_1(a) > u_1(x) + v_1(x)$ if and only if $u_2(a) > u_2(x) + v_2(x)$, and

iii) $u_1(x) + v_1(x) > u_1(y) + v_1(y) > u_1(a)$ if and only if $u_2(x) + v_2(x) > u_2(y) + v_2(y) > u_2(a)$.

This proposition also states that no recommendation choice is unique, and the ordinal rankings are uniquely identified as long as it matters for the choice.

4 Probabilistic Choice

In the last section, we provide a micro foundation for the baseline models in which recommendations affect deterministic choices through two different channels. In this section, we extend these ideas to probabilistic choice. One interpretation of randomness in choice data is that there is a group of individuals with unknown types, and the outside observer can only observe their aggregate behavior; this is the interpersonal interpretation. Another interpretation is the choices of a single individual in different situations; this is the intrapersonal interpretation. Our model can be interpreted as both intrapersonal and interpersonal probabilistic choices. In the classical choice-set variation framework, probabilistic choice data is well-studied. However, recommendations cannot be studied within this framework because the DM can choose outside of recommendation R. To incorporate this idea, we extend the definition of a (probabilistic) choice rule just as we did in our deterministic setting. Formally,

Definition 2. A choice rule ρ is a mapping from $X \times \mathcal{D}$ to $[0, 1]$ such that $\sum_{x\in X} \rho(x, R) = 1$.[12]

4.1 Parametric Probabilistic Choice

We first study two parametric models of recommendation for the two different channels with probabilistic data based on the Luce model. All the information about the models' predictions can be summarized by a finite set of parameters depending only on the alternatives. As is typically the case with parametric models, our parametric models offer three advantages over the non-parametric versions we study in the next section. First, our parametric models are very tractable, which is a desirable property for applications. We believe that our models are ready to use in applications because of their analytic and computational tractability. Second, the model possesses strong uniqueness properties. Third, it has sharp identification results for designing policies.

We make the following positivity assumption throughout this section to foster a comparison between the two channels in the parametric models. Notice that this positivity property, as argued by McFadden (1973), cannot be refuted based on any finite data set.

Assumption 1. For every x and R, $\rho(x, R) > 0$.

Preference Channel: In the RP model, the valuation of the recommended item is calculated by $u + v$. In terms of the Luce model, we assume that for each available alternative x, there are two alternative-specific parameters $u(x) > 0$ and $v(x) \geq 0$. Like the Luce model, u represents the crude measure of utility value. On the other hand, $v(x)$ captures the boost when alternative x is recommended. We sometimes use (u, v) as the primitive of the model for convenience. For notational simplicity, we write $f(A)$ as shorthand for $\sum_{x\in A} f(x)$ for any function f. We assume the DM pays attention to every alternative in the preference channel. Therefore, the probability that a recommended alternative is chosen equals its weight $(u+v)(x)$ divided by the total weight in the choice set. The probability that an alternative that is outside of the recommendation set is chosen is equal to its weight $u(x)$ divided by the total weight in the choice set. We summarize this into the following definition.

Definition 3. A choice rule ρ has a Luce RP representation if there exist functions $u : X \to \mathbb{R}_{++}$ and $v : X \to \mathbb{R}_+$ such that for $x \in X$,

$$\rho(x, R) = \begin{cases} \dfrac{u(x) + v(x)}{u(X) + v(R)} & \text{if } x \in R \\ \dfrac{u(x)}{u(X) + v(R)} & \text{otherwise} \end{cases} \tag{RP-Luce}$$

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

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

There are several interesting special cases of the model. First, one can imagine a case where the underlying utility u is constant across alternatives, and the only difference comes from the boost in utility v. The model is represented by $(\bar{u}, v)$, where $\bar{u}$ is a constant. In this case, the distribution of choices is uniform if there is no recommendation. Each recommendation set breaks this uniformity in favor of recommended alternatives. Second, the other possibility is that the gain in utility is constant across alternatives $(u, \bar{v})$, where $\bar{v}$ is a constant. In this case, $\rho(\cdot, X)$ becomes a convex combination of the uniform distribution and $\rho(\cdot, \emptyset)$. In one extreme ($\bar{v} = 0$), $\rho(x, R)$ is equal to $\rho(x, \emptyset)$ for all x and for all R. As $\bar{v}$ gets bigger, $\rho(\cdot, X)$ approaches the uniform distribution. Finally, it could be that the boost in utility v enjoys the same multiplicative factor relative to the underlying utility u. In this case, the model is represented by $(u, \lambda u)$, where $\lambda \geq 0$. Again, the size of λ determines the effect of recommendations. If λ is small, $\rho(x, R)$ is in close proximity of $\rho(x, \emptyset)$ for all x and for all R. All these special cases are one parameter extension of the Luce model capturing the effect of recommendations.

Attention Channel: In this model, each alternative x is represented by two parameters: $d(x) > 0$ and $u(x) > 0$.[13] The function d represents the likelihood of x being the pre-recommendation choice, hence $\sum_{x \in X} d(x) = 1$. The function u is a crude measure of the utility value, which extends preferences in the deterministic RA model to the probabilistic setting. An alternative with a high u value is chosen more often than an alternative with a low u value. We follow the Luce model to describe the choice probabilities given a default. To encompass our deterministic model, for a fixed default option, all non-recommended alternatives are chosen with zero probability except the default option. The probability 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:

$$W_{(a,A)}(x) = \begin{cases} \dfrac{u(x)}{u(A \cup a)} & \text{if } x \in A \cup a \\ 0 & \text{otherwise} \end{cases}$$

Note that $W_{(a,A)}$ is itself a parametric choice model where $\sum\limits_{x \in X} W_{(a,A)}(x) = 1$. According to this formulation, only the recommended alternatives and the pre-recommendation choice are chosen with positive probability. $W_{(a,A)}$ captures the randomness in preferences as in the Luce model. Note that the deterministic RA model is a limit case of this model since the deterministic model has no randomness in preferences.[14] We expand on this intuition by assuming that being the default option is probabilistic. Since $d(x)$ is the probability of x being the default option, the choice probability of any alternative is defined as a mathematical expectation: The probabilities

[13] We can relax to $d(x) \geq 0$ in the model. However, the positivity assumption implies $d(x) > 0$ in our model.

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

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 RA representation if there exist functions $u : X \to \mathbb{R}_{++}$ and $d : X \to (0,1)$ with $\sum_{x\in X} d(x) = 1$ such that for any $x \in X$,

$$\rho(x, R) = \sum_{a\in X} d(a) W_{(a,R)}(x) \tag{RA-Luce}$$

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

RA-Luce model inherently has two types of randomness. While d captures the randomness in pre-recommendation choice, u represents the randomness in preferences. Note that the odds of selecting a 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 $x \notin R$

$$\rho(x, R) = \frac{d(x)u(x)}{u(R \cup x)}$$

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 is fixed, and R represents the recommended alternatives. The effective weight of x becomes $d(x)u(x)$, which is strictly less than $u(x)$. This can be interpreted as the value of non-recommended alternatives being discounted while the recommended ones stay the same. For recommended alternatives, $x \in R$, our formula is more involved:

$$\rho(x, R) = \sum_{z\in X} \frac{d(z)u(x)}{u(R \cup z)}$$

There are two special cases of the model. The first one is that, similar to the RP case, the underlying utility u is constant across alternatives. In this case, the model is represented by $(\bar{u}, d)$, where $\bar{u}$ is a constant. Also, the recommended alternatives will be chosen with the same probability, while the non-recommended alternatives will have a different choice probability depending on d. Secondly, one can imagine a case where alternatives could be equally likely to be the default option. In this case, the model is represented by $(u, \frac{1}{|X|})$, and the choice probabilities are $\frac{1}{|X|}$ when there is no recommendation. Each recommendation set breaks this uniformity in favor of recommended alternatives. Note that there is no additional parameter in this case except u, which is the same as in the Luce model.

4.2 Behavioral Characterizations for Parametric Models

In this section, we discuss the behavioral implications of the RA-Luce and RP-Luce models. This will provide us with the empirical content of each model. These behavioral postulates will allow us to test and compare these models.

Since we extend the classical Luce model, as anticipated, some version of Luce's IIA appears in our characterization. Recall that Luce's IIA says that the odds of choosing one alternative over another do not depend on the feasible set. In the RA-Luce model, Luce's IIA holds for all recommended alternatives:

Recommended Luce-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 not 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. A modified version of Luce's Choice Axiom appears in our characterization. Our 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. We call this axiom R-Path Independence.

R-Path Independence. *For $x \notin R$ and $R \cup x \subseteq R'$, $\rho(x, R)\rho(R \cup x, R')$ is independent of R.*

As long as probabilities are strictly greater than zero, 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 the 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 two properties summarize the entire empirical content of the RA model.

Theorem 2 (Characterization)**.** Assume $\mathcal{D} = 2^X$. Then, ρ has an RA-Luce representation if and only if ρ satisfies Recommended Luce-IIA and R-Path Independence.

While this theorem provides the entire empirical content of the model, it requires the knowledge of complete data. This is very demanding in terms of the availability of data. We now demonstrate that we can provide a similar characterization of this model with limited data. Suppose $\mathcal{D}$ includes all possible recommendation sets with $|R| \leq 2$. Under limited data, we must impose a stronger axiom on the off-recommendation data. The reason why we need this axiom is that R-Path Independence is not strong enough in this limited domain. This concern does not exist if we have full data. To see why we need a stronger axiom, consider the following example in Table 1 with $X = \{a, b, c, d\}$. Notice that Recommended Luce-IIA and R-Path Independence are immediately satisfied.[15] However, none of these axioms restrict how $\rho(c, \{a, b\})$

[15] Since the data is 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 R-Path Independence.

behaves: R-Path Independence puts restriction on $\rho(c,\{a,b\})$ only if we also observe data on some other recommendations that include all a, b and c. Nonetheless, we might not observe a recommendation including all three alternatives due to the limited data assumption. Hence, to check for the model's validity under limited data, we need to impose a stronger axiom.

	$\{a,b\}$	$\{a,c\}$	$\{a,d\}$	$\{b,c\}$	$\{b,d\}$	$\{c,d\}$	$\{a\}$	$\{b\}$	$\{c\}$	$\{d\}$	$\emptyset$
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	1/120	5/12	1/12	5/12	1/12	5/12	1/8	1/8	3/8	1/8	1/4
d	19/120	1/12	5/12	1/12	5/12	5/12	1/8	1/8	1/8	3/8	1/4

Table 1. Probabilistic choice data satisfying Recommended Luce-IIA and R-Path Independence but does not have an RA-Luce representation under limited data.

The R-Path Independence axiom revolves around the fact that the choice probability of a not recommended alternative under recommendation set R is tightly related to its choice probability when recommended with the set R. It turns out this dependency can be made more explicit. We define the following shorthand: $r(z,x) := \frac{\rho(z,A)}{\rho(x,A)}$ for some A, including x and z. Note that Recommended Luce-IIA guarantees that $r(z,x)$ is well-defined. Each $r(z,x)$ captures exactly the choice ratio of x and z when both of them are recommended.

R-Independence. *For $x \notin R$, $\rho(x,R)(1+\sum_{z\in R} r(z,x))$ is independent of R.*

The intuition behind this axiom is that, within our model, 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 $\sum_{z\in R} r(z,x)$ increases and, in turn $\rho(x,R)$ decreases. Therefore, this axiom dictates that x must be chosen less as more alternatives are recommended when x is not recommended. Also, the rate that $\rho(x,R)$ also decreases depending 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 decreases less if x is chosen much more often than z when both are recommended.

It is clear that the example in Table 1 violates R-Independence. To see why, consider the choice probability of c under the recommendation set $\{a,b\}$ and $\emptyset$, we have $\rho(c,\emptyset) = \frac{1}{4} \neq \frac{1}{40} = \rho(c,\{a,b\})(1+r(a,c)+r(b,c))$. Moreover, we can also see that Recommended Luce-IIA and R-Independence imply R-Path Independence. 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 R-Path Independence.

Theorem 3. Let $\mathcal{D}$ includes all recommendation sets with $|R| \leq 2$. Then, ρ has an RA-Luce representation if and only if ρ satisfies Recommended Luce-IIA and R-Independence.

Theorem 3 provides a similar characterization with limited data. Thus, Theorem 3 weakens the data requirements of Theorem 2. Indeed, many models in decision theory require a similar rich data set, typically choices from all decision problems.

Characterization of RP-Luce: We now turn to the characterization of the RP-Luce model. Notice that the RP-Luce model is similar to the standard model except that the weights are multiplied by a factor if it is recommended. Therefore, the RP-Luce model satisfies a Luce axiom that is stronger than the Recommended Luce-IIA, which is the Strong Luce-IIA axiom.

Strong Luce-IIA. *For $x, y \notin R \Delta R'$,*

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

where Δ is the symmetric difference operator.

This property says that as long as x, y does not change its status across R and R', the Luce-IIA axiom should hold. Note that this Strong Luce-IIA implies every possible pair of Luce-IIA axiom. Firstly, note that if x and y are recommended across both R and R', then their status are constant across R and R'. Indeed, it is easy to check from the definition that $x, y \in R \cap R$ gives immediately that $x, y \notin R \Delta R'$, so that Recommended Luce-IIA is implied. A similar argument can produce the *Non-recommended Luce-IIA*, where x and y are both not recommended across R and R'. Interestingly, this property also connects x and y when one is recommended and the other is not, as a kind of *Cross-status Luce-IIA*.

One might suspect that this axiom alone is sufficient for RP-Luce. It turns out that we need an additional behavioral postulate. The idea is simple: Any not recommended item is chosen weakly more when the set of recommended items gets smaller. We call this property as R-Regularity.

R-Regularity. *For $x \notin R$, $\rho(x,R) \leq \rho(x, R \setminus y)$.*

Notice that this axiom is trivially satisfied in the standard Luce model since both sides of the inequality are equal to zero. With this axiom, we are able to state the characterization result.

Theorem 4. Let $|X| \geq 3$ and $\mathcal{D}$ include all recommendation sets with $|R| \leq 2$. Then, ρ has an RP-Luce representation if and only if ρ satisfies Strong Luce-IIA and R-Regularity.

As Theorem 3, Theorem 4 works under very mild data requirement: the size of the recommendation set is less than or equal to 2.[16]

4.3 Uniqueness and Identification for Parametric Models

In this section, we discuss the identification of the models. To achieve this, we first establish the fact that the parameters are unique in each model. Our first result shows that we can uniquely pin down the default option probabilities in the RA-Luce model. Also, the utility measure is unique up to a scaling factor.

[16] One caveat is that Theorem 4 requires that there are at least three available alternatives. The reason is that Strong Luce-IIA put no restriction on choice data if there are only two available alternatives. One can define another property, e.g. for $x, y \in R$, $\frac{\rho(x,R)}{\rho(y,R)} \frac{\rho(x,R\setminus\{x,y\})}{\rho(y,R\setminus\{x,y\})} = \frac{\rho(x,R\setminus y)}{\rho(y,R\setminus y)} \frac{\rho(x,R\setminus x)}{\rho(y,R\setminus x)}$. One can show that this property, along with Strong Luce-IIA and R-Regularity, are the necessary and sufficient conditions for RP-Luce with $|X| \geq 2$.

Proposition 3 (Uniqueness of RA-Luce)**.** Let (u_1, d_1) and (u_2, d_2) be two RA-Luce representations of the same choice data. Then $d_1 = d_2$, and $u_1 = \alpha u_2$ for some $\alpha > 0$.

Similar to the RA-Luce model, the RP model also uniquely pins down the recommendation factor and the utility measure is again unique up to a scaling factor.

Proposition 4 (Uniqueness of RP-Luce)**.** Let (u_1, v_1) and (u_2, v_2) be two RP-Luce representations of the same choice data. Then $u_1 = \alpha u_2$ and $v_1 = \alpha v_2$ for some $\alpha > 0$.

We now turn to the identification of the parameters. We first consider the RA-Luce model. Our first result illustrates that observations from only two recommendation sets are sufficient for unique identification.

Proposition 5 (Identification of RA-Luce)**.** Suppose ρ is RA-Luce. Let $\mathcal{D}$ include recommendation sets $\emptyset$ and $\{a\}$ for some $a \in X$, then we can fully identify the parameters of the model.

Note that RA-Luce requires only two data points to fully identify the model. The underlying reason behind the identification is as follows. Firstly, having no recommendation data allow us to identify $d(x)$. Secondly, suppose we have the recommendation data for one alternative. Since RA-Luce requires that the DM compare each default option with the recommended item, we can fully recover every w. Since the procedure is simple enough, we demonstrate the identification in the main text. Firstly, let $d(x) := \rho(x, \emptyset)$ for every $x \in X$. Then, we identify u. Note that u is unique up to a scaling factor; fix $a \in X$, and we let $u(a) := 1$. Note that, for $x \in X \setminus a$, we define $u(x) := \frac{\rho(x,\{a\})}{\rho(x,\emptyset)-\rho(x,\{a\})}$. Then, the parameters are fully identified.

On the other hand, RP-Luce requires more data to work with. Notice that the recommendation factor for an alternative can be "observed" only if the alternative is recommended on some occasions. Therefore, to retrieve this parameter, we must have data in which each alternative is recommended in some sets.

Proposition 6 (Identification of RP-Luce)**.** Suppose ρ is RP-Luce. Let $\mathcal{D}$ include all recommendation sets with $|R| \leq 1$, then we can fully identify the parameters of the model.

To see how identification in our model works, let $u(x) := \rho(x, \emptyset)$ for every $x \in X$. Then, we define $v(x) := \frac{\rho(x,\{x\})-\rho(x,\emptyset)}{1-\rho(x,\{x\})}$. Hence, we can fully identify the parameters. Notice that, in the above proposition, it is possible to obtain the same result by replacing the singleton recommendation data with some doubletons data where each alternative needs to be recommended at least once in one of the doubletons. This can reduce the data requirements for identification.

4.4 Behavioral Distinction, Similarity, and Intersection

In the deterministic framework, the RA and RP models can be distinguished only in one way but not the other. More specifically, we can only distinguish the RP model from the RA model when we observe a violation of the sandwich property. In the probabilistic framework, surprisingly, the distinction becomes both ways. This is due to the parametric assumptions.

It is easy to construct an example in which one probabilistic choice function has an RA-Luce representation but does not have an RP-Luce representation and vice versa.

On the other hand, although our two models share different characterizations, they share some common behavioral properties. Firstly, as one might immediately expect, both RA-Luce and RP-Luce satisfy Recommended Luce-IIA. This is because Strong Luce-IIA implies Recommended Luce-IIA. Secondly, as discussed in the previous section, we know that R-Independence implies R-Regularity. In fact, one can even check that R-Path Independence also implies R-Regularity. To see this, by R-Path Independence, we know $\rho(x,\emptyset)\rho(x,\{x,y\}) = \rho(x,\{y\})\rho(\{x,y\},\{x,y\})$. Hence, we know that $\frac{\rho(x,\emptyset)}{\rho(x,\{y\})} = \frac{\rho(\{x,y\},\{x,y\})}{\rho(x,\{x,y\})} > 1$. One can apply a similar argument to other sets.

One might wonder whether there are choice probabilities that satisfy all four axioms presented here. In other words, we wonder whether there is any choice behavior at the intersection of the RA and RP models. The answer is affirmative, and the choice probabilities are in a specific form as long as there are at least three alternatives ($|X| \geq 3$). In particular, given $R \neq \emptyset$, each recommended alternative is chosen with probability $\frac{1+|X|}{|X|(1+|R|)}$, and each non-recommended alternative is chosen with probability $\frac{1}{|X|(1+|R|)}$. For example, when $X = \{x,y,z\}$ and $R = \{x,y\}$, we have $\rho(x,R) = \rho(y,R) = 4/9$ and $\rho(z,R) = 1/9$. It is easy to check that it satisfies Strong Luce-IIA. To see that it satisfies R-Independence, note that we have the recommended Luce Ratio $r(z,x) = 1$. Then, when x is not included R, we have $\rho(x,R)(1+\sum_{z\in R} r(z,x)) = \frac{1}{|X|(1+|R|)}(1+|R|) = \frac{1}{|X|}$, which is independent of R. Hence, R-Independence is satisfied. Moreover, we can perform the identification exercise and represent the choice probabilities with the RA and RP models. For the RA model, it goes by the form $(\bar{u}, \frac{1}{|X|})$, where the default probability is uniform at $\frac{1}{|X|}$ and the underlying utility is a constant $\bar{u}$. For the RP model, it is represented by $(\bar{u}, |X|\bar{u})$, where again the underlying utility is a constant $\bar{u}$, and v is fixed at $|X|$.

4.5 Discrete Choice

In the following, we consider the model from the discrete choice perspective. In the framework of discrete choice, a decision maker's utility for an alternative x is defined as their underlying $u(x)$ plus an *error* term $\varepsilon(x)$, where $\varepsilon(x)$ is known as "random utility shock". The discrete choice model posits that 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. Formally, it requires 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 value within a set A given a (utility) function f,

$$\mathbf{E}(x,A;f) = \{\omega \in \Omega : f(x) + \varepsilon_\omega(x) \geq f(y) + \varepsilon_\omega(y) \text{ for all } y \in A\}$$

With this event definition, one can define the discrete choice model when $\varepsilon(x)$ is i.i.d. according to the extreme value distribution across $x \in X$ with noise parameter λ so that

$$\text{Prob}(x \text{ is chosen from } S) = \mathbb{P}(\mathbf{E}(x, S; u)) \tag{Logit}$$

It is well-known that the extreme value distribution gives rise to a closed-form solution

$$\mathbb{P}(\mathbf{E}(x, A; u)) = \frac{e^{u(x)/\lambda}}{\sum_{z \in A} e^{u(z)/\lambda}}$$

which shares the same explanatory power as the Luce model. In the recommendation framework, one can also utilize this discrete choice approach to obtain the Logit version of the RP and RA model. In the following, we will elaborate on these and show the equivalence between the Luce approach and the discrete choice approach, as in the standard framework.

Attention Channel: From the attention Channel, in order for an alternative to be chosen, an alternative need to 1) be inside the recommendation set R or be the default option a and 2) be the best alternative among $R \cup a$. Here, we assume independence between the distribution of default options and the random utility shock. Also, to avoid unnecessary extra notation, we will retain the probability measure $\mathbb{P}$ to ε, and designate a function d to denote the probability of being the default option. Therefore, one can define the logit version of the RA model

$$\rho(x, R) = \begin{cases} \displaystyle\sum_{a \in X} d(a)\mathbb{P}(\mathbf{E}(x, R \cup a; u)) & \text{if } x \in R \\ d(x)\mathbb{P}(\mathbf{E}(x, R; u)) & \text{if } x \notin R \end{cases} \tag{RA-Logit}$$

As we plug in the close-form solutions for the $\mathbb{P}$, it becomes clear that it is essentially the RA-Luce model. Therefore, the relationship between RA-Luce and RA-logit is analogous to the relationship between Luce and logit in the classical choice-set variation domain.

Preference Channel: From the preference channel, recommendation affects the valuation of the alternatives. We define the utility function as in Section 2 so that $U_R(x) = u(x) + v(x)$ if $x \in R$; $U_R(x) = u(x)$ if $x \notin R$. Since the preference channel assumes the DM pays attention to everything, the maximization problem is checking all $x \in X$. Therefore, similar to the standard framework, we can define the probabilistic choice as;

$$\rho(x, R) = \mathbb{P}(\mathbf{E}(x, R; U_R)) \tag{RP-Logit}$$

Plugging in the closed-form solution for $\mathbb{P}(\mathbf{E}(x, R; U_R))$, one can see that RP-Logit and RP-Luce share the same explanatory power over choice data.

4.6 Non-Parametric Probabilistic Choice

In the last few sections, we focused on two parametric models in which alternative-specific parameters describe the randomness of types. These models are analytically and computationally tractable but do not account for the possibility that tastes differ randomly between individuals.

In this section, we investigate the behavioral implications of non-parametric versions of our baseline models of recommendation. As in the standard Random Utility Model (RUM), the data is assumed to be generated from a group of different individuals making deterministic choices. In the standard framework, the Luce model can be captured in RUM by assuming independence in preference. In other words, the fraction of people who prefer a over b is the same regardless of their preference for another alternative c. By introducing the RUM version for our models in this probabilistic environment, we are also relaxing this independence built in the RA-Luce and RP-Luce models.

Before we define our models, we revisit the classical random utility model. RUM allows heterogeneous populations, and each type in the population is a utility maximizer. Let μ be a probability distribution on the set of utilities, $\mathcal{U}$. The probability distribution μ constitutes a probabilistic choice function π_μ such that

$$\pi_\mu(x, S) = \mu(\{u \in \mathcal{U} |\ x = \operatorname*{argmax}_{y \in S} u(y)\}) \tag{RUM}$$

The probability of x being chosen from S is the frequency of those types who choose x from S. We say that a probabilistic choice function, π, has a Random Utility representation if there exists μ such that $\pi = \pi_\mu$.

In the next two subsections, we investigate the behavioral implications of our RA and RP models in the non-parametric probabilistic choice environment. In both models, multiple representations might lead to the same choice behaviors. For example, in the RA model, (a, u) and (a, v) induce the same behavior as long as u and v are ordinally equivalent. Hence, we must define each choice type carefully in these models.

Attention Channel: In this framework, we assume that each decision-maker in the population belongs to the RA model. In other words, they can differ on their pre-recommendation choices and utility functions, (a, u), and (potentially) differ on their choice, denoted by $c_{(a,u)}$. To define the underlying type space, we let τ be a probability distribution over the set of the product of X and $\mathcal{U}$.[17] Therefore, probability distribution τ constitutes a probabilistic choice function ρ_τ such that

$$\rho_\tau(x, R) = \tau(\{(a, u) |\ x = c_{(a,u)}(R)\}) \tag{RA-RUM}$$

for every $R \in \mathcal{D}$. We say that a probabilistic choice function, ρ, has an RA-RUM representation if there exists τ on $\mathcal{T}$ such that $\rho = \rho_\tau$.

As in RUM, RA-RUM tests the hypothesis of a group of utility-maximizing individuals with limited attention. In RUM, individuals differ in terms of their tastes. In RA-RUM, not only their taste but also their pre-recommendation choices differ. Hence, RA-RUM has a much richer type space compared to RUM.

Preference Channel: Here, each decision-maker in the population belongs to the RP model.

[17] As in the deterministic case, we will assume these u are one-to-one so that there will be no cases of indifference.

In other words, each decision-maker's behavior is represented by a pair of utility functions (u, v), with choice denoted by $c_{(u,v)}$. To define a type space, we let μ be a probability distribution over the set of $\mathcal{U}^2$. The probability distribution μ constitutes a probabilistic choice function ρ_μ such that

$$\rho_\mu(x, R) = \mu(\{(u, v)|\ x = c_{(u,v)}(R)\}) \tag{RP-RUM}$$

for every $R \in \mathcal{D}$. We say that a probabilistic choice function, ρ, has an RP-RUM representation if there exists τ on $\mathcal{P}^\star$ such that $\rho = \rho_\mu$.

As in RA-RUM, RP-RUM tests the hypothesis of a group of utility-maximizing individuals affected by recommendations. In RA-RUM, consumers differ in terms of their tastes and default option and exhibit limited awareness. In PR-RUM, each decision-maker considers every alternative in the choice set. Therefore, the only departure from RUM we introduce here is that recommendation can change the ranking of recommended items. RP-RUM has a much richer type space compared to both RUM and RA-RUM.

It is a well-known fact that in the Luce model is a special case of the random utility model in the classical setup. In the recommendation environment, one might wonder whether RA-Luce belongs to RA-RUM and/or RP-Luce belongs to RP-RUM. It turns out that every choice rule ρ that has an RA-Luce (RP-Luce) representation has an RA-RUM (RP-RUM) representation. We provide a proof for this statement in Appendix.

4.7 Hasse Diagram

Within the RUM framework, Falmagne (1978) answered the question of whether individual utility maximization has any implications for aggregate data. For the characterization of the RUM, Falmagne (1978) utilizes a well-known concept in the literature: the Block-Marschak polynomials, named after Block and Marschak (1959)'s seminal work on the RUM. It has been shown that probabilistic choice data have a RUM representation if and only if its Block-Marschak polynomials are non-negative. Block and Marschak (1959) originally obtained the necessity. Falmagne (1978) showed that the non-negativity of the polynomials is also sufficient.

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

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

The Block-Marschak polynomials are defined with respect to the choice data ρ. Throughout the paper, we mostly skip denoting ρ and write $q(a, R)$ unless specified otherwise. q_ρ is defined for each recommended alternative. Interestingly, this definition can also be applied to non-recommended alternatives, and, as we shall see, it has an important role in both RA-RUM and

RP-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 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 considered 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 the 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 to a path starting from the empty set and ending at the grand set. For example, $\{a, b, c\} - \{b, c\} - \{c\} - \emptyset$ (with path $q(a, \{a, b, c\}) \to q(b, \{b, c\}) \to q(c, \{c\}))$ represents $u(a) > u(b) > u(c)$.[18] We first highlight that the above equality is no longer true in our model. But we discuss below how to recover similar equality and provide a similar visual representation for types in R-RUM.

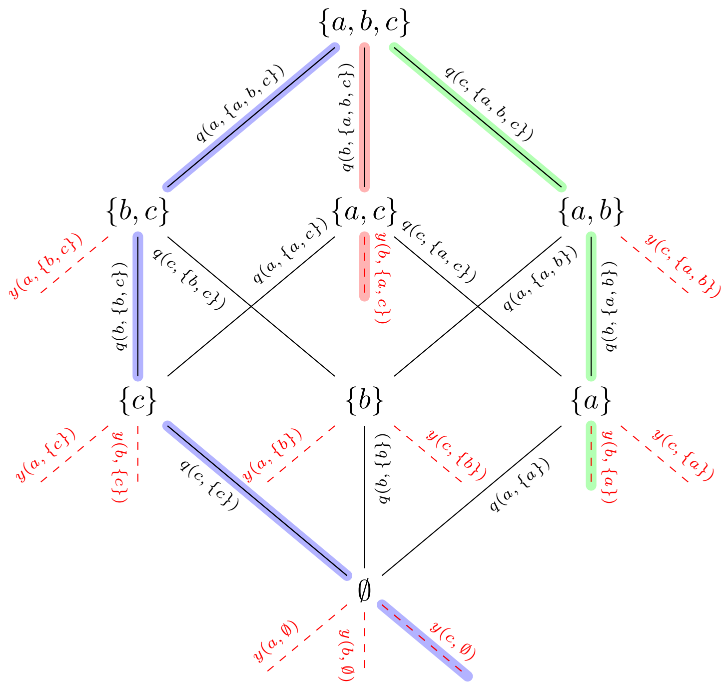

Figure 1. Hasse Diagram of RA-RUM: Each colored path represents a specific type: Type (c, u_1) where $u_1(a) > u_1(b) > u_1(c)$ highlighted by the blue (the longest) path, Type (b, u_2) where $u_2(c) > u_2(b) > u_2(a)$ highlighted by the green path, and Types (b, u_3) and (b, u_4) where $u_3(b) > u_3(a) > u_3(c)$ and $u_4(b) > u_4(c) > u_4(a)$ highlighted by the red (the shortest) path.

Compared to RUM, RA-RUM and RP-RUM have two sets of BM conditions: one for rec-

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

ommended 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 denote nodes and flows with the same notation. $y_\rho(a, \{c\})$ denotes both the phantom node and the flow to that node. Interestingly, if we also take into account y's, we recover the equality of the inflow and outflow of all black and red lines.[19] 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 RA-RUM and RP-RUM by a path in the new Hasse diagram. Firstly, we consider RA-RUM. Each type corresponds to a path starting from the grand set and ending at a phantom node. For example, the path $q(a, \{a, b, c\}) \to q(b, \{b, c\}) \to q(c, \{c\}) \to y(c, \emptyset)$ represents $u_1(a) > u_1(b) > u_1(c)$ with c being the pre-recommendation choice, hence the type is (c, u_1) (the longest (blue) path in Figure 1). In this path, the leakage appears at node $\emptyset$, which implies the pre-recommendation choice has the minimum utility according to u_1. On the other hand, $q(c, \{a, b, c\}) \to q(b, \{a, b\}) \to y(b, \{a\})$ represents type (b, u_2) where $u_2(c) > u_2(b) > u_2(a)$ (the green path in Figure 1). In this path, the leakage appears at node $\{a\}$, which implies the pre-recommendation choice must be the alternative just above a in u_2, which is b. Each of these two paths corresponds to a unique type. However, the path $q(b, \{a, b, c\}) \to y(b, \{a, c\})$ corresponds to two types (b, u_3) and (b, u_4) where $u_3(b) > u_3(a) > u_3(c)$ and $u_4(b) > u_4(c) > u_4(a)$, respectively. Notice that these two types cannot be distinguished because they always choose b. In other words, we cannot identify the relative ranking of a and c, which does not matter for the choices. This is a trivial non-uniqueness of RA-RUM. Note that throughout these examples, every leakage path y follows its immediate preceding q (each path flows straight at the last node, *i.e.,* $q(a, A \cup a) \to y(a, A)$ for $a \notin A$). If a path ends with $q(a, A\cup a) \to y(a, A)$, the corresponding pre-recommendation choice is $a = c(\emptyset)$, and a yields a higher utility compared to all alternatives in A. As it turns out, all types in RA-RUM can be represented by such paths.

The observations above can be generalized. If we have a path $q(a_1, A_0) \to \cdots \to q(a_k, A_{k-1}) \to y(a_k, A_k)$, where $A_0 = X$, and $A_i = A_{i-1} \setminus a_{i-1}$ for $i \le k$, then any type (a, u) represented by the path must obey the following properties:

1. $a = a_k$,
2. $u(a_i) > u(a_j)$ for $i < j \le k$,
3. $u(a) > \max_{x\in A_k} u(x)$.

The first property says that the pre-recommendation choice must be the last alternative in the sequence, i.e., $c_{(a,u)}(\emptyset) = a_k$. The second condition states u is decreasing in i in $X \setminus A_k$.

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

In addition, the last condition implies that $c_{(a,u)}(A_k)$ must be a_k. There must be no variation in the behavior of choice of any type (a, u) represented by the path. When we examine the uniqueness properties of RA-RUM, we will be able to make use of this discussion.

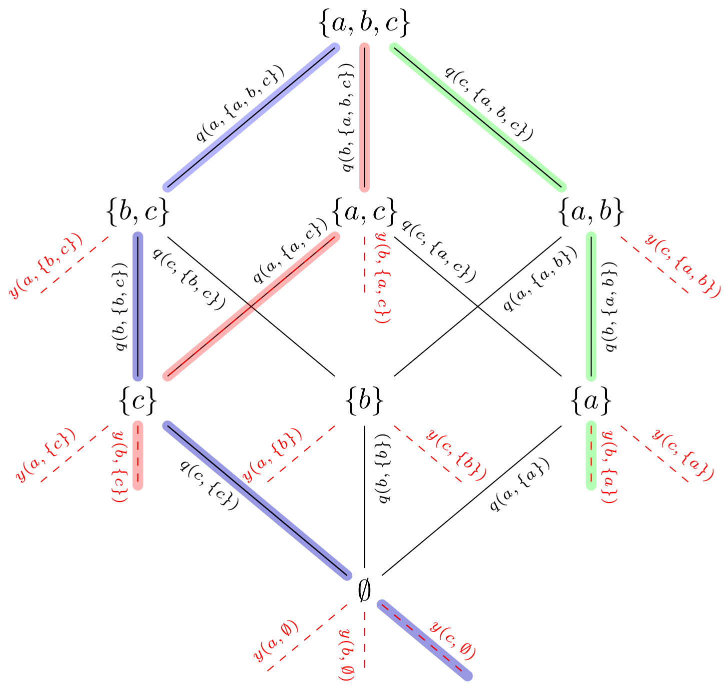

Figure 2. Hasse Diagram of RP-RUM: Each colored path represents a specific set of types.

On the contrary, in RP-RUM, every possible path corresponds to a type (Figure 2). Firstly, it allows paths that are permitted in RA-RUM. For example, $q(a, \{a, b, c\}) \to q(b, \{b, c\}) \to q(c, \{c\}) \to y(c, \emptyset)$ (the blue path in Figure 2) represents the choice type $c_{(u_1,v_1)}$, where $u_1(a) + v_1(a) > u_1(b) + v_1(b) > u_1(c) + v_1(c) > u_1(c) > u_1(b) > u_1(a)$. Even though this path yields a unique type in RA-RUM, this is no longer true in RP-RUM. Indeed, it also represents the type (u_2, v_2), where $u_2(a) + v_2(a) > u_2(b) + v_2(b) > u_2(c) + v_2(c) > u_2(c) > u_2(a) > u_2(b)$. There are several similarities between these types: i) $u_1 + v_1$ and $u_2 + v_2$ induce the same ordering, ii) c is the worst for $u_1 + v_1$ and $u_2 + v_2$, and iii) the lower counter set of c with respect to u_1 and u_2 are the same. Notice $c_{(u_1,v_1)}$ and $c_{(u_2,v_2)}$ are inherently indistinguishable even in the deterministic model. This is a trivial non-uniqueness of RP-RUM.

Moreover, RP-RUM allows paths that are not permitted in RA-RUM. For instance, RP-RUM allows the path $q(b, \{a, b, c\}) \to q(a, \{a, c\}) \to y(b, \{c\})$ (the red path in Figure 2). This path captures the choice type $c_{(u,v)}$, where $u(b)+v(b) > u(a)+v(a) > u(b) > u(a) > u(c)+v(c) > u(c)$. Note that $c_{(u,v)}(\{a, b\}) = c_{(u,v)}(\emptyset) = b$ but $c_{(u,v)}(\{a\}) = a$. Hence Sandwich Property is violated. That is why this path is not permitted in RA-RUM. Note that this path ends with $q(a, \{a, c\}) \to y(b, \{c\})$, which is not allowed in RA-RUM. Here, this implies that the corresponding pre-recommendation choice is $b = c(\emptyset)$, and b yields the highest u-value in X. Therefore if this path represents any type (u, v), then i) b must be the u-maximizer, ii) $u(b)+v(b) > u(a)+v(a) > u(b)$, and iii) $u(b) > \max\{u(a), u(c) + v(c)\}$. Again, all types represented by the same path induce the same choice behavior. Hence, paths are the only object we need to summarize the choices in RP-RUM.

We can generalize the observations above. If we have a path $q(a_1, A_0) \to \cdots \to q(a_k, A_{k-1}) \to y(b, A_k)$ where $b = a_i$ for some $i \leq k$, $A_0 = X$, and $A_i = A_{i-1} \setminus a_{i-1}$ for $i \leq k$, then any type (u, v) represented by the path must obey the following properties:

1. $u(b) = \max_{x \in X} u(x)$,
2. $u(a_i) + v(a_i) > u(a_j) + v(a_j)$ for $i < j \leq k$,
3. $\min_{x \in X \setminus A_k} u(x) + v(x) > u(b) > \max_{x \in A_k} u(x) + v(x)$.

The first property says that b must be the u-maximizer in X, i.e, $c_{(u,v)}(\emptyset) = b$. The second condition identifies the ordinal ranking of $u + v$, which is decreasing in i. In addition, the last condition implies that $c_{(u,v)}(A_k)$ must be b. A combination of all conditions implies $c_{(u,v)}(A_{i-1}) = a_i$. Any type (u, v) represented by the path must induce the same choice behavior. This discussion will be useful when we investigate the uniqueness properties of RP-RUM.

4.8 Behavioral Characterization for Non-Parametric Choice

We now introduce behavioral postulates for RP-RUM and RA-RUM. The first axiom is the non-negativity of BM polynomials in the recommendation domain. This non-negativity is closely related to the non-negativity of BM polynomials in the standard domain. In the standard domain, the non-negativity of the BM polynomials means that items must be chosen marginally more if there are fewer available items. Here, the non-negativity of $y's$ and $q's$ means that both recommended and non-recommended alternatives must be chosen marginally more if there are fewer recommended products.

Non-negativity of BM. *For $a \in R$, $q(a, R) \geq 0$ and $y(a, R \setminus a) \geq 0$.*

There is another important property in this domain. Notice that the positive recommendation assumption suggests that the difference between $\rho(a, R)$ and $\rho(a, R \setminus a)$ must be weakly positive. Here, we look into a finer detail of this effect. Notice that the difference $q(a, R) - y(a, R \setminus a)$ captures the idea of the marginal change in the magnitude of the positive recommendation effect. To see this, one can interpret this axiom as the comparison between the marginal gain from regularity in choice probabilities of alternatives when they are recommended and not recommended. In particular, this behavioral postulate says the marginal gain will always be higher if recommended.

Positive Marginal Recommendation. *For $a \in R$, $q(a, R) \geq y(a, R \setminus a)$.*

We show that the non-negativity of BM polynomials alone fully characterizes RP-RUM. To characterize RA-RUM, we need an additional condition: Positive Marginal Recommendation. Notice that by requiring qs' greater than ys', the non-negativity of qs' is immediately implied.

Since the two models are general, we need a rich data set for characterization. Therefore, we assume that $\mathcal{D} = 2^X$ for the next theorem.[20]

Theorem 5 (Characterization)**.** Let $\mathcal{D} = 2^X$. Then,
a) ρ has an RP-RUM representation if and only if ρ satisfies Non-negativity of BM;
b) ρ has an RA-RUM representation if and only if ρ satisfies Non-negativity of BM and Positive Marginal Recommendation.

The sufficiency proof of the theorem is constructive. We provide an algorithm to compute a full distribution of types in the RA-RUM and RP-RUM representation. The algorithm is certainly helpful in applications if one would like to have an estimate of the type distribution. However, just as the RUM in the standard choice domain, in general, there is no unique distribution over types that can explain the choice data. We explore more on this in the next subsection.

4.9 Uniqueness and Identification for Non-Parametric Choice

In this section, we elaborate on to what extent the type distribution can be identified. We first look at the uniqueness result for RA-RUM. Let $L_u(a)$ represent the strict lower contour set of a according to u. We state the uniqueness result in the following proposition.

Proposition 7 (Uniqueness and identification of RA-RUM)**.** Let τ^1 and τ^2 be two RA-RUM representations of the same choice data. Then for every $R \subseteq X$, $b \notin R$ and $i = 1, 2$,
i) $y_\rho(b, R) = \tau^i\Big(\{(b, u) | \min_{x \in X \setminus R \cup b} u(x) > u(b) > \max_{x \in R} u(x)\}\Big)$,
ii) $q_\rho(b, R \cup b) = \tau^i\Big(\{(a, u) | \min_{x \in X \setminus R \cup b} u(x) > u(b) > \max_{x \in R} u(x) \text{ and } a \in R \cup b\}\Big)$.

The first property in Proposition 7 says that $y_\rho(b, R)$ must be equal to the probability that b is the default option while b is exactly and only better than the set of alternatives R. On the other hand, the second property says that $q_\rho(b, R)$ must be equal to the probability that b is exactly and only better than the set of alternatives R and the default is within $R \cup b$. Note that these two conditions imply that

$$q_\rho(b, R \cup b) - y_\rho(b, R) = \tau^i\Big(\{(a, u) | \min_{x \in X \setminus R \cup b} u(x) > u(b) > \max_{x \in R} u(x) \text{ and } a \in R\}\Big)$$

where the LHS is exactly the object of Positive Marginal Recommendation. From here, one can immediately see that $q's$ must be greater than its respective $y's$. Notice that this difference captures the fraction of people who rank b exactly above R while their default is in R. Surely, this fraction of people would switch to b if b is included in the recommendation set.

We now look at the uniqueness result for RP-RUM.

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

Proposition 8 (Uniqueness and identification of RP-RUM). Let μ^1 and μ^2 be two RP-RUM representations of the same choice data. Then for every $R \subseteq X$, $b \in X \setminus R$ and $i = 1, 2$,
i) $y_\rho(b, R) = \mu^i\Big(\{(u, v)| \min_{x \in X\setminus R} u(x) + v(x) > u(b) > \max\{\max_{x\in R} u(x) + v(x), \max_{x\in X\setminus b} u(x)\}\}\Big)$,
ii) $q_\rho(b, R\cup b) = \mu^i\Big(\{(u, v)| \min_{x\in X\setminus R} u(x)+v(x) > u(b)+v(b) > \max\{\max_{x\in R} u(x)+v(x), \max_{x\in X} u(x)\}\}\Big)$.

Notice that the uniqueness result of the RP-RUM is closely related to the uniqueness result of the RA-RUM model. Firstly, here $y_\rho(b, R)$ represents all types that assign higher (lower) $u+v$ value to all alternatives in $X \setminus R$ (R) than $u(b)$. Notice that $X \setminus R$ also includes b. In addition, for these types, b attains the maximum u-value in X.

On the other hand, $q_\rho(b, R \cup b)$ represents all types that assign higher (lower) $u + v$ value to all alternatives in $X \setminus R$ (R) than $u(b) + v(b)$. Notice that $X \setminus R$ also includes b. Therefore, it also captures the different possibilities of $u(b) + v(b)$ above b. We can see from this result why Positive Marginal Recommendation is not satisfied.

The next example illustrates our identification and characterization results.

Example 2. Consider a simple parametric example where $X = \{a, b\}$ and we observe the following data.

$\rho_\lambda(\cdot, R)$	$\{a, b\}$	$\{a\}$	$\{b\}$	$\emptyset$
a	λ	0.7	0.1	0.6
b	$1 - \lambda$	0.3	0.9	0.4

To answer whether this choice data has RA-RUM or RP-RUM representations, we first calculate the corresponding q_ρ and y_ρ, which are depicted in the following Hasse diagram in Figure 3. If λ is between 0.1 and 0.7, all q_ρ and y_ρ are non-negative. Hence Non-negativity of BM is satisfied for those λs. By Theorem 5, this data has an RP-RUM representation as long as $\lambda \in [0.1, 0.7]$. Positive Marginal Recommendation is only satisfied when λ equals 0.2. Again by Theorem 5, this data has an RA-RUM representation for $\lambda = 0.2$. While a characterization theorem can inform of the existence or non-existence of a particular representation, such a theorem is not useful in finding such a representation. We can use our identification results and the corresponding Hasse diagram to do that.

Since our data has an RP-RUM representation for all $\lambda \in [0.1, 0.7]$, there exists μ_λ representing ρ_λ. For illustration purposes, take $\lambda = 0.4$. We can use Proposition 8 to identify $\mu_{0.4}$. We first identify possible paths and their corresponding weights. Hasse diagram shows that we must assign at least a flow of 0.1 to the path $q(a, \{a, b\}) \to y(a, \{b\})$ and the flow of 0.3 to the path $q(b, \{a, b\}) \to y(b, \{a\})$. This means that we identify two types: $u_1(a)+v_1(a) > u_1(a) > u_1(b)+v_1(b) > u_1(b)$ and $u_2(b)+v_2(b) > u_2(b) > u_2(a)+v_2(a) > u_2(a)$. Hence, $\mu_{0.4}(\{(u_1, v_1)\}) = 0.1$ and $\mu_{0.4}(\{(u_2, v_2)\}) = 0.3$. In addition, we must assign a flow of 0.2 to the path $q(a, \{a, b\}) \to q(b, \{b\}) \to y(a, \emptyset)$. This path corresponds type (u_3, v_3) such that $u_3(a) + v_3(a) > u_3(b) + v_3(b) > u_3(a) > u_3(b)$. Note that $c_{(u_3,v_3)}(\{a, b\}) = c_{(u_3,v_3)}(\emptyset) = a$ but $c_{(u_3,v_3)}(\{b\}) = b$. Hence Sandwich Property is violated. Since we must have this type in

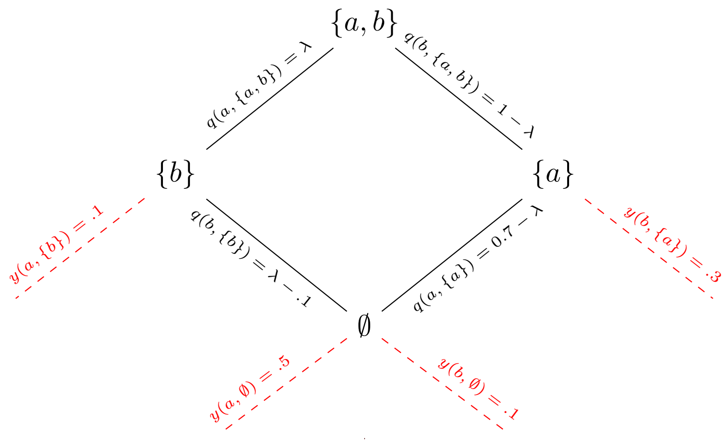

Figure 3. Hasse Diagram of Example 2.

Path	Type	$\mu_{0.4}$	$\mu'_{0.4}$
$q(a,\{a,b\}) \to y(a,\{b\})$	$u_1(a)+v_1(a) > u_1(a) > u_1(b)+v_1(b) > u_1(b)$	0.1	0.1
$q(b,\{a,b\}) \to y(b,\{a\})$	$u_2(b)+v_2(b) > u_2(b) > u_2(a)+v_2(a) > u_2(a)$	0.3	0.3
$q(b,\{a,b\}) \to q(a,\{a\}) \to y(b,\emptyset)$	$u_3(a)+v_3(a) > u_3(b)+v_3(b) > u_3(a) > u_3(b)$	0.2	0.3
$q(b,\{a,b\}) \to q(a,\{a\}) \to y(a,\emptyset)$	$u_4(b)+v_4(b) > u_4(a)+v_4(a) > u_4(a) > u_4(b)$	0.3	0.2
$q(a,\{a,b\}) \to q(b,\{b\}) \to y(b,\emptyset)$	$u_5(a)+v_5(a) > u_5(b)+v_5(b) > u_5(b) > u_5(a)$	0.1	0
$q(a,\{a,b\}) \to q(b,\{b\}) \to y(a,\emptyset)$	$u_6(b)+v_6(b) > u_6(a)+v_6(a) > u_6(b) > u_6(a)$	0	0.1

Table 2. Two RP-RUM representations for $\rho_{0.4}$

the support of $\mu_{0.4}$, this data cannot be represented by RA-RUM. The rest of the flows can be attached to $q(b,\{a,b\}) \to q(a,\{a\}) \to y(a,\emptyset)$ and $q(a,\{a,b\}) \to q(b,\{b\}) \to y(b,\emptyset)$. They correspond to (u_4,v_4) and (u_5,v_5) where $\mu_{0.4}(\{(u_4,v_4)\}) = .3$ and $\mu_{0.4}(\{(u_5,v_5)\}) = .1$ where $u_4(b)+v_4(b) > u_4(a)+v_4(a) > u_4(a) > u_4(b)$ and $u_5(a)+v_5(a) > u_5(b)+v_5(b) > u_5(b) > u_5(a)$. We should note that $\mu_{0.4}$ is not unique. The next table provides an alternative representation for the same choice data.

We now illustrate the identification for RA-RUM. Note that we have an RA-RUM representation only when $\lambda = 0.2$. Let $\mu_{0.2}$ representing $\rho_{0.2}$. We can use Proposition 7 to identify $\mu_{0.2}$. We first identify possible paths and their corresponding weights. Hasse diagram shows that we must assign at least a flow of 0.1 to the path $q(a,\{a,b\}) \to y(a,\{b\})$ and the flow of 0.3 to the path $q(b,\{a,b\}) \to y(b,\{a\})$. This means that we identify two types: (a,u_1) and (b,u_2) where $u_1(a) > u_1(b)$ and $u_2(b) > u_2(a)$. Hence, $\mu_{0.2}(\{(a,u_1)\}) = 0.1$ and $\mu_{0.2}(\{(b,u_2)\}) = 0.3$. In addition, we must assign a flow of 0.1 to the path $q(a,\{a,b\}) \to q(b,\{b\}) \to y(b,\emptyset)$. This path corresponds to Type (b,u_1). The rest of the flows must be attached to $q(b,\{a,b\}) \to q(a,\{a\}) \to y(a,\emptyset)$, which corresponds to Type (a,u_2). Hence $\mu_{0.2}(\{(b,u_1)\}) = 0.1$ and $\mu_{0.2}(\{(a,u_2)\}) = 0.5$. We should note that $\mu_{0.2}$ is unique.

$\lrcorner$

We finish this section by establishing the relationships between Luce versus RUM versions of our models. It is a well-known fact that the Luce model is a special case of the random utility model. In the recommendation environment, one might wonder whether RA-Luce belongs to RA-RUM and/or RP-Luce belongs to RP-RUM. It turns out that it is, which we state in the following.

Theorem 6.

a) Every choice rule ρ with an RA-Luce representation has an RA-RUM representation.

b) Every choice rule ρ with an RP-Luce representation has an RP-RUM representation.

5 Conclusion

Recommendations are abundant and prevalent in our lives. In this paper, we consider two different channels through which recommendations can affect choices: attention and preference channels. We first consider two deterministic models separately capturing these channels. We show that the deterministic models share the same behavioral trait: Independence of Irrelevant Recommended Alternatives. Moreover, one can distinguish RP from RA if they observe choice behavior that involves a violation of Sandwich Property. Supported by our deterministic model of behavior, we introduce probabilistic choice models, which can be interpreted as coming from aggregate choice data. We propose parametric versions of the models for tractability and applicability, namely the RA-Luce and RP-Luce models. We show that we can fully distinguish these two models from characterization. Lastly, to enhance the model's explanatory power, we incorporate the idea of the Random Utility Model and propose the RA-RUM and RP-RUM model. We fully characterize these two models and show that each closely connects to the classic well-known standard probabilistic choice models.

While our models investigate two important channels through which a recommendation can affect choice, we expect 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 include developing more elaborate and complex models of recommendation, including strategic recommendations, limited consideration, status quo, behavioral search, satisficing, and temptation. Therefore, we believe this paper also paves a palpable path for fruitful future research and applications where we can apply the economic wisdom accumulated throughout the years for the standard models to this setting.

References

Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2013). Do recommender systems manipulate consumer preferences? A study of anchoring effects. *Information Systems Research*, *24*(4).

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.

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.

Bairathi, M., Zhang, X., & Lambrecht, A. (2022). The value of platform endorsement. *Working Paper*.

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.

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).

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

Cosley, D., Lam, S. K., Albert, I., Konstan, J. A., & Riedl, J. (2003). Is seeing believing? How recommender interfaces affect users' opinions. *Conference on Human Factors in Computing Systems - Proceedings*.

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.

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.

Guney, B., Richter, M., & Tsur, M. (2018). Aspiration-based choice. *Journal of Economic Theory*, *176*.

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.

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

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).

Ke, S., Wu, B., & Zhao, C. (2021). Learning from a Black Box. *Working Paper*.

Kovach, M., & Tserenjigmid, G. (2022a). Behavioral foundations of nested stochastic choice and nested logit. *Journal of Political Economy*, *130*(9), 2411–2461.

Kovach, M., & Tserenjigmid, G. (2022b). The focal luce model. *American Economic Journal: Microeconomics*, *14*(3), 378–413.

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.

Manzini, P., & Mariotti, M. (2007). Sequentially rationalizable choice. *American Economic Review*, *97*(5), 1824–1839.

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. (2013). Choice by iterative search. *Theoretical Economics*, *8*(3), 701–728.

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.

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*, *67*(11), 6915–6933.

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

6 Appendix

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

Proof of Theorem 1(a)

Proof. Since u and v are ordinal, we would prove the theorem using the ordinal version of the model. In particular, we translate u and v into a preference ordering (i.e., a linear order) $\succ^\star$ on $X \cup X^*$, where X^* contains typical element x^* which represents the recommended version of x. Since $v \geq 0$, for each alternative x, it must be that $x^* \succ^\star x$.

We use σ as an arbitrary element of $X \cup X^*$. For every set R, we define R_x as x^* if x belongs to R, otherwise R_x is simply x. We now define a binary relation on $X \cup X^*$ for any two distinct alternatives x and y:

$$(R_x, R_y) \in P \text{ if } x = c(R).$$

Claim 1. *P is asymmetric.*

Proof. For two distinct σ_1 and σ_2, if $(\sigma_1, \sigma_2) \in P$ then $(\sigma_2, \sigma_1) \notin P$. Assume not, there exists R_1 and R_2 such that $x = c(R_1)$ and $y = c(R_2)$. There are four cases depending on whether x or y is recommended. If x is recommended, then $\sigma_1 = x^*$, x otherwise. Similarly, σ_2 is y if it is not recommended, y^* otherwise.

First we assume that $x, y \in R_1 \cap R_2$, then by IIRA(1), we must have $x = c(\{x, y\}) = y$, a contradiction. Now assume $x \in R_1 \cap R_2$ and $y \notin R_1 \cup R_2$, then IIRA(1) implies $x = c(\{x, y\})$ whereas IIRA(2) implies $y = c(\{x, y\})$, a contradiction. The third case where $x \notin R_1 \cap R_2$ and $y \in R_1 \cup R_2$ is identical to the second case. Finally, assume $x, y \notin R_1 \cap R_2$. Then applying IIRA(2) twice yields $x = c(\{x, y\}) = y$, a contradiction. Therefore, P is asymmetric. ■

Let a be $c(\emptyset)$. Given the definition P, we reveal that a is preferred to any alternative in $X \setminus a$. Moreover, we cannot reveal the relative ranking of these alternatives. Formally, for $x, y \neq a$, aPx and aPy then $(x, y) \notin P$ or $(y, x) \notin P$. Notice that a might be revealed to be better than alternatives in X^*. Especially, if $a = c(R)$, then aPy^* for all $y \in R$. The same intuition applies to these alternatives too. Hence, P is incomplete for the lower contour set of a.

On the other hand, P is complete in the upper contour set of a. To show this, assume x^*Pa and y^*Pa for $x, y \neq a$. In other words, there exist $R \ni x$ and $R' \ni y$ such that $x = c(R)$ and $y = c(R')$ where $a \notin R \cup R'$. If $x \in R'$ or $y \in R$, we would reveal y^*Px^* or x^*Py^*, respectively. Assume not. Then consider $\{x, y\}$ as the recommended set. First, $c(\{x, y\}) \in \{x, y\}$. If not, $a = c(\{x\})$ by IIRA(2). Then by IIRA(1), x cannot be chosen from $c(R)$. Hence, either y^*Px^* or x^*Py^*, P is complete for the upper contour set of a.

Claim 2. *If x^*Py^* and $x^* \neq a$ then x^*Pa.*

Proof. x^*Py^* implies that there exists R such that $x, y \in R$ and $x = c(R)$. Since $\{x\} \subset R$, IIRA(1) implies $c(\{x\}) = x$ implying x^*Pa. ■

Note that our definition of P does not relate a and a^* since it requires two distinct alternatives. However, it is possible that we can conclude that a^*Pa. To see this, assume $a = c(R)$ where $a, x \in R$ and $x = c(R')$ where $a \notin R'$. The former implies a^*Px^* and the former yields x^*Pa. Hence we infer that a is ranked strictly higher when recommended.

Claim 3. *P is transitive in the strict upper contour set set of a.*

Proof. Assume that $x^*Py^*Pz^*$ for three distinct x, y, and z. There exists R and R' such that (i) $x = c(R)$ and $y = c(R')$, and ii) $\{x, y\} \subset R$ and $\{y, z\} \subset R'$. IIRA(1) implies that $c(\{x, y\}) = x$ and $c(\{y, z\}) = y$. Then we consider $c(\{x, y, z\})$. We must have $c(\{x, y, z\}) \in \{x, y, z\}$ by IIRA(2). It cannot be z since IIRA(1) implies $z = c(\{y, z\})$. It cannot be y since IIRA(1) implies $y = c(\{x, y\})$ which is a contradiction. Hence, we have $x = c(\{x, y, z\})$, which gives x^*Pz^*. ■

Take any completion of the transitive closure of P such that x^*Px for x and call it $\succ^\star$. Note that the transitive closure would immediately imply x^*Px if there exists $R \neq \emptyset$ such that $c(R) = x$ by using Claim 2 and the definition of P by the data $c(\emptyset) = a$. For x that is never chosen, x^*Px is trivial.

It is routine to show that the representation holds by choosing u and v appropriately. Note that $v \geq 0$ by the fact of $x^* \succ^\star x$. ■

Proof of Theorem 1(b)

Proof. Since u is purely ordinal, in the proof, we would translate u into a preference ordering (i.e., a linear order) $\succ$ on X. We first identify the default option a. We set $a := c(\emptyset)$. If $c(R) \notin R$ then by IIRA(2), we have $c(R) = a$. Hence, a is unique. For every distinct $x, y \in R \cup a$, we write

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

Claim 4. *P is asymmetric*

Proof. For two distinct x and y, if $(x, y) \in P$ then $(y, x) \notin P$. Assume not. Then there exists R and R' such that $x = c(R)$ and $y = c(R')$ and $\{x, y\} \subset (R \cap R') \cup a$. If $x, y \neq a$, then by IIRA(1), we must have $x = c(\{x, y\}) = y$. On the other hand, if $x = a$, then $y \neq a$. By IIRA(1), we have $c(\{y\}) = y$. However, by Sandwich Property, since $\emptyset \subseteq \{y\} \subseteq R$, we must have $x = c(\{y\})$. A contradiction must arise regardless. Therefore, P is asymmetric. ■

Note that if the default option is preferred to two distinct alternatives, we cannot reveal the relative ranking of these alternatives. In other words, for $x, y \neq a$, 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, $c(\{x, y\}) \in \{x, y\}$. If not, $a = c(\{x\})$ by Sandwich Property. Then by IIRA(1), x cannot be chosen from $c(R)$. Hence, either xPy or yPx, P is complete for the upper counter set of a.

Claim 5. *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 $\{x\} \subset R$, IIRA(1) implies $c(\{x\}) = x$ implying xPa. ■

Claim 6. *If $xPyPz$ then xPz.*

Proof. First, note that x cannot be a since P is silent for the lower counter set of a. If z is equal to a, by Claim 5, 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 IIRA(1), Sandwich Property. To see this, suppose $c(\{x, z\}) \notin \{x, z\}$, then $a = c(\{x, z\})$ by IIRA(2). However, by Sandwich Property, we must have $c(\{x\}) = a$. On the other hand, if $z = c(\{x, z\})$, by IIRA(1), we must have $c(\{z\}) = z$. Contradictions in either way. Finally, we assume that x, y, z are distinct from a. Then we consider $c(\{x, y, z\})$. We must have $c(\{x, y, z\}) \in \{x, y, z\}$ by Sandwich Property. It cannot be z since IIRA(1) implies $z = c(\{y, z\})$ which contradicts Claim 4. Finally, it cannot be y since IIRA(1) implies $y = c(\{x, y\})$ which contradicts Claim 4. Hence, we have $x = c(\{x, y, z\}$, which gives xPz. ■

Take any completion of P and call it $\succ$. It is routine to show that the representation holds by choosing u appropriately. ■

Proof of Proposition 1

Proof. Suppose (a_1, u_1) and (a_2, u_2) represents the same choice rule. For i), note that it is immediate that $a_1 = c(\emptyset) = a_2$. For ii) Suppose not. There exists b such that $u_1(a) > u_1(b)$ but $b_2(b) > u_2(a)$. Then, we know that $c_{(a,u_1)}(\{b\}) = a \neq b = c_{(a,u_2)}(\{b\})$. Contradiction. For iii), suppose not, there exists $x, y \in X$ such that $u_1(x) > u_1(y) > u_1(a)$ but $u_2(y) > u_2(x) > u_2(a)$. Then, we have $c_{(a,u_1)}(\{x, y\}) = x \neq y = c_{(a,u_2)}(\{x, y\})$. Contradiction arises. ■

Proof of Proposition 2

Proof. Suppose (u_1, v_1) and (u_2, v_2) represent the same choice rule. For i), note that it is immediate that $\text{argmax}_{x\in X} u_1(x) = c(\emptyset) = \text{argmax}_{x\in X} u_2(x)$. We call it a. For ii), suppose not, there exists b such that $u_1(a) > u_1(b)+v_1(b)$ and $u_2(b)+v_2(b) > u_2(a)$. Then, we know that $c_{(u_1,v_1)}(\{b\}) = a \neq b = c_{(u_2,v_2)}(\{b\})$. Contradiction arises. For iii), suppose not, there exists $x, y \in X$ such that $u_1(x)+v_1(x) > u_1(y)+v_1(y) > u_1(a)$ but $u_2(y)+v_2(y) > u_2(x)+v_2(x) > u_2(a)$. Then, we have $c_{(u_1,v_1)}(\{x,y\}) = x \neq y = c_{(u_2,v_2)}(\{x,y\})$. Contradiction arises. ■

Proof of Theorem 2

Proof. We first prove the necessity of the axioms. Suppose the model is correct. We prove the necessity of Recommended Luce-IIA. Note that, for $x, y \in B$.

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

Since B is arbitrary, it immediately implies Recommended Luce-IIA. We then prove the necessity of R-Path Independence. Let $x \in B$ and $x \notin A \subseteq B$, we show have $\rho(x, A)\rho(A \cup x, B) = \rho(x, \emptyset)\rho(x, B)$. To see this, we have $\rho(x, A)\rho(A \cup x, B) - \rho(x, \emptyset)\rho(x, B) = d(x)[\frac{u(x)}{u(A\cup x)}[\rho(x, B) + \rho(A, B)] - \rho(x, B)]$ $= \frac{d(x)}{u(A\cup x)}[u(x)\rho(A, B)-u(A)\rho(x, B)]$, where $u(x)\rho(A, B)-u(A)\rho(x, B) = u(x)[\sum_{y\in B}\frac{d(y)}{u(B)}u(A)+\sum_{y\notin B}\frac{d(y)}{w(B\cup y)}u(A)]- u(A)[\sum_{y\in B}\frac{d(y)}{u(B)}u(x) + \sum_{y\notin B}\frac{d(y)}{w(B\cup y)}u(x)] = 0$. Hence, since A is arbitrary, R-Path Independence is proven.

For sufficiency, we define $d(x) := \rho(x, \emptyset) \geq 0$ and $u(x) := \rho(x, X) > 0$. First, R-Path Independence implies that $\rho(x, A)\rho(A \cup x, X) = \rho(x, \emptyset)\rho(x, X)$. Hence we have representation for non-recommended alternative, *i.e.* for $x \notin A$, $\rho(x, A) = \rho(x, \emptyset)\frac{\rho(x,X)}{\rho(A\cup x,X)} = d(x)\frac{u(x)}{\sum\limits_{z\in A\cup x} u(z)} = d(x)W_x(x, A)$. For recommended alternatives, we first make the following claim.

Claim 7. *R-Path Independence 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 R-Path Independence, we have, for every $x \notin X\setminus A$, $\rho(x, A)(\rho(x, X)+\rho(A, X)) = \rho(x, \emptyset)\rho(x, X) \iff \rho(x, A)+\frac{\rho(x,A)}{\rho(x,X)}\rho(A, X) = \rho(x, \emptyset)$. Summing all $x \notin A$ on both side, we have $\sum_{x\notin A}(\rho(x, A) + \frac{\rho(x,A)}{\rho(x,X)}\rho(A, X)) = \sum_{x\notin A}\rho(x, \emptyset) \iff 1 - \rho(A, A) + \rho(A, X)\sum_{x\notin A}\frac{\rho(x,A)}{\rho(x,X)} = 1 - \rho(A, \emptyset) \iff \rho(A, A) - \rho(A, \emptyset) = \rho(A, X)\sum_{y\notin A}\frac{\rho(y,A)}{\rho(y,X)}$. ■

By Recommended Luce-IIA, for $x, y \in A$ we have $\frac{\rho(y,A)}{\rho(x,A)} = \frac{\rho(y,X)}{\rho(x,X)}$. By summing all $y \in A$, we have $\frac{\rho(A,A)}{\rho(x,A)} = \frac{\rho(A,X)}{\rho(x,X)} \iff \rho(x, A) = \frac{\rho(A,A)\rho(x,X)}{\rho(A,X)}$. Hence, for $x \in A$, by Claim 7 and expanding on $\rho(A, A)$

$$\begin{aligned}\rho(x, A) &= \frac{\rho(x, X)}{\rho(A, X)}\Big[\rho(A, \emptyset) + \rho(A, X)\sum_{y\notin A}\frac{\rho(y, A)}{\rho(y, X)}\Big] = \frac{\rho(x, X)}{\rho(A, X)}\Big[\rho(A, \emptyset) + \rho(A, X)\sum_{y\notin A}\frac{\rho(y, \emptyset)}{\rho(A \cup y, X)}\Big] \\ &= \rho(x, X)\Big[\frac{\rho(A, \emptyset)}{\rho(A, X)} + \sum_{y\notin A}\frac{\rho(y, \emptyset)}{\rho(A \cup y, X)}\Big] = \sum_{y\in A}\frac{d(y)u(x)}{\sum\limits_{z\in A} u(z)} + \sum_{y\notin A}\frac{d(y)u(x)}{\sum\limits_{z\in A\cup y} u(z)} = \sum_{a\in X} d(a)W_a(x, A)\end{aligned}$$

Where the second equality is due to $\rho(y, A) = \rho(y, \emptyset)\frac{\rho(y,X)}{\rho(A\cup y,X)}$ for $y \notin A$ and the fourth equality is due to construction of d and u. The proof is complete. ■

Proof of Theorem 3

Proof. The proof for the necessity of Recommended Luce-IIA is proven in Theorem 2. We prove the necessity of R-Independence. Suppose the model is correct, for $x \notin A$, $\frac{\rho(x,\emptyset)}{\sum_{z\in A\cup x} r(z,x)} = \frac{d(x)}{\sum_{z\in A\cup x}\frac{u(z)}{u(x)}} =$

$\frac{d(x)u(x)}{\sum_{z\in A\cup x} u(z)} = \rho(x, A)$, where the first equality given is by the necessity proof of Recommended Luce-IIA. Since A is arbitrary, it is proven.

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

Claim 8. *For any $x, y \in A$ with $|A| \leq k-1$, we have $\frac{\rho(x,A)}{\rho(y,A)} = \frac{u(x)}{u(y)}$. And similarly, for any $x \in A$ and $B \subseteq A$ we have $\frac{\rho(x,A)}{\rho(B,A)} = \frac{u(x)}{u(B)}$*

Proof. Note that, firstly, for any set $A \supseteq \{x, z_0\}$, we have $u(x) = \frac{\rho(x,A)}{\rho(z_0,A)}$ by Recommended Luce-IIA. Then, for any $x, y \in A$, we have $\frac{\rho(x,A)}{\rho(y,A)} = \frac{\rho(x,A\cup z_0)}{\rho(y,A\cup z_0)} = \frac{\rho(x,A\cup z_0)\rho(z_0,A\cup z_0)}{\rho(z_0,A\cup z_0)\rho(y,A\cup z_0)} = \frac{u(x)}{u(y)}$. Hence, the first part is proven. The second part is immediate. ■

We first show that the representation holds for the off-recommendation set. Let $x \notin A$, by R-Independence and the Claim 8, we have $\rho(x, A) = \frac{\rho(x,\emptyset)}{1+\sum_{z\in A} r(z,x)} = \frac{d(x)}{\sum_{z\in A\cup x} \frac{u(z)}{u(x)}} = \frac{d(x)u(x)}{u(A\cup x)}$. Hence, the representation holds for off-recommendation alternative.

Then, we show that the representation holds for recommended items (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 $\rho(x, A) + \sum_{y\in A\setminus x} \rho(y, A) + \sum_{y\in X\setminus A} \rho(y, A) = 1 \iff = \rho(x, A) + \sum_{y\in A\setminus x} \rho(x, A)\frac{\rho(y,\{x,y\})}{\rho(x,\{x,y\})} = 1 - \sum_{y\in X\setminus A} \rho(y, A)$ (By Recommended Luce-IIA) $\iff \rho(x, A) + \sum_{y\in A\setminus x} \rho(x, A)\frac{u(y)}{u(x)} = 1 - \sum_{y\in X\setminus A} \rho(y, A)$ (By Claim 8) $\iff \rho(x, A)\frac{u(A)}{u(x)} = 1 - \sum_{y\in X\setminus A} \rho(y, A) \iff \rho(x, A) = \frac{u(x)}{u(A)}\left[1 - \sum_{y\in X\setminus A} \frac{d(y)u(y)}{u(A\cup y)}\right]$ (By construction of $\rho(y, A)$ for $y \notin A$). Then, it remains to show that $\frac{u(x)}{u(A)}\left[1 - \sum_{y\in X\setminus A} \frac{d(y)u(y)}{u(A\cup y)}\right] = u(x)\left[\frac{d(A)}{u(A)} + \sum_{y\in X\setminus A} \frac{d(y)}{u(A\cup y)}\right]$. Note that $\frac{u(x)}{u(A)}\left[1 - \sum_{y\in X\setminus A} \frac{d(y)u(y)}{u(A\cup y)}\right] - u(x)\left[\frac{d(A)}{u(A)} + \sum_{y\in X\setminus A} \frac{d(y)}{u(A\cup y)}\right] = u(x)\left[\frac{1-d(A)}{u(A)} - \sum_{y\in X\setminus A} [\frac{d(y)u(y)}{u(A)u(A\cup y)} + \frac{d(y)}{u(A\cup y))}]\right] = u(x)\left[\sum_{y\in X\setminus A} d(y)[\frac{1}{u(A)} - \frac{u(y)}{u(A)u(A\cup y)} - \frac{1}{u(A\cup y)}]\right] = 0$. Hence, we have shown that, for $x \in A$, $\rho(x, A) = u(x)\left[\frac{d(A)}{u(A)} + \sum_{y\in X\setminus A} \frac{d(y)}{u(A\cup y)}\right]$. By re-arrangement, one can see that it is the representation for $\rho(x, A)$ where $x \in A$. It is proven. ■

Proof of Theorem 4

Proof. The necessity proof is straightforward. In the following, we will adopt three IIA implied by Strong Luce-IIA. Namely, Recommended Luce-IIA, Non-recommended Luce-IIA and Cross-status Luce-IIA, which are discussed in the main text.

For the sufficiency, we will prove it with two different cases. The first case is that $|X| \geq 3$. we let $u(x) := \rho(x, \emptyset)$. We designate an anchoring element z_0 such that for every $x \in X \setminus z_0$, we let $l(x|R, z_0) := \frac{\rho(x,R)}{\rho(z_0,R)}\rho(z_0, \emptyset)$ for some R such that $x \in R$ and $z_0 \notin R$. Firstly, note that denominator is greater than 0 by the assumption that choice probability is positive.

Claim 9. *For $R \neq X$, $y, z \notin R$ and $x \in R$, $l(x|\{x\}, z) = l(x|R, y)$.*

Proof. To see this, let $y \notin R$, by Non-recommended Luce-IIA, we have $\rho(y, \{x\}) = \frac{\rho(z,\{x\})}{\rho(z,\emptyset)}\rho(y, \emptyset)$, which we substitute into $\frac{\rho(x,\{x\})}{\rho(x,R)} = \frac{\rho(y,\{x\})}{\rho(y,R)}$ (by Cross-status Luce-IIA). By simplifying, we get $l(x|R, z) := \frac{\rho(x,\{x\})}{\rho(z,\{x\})}\rho(z, \emptyset) = \frac{\rho(x,R)}{\rho(y,R)}\rho(y, \emptyset) := l(x|R, y)$. ■

Hence, the definition of l is independent of the anchoring element z_0. We will write $l(x)$ to replace $l(x|R, z_0)$ for every $x \notin z_0$. Also, we can define $l(z_0)$ which any other arbitrary anchor.

Claim 10. *1) For $x, y \notin R$, $\frac{\rho(x,R)}{\rho(y,R)} = \frac{u(x)}{u(y)}$. 2) For $x, y \in R$, $\frac{\rho(x,R)}{\rho(y,R)} = \frac{l(x)}{l(y)}$. 3) For $x \in R$ and $y \notin R$, $\frac{\rho(x,R)}{\rho(y,R)} = \frac{l(x)}{u(y)}$.*

Proof. To see 1) holds, note that Non-recommended Luce-IIA implies $\frac{\rho(x,R)}{\rho(y,R)} = \frac{\rho(x,\emptyset)}{\rho(y,\emptyset)} = \frac{u(x)}{u(y)}$. To see 2) holds, note that for $R \neq X$ and some $z \notin R$, $\frac{\rho(x,R)}{\rho(y,R)} = \frac{\rho(x,R)}{\rho(y,R)}\frac{\rho(z,\emptyset)}{\rho(z,\emptyset)}\frac{\rho(z,R)}{\rho(z,R)} = \frac{l(x)}{l(y)}$. For $R = X$, we use Recommended Luce-IIA so that $\frac{\rho(x,X)}{\rho(y,X)} = \frac{\rho(x,\{x,y\})}{\rho(y,\{x,y\})} = \frac{\rho(x,\{x,y\})}{\rho(y,\{x,y\})}\frac{\rho(z,\emptyset)}{\rho(z,\emptyset)}\frac{\rho(z,\{x,y\})}{\rho(z,\{x,y\})} = \frac{l(x)}{l(y)}$ for some $z \notin \{x,y\}$. To see 3) holds, $\frac{\rho(x,R)}{\rho(y,R)} = \frac{\rho(x,R)}{\rho(y,R)}\frac{\rho(y,\emptyset)}{\rho(y,\emptyset)} = \frac{l(x)}{u(y)}$. The proof is complete. ■

Lastly, for $x \in R$, taking summation over all choice ratio of every elements in X with $\rho(x,R)$ being at the denominator, we have $\sum_{y\in X}\frac{\rho(y,R)}{\rho(x,R)} = \frac{\sum_{y\in R} l(x)+\sum_{y\in R^c} u(x)}{l(x)} \iff \rho(x,R) = \frac{l(x)}{\sum_{y\in R} l(x)+\sum_{y\in R^c} u(x)}$. One can apply the same argument analogously to the case that $x \notin R$. Then, we define $v(x) := l(x)-u(x)$. The representations hold if $l(x) \geq u(x)$ for every $x \in X$. To see this, We check $\rho(y,\{x\})$ and $\rho(y,\emptyset)$ for some $y \neq x$. It is easy to show that $l(x) \geq u(x)$ if and only if $\rho(y,\{x\}) = \frac{u(y)}{\sum_{z\in X\setminus x} u(z)+l(x)} \geq \frac{u(y)}{\sum_{z\in X\setminus x} u(z)+u(x)} = \rho(y,\emptyset)$ which is given by R-Regularity. Hence, the proof is complete. ■

Proof of Proposition 3

Proof. Suppose that (u_1,d_1) and (u_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 $\frac{u_1(x)}{\sum_{x\in X} u_1(x)} = \rho(x,X) = \frac{u_2(x)}{\sum_{x\in X} u_2(x)}$. Hence, $u_1 = \frac{\sum_{x\in X} u_1(x)}{\sum_{x\in X} u_2(x)} u_2$, where $\frac{\sum_{x\in X} u_1(x)}{\sum_{x\in X} u_2(x)} > 0$ by definition. ■

Proof of Proposition 4

Proof. Suppose that (u_1,v_1) and (u_2,v_2) represent the same choice rule. Then, by definition, for every $x \in X$, $\frac{u_1(x)}{u_1(X)} = \rho(x,\emptyset) = \frac{u_2(x)}{u_2(X)}$. Therefore, we get $u_1(x) = \frac{u_1(X)}{u_2(X)} u_2(x)$, where $\frac{u_1(X)}{u_2(X)} > 0$ and is independent of x. Also, for every $x \in X$, we have $\frac{u_1(z)}{u_1(X)+v_1(x)} = \rho(z,\{x\}) = \frac{u_2(z)}{u_2(X)+v_2(x)}$. Hence, by putting $u_1 = \alpha u_2$, we have $\frac{\alpha u_2(z)}{\alpha u_2(X)+v_1(x)} = \frac{u_2(z)}{u_2(X)+v_2(x)}$. Then, we get $v_1(x) = \alpha v_2(x)$. ■

Proof of Lemma 1

To prove Theorem 5, we need to first prove the following Lemma.
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 prove by strong induction by "stepping down". For $R = X\setminus\{x\}$, we have RHS $= q(x,X) = \rho(x,X)$. LHS $= \sum_{a\in X\setminus\{x\}} q(a, X\setminus\{x\}) + \sum_{a\notin X\setminus\{x\}} y(a, X\setminus\{x\}) = \sum_{a\in X\setminus\{x\}} q(a, X\setminus\{x\}) + \rho(x, X\setminus\{x\}) = \sum_{a\in X\setminus\{x\}}\Big[\rho(a, X\setminus\{x\}) - \rho(a,X)\Big] + \rho(x, X\setminus\{x\}) = \sum_{a\in X}\rho(a, X\setminus\{x\}) - \sum_{a\in X\setminus\{x\}}\rho(a,X) = 1 - \sum_{a\in X\setminus\{x\}}\rho(a,X) = \rho(x,X)$. The first step is proven. Suppose that equality holds for size of $k+1, k+2, ...N-1$. Let $|R|= k$, LHS-RHS$= \sum_{a\in R} q(a,R) + \sum_{a\notin R} y(a,R) - \sum_{b\notin R} q(b, R\cup b)$
$= \sum_{a\in R}\Big(\rho(a,R) - \sum_{B\supset R} q(a,B)\Big) + \sum_{a\notin R}\Big(\rho(a,R) - \sum_{a\notin B\supset R} y(a,B)\Big) - \sum_{b\notin R} q(b, R\cup b)$
$= \sum_{a\in X}\rho(a,R) - \Big[\sum_{a\in R}\sum_{B\supset R} q(a,B) + \sum_{a\notin R}\sum_{a\notin B\supset R} y(a,B) + \sum_{b\notin R} q(b, R\cup b)\Big]$. Since $\sum\limits_{a\in X}\rho(a,R) = 1$, 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

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

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

=...(repetitively applying induction hypothesis)

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

where the second-to-last equality is given by the definition of q. Hence, it is proven. ■

Proof of Theorem 5(a)

Proof. Since u and v are ordinal, we would prove the theorem using the ordinal version of the model as in Proof of Theorem 1(a), where each agent (u, v) is denoted with $\succ^\star$.

For the necessity proof, we suppose the data follows the model. We introduce the following notation, for $b \in X$ and $A \subseteq X\setminus b$, $M_y(b,A) := \mu\Big(\{\succ^\star|(X\setminus b)\cup A^* = L_{\succ^\star}(b)\}\Big)$ and $M_q(b, A\cup b) := \mu\Big(\{\succ^\star|X\cup A^* = L_{\succ^\star}(b^*)\}\Big)$, where $L_{\succ^\star}(a)$ is the strict lower contour set of a according to $\succ^\star$.

Claim 11. *For $b \in X$ and $A \subseteq X \setminus b$, i) $M_y(b,A) = y(b,A)$, ii) $M_q(b,A) = q(b, A\cup b)$*

Proof. We prove by strong induction by "stepping down". For $A = X \setminus \{x\}$, we have, $y(x, X\setminus\{x\}) = \rho(x, X\setminus\{x\}) = \mu\Big(\{\succ^\star|(X\setminus\{x\})\cup(X\setminus\{x\})^* = L_{\succ^\star}(x)\}\Big) = M_y(x, X\setminus\{x\})$. 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 $|A|= k$, $y(x,A) = \rho(x,A) - \sum_{x\notin B\supset A} y(x,B) = \sum_{x\notin B\supseteq A} M_y(x,B) - \sum_{x\notin B\supset A} M_y(x,B) = M_y(x,A)$, where the second-to-last equality is given by Definition and induction hypothesis. Hence, the proof is complete for i). For ii), we prove again by strong induction by "stepping down". For $A = X\setminus\{x\}$, we have $q(x,X) = \rho(x,X) = \mu\Big(\{\succ^\star|X\cup(X\setminus\{x\})^* = L_{\succ^\star}(x)\}\Big) = M_q(x, X\setminus\{x\})$. 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 $|A|= k$. $q(x,A) = \rho(x,A) - \sum_{x\notin B\supset A} q(x,B) = \sum_{x\notin B\supseteq A} M_q(x,B) - \sum_{x\notin B\supset A} M_q(x,B) = M_q(x,A)$ where the second-to-last equality is given by Definition and induction hypothesis. Hence, the proof is complete for ii). ■

From this claim, we immediately show Non-negativity of BM, since μ are by definition positive.

For the sufficiency proof, 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 will note $\pi_{R^*}\sigma$ as the extended sequence for some $\sigma \in X\cup X^*$. we first construct $G(a^*) := q(a,X)$ for all $a^* \in X^*$. Then, we construct, recursively,

$$G(\pi_{R^*}\sigma)=\begin{cases} G(\pi_{R^*})\dfrac{y(z,X\setminus R)}{\sum_{\pi'_{R^*}\in\Pi_{R^*}}G(\pi_{R^*})} & \text{if } \sigma=z\in R\\ G(\pi_{R^*})\dfrac{q(z,X\setminus R)}{\sum_{\pi'_{R^*}\in\Pi_{R^*}}G(\pi_{R^*})} & \text{if } \sigma=z^*\in(X\setminus R)^* \end{cases}$$

Note that $G\geq 0$ by Non-negativity of BM. We then prove the following properties for G.

Claim 12. *For $R\subseteq X$, we have. i) $\sum_{\pi_{R^*}\in\Pi_{R^*}}G(\pi_{R^*}\sigma)=y(x,X\backslash R)$ for $\sigma=x\in R$, ii) $\sum_{\pi_{R^*}\in\Pi_{R^*}}G(\pi_{R^*}\sigma)=q(x,X\setminus R)$ for $\sigma=x^*\in(X\setminus R)^*$, iii) $\sum_{\pi_{R^*}\in\Pi_{R^*}}G(\pi_{R^*})=\sum_{x\in R}q(x,x\cup(X\setminus R))$, iv) $G(\pi_{R^*})=\sum_{\sigma\in R\cup(X\setminus R)^*}G(\pi_{R^*}\sigma)$ for all π_{R^*},*

Proof. Note that i) and ii) are straight forward. For iii), we have $LHS=\sum_{\pi_{R^*}\in\Pi_{R^*}}G(\pi_{R^*})=\sum_{z\in R}\sum_{\pi_{(R\setminus z)^*}\in\Pi_{(R\setminus z)^*}}G(\pi_{(R\setminus z)^*}z^*)=\sum_{z\in R}q(z,x\cup(X\setminus R))=RHS$ (By (ii)). For iv), we have $RHS=\sum_{\sigma\in R\cup(X\setminus R)^*}G(\pi_{R^*}\sigma)=\frac{G(\pi_{R^*})}{\sum_{\pi'_{R^*}\in\Pi_{R^*}}G(\pi'_{R^*})}(\sum_{z\in R}y(z,X\setminus R)+\sum_{z\in X\setminus R}q(z,X\setminus R))$ $=\frac{G(\pi_{R^*})}{\sum_{\pi'_{R^*}\in\Pi_{R^*}}G(\pi'_{R^*})}(\sum_{z\in R}q(z,z\cup(X\setminus R))$ (By Lemma 1) $=G(\pi_{R^*})=LHS$(By (iii)) ■

We then define each individual weight. We will consider permutation on $X\cup X^*$. We let π denote each permutation in $X\cup X^*$. Note that we only consider permutation where x^* comes before x. Then, we perform a "truncation" on π. Firstly, we truncate π up to the first element in X appears in the sequence. For example, for $X=\{x,y,z\}$, we will truncate the following sequence in the following way.

$$\underbrace{x^*y^*x}_{\pi^t}z^*zy$$

Here, since the type space of RP-RUM is huge and many types are inherently indistinguishable from other, we mainly assign the weights on a group of type whose behavior are the same. Then, we allocate the weight evenly across the group. We define, for every $\pi_{R^*}\in\Pi_{R^*}$ and $x\in X$, for every π such that $\pi^t=\pi_{R^*}x$,

$$\hat{\mu}_\pi:=G(\pi_{R^*}x)*\frac{1}{|\{\pi:\pi^t=\pi_{R^*}x\}|}.$$

Firstly, note that $\hat{\mu}\geq 0$ due to the fact that $G\geq 0$. Claim 12(iv) ensures that $\hat{\mu}$ is additive. In the following, we introduce the following notation. Let $L_\pi(\sigma)$ be the lower contour set of σ according to π. We define $\hat{M}_y(b,A):=\hat{\mu}\Big(\{\pi|(X\setminus b)\cup A^*=L_\pi(b)\}\Big)$ and $\hat{M}_q(b,A\cup b):=\hat{\mu}\Big(\{\pi|X\cup A^*=L_\pi(b^*)\}\Big)$. Since the weights are constructed, we now check for the choice data. For $x\notin R$, $\hat{\rho}(x,R)=\sum_{x\notin B\supseteq R}\hat{M}_y(x,B)$ (by definition) $=\sum_{x\notin B\supseteq R}\sum_{\pi_{(X\setminus B)^*}\in\Pi_{(X\setminus B)^*}}G(\pi_{(X\setminus B)^*}x)$ (by construction) $=\sum_{x\notin B\supseteq R}y(x,B)$ (By Claim 12(i)) $=\rho(x,R)$. On the other hand, for $x\in R$, $\hat{\rho}(x,R)=\sum_{A\supseteq R}\hat{M}_q(x,A)$ (by definition) $=\sum_{A\supseteq R}\sum_{\pi_{(X\setminus A)^*}\in\Pi_{(X\setminus A)^*}}G(\pi_{(X\setminus A)^*}x)$ (by construction) $=\sum_{A\supseteq R}q(x,A)$ (By Claim 12(ii)) $=\rho(x,R)$. Hence, the constructed weights explain the data. Since it explains the data, it is immediately that $\sum\limits_{\pi\in\Pi}\hat{\mu}_\pi=1$. The sufficiency proof is complete. ■

Proof of Theorem 5(b)

Proof. Since u is ordinal, we would prove again the theorem using the ordinal version of the model as in Proof of Theorem 1(b), where each agent (a,u) is denoted with $(a,\succ)$. For the necessity proof, we suppose the data follows the model. We introduce the notation, for $a,b\in X$ and $b\notin A$, $T(a,b,A):=\tau\Big(\{(a,\succ)\in\mathcal{T}:A=L_\succ(b)\}\Big)$, where $L_\succ(a)$ is the strict lower contour set of a according to $\succ$.

Claim 13. *For $x\notin A$, i) $T(x,x,A)=y(x,A)$, ii) $\sum\limits_{a\in A\cup\{x\}}T(a,x,A)=q(x,A\cup\{x\})$*

Proof. For i), we prove by strong induction by "stepping down". For $A=X\backslash\{x\}$, we have, $y(x,X\backslash\{x\})=\rho(x,X\setminus\{x\})$ (By definition of y)$=\tau\Big(\{(x,\succ):X\setminus\{x\}=L_\succ(x)\}\Big)=T(x,x,X\setminus\{x\})$ 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 $|A|=k$, we have

$y(x,A) = \rho(x,A) - \sum\limits_{x\notin B\supset A} y(x,B) = \sum\limits_{x\notin B\supseteq A} T(x,x,B) - \sum\limits_{x\notin B\supset A} T(x,x,B)$ (by Definition and induction hypothesis) $= T(x,x,A)$. 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, $q(x,X) = \rho(x,X)$ (By definition) $= \sum\limits_{a\in A\cup\{x\}} \tau\Big(\{(a,\succ) : X\setminus\{x\} = L_{\succ}(x)\}\Big) = \sum\limits_{a\in X} T(a,x,X\setminus\{x\})$. 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 $|A|= k$, we have $q(x,A\cup\{x\}) = \rho(x,A\cup\{x\}) - \sum\limits_{B\supset A\cup\{x\}} q(x,B) = \sum\limits_{B\supseteq A\cup\{x\}}\sum\limits_{a\in B} T(a,x,B\setminus\{x\}) - \sum\limits_{B\supset A\cup\{x\}}\sum\limits_{a\in B} T(a,x,B\setminus\{x\})$ (by Definition and induction hypothesis) $= \sum\limits_{a\in A\cup\{x\}} T(a,x,A)$. Hence, the proof is complete for ii). ∎

From this claim, we immediately show Non-negativity of BM and Positive Marginal Recommendation since for $x\notin A$, we have for y, $y(x,A) = T(x,x,A) \geq 0$; for q, $q(x,A\cup\{x\}) - y(x,A) = \sum\limits_{a\in A\cup\{x\}} T(a,x,A) - T(x,x,A) = \sum\limits_{a\in A} T(a,x,A) \geq 0$. The necessity proof is complete.

For the sufficiency proof, we make the following claim for later use.

Claim 14. *Positive Marginal Recommendation and Lemma 1 implies for all a, $q(a,a) = y(a,\emptyset)$.*

Proof. One can prove by contradiction. Suppose not, there exists a such that $q(a,a) > y(a,\emptyset)$. Due to Lemma 1, then there exists b such that $q(b,b) < y(b,\emptyset)$. ∎

Then, we need to introduce the following 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 $\pi\in\Pi$. Then, we 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 Non-negativity of BM and Positive Marginal Recommendation. We then make the two following claims.

Claim 15. *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. $\sum_{b\in X\setminus A\cup R\cup\{a\}} F(a,b\pi_R a,A) = \sum_{b\in X\setminus R\cup A\cup\{a\}} \frac{F(a,\pi_R a,A)(q(b,R\cup A\cup\{a,b\})-y(b,R\cup A\cup\{a\}))}{\sum_{x\in R\cup A\cup\{a\}} q(x,R\cup A\cup\{a\})} =$
$F(a,\pi_R a,A)\frac{\sum_{b\in X\setminus R\cup A\cup\{a\}}(q(b,R\cup A\cup\{a,b\})-y(b,R\cup A\cup\{a\}))}{\sum_{x\in R\cup A\cup\{a\}} q(x,R\cup A\cup\{a\})} = F(a,\pi_R a,A)$ (by Lemma 1). ∎

Claim 16. *For every non-empty A and $x\notin A$, i) $\sum\limits_{a\in A}\sum\limits_{B\subseteq A\setminus\{a\}}\sum\limits_{\pi_B\in\Pi_B} F(a,\pi_B a,A\setminus(B\cup a)) = \sum\limits_{b\in A} q(b,A)$, ii) $\sum\limits_{a\in A}\sum\limits_{B\subseteq A\setminus\{a\}}\sum\limits_{\pi_B\in\Pi_B} F(a,x\pi_B a,A\setminus(B\cup a)) = q(x,A\cup\{x\}) - y(x,A)$*

Proof. 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 $\sum\limits_{a\in A}\sum\limits_{B\subseteq A\setminus\{a\}}\sum\limits_{\pi_B\in\Pi_B} F(a,\pi_B a,A\setminus(B\cup a)) = F(a,a,\emptyset) = y(a,\emptyset) = q(a,a)$ (by Claim 14). For ii), we have $\sum\limits_{a\in A}\sum\limits_{B\subseteq A\setminus\{a\}}\sum\limits_{\pi_B\in\Pi_B} F(a,x\pi_B a,A\setminus(B\cup a)) = F(a,xa,\emptyset) = \frac{F(a,a,\emptyset)(q(x,\{x,a\})-q(x,\{a\})}{q(a,a)} = q(x,\{x,a\}) - q(x,\{a\})$ (by Claim 14). 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, $\sum\limits_{a\in A}\sum\limits_{B\subseteq A\setminus\{a\}}\sum\limits_{\pi_B\in\Pi_B} F(a,\pi_B a,A\setminus(B\cup a)) = \sum\limits_{a\in A} F(a,a,A\setminus\{a\}) + \sum\limits_{x\in A}\sum\limits_{a\in A\setminus\{x\}}\sum\limits_{B\subseteq(A\setminus\{x\})\setminus\{a\}}\sum\limits_{\pi_B\in\Pi_B} F(a,x\pi_B a,(A\setminus\{x\})\setminus(B\cup a)) = \sum\limits_{a\in A} y(a,A\setminus\{a\}) + \sum\limits_{x\in A}[q(x,(A\setminus$

$\{x\})\cup\{x\})-y(x,A\setminus\{x\})] = \sum_{a\in A} q(a,A)$, where the second equality is given by construction and induction hypotheses of (ii) where $|A\setminus\{x\}|= k-1$. Hence, it is confirmed that $i)$ is true for size of k. Then, for $ii)$, $\sum_{a\in A}\sum_{B\subseteq A\setminus\{a\}}\sum_{\pi_B\in\Pi_B} F(a, x\pi_B a, A\setminus(B\cup a)) = \frac{q(x,A\cup\{x\})-y(x,A)}{\sum_{b\in A} q(b,A)} \sum_{a\in A}\sum_{B\subseteq A\setminus\{a\}}\sum_{\pi_B\in\Pi_B} F(a,\pi_B a, A\setminus(B\cup a)) = \frac{q(x,A\cup\{x\})-y(x,A)}{\sum_{b\in A} q(b,A)} \sum_{b\in A} q(b,A) = q(x, A\cup\{x\}) - y(x,A)$, where the second equality is given by the fact (i) is true for size of A to be k. 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, RA-RUM type space is big and many types are behaviorally indistinguishable from other types. We main assign weight on a group of types, and distribute even weight across them. We define

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

Firstly, note that $\hat{\tau}_{a,\pi}\geq 0$ due to the fact that $F\geq 0$. Claim 15 ensures that $\hat{\tau}$ is additive. In the following, we introduce the notation, for $a,b\in X$ and $b\notin A$, $\hat{M}(a,b,A) = \sum\{\hat{\tau}_{a,\pi} :\ A = L_\pi(b)\}$.

Claim 17. *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 15, 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 15, 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 16(ii), it is proven. ■

Since the weight are constructed, we have, for $x\in R$, $\hat{\rho}(x,R) = \sum_{A\supseteq R}\sum_{a\in A}\hat{M}(a,x,A\setminus\{x\})$ (by Definition) $\sum_{A\supseteq R}\left[\hat{M}(x,x,A\setminus\{x\}) + \sum_{a\in A\setminus\{x\}}\hat{M}(a,x,A\setminus\{x\})\right] = \sum_{A\supseteq R}\left[y(x,A\setminus\{x\}) + q(x,A) - y(x,A\setminus\{x\})\right]$ (By Claim 17) $= \sum_{A\supseteq R} q(x,A) = \rho(x,R)$. On the other hand, for $x\notin R$, $\hat{\rho}(x,R) = \sum_{a\notin B\supseteq R}\hat{M}(x,x,B)$ (by Definition) $= \sum_{a\notin B\supseteq R} y(x,B)$ (By Claim 17(i)) $= \rho(x,R)$. Hence, the constructed weights explain the data. Since it explains the data, it is immediately that $\sum_{a\in X}\sum_{\pi\in\Pi}\hat{\tau}_{a,\pi} = 1$. The sufficiency proof is complete. ■

Proof of Proposition 7

Proof. It is basically proven in Claim 13. To see this, note that using the primitive of $(b,\succ)$, we defined $T(a,b,A) := \tau\Big(\{(a,\succ)\in\mathcal{T} : A = L_\succ(b)\}\Big)$ where $L_\succ(a)$ is the strict lower contour set of a according to $\succ$ before Claim 13. Also, we prove in Claim 13 that for $x\notin A$, i) $T(x,x,A) = y_\rho(x,A)$ and ii) $\sum_{a\in A\cup\{x\}} T(a,x,A) = q_\rho(x,A\cup\{x\})$. Therefore, for Proposition 7 (i), we have $y_\rho(b,R) = T(b,b,R) = \tau(\{(b,\succ)\in: R = L_\succ(b)\})$. In other words, it captures all the types $(b,\succ)$ that have pre-recommendation choice as b and b gives a higher utility than all elements in R and gives less utility than all elements in $R\setminus(R\cup b)$. Therefore, it gives $\tau\Big(\{(b,u)|\min_{x\in X\setminus R\cup b} u(x) > u(b) > \max_{x\in R} u(x)\}\Big)$. Similar arguments can be made for Proposition 7(ii) using Claim 13(ii). ■

Proof of Proposition 8

Proof. It is basically proven in Claim 11. To see this, using the primitive $\succ^\star$, we defined before Claim 11 that for $a\in X$ and $A\subseteq X\setminus a$, $M_y(b,A) := \mu\Big(\{\succ^\star|(X\setminus b)\cup A^* = L_{\succ^\star}(b)\}\Big)$ and $M_q(b,A\cup b) := \mu\Big(\{\succ^\star|X\cup A^* = L_{\succ^\star}(b^*)\}\Big)$ and proved in Claim 11 that For $a\in X$ and $A\subseteq X\setminus a$, i) $M_y(b,A) = y(b,A)$ and ii) $M_q(b,A) = q(b,A\cup b)$. Therefore, for Proposition 8(i), we have $y_\rho(b,R) = M_y(b,R) = \mu(\{\succ^\star|(X\setminus b)\cup R^* =$

$L_{\succ^\star}(b)\}$). In other words, it captures all the types $\succ^\star$ where b is better than all alternatives in R^* (the recommended version) and all alternatives in $X \setminus b$ (the regular version) , and is worse than all alternatives in $(X \setminus R)^*$ (the recommended version). Therefore, it gives $y_\rho(b,R) = \mu^i\Big(\{(u,v)| \min_{x\in X\setminus R} u(x)+v(x) > u(b) > \max\{\max_{x\in R} u(x)+v(x), \max_{x\in X\setminus b} u(x)\}\}\Big)$. Similar arguments can be made for Proposition 8(ii) using Claim 11(ii). ■

Proof of Theorem 6(a)

Proof. It suffices to show that it satisfies Non-negativity of BM. We first state the following fact, due to the the well known fact that any Luce is RUM.

Fact 1. *Let $k_i \geq 0$ for $i = 1, ..., N$, and we denote $\mathcal{N} = \{1, .., N\}$ Then, for any $K > 0$, we have* $\sum_{A\subseteq \mathcal{N}} (-1)^{|A|} \frac{1}{K+\sum_{i\in A} k_i} \geq 0$

We first consider $y(x,R)$. $y(x,R) = \sum_{x\notin B\supseteq R} (-1)^{|B\setminus R|} \rho(x,B) = \sum_{x\notin B\supseteq R} (-1)^{|B\setminus R|} \frac{u(x)}{u(X)+v(B)} = u(x) \sum_{x\notin B\supseteq R} (-1)^{|B\setminus R|} \frac{1}{u(X)+v(R)+v(B\setminus R)} = u(x) \sum_{A\subseteq X\setminus(R\cup x)} (-1)^{|A|} \frac{1}{u(X)+v(R)+v(A)}$. We use the above fact by letting $u(X)+v(R) = K > 0$ with $v(x) \geq 0$. Then, it is immediate that $y(x,R) \geq 0$. Analogously, one can also see that $q(x,R) \geq 0$. ■

Proof of Theorem 6(b)

Proof. It suffices to show that it satisfies Non-negativity of BM and Positive Marginal Recommendation. For Non-negativity of BM, we prove it by using standard results from the relationship between RUM and Luce model. Notice that, for $x \notin R$, $y(x,R) = \sum_{x\notin B\supseteq R} (-1)^{|B\setminus R|} \rho(x,B) = \sum_{x\notin B\supseteq R} (-1)^{|B\setminus R|} d(x) \frac{u(x)}{u(B\cup x)}$ $= d(x) \sum_{A\supseteq R\cup x} (-1)^{|A\setminus(R\cup x)|} \frac{u(x)}{u(A)}$. Since every Luce model has a RUM representation. Hence, the term $\sum_{A\supseteq R\cup x} (-1)^{|A\setminus(R\cup x)|} \frac{u(x)}{u(A)}$ is guaranteed to be non-negative since it is exactly the standard block Marschak polynomials of a Luce model. For Positive Marginal Recommendation, we first make the following auxiliary claim.

Claim 18. *For every $A \subset X$ and $z \notin A$, $\sum_{B\supseteq A} (-1)^{|B\setminus A|} \frac{1}{u(B\cup z)} = 0$*

Proof. $\sum_{B\supseteq A} (-1)^{|B\setminus A|} \frac{1}{u(B\cup z)} = \sum_{B\supseteq A} (-1)^{|B\setminus A|} \frac{1}{u(B\cup z)} = \sum_{z\notin B\supseteq A} (-1)^{|B\setminus A|} \frac{1}{u(B\cup z)} + \sum_{z\in B\supseteq A} (-1)^{|B\setminus A|} \frac{1}{u(B\cup z)}$ $= \sum_{C\supseteq A\cup z} (-1)^{|C\setminus A\cup z|} \frac{1}{w(C)} + \sum_{B\supseteq A\cup z} (-1)^{|B\setminus(A\cup z)|+1} \frac{1}{u(B)} = 0$ ■

Then, we make the following claim. To prove this, we utilize the following expression of $\rho(x,A)$ for $x \in A$, $\rho(x,A) = \rho(x, A\setminus x) + \sum_{z\in A\setminus x} \frac{u(x)}{u(z)} \rho(z, A\setminus z) + \sum_{z\notin A} \frac{u(x)}{u(z)} \rho(z,A)$.

Claim 19. *For every $x \in A$, $q(x,A) = y(x, A\setminus x) + \sum_{z\in A\setminus x} \frac{u(x)}{u(z)} y(z, A\setminus z)$*

Proof. $q(x,A) = \sum_{B\supseteq A} (-1)^{|B\setminus A|} \Big[\rho(x, B\setminus x) + \sum_{z\in B\setminus x} \frac{u(x)}{u(z)} \rho(z, B\setminus z) + \sum_{z\notin B} \frac{u(x)}{u(z)} \rho(z,B)\Big]$
$= \sum_{B\supseteq A} (-1)^{|B\setminus A|} \Big[\rho(x, B\setminus x) + \sum_{z\in A\setminus x} \frac{u(x)}{u(z)} \rho(z, B\setminus z) + \sum_{z\in B\setminus A} \frac{u(x)}{u(z)} \rho(z, B\setminus z) + \sum_{z\notin B} \frac{u(x)}{u(z)} \rho(z,B)\Big]$
$= y(x, A\setminus x) + \sum_{z\in A\setminus x} \frac{u(x)}{u(z)} y(z, A\setminus z) + \sum_{B\supseteq A} (-1)^{|B\setminus A|} \Big[\sum_{z\in B\setminus A} \frac{u(x)}{u(z)} \rho(z, B\setminus z) + \sum_{z\notin B} \frac{u(x)}{u(z)} \rho(z,B)\Big]$
where the last inequality is by applying the definition of y on the first two terms. It remains to show that the last sum is zero. Note that
$\sum_{B\supseteq A} (-1)^{|B\setminus A|} \Big[\sum_{z\in B\setminus A} \frac{u(x)}{u(z)} \rho(z, B\setminus z) + \sum_{z\notin B} \frac{u(x)}{u(z)} \rho(z,B)\Big] = u(x) \sum_{B\supseteq A} (-1)^{|B\setminus A|} \sum_{z\in X\setminus A} \frac{1}{u(z)} \rho(z, B\setminus z) = u(x) \sum_{B\supseteq A} (-1)^{|B\setminus A|} \sum_{z\in X\setminus A} \frac{1}{u(z)} \frac{u(z)d(z)}{u(B\cup z)} = u(x) \sum_{z\in X\setminus A} d(z) \sum_{B\supseteq A} (-1)^{|B\setminus A|} \frac{1}{u(B\cup z)}$ (By switching summation sign) $= 0$ (by Claim 18). ■

Hence, Claim 19 is proven. Hence, to show that Positive Marginal Recommendation is satisfied, we have $q(x,A) - y(x, A\setminus x) = \sum_{z\in A\setminus x} \frac{u(x)}{u(z)} y(z, A\setminus z) \geq 0$. The proof is complete. ■