

# Random Collection\*

Tri Phu Vu<sup>†</sup>

University of Maryland

November 21, 2025

## Abstract

This paper studies choice situations in which a decision maker can choose multiple alternatives. Given a menu of available options, the decision maker selects a subset of the menu with certain probabilities. We employ an axiomatic approach to characterize various parametric models in the literature. Our results elucidate the implications of the functional form assumptions and shed light on the distinctions between models. The behavioral postulates offer simple tools for testing and falsifying the choice procedures used by the decision maker and reveal a close connection between models that are seemingly unrelated.

Keywords: Stochastic choice correspondence, choice bundles, choice correspondence, consideration set.

JEL classification: D01, D11, D91.

---

\*I am indebted to Yusufcan Masatlioglu, Emel Filiz-Ozbay, and Erkut Y. Ozbay for their continuous guidance, encouragement, and support. I want to thank Marina Agranov, Victor H. Aguiar, Yunus C. Aybas, Christopher P. Chambers, Kfir Eliaz, Andrew Ellis, Onur Kesten, Matthew Kovach, Marco Mariotti, Kirby Nielsen, Quang Pham, Daniel Reck, Gerelt Tserenjigmid, Daniel Vincent, Kemal Yildiz, and seminar audiences at the University of Maryland, Caltech Summer Program, and BRIC XI for helpful discussions and comments.

<sup>†</sup>Department of Economics, 3114 Tydings Hall, 7343 Preinkert Dr., College Park, MD 20742.  
Email: tvuphu@umd.edu.

# 1 Introduction

Decision makers frequently encounter choices with many options. The standard approach to modeling choice behavior assumes that decision makers select a single alternative each time they face a choice problem. In many environments, however, it is natural for decision makers to choose a collection of options rather than just one. Examples are abundant. Platforms often select a set of products to recommend to their customers. Supermarkets regularly choose an assortment of products to offer. Coaches in soccer or basketball typically pick a group of players for the starting lineup. Investors select a portfolio of financial assets to invest in. Investigating the behavior of decision makers in such situations is crucial, as it enables more accurate modeling of choices that naturally involve selecting multiple options.

This paper studies choice problems where decision makers can choose a collection of alternatives, allowing for variation in the selected collection. Formally, given a set  $S$  of feasible options, decision makers can choose an arbitrary subset  $T$  of  $S$  with a probability  $\mu(T, S)$ , where  $\mu$  represents a probability measure over the power set of  $S$ . Various examples and interpretations of  $\mu$  exist; we present several below.

- **Product recommendation:** A platform uses an algorithm based on individual characteristics, purchase history, popularity of the products, etc., to recommend some products to customers from the set of products it offers. In online environments, Amazon’s “Amazon’s Choice,” Spotify’s “Editor’s Picks,” and Disney+’s “Fan Favorites” are typical examples. In physical settings, [Kawaguchi et al. \(2021\)](#) illustrate that some vending machines in Japan can utilize facial recognition technology to analyze a customer’s age and gender and provide drink recommendations based on these demographic factors. Because the underlying data varies across customers and over time, the platform (or the vending machine) recommends each collection  $T$  with probability  $\mu(T, S)$ .
- **Assortment choice:** A supermarket selects a collection of items to offer to its customers from the set of all products it can provide. Due to limited shelf space, varying items’ perishability, and fluctuating demand, the collection of items the supermarket selects varies over time. Consequently, the supermarket offers a set  $T$  of products with probability  $\mu(T, S)$ . Such random assortments can arise from the supermarket optimizing its expected revenue ([Ma, 2023](#)).

- **Approval behavior:** In online shopping, customers often add a subset of the available products to a wishlist or shopping cart for later consideration. Such actions, referred to as approval behavior (Manzini et al., 2024), also include adding videos to a playlist on YouTube or marking places as “Want to go” on Google Maps. In these cases,  $\mu(T, S)$  denotes the probability that the decision maker approves a collection  $T$  of products for later use.

In the examples above,  $\mu(T, S)$  represents the choices of a single decision maker in varying situations (intrapersonal). Depending on the context,  $\mu(T, S)$  can also be interpreted as the choices of different decision makers within a population (interpersonal). For instance, in the Japanese vending machine example, the analyst may observe data from multiple vending machines, and  $\mu(T, S)$  may indicate the proportion of vending machines recommending a set  $T$  of products.

Various functional forms of  $\mu$  have been proposed in the literature, each grounded in distinct underlying narratives and offering different interpretations. Rather than introducing a new model, this paper undertakes a systematic examination of well-known parametric models of  $\mu$  listed in Table 1. Several models in Table 1 were originally proposed to study the formation of consideration sets.<sup>1</sup> However, as we will thoroughly illustrate in Section 2, these parametric functional forms can also be applied to investigate choice situations that naturally involve selecting multiple options, such as supermarket assortment problems or product recommendations.

Despite the widespread use and numerous applications of the models of  $\mu$  in Table 1, little is known about their behavioral implications. Additionally, it is unclear how to distinguish these models from each other, as they often have complicated functional forms and are motivated by distinct stories. This paper provides behavioral characterizations of several formulations of  $\mu$  in Table 1 and highlights a close connection between well-known but seemingly unrelated models. Our results offer a better understanding of the implications of functional form assumptions, allowing the analyst to differentiate between alternative models of  $\mu$  and to identify the most suitable model for their specific application.

First, many formulations of  $\mu$ , despite their cumbersome functional forms, have simple and intuitive behavioral foundations, with several characterized by a single

---

<sup>1</sup>Consideration set formation is another interpretation of  $\mu$ . For this interpretation, see Manzini and Mariotti (2014) and Cattaneo et al. (2020), among others.

Models	Papers	Factors impacting $\mu$
<b>List A</b>		
Logit	<a href="#">Brady and Rehbeck (2016)</a>	Importance (or salience) of sets of options
Nested stochastic choice	<a href="#">Kovach and Tserenjigmid (2022)</a>	Nests (a partition of the grand set)
Nested logit	<a href="#">McFadden (1978)</a>	Nests and the DM’s preference
<b>List B</b>		
Independent Choice	<a href="#">Manzini and Mariotti (2014)</a>	Importance (or salience) of individual options
<b>List C</b>		
Random categorization	<a href="#">Aguiar (2017)</a>	Salience of categories of options
Elimination by aspects	<a href="#">Tversky (1972)</a>	Salience of attributes of options
Attribute rule	<a href="#">Gul et al. (2014)</a>	Salience of attributes of options
Random reference model	<a href="#">Kibris et al. (2024)</a>	Salience of individual options

Table 1: Models of  $\mu$  and their underlying motivations

*Notes.* Models are divided into lists A, B, and C based on their behavioral characterizations.

axiom. While functional forms facilitate empirical applications, it is often challenging to tell whether two parametric formulations are equivalent, related, or distinct, especially when they have complicated forms. Identifying behavioral foundations allows us to perform such comparisons more easily. We show that several models of  $\mu$  that are conceptually distinct in their underlying stories are strikingly similar in behavioral predictions. At the same time, formulations that look alike intersect only in special cases. Second, the behavioral postulates identified in the paper offer a set of testable predictions that allow for empirical verification of these models. An outside analyst with access to data on  $\mu$  can utilize these behavioral implications to identify the underlying data-generating process and potentially falsify models.<sup>2</sup>

Our characterizations of various models of  $\mu$  identify two main behavioral postulates and their variants: Independence of Irrelevant Sets (IIS) and Relative Additivity. IIS conceptually resembles the Independence of Irrelevant Alternatives (IIA) axiom in the probabilistic choice literature, but differs in the domain on which it operates. IIS states that the relative choice probability of two collections is independent of the

<sup>2</sup>For instance, a researcher interested in understanding the formation of consideration sets might apply the behavioral postulates identified in our paper to the data to investigate how decision makers choose their consideration sets. Such data is available; see, for instance, [Ellis et al. \(2025\)](#).

presence or absence of other sets:

$$\frac{\mu(T, S)}{\mu(T', S)} = \frac{\mu(T, S')}{\mu(T', S')} \quad \text{for all non-empty } T, T' \subseteq S \cap S'.$$

Section 2 shows that IIS and its variants appear in the characterizations of  $\mu$  in Logit, Independent Choice, and Nested Stochastic Choice. Meanwhile, roughly speaking, Relative Additivity states that adding a new option to the choice set has a constant effect on the choice probabilities of a given collection. That is, for a fixed  $S \ni x$ ,

$$\frac{\mu(T, S \setminus \{x\})}{\mu(T, S) + \mu(T \cup \{x\}, S)} \quad \text{is independent of } T \text{ for all non-empty } T \subseteq S \setminus \{x\}.$$

To understand Relative Additivity, suppose  $S \setminus \{x\}$  is the initial choice set and  $T$  is the chosen collection. Adding  $x$  to the choice set can affect  $T$  in two different ways. On the one hand,  $x$  may have only an indirect effect, in which case it does not alter the chosen collection but only affects its selection probability. In that situation,  $T$  is selected with probability  $\mu(T, S)$ . On the other hand,  $x$  may have a direct impact and be added to the chosen collection, in which case  $T \cup \{x\}$  is selected with probability  $\mu(T \cup \{x\}, S)$ . Thus, the sum  $\mu(T, S) + \mu(T \cup \{x\}, S)$  can be interpreted as the total choice probability associated with  $T$  under the expanded choice set. The ratio  $\frac{\mu(T, S \setminus \{x\})}{\mu(T, S) + \mu(T \cup \{x\}, S)}$  then captures the relative probability of choosing a collection before and after adding the new option. Relative Additivity states that this relative choice probability is independent of  $T$ . In Section 2, we show that Relative Additivity and its variants appear in the characterizations of  $\mu$  in Random Categorization, Elimination by Aspects, Attribute Rule, and Random Reference Model.

Figure 1 summarizes the connections among various models of  $\mu$ . We identify the following key relationships. First, Random Reference Model is notably distinct from other frameworks: it intersects with some of them only in a special case of  $\mu$  referred to as singleton- $\mu$ . This  $\mu$ , formally defined in Section 3, is related to the Luce model in the probabilistic choice literature. Second, many models of  $\mu$  intersect at the singleton- $\mu$ , but not at the deterministic-with-full-choice- $\mu$ , where the whole choice set is always selected with certainty. Third, the endogenous versions of the formulations of  $\mu$  in Elimination by Aspects (static version, see Section 2.5.1), Attribute Rule, and Random Categorization are all equivalent. Finally, the formulation of  $\mu$  in Independent Choice is equivalent to the intersection of Logit and Random Cat-

egorization, a result also obtained by Kovach and Suleymanov (2023) in a different setting.

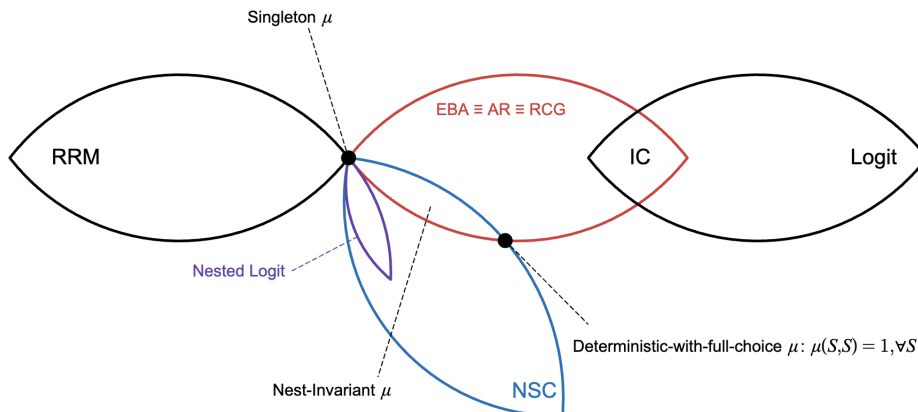


Figure 1: Relationship among various models of  $\mu$

*Notes.* Label legend: Logit (Brady and Rehbeck, 2016), IC—Independent Choice (Manzini and Mariotti, 2014), EBA—endogenous (static) Elimination By Aspects (Tversky, 1972), AR—endogenous (first stage) Attribute Rule (Gul et al., 2014), RCG—Random Categorization (Aguiar, 2017), NSC—Nested Stochastic Choice (Kovach and Tserenjigmid, 2022), RRM—Random Reference Model (Masatlioglu and Ok, 2014; Kibris et al., 2024). The definitions of deterministic-with-full-choice  $\mu$ , singleton- $\mu$ , and nest-invariant  $\mu$  are given in Section 3. Singleton- $\mu$  is the intersection of RRM, NSC, and EBA/AR/RCG. Singleton- $\mu$  is related to the Luce model in the probabilistic choice literature. Nest-invariant  $\mu$  is the intersection of EBA/AR/RCG and NSC. Nest-invariant  $\mu$  is a subset of NSC satisfying the probabilistic version of the attention filter condition studied in Masatlioglu et al. (2012). We assume that the chosen collection is non-empty in all models of  $\mu$  in the figure.

The paper proceeds as follows. Section 2 introduces various models of  $\mu$  and provides their behavioral characterizations. Section 3 analyzes the relationships among these models. Section 4 reviews the related literature. Appendix A provides characterizations of additional models. All proofs are delegated to Appendix B.

## 2 Models and Behavioral Characterizations

### 2.1 Primitives

Let  $X$  be the finite set of all alternatives. We call  $X$  the grand set and assume that there are at least three distinct alternatives in  $X$ . Typical elements of  $X$  are denoted by  $x, y, z$ . Let  $\mathcal{X}$  be the set of all non-empty subsets of  $X$ . Let  $S \in \mathcal{X}$  denote a choice set. For notational simplicity, given a set  $S$  and an element  $x$ , we write  $S \cup x$  instead of  $S \cup \{x\}$  and  $S \setminus x$  instead of  $S \setminus \{x\}$ . Denote the sets of non-negative and positive real numbers by  $\mathbb{R}_+$  and  $\mathbb{R}_{++}$ , respectively.

Let  $\mu$  denote a stochastic choice correspondence from  $\mathcal{X} \times \mathcal{X}$  to  $[0, 1]$ , meaning that  $\mu$  satisfies three properties: (i)  $\mu(T, S) \in [0, 1]$  for all non-empty  $T \subseteq S$  and for all  $S \in \mathcal{X}$ , (ii)  $\sum_{T: T \subseteq S, T \neq \emptyset} \mu(T, S) = 1$  for all  $S \in \mathcal{X}$ , and (iii)  $\mu(T, S) = 0$  whenever  $T \not\subseteq S$ . In this definition,  $\mu(T, S)$  represents the probability of choosing a non-empty collection  $T$  from choice set  $S$ .<sup>3</sup> We say that  $\mu$  has full support if  $\mu(T, S) > 0$  for all pairs  $(T, S)$  such that  $T \subseteq S \subseteq X$  and  $T$  is non-empty.

## 2.2 Logit

In [Brady and Rehbeck \(2016\)](#), each collection  $T$  is associated with a numerical value denoted by  $\pi(T)$ , where  $\pi$  is a mapping from  $\mathcal{X}$  to  $\mathbb{R}_{++}$ . [Brady and Rehbeck \(2016\)](#) interpret  $\pi(T)$  as the likelihood that  $T$  is feasible in  $X$ . Depending on the context, the number  $\pi(T)$  may represent the value, importance, salience, or weight of  $T$  from the DM’s perspective. For instance, in the product recommendation example,  $\pi(T)$  may correspond to the platform’s benefit from recommending a set  $T$  of products to its customers. In the case of supermarket assortment choice,  $\pi(T)$  may denote the sales revenue generated by items in  $T$ . Given a choice set  $S$ , the probability of choosing a non-empty  $T \subseteq S$  is proportional to its value according to a logit formula:

$$\mu_{\text{LG}}(T, S) = \frac{\pi(T)}{\sum_{T': T' \subseteq S, T' \neq \emptyset} \pi(T')} \quad \text{for all non-empty } T \subseteq S.$$

The logit formulation of  $\mu_{\text{LG}}$  is reminiscent of the Luce model in probabilistic choice theory ([Luce, 1959](#)). The Luce model and  $\mu_{\text{LG}}$ , however, are distinct as they operate in two different domains.

It is well known that the Luce model can be characterized by a single axiom, Independence of Irrelevant Alternatives (IIA), which states that the ratio between two choice probabilities does not depend on the presence or absence of other options in the choice set. As  $\mu_{\text{LG}}$  resembles the Luce model, a similar version of IIA, called Independence of Irrelevant Sets, can fully characterize  $\mu_{\text{LG}}$ . The Independence of Irrelevant Sets (IIS), formalized in Axiom 1 below, states that the relative probability of choosing two collections is independent of the presence or absence of other collections. Note that IIS and IIA share similar underlying intuitions but differ in their domains of application.

---

<sup>3</sup>Several models of  $\mu$  examined in the paper allow the chosen collection to be empty; the decision maker is often assumed to select a default option in this case. [Appendix A.1](#) provides characterizations of these models when  $T$  can be empty.

**Axiom 1** (Independence of Irrelevant Sets - IIS).

$$\frac{\mu(T, S)}{\mu(T', S)} = \frac{\mu(T, S')}{\mu(T', S')} \quad \text{for all non-empty } T, T' \subseteq S \cap S'$$

as long as the two ratios are well defined and positive.

**Proposition 1** (Characterization of Logit  $\mu$ ). A full-support  $\mu$  has a Logit representation if and only if it satisfies IIS.

Proposition 1 states that the logit formulation of  $\mu$  in Brady and Rehbeck (2016) is captured by a simple behavioral postulate. The necessity part in Proposition 1 is straightforward. For the sufficiency direction, the proof is constructive and follows an approach similar to that used in the characterization of the Luce model.

### 2.3 Random Categorization

Consider a decision maker (hereafter DM) who divides the grand set into different categories. Each category represents a bundle of alternatives. The categories can overlap with each other. The DM selects a particular category and subsequently chooses a collection by including all available options that belong to the category she considers. In simple terms, the chosen collection is the intersection of the choice set and the considered category. Aguiar (2017) calls this model Random Categorization (RCG).

Formally, Aguiar (2017) assumes that each category  $C$  is a subset of the grand set  $X$ . Each category  $C$  has an associated probability  $m(C)$ , where  $m: \mathcal{X} \rightarrow [0, 1]$  is a probability distribution over  $\mathcal{X}$ .<sup>4</sup> The decision maker first considers a category  $C$ , which happens with probability  $m(C)$ . If there is nothing for the decision maker to choose from, i.e.,  $C \cap S = \emptyset$ , the DM redraws another category until finding one category  $C'$  such that  $C' \cap S \neq \emptyset$ . For this process to end in a finite number of steps, each option  $x \in X$  is assumed to belong to at least one category  $C$  with  $m(C) > 0$ . The probability of choosing a non-empty  $T \subseteq S$  is the sum over all categories  $C$  for which the intersection of  $S$  and  $C$  equals  $T$ , normalized by the probability of all

---

<sup>4</sup>Aguiar (2017) assumes that the collection of categories is a strict subset of  $\mathcal{X}$ . However, one can easily extend the collection of categories to  $\mathcal{X}$  by simply assigning zero weight to all non-empty subsets of  $X$  that do not belong to the original set of categories.

categories overlapping with  $S$ :

$$\mu_{\text{RCG}}(T, S) = \frac{\sum_C m(C) \mathbb{1}(T = C \cap S)}{\sum_{C: S \cap C \neq \emptyset} m(C)} \quad \text{for all non-empty } T \subseteq S.$$

Note that the normalization term,  $\sum_{C: S \cap C \neq \emptyset} m(C)$ , arises from the non-emptiness of the chosen collection. When the chosen collection is allowed to be empty, this term disappears.<sup>5</sup>

The RCG model is applicable to the assortment choice of supermarkets or product recommendations by platforms mentioned earlier. For instance, in the supermarket assortment choice example, each category  $C$  may represent a product category (fruits, vegetables, etc.), and  $m(C)$  may correspond to the probability that items in  $C$  are in stock. The supermarket would like to offer items in  $C$  to customers, but only items in  $S$  are available; hence, the supermarket offers everything in  $C \cap S$ . In the product recommendation example, each category  $C$  may also represent a product category.

The RCG model is characterized by two simple axioms: Positivity-1 and Relative Additivity. The former concerns the coverage of chosen collections, whereas the latter links choice probabilities across two choice sets of different sizes.

**Axiom 2 (Positivity-1).** For all  $(x, S) \in X \times \mathcal{X}$  with  $x \in S$ , there exists  $T \ni x$  such that  $\mu(T, S) > 0$ .

Positivity-1 arises from the assumption that each option  $x$  belongs to at least one category  $C$  with  $m(C) > 0$ . To elaborate, suppose  $x \in S$ . Let  $C$  be an arbitrary category that includes  $x$  with  $m(C) > 0$ . Then  $T = C \cap S$  is chosen with positive probability from choice set  $S$ . Clearly,  $T$  contains  $x$ . Note that Positivity-1 is trivially satisfied when  $\mu$  has full support.

**Axiom 3 (Relative Additivity).** For all  $x, T, T'$ , and  $S$  such that  $x \in S$  and  $T, T' \subseteq S \setminus x$  and  $T, T' \neq \emptyset$ ,

$$\mu(T, S \setminus x)[\mu(T', S) + \mu(T' \cup x, S)] = \mu(T', S \setminus x)[\mu(T, S) + \mu(T \cup x, S)].$$

To understand Relative Additivity, consider some of its implications. When  $\mu$  has

---

<sup>5</sup>Appendix A.1 provides a characterization of  $\mu_{\text{RCG}}$  when the chosen collection can be empty.

full support, Relative Additivity states that

$$\frac{\mu(T, S \setminus x)}{\mu(T, S) + \mu(T \cup x, S)} = \frac{\mu(T', S \setminus x)}{\mu(T', S) + \mu(T' \cup x, S)} \quad \text{for all non-empty } T, T' \subseteq S \setminus x.$$

Consequently, the ratio  $\frac{\mu(T, S \setminus x)}{\mu(T, S) + \mu(T \cup x, S)}$  is independent of  $T$ . To understand this ratio, suppose the initial choice set is  $S \setminus x$  and  $T$  is the chosen collection. Adding a new item  $x$  to  $S \setminus x$  may impact  $T$  in two distinct ways. When  $x$  has an indirect effect, the chosen collection remains the same and  $T$  is selected with updated probability  $\mu(T, S)$ . Alternatively,  $x$  may have a direct effect and be added to the chosen collection, in which case  $T \cup x$  is selected with probability  $\mu(T \cup x, S)$ . Thus, the sum  $\mu(T, S) + \mu(T \cup x, S)$  can be seen as the total choice probability related to  $T$  under the expanded choice set. Hence, the ratio  $\frac{\mu(T, S \setminus x)}{\mu(T, S) + \mu(T \cup x, S)}$  captures the relative probability of choosing a collection before and after adding the new option. Relative Additivity states that this relative probability is independent of the collection chosen initially. The multiplicative form of Relative Additivity in Axiom 3 accounts for the possibility that the ratio  $\frac{\mu(T, S \setminus x)}{\mu(T, S) + \mu(T \cup x, S)}$  might not be well defined when  $\mu$  does not have full support.

Another implication of Relative Additivity is that adding a new alternative to the choice set (weakly) decreases the choice probability of an existing collection. That is,  $\mu(T, S) \leq \mu(T, S \setminus x)$  for all non-empty  $T \subseteq S \setminus x$  with  $x \in S$ . This is the monotonicity condition of  $\mu$  studied in Cattaneo et al. (2020). Consequently, Relative Additivity is sufficient for monotonicity.

Finally, note that Relative Additivity arises from an independence structure between what is chosen (collection  $T$ ) and how its probability is assigned (via category  $C$ ). In RCG, the chosen collection is simply an outcome of the category selection. The probability of choosing  $T$  depends entirely on the probability distribution over categories and is independent of any properties of  $T$ . Once a category is drawn, there is no additional weight applied directly to the chosen set. This independence structure gives rise to Relative Additivity. In general, Relative Additivity emerges in models that exhibit this independence structure and disappears when the structure is absent (see Section 2.5 for details).

We are now ready to state a characterization of RCG. Proposition 2 establishes that the RCG model is characterized by Axioms 2-3.

**Proposition 2 (Characterization of RCG).**  $\mu$  has an RCG representation if and only if it satisfies Positivity-1 and Relative Additivity.

Proposition 2 allows the researcher to test and falsify the RCG model by checking two simple conditions. The proof of Proposition 2 is constructive and uses induction based on the number of alternatives in the choice set. First, for each category  $C \in \mathcal{X}$ , we define its probability as  $m(C) = \mu(C, X)$ . By construction,  $m$  is a probability distribution over categories, and  $\mu$  has an RCG representation at  $X$ . Second, we use Relative Additivity to show that  $\mu$  has an RCG representation at choice sets having  $|X| - 1, \dots, 2$  elements by induction. The induction step utilizes the fact that Relative Additivity connects choice probabilities at two choice sets of different sizes ( $S$  and  $S \setminus x$ ), so we can apply the representation at a larger choice set to obtain the representation in a smaller choice set.

## 2.4 Independent Choice

In [Manzini and Mariotti \(2014\)](#), the choice probability of a collection depends on the probabilities of choosing options within the collection. Each option  $x \in X$  has an independent parameter  $\gamma(x) \in (0, 1)$ , which represents the probability of that option being selected. In the supermarket assortment example,  $\gamma(x)$  may correspond to the probability that product  $x$  is in stock. In the product-recommendation setting,  $\gamma(x)$  may denote the probability that a similar customer buys  $x$ .

For a given choice set  $S$ , a non-empty subset  $T$  of  $S$  is chosen if and only if the DM selects every alternative in  $T$  while ignoring all options in  $S \setminus T$ . The probability of choosing  $T$  is defined as

$$\mu_{\text{IC}}(T, S) = \frac{\prod_{x \in T} \gamma(x) \prod_{y \in S \setminus T} (1 - \gamma(y))}{1 - \prod_{z \in S} (1 - \gamma(z))} \quad \text{for all non-empty } T \subseteq S.$$

In the formulation of  $\mu_{\text{IC}}(T, S)$ , the normalization term,  $1 - \prod_{z \in S} (1 - \gamma(z))$ , represents the probability of drawing a non-empty subset of  $S$ . It arises from the non-emptiness of the chosen collection. In the decision-making process, this normalization ensures that if the DM draws an empty collection as her choice, which occurs with probability  $\prod_{z \in S} [1 - \gamma(z)]$ , she will redraw another collection (with replacement) until a non-empty collection is selected.<sup>6</sup>

---

<sup>6</sup>When the chosen collection is allowed to be empty, the normalization term disappears. [Appendix A.1](#) provides a characterization of  $\mu_{\text{IC}}$  in this case.

The three formulations of  $\mu$  introduced, Logit, Random Categorization, and Independent Choice, differ in their functional forms and underlying narratives. One may wonder whether the three models are related in a particular way. Interestingly, there is a notable connection between them: Independent Choice is equivalent to the intersection of Logit and Random Categorization. In Proposition 3 below, we present a behavioral characterization of Independent Choice (IC). We show that a full-support  $\mu$  has an Independent Choice representation if and only if it satisfies IIS and Relative Additivity. Consequently, the relationship between the three models follows directly from their behavioral foundations.

**Proposition 3 (Characterization of IC).** A full-support  $\mu$  has an Independent Choice representation if and only if it satisfies IIS and Relative Additivity.

Proposition 3 indicates that verifying IIS and Relative Additivity is sufficient to determine whether a full-support  $\mu$  has an Independent Choice representation. Proposition 3, together with Propositions 1-2, sheds light on the connection between Logit, Random Categorization, and Independent Choice. Note that the relationship among the three models was first documented in Kovach and Suleymanov (2023) when studying reference-dependent probabilistic choice in a different setting.<sup>7</sup> Kovach and Suleymanov (2023) do not investigate the behavioral implications of the three models and use their functional forms to identify the relationship. We employ an alternative approach by using their behavioral foundations. Our axiomatic approach clarifies the implications of the functional form assumptions and guides the analyst in selecting the most suitable model for their application of interest. Moreover, the relationships among different models of  $\mu$  often become evident through their behavioral characterizations. As we will show in Section 3, our approach is particularly useful for understanding the connections between various models of  $\mu$  when they exhibit more complex functional forms and are motivated by highly distinct narratives.

The proof of sufficiency in Proposition 3 proceeds as follows. First, we use the IIS axiom and apply Proposition 1 to obtain a Logit representation of  $\mu$ , allowing us to express  $\mu(T, S) = \frac{\pi(T)}{\sum_{T': T' \subseteq S, T' \neq \emptyset} \pi(T')}$  for some  $\pi: \mathcal{X} \rightarrow \mathbb{R}_{++}$ . Next, we apply Relative Additivity to this functional form and, through induction, demonstrate that the Logit

---

<sup>7</sup>More precisely, they document the relationship between the three models when the chosen collection is allowed to be empty. Our result here applies to the case when the selected collection must be non-empty. In Appendix A.1, we provide characterizations of RCG and IC when the DM can select an empty set. Through the behavioral characterizations, we also re-establish their result.

representation can be rewritten as an Independent Choice representation by defining  $\gamma(x) = \frac{\pi(X)}{\pi(X) + \pi(X \setminus x)}$  for all options  $x$ .

The proof of Proposition 3 also sheds light on the relationship among the primitives in the Logit, Random Categorization, and Independent Choice models. Specifically, given the normalization  $\sum_{T': \emptyset \neq T' \subseteq X} \pi(T') = 1$  in the Logit model, we show that the probability measure over categories in the RCG representation must correspond to the  $\pi$  function in the Logit representation:  $m(C) = \pi(C)$  for all  $C \in \mathcal{X}$ . Consequently, the equations  $\gamma(x) = \frac{\pi(X)}{\pi(X) + \pi(X \setminus x)}$  for all  $x \in X$  and  $m(C) = \pi(C)$  for all  $C \in \mathcal{X}$  jointly characterize the interconnection between the primitives in the three models.

## 2.5 Saliency Models

In this section, we investigate three models that incorporate the saliency of various aspects of the choice environment into the decision-making process. These models differ in the type of saliency they consider: (1) saliency of attributes in Elimination By Aspects (Tversky, 1972), (2) saliency of alternatives in Random Reference Model (Kibris et al., 2024), and (3) saliency of nests in Nested Stochastic Choice (Kovach and Tserenjigmid, 2022).

### 2.5.1 Saliency of Attributes

Many products in real life have observable attributes. For instance, a typical soft drink can be differentiated from others by its price, size, design, calorie content, and intended use. Consider a DM who pays attention to a particular attribute and ignores options that do not possess that attribute when making choices. To illustrate, when recommending soft drinks on hot days, a vending machine may prioritize refreshing and hydrating options, such as sparkling water and lemonade, while neglecting heavier soft drinks like colas and sodas. Tversky (1972) calls this model Elimination By Aspects (EBA). Throughout the paper, we focus on the so-called static version of EBA, which is characterized by a one-time elimination of alternatives, as the original EBA model developed by Tversky (1972) allows for a sequence of eliminations until only one option remains. This focus allows the decision makers to choose a collection of options and facilitates comparisons across models, as all other formulations of  $\mu$  examined in this paper are also static. With a slight abuse of terminology, we will refer to the static EBA model simply as EBA. When clarification is necessary, we will refer to the original EBA model as sequential EBA.

Formally, let  $\{i, j, k, \dots\}$  denote the finite set of all observable attributes. Let  $\omega_i > 0$  be the level of attention that attribute  $i$  receives. It measures the salience (or value) of aspect  $i$ . Without loss of generality, assume that  $\sum_i \omega_i = 1$ . Let  $A_i$  be the set of options having aspect  $i$ . Each option has at least one attribute. Suppose aspect  $i$  is salient. All options without aspect  $i$  fail to receive attention. The DM chooses all feasible options having aspect  $i$ ; hence, the selected collection is  $A_i \cap S$ , with  $S$  being the choice set. When  $A_i \cap S = \emptyset$ , the DM switches attention to a new attribute until finding some aspect  $j$  such that  $A_j \cap S \neq \emptyset$ . The probability of choosing a non-empty  $T \subseteq S$  is the normalized sum over all attributes  $i$  satisfying  $T = A_i \cap S$ .<sup>8</sup>

$$\mu_{\text{EBA}}(T, S) = \sum_i \frac{\omega_i \mathbb{1}(T = A_i \cap S)}{\sum_{j: A_j \cap S \neq \emptyset} \omega_j} \quad \text{for all non-empty } T \subseteq S. \quad (1)$$

In the supermarket assortment choice example, aspect  $i$  may represent the shelf life of a product. The collection  $A_i$  could then denote the group of items with similar shelf life. As only products in  $S$  are available, the supermarket offers items in  $A_i \cap S$  to its customers.

To provide a characterization of EBA, we first document a connection between EBA and RCG studied in Section 2.3. In Remark 1, we demonstrate that EBA is nested in RCG by showing that one can obtain a representation for the latter from a representation of the former by appropriately defining the probability distribution over the categories.

**Remark 1.** Let the distribution over categories  $m: \mathcal{X} \rightarrow [0, 1]$  in the RCG model be such that  $m(C) = \sum_{i: A_i = C} \omega_i$ . For all non-empty  $S \subseteq X$  and all non-empty  $T \subseteq S$ ,

$$\mu_{\text{EBA}}(T, S) = \mu_{\text{RCG}}(T, S).$$

Remark 1 follows from the observation that each category  $C$  in the RCG model can be interpreted as the set of options possessing a specific attribute in the EBA model. If  $m(C) = \sum_{i: A_i = C} \omega_i$ , i.e., the probability of drawing category  $C$  equals the total consideration probability of all attributes  $i$  whose collection of options having

---

<sup>8</sup>In Appendix A.2, we show that the static EBA model described here is identical to the first stage of the choice procedure (after reformulation) in the Attribute Rule model introduced by Gul et al. (2014). Our equivalence result sheds light on the differences in choice behaviors between the original EBA model and the Attribute Rule model noted by Gul et al. (2014). See Appendix A.2 for details.

aspect  $i$  is exactly  $C$ , then the two models generate identical choice behavior.

Given that RCG nests EBA as a particular case, it follows from the characterization of  $\mu_{\text{RCG}}$  in Proposition 2 that Positivity-1 (Axiom 2) and Relative Additivity (Axiom 3) are necessary for the EBA model. The two axioms, however, are not sufficient for a characterization of EBA. Note that in the EBA model, the DM first selects a particular attribute and subsequently chooses a collection containing all feasible options possessing that attribute. Hence, the decision maker's choices depend on the set of attributes, which is assumed to be observable and exogenous. This exogeneity assumption of attributes imposes a stronger restriction on  $\mu$  than Positivity-1. We refer to this stronger condition as Positivity-2 (Axiom 4). Note that Positivity-2 is nested in Positivity-1, as the former implies the latter.

**Axiom 4 (Positivity-2).** For all non-empty  $T \subseteq S$ ,  $\mu(T, S) > 0$  if and only if there exists an attribute  $i$  such that  $T = A_i \cap S$ .

Positivity-2 follows from the definition of the EBA model. Positivity-2 states that the DM selects a collection with positive probability if and only if it corresponds to the set of all feasible options possessing a specific attribute. By combining Positivity-2 and Relative Additivity, we obtain a characterization of EBA. This is established in Proposition 4.

**Proposition 4 (Characterization of EBA).**  $\mu$  has an EBA representation if and only if it satisfies Positivity-2 and Relative Additivity.

Proposition 4 allows the researcher to verify whether the decision maker's choice aligns with the elimination of aspects model. The proof of Proposition 4 follows from the characterization of  $\mu_{\text{RCG}}$  in Proposition 2.

We complete this section by commenting on the exogeneity assumption of attributes in EBA. The model assumes that attributes are exogenously given and observed by the outside analyst. Such an assumption can be overly restrictive in certain contexts, as the outside analyst may not directly observe the attributes. Hence, relaxing the exogeneity assumption by endogenizing attributes can improve the practical relevance of the model. Additionally, it allows us to compare EBA to other models that do not have attributes in the choice environment.

To investigate the EBA model when attributes are endogenous, we first define an endogenous EBA model in Definition 1.

**Definition 1 (Endogenous EBA).**  $\mu$  has an endogenous EBA representation if there exists a finite collection of attributes  $\{i, j, k, \dots\}$ , with each option having at least one attribute, such that  $\mu$  has an EBA representation with  $\{i, j, k, \dots\}$  being the set of attributes.

Note that as the attributes in the ordinary sense are not observed, the term *attributes* in Definition 1 does not possess a specific meaning beyond functioning as a label. In particular, the outside analyst can define *artificial aspects*  $\{i, j, k, \dots\}$  and the associated set of options  $A_i$  possessing attribute  $i$  to test whether the dataset adheres to the elimination of the defined artificial aspects.

In environments where attributes are endogenous, Remark 2 shows that EBA and RCG are equivalent. Note that when attributes are observable, Remark 1 states that EBA is strictly nested in RCG, with the strictness coming from the exogeneity of attributes. Endogenizing the attributes eliminates the condition imposed by their exogeneity; hence, EBA corresponds to RCG without this additional restriction.

**Remark 2.**  $\mu$  has an endogenous (static) EBA representation if and only if it has an RCG representation.

Together with Proposition 2, Remark 2 implies that verifying Positivity-1 and Relative Additivity is sufficient to determine whether the decision maker's choice adheres to the one-time elimination of aspects, where the aspects can be defined arbitrarily by the analyst.

## 2.5.2 Salience of Alternatives

Many studies in psychology and economics have documented that the decision maker's behaviors often depend on what she perceives as a reference point (Samuelson and Zeckhauser, 1988; Knetsch, 1989; Kahneman, 1992).<sup>9</sup> There are two main views on modeling these reference-dependent behaviors. The first view assumes that the reference point directly impacts the DM's evaluation and explicitly incorporates the reference point into the DM's utility function. In the second view, the reference point does not affect the DM's utility function but rather influences what the DM perceives and is willing to select.<sup>10</sup>

---

<sup>9</sup>Reference points can be default options, initial endowments, entitlements, or past choices.

<sup>10</sup>Note that the two views are not mutually exclusive.

The Random Reference Model (RRM) in this section adopts the second view above and is based on [Masatlioglu and Ok \(2014\)](#) and [Kibris et al. \(2024\)](#). There exist multiple reference points. The reference point is endogenously determined by a random process. The RRM restricts the reference point to be an option within the choice set, i.e., a reference point must be a feasible alternative. Each option  $x \in X$  has an associated value  $s_x > 0$ , which may represent its salience, value, or importance. The probability of an option being the reference point depends on its salience relative to the total salience of all feasible alternatives. Given choice set  $S$ , option  $x \in S$  is the reference point with probability  $\frac{s_x}{\sum_{y \in S} s_y}$ . When alternative  $x$  serves as the reference point, it imposes a constraint on the DM through a constraint set  $\mathcal{Q}(x) \subseteq X$ . The set  $\mathcal{Q}(x)$  is interpreted as the collection of options that are appealing from the perspective of  $x$ . It may correspond to the set of options that  $x$  reminds the decision maker of, the set of options that are similar to  $x$ , or the set of options that are unambiguously better than  $x$ . Options that are not appealing from the perspective of  $x$  fail to receive attention. The DM selects all feasible alternatives within her constraint set. Given choice set  $S$ , the probability of choosing a non-empty  $T \subseteq S$  is defined as

$$\mu_{\text{RRM}}(T, S) = \sum_{x \in S} \frac{s_x}{\sum_{y \in S} s_y} \mathbb{1}(T = \mathcal{Q}(x) \cap S) \quad \text{for all non-empty } T \subseteq S. \quad (2)$$

In the supermarket assortment choice example, the reference point  $x$  may represent a product currently in high demand, and the constraint set associated with  $x$ ,  $\mathcal{Q}(x)$ , may include complementary items or products frequently purchased together with  $x$ . In the case of Japanese vending machines, as the machine can scan the current customer's face and recall similar past customers from its database, the reference point  $x$  may represent the most similar past purchase. The set  $\mathcal{Q}(x)$  may then represent the collection of drinks similar to  $x$ .

Regarding the constraint set, [Masatlioglu and Ok \(2014\)](#) assume that each option must be appealing from the perspective of itself, i.e., each alternative  $x$  is always included in its constraint set  $\mathcal{Q}(x)$ . This property is related to experimental findings that one alternative becomes more desirable when it serves as the reference point ([Knettsch, 1989](#)). We further assume that the constraint sets associated with two different options are distinct:  $\mathcal{Q}(x) \neq \mathcal{Q}(y)$  when  $x \neq y$ . This assumption allows the decision maker to possibly make different choices under two reference points, even when they are equally salient. The property is related to an experimental finding that

individuals often exhibit distinct behaviors under different reference points (Sullivan and Kida, 1995; Lin et al., 2006).

The functional form of  $\mu_{\text{RRM}}$  resembles that in both EBA and RCG. All three models formulate the decision maker's choice as a collection of feasible options satisfying certain properties. In EBA, these options possess the salient attribute. In RCG, they fall into the same category. In RRM, they belong to a constraint set imposed by the salient alternative (the reference point). Given the similarities in how the three models formalize choices, as we will show, all behavioral postulates in RRM are connected to those in EBA and RCG. However, the conditions imposed on the constraint sets and the feasibility of the reference point distinguish RRM from EBA and RCG.<sup>11</sup>

To provide a characterization of  $\mu_{\text{RRM}}$ , we adopt a *revealed constraint approach*. We first assume that  $\mu$  has an RRM representation and identify the constraint sets from  $\mu$ . We subsequently utilize the identified constraint sets in behavioral postulates. Identifying the constraint sets uses choice probabilities from binary choice sets. Table 2 presents the probabilities of choosing different collections at choice set  $\{x, y\}$  under four possible scenarios regarding the constraint sets.

Scenarios	$\mu_{\text{RRM}}(\{x, y\}, \{x, y\})$	$\mu_{\text{RRM}}(\{x\}, \{x, y\})$	$\mu_{\text{RRM}}(\{y\}, \{x, y\})$
$x \in \mathcal{Q}(y)$ and $y \in \mathcal{Q}(x)$	1	0	0
$x \notin \mathcal{Q}(y)$ and $y \notin \mathcal{Q}(x)$	0	$\frac{s_x}{s_x + s_y}$	$\frac{s_y}{s_x + s_y}$
$x \notin \mathcal{Q}(y)$ and $y \in \mathcal{Q}(x)$	$\frac{s_x}{s_x + s_y}$	0	$\frac{s_y}{s_x + s_y}$
$x \in \mathcal{Q}(y)$ and $y \notin \mathcal{Q}(x)$	$\frac{s_y}{s_x + s_y}$	$\frac{s_x}{s_x + s_y}$	0

Table 2: Probabilities of choosing different collections in choice set  $\{x, y\}$  in RRM

Suppose  $\mu$  has an RRM representation. From Table 2, it is immediate that  $y$  belongs to the constraint set associated with  $x$  if and only if the DM selects the singleton set  $\{x\}$  from binary choice set  $\{x, y\}$  with zero probability. To understand why this property holds under the RRM, regarding the if part, note that if  $y \notin \mathcal{Q}(x)$ , then the DM chooses  $\{x\}$  from choice set  $\{x, y\}$  when the reference point is  $x$ . This implies that  $\mu(\{x\}, \{x, y\}) > 0$ , which contradicts the assumption that  $\mu(\{x\}, \{x, y\}) = 0$ . For the only-if part, by definition of the RRM, the DM chooses  $\mathcal{Q}(x) \cap \{x, y\}$  or  $\mathcal{Q}(y) \cap \{x, y\}$  when  $x$  or  $y$  serves as the reference point, respectively.

<sup>11</sup>In Section 3, we show that RRM and RCG intersect only in a particular case where the decision maker always selects singleton sets.

The assumption that  $y \in \mathcal{Q}(x)$  implies  $y \in \mathcal{Q}(x) \cap \{x, y\}$  and the fact that  $y$  is included in its constraint set implies  $y \in \mathcal{Q}(y) \cap \{x, y\}$ . Hence, regardless of which option serves as the reference point, the DM never chooses  $\{x\}$ . It follows that the choice probability of  $\{x\}$  in choice set  $\{x, y\}$  must be equal to 0.

To formalize the observation above, define the *revealed constraint function*  $\mathcal{Q}^R: X \rightarrow \mathcal{X}$  as follows. For each  $x \in X$ ,  $\mathcal{Q}^R(x)$  includes  $x$  and all options  $y \neq x$  for which  $\mu(\{x\}, \{x, y\}) = 0$ . Mathematically,

$$\mathcal{Q}^R(x) = \{x\} \cup \{y : y \in X, y \neq x \text{ such that } \mu(\{x\}, \{x, y\}) = 0\} \quad \text{for all } x \in X.$$

The revealed constraint function  $\mathcal{Q}^R$  corresponds to the underlying constraint function if  $\mu$  has an RRM representation.

We are now ready to state the behavioral postulates of RRM. The first axiom (Axiom 5) follows from the assumption that the constraint sets associated with two different alternatives are distinct.

**Axiom 5 (Distinct Constraint Sets).**  $\mathcal{Q}^R(x) \neq \mathcal{Q}^R(y)$  when  $x \neq y$ .

The second and third axioms (Axioms 6 and 7) are relatively similar to those in EBA and RCG, which follow from the similarity in the formulation of  $\mu$  across the three models.

Axiom 6 states that the DM chooses a collection  $T$  with a positive probability if and only if there exists a reference point  $x$  such that  $T$  is identical to the collection of all feasible alternatives that belong to the constraint set associated with  $x$ . Axiom 6 is a direct consequence of the definition of RRM. Compared to RCG and EBA models, Axiom 6 is similar to Axiom 2 and Axiom 4.

**Axiom 6 (Positivity-3).** For all non-empty  $T \subseteq S$ ,  $\mu(T, S) > 0$  if and only if there exists some  $x \in S$  such that  $T = \mathcal{Q}^R(x) \cap S$ .

Axiom 7 below is a slightly modified variant of Relative Additivity in EBA and RCG models.

**Axiom 7 (Relative Additivity-1).** For all non-empty  $T, T' \subseteq S \setminus x$  and  $x \in S$  such that  $T, T' \neq \mathcal{Q}^R(x) \cap (S \setminus x)$

$$\mu(T, S \setminus x)[\mu(T', S) + \mu(T' \cup x, S)] = \mu(T', S \setminus x)[\mu(T, S) + \mu(T \cup x, S)].$$

Axiom 7 deviates from the original Relative Additivity by narrowing down its applicability through explicitly requiring  $T, T'$  to satisfy an additional condition:  $T, T' \neq \mathcal{Q}^R(x) \cap (S \setminus x)$ . This additional constraint arises from the assumption that the reference point must be feasible and included in the associated constraint set. Specifically, the RRM allows  $x$  to be a reference point under choice set  $S \ni x$ , but it cannot be the reference point under  $S \setminus x$  because  $x$  is no longer available. To understand how this assumption affects Relative Additivity, consider a scenario where  $T = \mathcal{Q}^R(x) \cap (S \setminus x)$ , so the condition in Axiom 7 fails to hold. It follows that  $T \cup x = \mathcal{Q}^R(x) \cap S$  because  $x \in \mathcal{Q}^R(x)$ . Consequently, the reference point  $x$  influences the probability of choosing  $T \cup x$  under  $S$ . However, there are no similar impacts under the choice set  $S \setminus x$  because  $x$  cannot serve as a reference point when it is no longer feasible. Hence, Relative Additivity fails to hold, and that explains why the condition  $T, T' \neq \mathcal{Q}^R(x) \cap (S \setminus x)$  is present in Axiom 7.

The next behavioral postulate imposes a restriction on  $\mu$  when  $T = \mathcal{Q}^R(x) \cap (S \setminus x)$ , a case not covered by Axiom 7. It is also a variant of Relative Additivity.

**Axiom 8 (Relative Additivity-2).** Suppose  $x \in S$  and  $T = \mathcal{Q}^R(x) \cap (S \setminus x)$  is non-empty. For all non-empty  $T' \subseteq S \setminus x$  such that  $T' \neq T$

$$\mu(T, S \setminus x)[\mu(T', S) + \mu(T' \cup x, S)] = \mu(T', S \setminus x) \left[ \mu(T, S) + \mu(T \cup x, S) - \underbrace{\frac{\mu(\mathcal{Q}^R(x), X)}{\sum_{y \in S} \mu(\mathcal{Q}^R(y), X)}}_{\text{adjustment term}} \right]. \quad (3)$$

Axiom 8 primarily differs from the original Relative Additivity and Axiom 7 through the inclusion of an additional term,  $-\frac{\mu(\mathcal{Q}^R(x), X)}{\sum_{y \in S} \mu(\mathcal{Q}^R(y), X)}$ , on the right-hand side of equation (3). This adjustment term, in absolute value, represents the probability of  $x$  being the reference point under  $S$ , given that the salience of  $x$  is identified as  $s_x = \mu(\mathcal{Q}^R(x), X)$ . Including this adjustment term, as in Axiom 8, neutralizes the influence of  $x$  on the choice probability of  $T \cup x$  when serving as the reference point at  $S$ .

We are now ready to state a characterization of RRM.

**Proposition 5 (Characterization of RRM).**  $\mu$  has an RRM representation if and only if it satisfies Axioms 5-8.

Proposition 5 provides the analyst with a practical means to test whether  $\mu$  has an RRM representation by validating four simple conditions. The proof of Proposition 5

proceeds as follows. We first define the constraint function  $\mathcal{Q}$  as the *revealed constraint function*  $\mathcal{Q}^R$  and the salience of option  $x$  as  $s_x = \mu(\mathcal{Q}^R(x), X)$ . Axioms 5 and 6 guarantee that  $\{\mathcal{Q}(x)\}_x$  and  $\{s_x\}_x$  are well defined. We subsequently show that  $\mu$  has an RRM representation given the defined constraint sets and salience parameters. The proof uses induction based on the number of alternatives in the choice set. The proof of Proposition 5 also provides a useful identification result, demonstrating that the constraint function is uniquely identified, and the salience parameters are unique up to uniform scaling.

### 2.5.3 Salience of Nests

Nested logit is one of the most common parametric frameworks for analyzing probabilistic choice. Kovach and Tserenjigmid (2022) generalize the nested logit model by introducing a Nested Stochastic Choice (NSC).<sup>12</sup> In an NSC, there exist pairwise disjoint nests  $N_1, N_2, \dots, N_q$  ( $q \geq 1$ ) that partition the grand set. Each nest represents a collection of alternatives sharing some common characteristics. Every nest has an associated value. The DM first draws a salient nest and subsequently selects a collection comprising all feasible options within the salient nest. The DM's choice over nests is governed by a Luce formula through a weighting (or salience) function  $\sigma: \mathcal{X} \rightarrow \mathbb{R}_+$ . The function  $\sigma$  assigns a weight  $\sigma(T)$  to each collection  $T$ , where the weight  $\sigma(T)$  may represent the salience, value, or importance of  $T$ . Kovach and Tserenjigmid (2022) further assume that  $\sigma(\emptyset) = 0$  and  $\sigma(T) > 0$  whenever  $T \neq \emptyset$ . Given choice set  $S$ , the probability of choosing a non-empty  $T \subseteq S$  is defined as

$$\mu_{\text{NSC}}(T, S) = \sum_{i=1}^q \frac{\sigma(N_i \cap S)}{\sum_{j=1}^q \sigma(N_j \cap S)} \mathbb{1}(T = N_i \cap S) \quad \text{for all non-empty } T \subseteq S. \quad (4)$$

In the supermarket assortment choice example, each nest may represent a distinct product category, such as meat, fish, frozen vegetables, or dairy, while  $\sigma(T)$  may indicate the sales of the items in  $T$ . In the Japanese vending machine context, each nest may correspond to a specific product brand, such as Coke or 7UP, with  $\sigma(T)$  denoting the within-day revenue of drinks in  $T$ .

The formulation of  $\mu$  in equation (4) looks similar to that in EBA, RCG, or RRM. As each nest in NSC is a collection of alternatives, the notion of a salient nest in NSC is conceptually comparable to the notions of a category in RCG, a set of options

---

<sup>12</sup>See Appendix A.3 for the relationship among nested logit and other models in our framework.

possessing a specific attribute in EBA, or a constraint set in RRM. The primary distinction between NSC and these three models lies in how the weighting functions impact choice probabilities. In NSC, the probability of choosing a collection depends on the collection’s salience value, as  $\sigma(T)$  explicitly appears in  $\mu_{\text{NSC}}(T, S)$ . In contrast, in EBA, RCG, or RRM, the probability of selecting a collection is independent of its salience, properties, or structure. Instead, this probability depends on category-specific weight (in RCG), attribute-specific weight (in EBA), or item-specific weight (in RRM). More generally, EBA, RCG, and RRM exhibit an independence structure between what is chosen and how its probability is assigned. Meanwhile, in NSC, such an independence structure is absent. As we will show in Section 3, this fundamental distinction sets NSC apart from the other three models.

Because of the way in which the probability of choosing a collection is assigned, NSC is behaviorally closest to the Logit model of Brady and Rehbeck (2016). Intuitively, this similarity arises from the fact that Brady and Rehbeck (2016) adopt a logit functional form, whereas NSC is a generalization of the nested logit. In Section 2.2, we showed that a full-support Logit  $\mu$  is characterized by IIS (Axiom 1). The primary behavioral postulate in NSC, Path-Independence of Irrelevant Sets (PIIS, Axiom 9), is conceptually analogous to IIS. Broadly speaking, PIIS states that  $\frac{\mu(T, S)}{\mu(T^*, S)} \frac{\mu(T^*, S')}{\mu(T', S')}$  is independent of the collection  $T^*$  (the “path”) and choice sets  $S, S'$  whenever the two ratios are well defined and positive.

**Axiom 9 (Path-Independence of Irrelevant Sets, PIIS).** For arbitrary  $T$  and  $T'$ ,

$$\frac{\mu(T, S)}{\mu(T^*, S)} \frac{\mu(T^*, S')}{\mu(T', S')} \quad \text{is independent of the choice of } (T^*, S, S')$$

as long as the two ratios are well defined and positive.

PIIS constitutes a generalization of IIS. To elaborate, whenever two collections  $T$  and  $T'$  are selected with strictly positive probabilities in both choice sets  $S$  and  $S'$ , applying PIIS with  $(T^*, S, S') = (T, S, S')$  and  $(T^*, S, S') = (T', S, S')$  yields IIS. This scenario, however, does not always occur in the NSC model because the decision maker may choose certain collections with zero probability. Consequently, PIIS can be interpreted as a generalization of IIS that accommodates situations where  $\mu$  does not have full support.

Besides PIIS, characterizing NSC requires two additional behavioral postulates

(Axioms 10-11). Axiom 10 follows from the partition structure of the nests in the model's definition. Given observed data on  $\mu$ , let  $N_1^R, N_2^R, \dots, N_q^R$  be all pairwise distinct collections selected with positive probability in the grand set ( $q \geq 1$ ). These collections are the *revealed nests*, i.e., the underlying nests if  $\mu$  has an NSC representation. Axiom 10 states that these *revealed nests* form a partition of the grand set.

**Axiom 10 (Partition).**  $N_i^R \cap N_j^R = \emptyset$  for all  $i, j$  such that  $i \neq j$ . Furthermore,  $\bigcup_{i=1}^q N_i^R = X$ .

Axiom 11, presented below, is similar to Positivity-1 (Axiom 2) in RCG, Positivity-2 (Axiom 4) in EBA, or Positivity-3 (Axiom 6) in RRM. Axiom 11 states that the DM chooses a collection with a strictly positive probability if and only if it corresponds to the set of feasible options within some *revealed nest*. This behavioral postulate is also a consequence of the model's definition.

**Axiom 11 (Positivity-4).** For all non-empty  $T \subseteq S$ ,  $\mu(T, S) > 0$  if and only if there exists some  $i$  such that  $T = N_i^R \cap S$ .

We are now ready to state a characterization of NSC. Proposition 6 establishes that NSC is characterized by Axioms 9-11.

**Proposition 6 (Characterization of NSC).**  $\mu$  has an NSC representation if and only if it satisfies Axioms 9-11.

Proposition 6 implies that determining whether  $\mu$  has an NSC representation is equivalent to testing whether it satisfies three conditions in Axioms 9-11. The proof of sufficiency in Proposition 6 proceeds as follows. Define the nests as *the revealed nests*. Obtaining an NSC representation is trivial when there is only one revealed nest. When at least two distinct revealed nests exist, we construct the weighting function  $\sigma$  using a cross-nest strategy. Fix two arbitrary options in two different revealed nests:  $x_1 \in N_1^R$  and  $x_2 \in N_2^R$ . Assign  $\sigma(\{x_1\}) = 1$  and let  $\sigma(\{x_2\}) = \frac{\mu(\{x_2\}, \{x_1, x_2\})}{\mu(\{x_1\}, \{x_1, x_2\})}$ . When  $T$  and  $\{x_1\}$  are in different revealed nests, define  $\sigma(T)$  through  $\sigma(\{x_1\})$  by setting  $\sigma(T) = \frac{\mu(T, T \cup x_1)}{\mu(\{x_1\}, T \cup x_1)} \sigma(\{x_1\})$ . When  $T$  and  $\{x_1\}$  are in the same revealed nest, we use the fact that  $T$  and  $\{x_2\}$  are in two different nests and define  $\sigma(T)$  through  $\sigma(\{x_2\})$  by setting  $\sigma(T) = \frac{\mu(T, T \cup x_2)}{\mu(\{x_2\}, T \cup x_2)} \sigma(\{x_2\})$ . Using Axioms 9-11, we prove that the  $\sigma$  function is well defined and that  $\mu$  has an NSC representation given  $\sigma$  and

the revealed nests. The proof of Proposition 6 also indicates that if  $\mu$  has an NSC representation, then the nests can be uniquely identified and the weighting function is unique up to uniform scaling.

### 3 Relationship Between Models

In this section, we investigate the relationship between various models of  $\mu$  studied in Section 2. We first define three particular cases of  $\mu$ . In Definition 2, we say that  $\mu$  is deterministic-with-full-choice if the decision maker selects everything in the choice set with certainty. In the consideration set literature, this specification of  $\mu$  is often referred to as full attention.

**Definition 2.**  $\mu$  is deterministic-with-full-choice if  $\mu(S, S) = 1$  for all non-empty  $S \subseteq X$ .

In Definition 3 below, we introduce a singleton representation of  $\mu$ . Under a singleton- $\mu$ , the DM chooses a collection with positive probability if and only if it contains only one option (part (i)). Additionally, the relative choice probabilities are independent of the choice set (part (ii)). Note that the singleton- $\mu$  and the Luce model in the probabilistic choice literature are distinct, as they operate in two different domains. However, one can obtain the Luce model from a singleton- $\mu$  by appropriately defining choice probabilities.<sup>13</sup>

**Definition 3 (Singleton- $\mu$ ).**  $\mu$  admits a singleton representation if

- (i)  $\mu(\{x\}, S) > 0$  for all  $x \in S$ , for all  $S \subseteq X$ , and  $\mu(T, S) = 0$  whenever  $|T| \geq 2$ ;
- (ii)  $\frac{\mu(\{x\}, S)}{\mu(\{y\}, S)} = \frac{\mu(\{x\}, S')}{\mu(\{y\}, S')}$  for all  $S, S' \supseteq \{x, y\}$ .

Definition 4 below presents a particular case of the NSC model, called nest-invariant  $\mu$ , where the weighting function remains constant within the same nest:  $\sigma(T) = \sigma(T')$  whenever  $T, T' \neq \emptyset$  and  $T, T' \subseteq N_i$  for some  $i$ . This scenario may correspond to an environment where every collection of items within the same nest is equally salient, or to a decision maker who values such collections equally.

**Definition 4 (Nest-invariant  $\mu$ ).**  $\mu$  has a nest-invariant representation if  $\mu$  has an NSC representation with a nest-invariant weighting function.

---

<sup>13</sup>To elaborate, suppose  $\mu$  has a singleton representation. Let  $\rho: X \times \mathcal{X} \rightarrow [0, 1]$  be a stochastic choice function. For all non-empty  $S \subseteq X$ , define  $\rho(x, S) = 0$  if  $x \notin S$  and  $\rho(x, S) = \mu(\{x\}, S)$  if  $x \in S$ . It is straightforward to verify that  $\rho$  has a Luce representation.

The nest-invariant  $\mu$  encompasses both the deterministic-with-full-choice- $\mu$  and the singleton- $\mu$  as special cases. Specifically, the deterministic-with-full-choice  $\mu$  corresponds to a nest-invariant  $\mu$  wherein there exists a single nest that is identical to the grand set. Meanwhile, the singleton- $\mu$  is equivalent to a nest-invariant  $\mu$  where there are  $|X|$  nests, each consisting of a distinct option.

Although the nest-invariant  $\mu$  is defined in terms of the model's unobserved primitive (the weighting function), it can also be characterized solely by the observed data. The nest-invariant  $\mu$  satisfies a property called the probabilistic attention filter. This property specifies conditions under which adding a new item to the choice set does not affect the choice probabilities of a collection chosen initially. It is a generalization of the attention filter condition studied in Masatlioglu et al. (2012) to a probabilistic setting.<sup>14</sup>

**Definition 5 (Probabilistic attention filter).** Take  $x \in S$  and non-empty  $T \subseteq S \setminus x$ . Then  $\mu(T, S) = \mu(T, S \setminus x)$  if  $\mu(\{x\}, S) = 0$ ,  $\mu(T, S) > 0$ , and  $\mu(T, S \setminus x) > 0$ .

To understand the probabilistic attention filter property, consider a scenario where the initial choice set is  $S \setminus x$  and  $T \neq \emptyset$  is the chosen collection, which implies that  $\mu(T, S \setminus x) > 0$ . The property states that adding  $x$  to the choice set does not impact the choice probability of  $T$  provided that two conditions are satisfied: (i)  $\{x\}$  is not selected in the expanded choice set, and (ii)  $T$  is still chosen with positive probability in the expanded choice set. Intuitively, these conditions ensure that the introduction of  $x$  does not shift the decision maker's focus away from  $T$ ; hence, the choice probability of  $T$  remains unchanged. We will show in Proposition 8 that the nest-invariant  $\mu$  is equivalent to an NSC that satisfies the probabilistic attention filter condition. Consequently, the nest-invariant  $\mu$  is characterized by three axioms in the NSC model (Axioms 9-11) along with the probabilistic attention filter property.

We are now ready to provide the connections between different models of  $\mu$ .

**Relationship between Logit, IC, and other models.** As mentioned earlier in the discussion following Proposition 3, the intersection of Logit and RCG is identical to IC. Additionally, Remark 2 states that RCG and endogenous (static) EBA are equivalent. Therefore, the intersection of Logit and endogenous EBA is also identical

---

<sup>14</sup>The deterministic attention filter condition states that choice is unaffected by an overlooked alternative:  $\Gamma(S) = \Gamma(S \setminus x)$  if  $x \in S \setminus \Gamma(S)$  and  $x \in S$ . Here,  $\Gamma(S)$  denotes the (deterministic) chosen collection from choice set  $S$ .

to IC. This result implies that the intersection of two models of  $\mu$  that have proven useful in capturing individuals' behavior has its own distinct appeal.

Regarding other formulations of  $\mu$ , note that Logit requires  $\mu$  to have full support. In both RRM and NSC, this full-support condition fails to hold, as the DM always chooses some collections with zero probability. For instance, in NSC, two distinct and overlapping collections cannot be selected with positive probability from the same choice set. Consequently, Logit  $\mu$  is disjoint from RRM and NSC. Additionally, since IC is nested in Logit following Propositions 1-3, IC is also disjoint from RRM and NSC. These results are stated in Remark 3.

**Remark 3 (Logit, IC, and other models).** Suppose  $\mu$  has an RRM or NSC representation. Then it does not have a Logit or an IC representation.

**Relationship between RRM and other models.** Proposition 7 below summarizes the connections between RRM and other formulations of  $\mu$ . It establishes that RRM is sharply distinct from these models, as RRM intersects other models only in a particular case where  $\mu$  has a singleton representation.

**Proposition 7 (RRM and other models).** The following are equivalent:

- (i)  $\mu$  has a singleton representation
- (ii)  $\mu$  has both RRM and endogenous EBA (or RCG) representations
- (iii)  $\mu$  has both RRM and NSC representations.

Two comments follow from Proposition 7. First, although conceptually similar in terms of the underlying choice procedure and behavioral postulates, RRM intersects endogenous EBA (or RCG) only in a particular case when  $\mu$  has a singleton representation. The strong distinction between RRM and these two models primarily comes from the restrictions imposed on the reference points and constraint sets in RRM. When these restrictions are relaxed, the intersection of these models enlarges.

Second, Proposition 7 implies that RRM and NSC intersect when the constraint sets in RRM and the collection of nests in NSC share a common structure. Specifically, in the RRM, each constraint set must consist of a single option. Similarly, each nest in the NSC model also comprises only one alternative. Consequently, at the intersection of RRM and NSC, there is a one-to-one mapping that links each nest to an identical constraint set. To understand why this result holds, note that, by definition,  $\mu(\{x\}, S) > 0$  for all  $x$  and  $S \ni x$  when  $\mu$  has a singleton representation.

On the one hand, in RRM, the constraint set associated with option  $x$  cannot include another alternative  $y \neq x$ . Otherwise, it follows from Table 2 (see Section 2.5.2) that  $\mu(\{x\}, \{x, y\}) = 0$ , which contradicts the definition of a singleton- $\mu$ . On the other hand, in NSC, if there exists a nest  $N_i$  consisting of at least two distinct alternatives  $x$  and  $y$ , the DM will select  $\{x\}$  with zero probability at choice set  $\{x, y\}$ , which also contradicts the singleton- $\mu$  definition.

**Relationship between NSC and other models.** Proposition 8 below summarizes the connections between NSC and other formulations of  $\mu$ . The link between NSC and RRM is omitted as it is detailed in Proposition 7. Overall, Proposition 8 indicates that NSC is more closely aligned with other models than RRM, as the intersection of NSC and a given model always nests the intersection of RRM and that model.

**Proposition 8 (NSC and other models).** The following are equivalent:

- (i)  $\mu$  has a nest-invariant representation
- (ii)  $\mu$  has NSC and endogenous EBA (or RCG) representations
- (iii)  $\mu$  has an NSC representation and satisfies the probabilistic attention filter property.

Two remarks are in order. First, note that nest-invariant  $\mu$  includes deterministic-with-full-choice- $\mu$  and singleton- $\mu$  as two particular cases. Consequently, Proposition 8 implies that the intersection of NSC and endogenous EBA (or RCG) nests the intersection of NSC and RRM (see also Proposition 7). Second, at the intersection of NSC and endogenous EBA, the attributes in the EBA model are mutually exclusive, meaning that each option has exactly one attribute. Similarly, at the intersection of NSC and RCG, the categories in RCG are necessarily non-overlapping. These results follow directly from the partition structure of the collection of nests in the NSC model.

In summary, by combining the results in Propositions 1-3, Propositions 7-8, Remarks 2-3, and Corollary 2, we establish the connections among all models of  $\mu$  examined in the paper (see Figure 1 for their relationships).

## 4 Related Literature

Our paper builds on the existing literature on random consideration set formation, pioneered by Manzini and Mariotti (2014) and Cattaneo et al. (2020), among others. Our work contributes to this literature by demonstrating the link between various

parametric models of random consideration set formation and by offering a practical tool, based on behavioral postulates, to test and potentially falsify these models.

As the  $\mu$  object studied in the paper is a stochastic choice correspondence, our work contributes to the literature axiomatizing choice correspondences. Most papers in this literature focus on deterministic environments.<sup>15</sup> Our paper belongs to the small set of studies axiomatizing stochastic choice correspondences (see also [Barberá and Pattanaik \(1986\)](#)). Additionally, the decision maker in our framework deliberately chooses multiple alternatives each time she faces a choice problem. Meanwhile, the literature on deterministic choice correspondences typically assumes that the DM selects a single option, and a choice correspondence arises from factors such as indifference, incompleteness, multiple preferences ([Rubinstein, 2012](#); [Balakrishnan et al., 2022](#)), or frames ([Salant and Rubinstein, 2008](#)).

The work closest to ours is [Kovach and Suleymanov \(2023\)](#). Their paper assumes that a reference point affects the formation of  $\mu$  and interprets  $\mu$  as a distribution of consideration sets. Their focus differs fundamentally from ours: we investigate choices that involve selecting multiple options, whereas their study analyzes how reference-dependent attention affects the choice probabilities of individual options. Moreover, their paper does not provide a behavioral characterization of any model of  $\mu$  examined in our study.

Another broadly related paper is [Manzini et al. \(2024\)](#), which proposes a model of approval in which decision makers approve a collection of options for later use from a ranked list of items. [Manzini et al. \(2024\)](#) introduce a new model rather than providing a systematic examination of well-known models of  $\mu$ , as we do. Additionally, items in our setting are not necessarily ordered.

## 5 Conclusion

We study choice situations that naturally involve selecting multiple options. We systematically investigate various parametric and stochastic models in the literature by providing their behavioral foundations and identifying their relationships. Theoretically, our results indicate that popular models, though based on distinct underlying narratives, share similar behavioral foundations. These theoretical results help the

---

<sup>15</sup>Recent studies axiomatize deterministic choice correspondences in different contexts. See, for instance, [Aizerman and Aleskerov \(1995\)](#); [Masatlioglu and Ok \(2005\)](#); [Salant and Rubinstein \(2008\)](#); [Stoye \(2015\)](#), among others.

analyst better understand the implications of the functional form assumptions and distinguish between different models of  $\mu$ . Empirically, our behavioral postulates facilitate the identification and testing of the choice procedures used by decision makers. Business managers, for instance, can use our behavioral characterizations to investigate how customers form their consideration sets, which is essential for analyzing their choices.

## References

- AGUIAR, V. H. (2017): “Random Categorization and Bounded Rationality,” *Economics Letters*, 159, 46–52.
- AIZERMAN, M. A. AND F. T. ALESKEROV (1995): *Theory of Choice*, vol. 38 of *Studies in Mathematical and Managerial Economics*, Amsterdam: North-Holland / Elsevier.
- BALAKRISHNAN, N., E. A. OK, AND P. ORTOLEVA (2022): “Inference of Choice Correspondences,” *Working Paper*.
- BARBERÁ, S. AND P. K. PATTANAIK (1986): “Falmagne and the rationalizability of stochastic choices in terms of random orderings,” *Econometrica: Journal of the Econometric Society*, 707–715.
- BRADY, R. L. AND J. REHBECK (2016): “Menu-Dependent Stochastic Feasibility,” *Econometrica*, 84, 1203–1223.
- CATTANEO, M. D., X. MA, Y. MASATLIOGLU, AND E. SULEYMANOV (2020): “A random attention model,” *Journal of Political Economy*, 128, 2796–2836.
- ELLIS, K., E. FILIZ-OZBAY, AND E. OZBAY (2025): “Consideration and Choice,” *Working Paper*.
- GUL, F., P. NATENZON, AND W. PESENDORFER (2014): “Random Choice as Behavioral Optimization,” *Econometrica*, 82, 1873–1912.
- KAHNEMAN, D. (1992): “Reference Points, Anchors, Norms, and Mixed Feelings,” *Organizational Behavior and Human Decision Processes*, 51, 296–312.

- KAWAGUCHI, K., K. UETAKE, AND Y. WATANABE (2021): “Designing context-based marketing: Product recommendations under time pressure,” *Management Science*, 67, 5642–5659.
- KIBRIS, Ö., Y. MASATLIOGLU, AND E. SULEYMANOV (2024): “A Random Reference Model,” *American Economic Journal: Microeconomics*, 16, 155–209.
- KNETSCH, L. (1989): “The Endowment Effect and Evidence of Nonreversible Indifference Curves,” *American Economic Review*, 79, 1277–1284.
- KOVACH, M. AND E. SULEYMANOV (2023): “Reference dependence and random attention,” *Journal of Economic Behavior & Organization*, 215, 421–441.
- KOVACH, M. AND G. TSERENJIGMID (2022): “Behavioral foundations of nested stochastic choice and nested logit,” *Journal of Political Economy*, 130, 2411–2461.
- LIN, C.-H., W.-H. HUANG, AND M. ZEELLENBERG (2006): “Multiple reference points in investor regret,” *Journal of Economic Psychology*, 27, 781–792.
- LUCE, R. D. (1959): *Individual choice behavior*, Wiley, New York.
- MA, W. (2023): “When is assortment optimization optimal?” *Management Science*, 69, 2088–2105.
- MANZINI, P. AND M. MARIOTTI (2014): “Stochastic Choice and Consideration Sets,” *Econometrica*, 82, 1153–1176.
- MANZINI, P., M. MARIOTTI, AND L. ÜLKÜ (2024): “A model of approval with an application to list design,” *Journal of Economic Theory*, 217, 105821.
- MASATLIOGLU, Y., D. NAKAJIMA, AND E. Y. OZBAY (2012): “Revealed Attention,” *American Economic Review*, 102, 2183–2205.
- MASATLIOGLU, Y. AND E. A. OK (2005): “Rational Choice with Status Quo Bias,” *Journal of Economic Theory*, 121, 1–29.
- (2014): “A Canonical Model of Choice with Initial Endowments,” *Review of Economic Studies*, 81, 851–883.

- McFADDEN, D. (1978): “Modelling the choice of residential location,” in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlqvist and J. W., North Holland.
- RUBINSTEIN, A. (2012): *Lecture notes in microeconomic theory: the economic agent*, Princeton University Press.
- SALANT, Y. AND A. RUBINSTEIN (2008): “(A, f): Choice with Frames,” *Review of Economic Studies*, 75, 1287–1296.
- SAMUELSON, W. AND R. ZECKHAUSER (1988): “Status Quo Bias in Decision Making,” *Journal of Risk and Uncertainty*, 1, 7–59.
- STOYE, J. (2015): “Choice theory when agents can randomize,” *Journal of Economic Theory*, 155, 131–151.
- SULLIVAN, K. AND T. KIDA (1995): “The effect of multiple reference points and prior gains and losses on managers’ risky decision making,” *Organizational Behavior and Human Decision Processes*, 64, 76–83.
- TVERSKY, A. (1972): “Elimination by aspects,” *Psychological Review*, 79, 281–299.

## Appendix A Additional Models

### A.1 Logit, RCG, and IC with empty chosen collection

In the main body of the paper, we investigated Logit, RCG, and IC under the assumption that the chosen collection is non-empty. In this section, we relax this assumption and provide characterizations for the three models when the chosen collection is allowed to be empty. We add a superscript  $^o$  to the names of these models to differentiate them from the models studied in the main body of the paper. Table 3 provides the functional forms of the three models when the chosen collection can be empty.

Models	Functional forms when the chosen collection can be empty
Logit $^o$	$\mu_{\text{LG}}^o(T, S) = \frac{\pi(T)}{\sum_{T': T' \subseteq S} \pi(T')} \quad \text{for all } T \subseteq S$
RCG $^o$	$\mu_{\text{RCG}}^o(T, S) = \sum_C m(C) \mathbb{1}(T = S \cap C) \quad \text{for all } T \subseteq S$
IC $^o$	$\mu_{\text{IC}}^o(T, S) = \prod_{x \in T} \gamma(x) \prod_{y \in S \setminus T} (1 - \gamma(y)) \quad \text{for all } T \subseteq S$

Table 3: Logit, RCG, and IC with empty chosen collections

**Characterization of RCG $^o$ .** Under the new assumption that the chosen collection can be empty, RCG $^o$  is characterized by a single behavioral postulate, Additivity (Axiom 12). Additivity states that removing an alternative  $x$  from the choice set increases the choice probability of any collection  $T$  that excludes  $x$ , with the increase being precisely equal to the likelihood that  $T \cup x$  is the chosen collection in the original choice set.

**Axiom 12 (Additivity).** For all  $x, T, S$  such that  $x \in S$  and  $T \subseteq S \setminus x$

$$\mu(T, S \setminus x) - \mu(T, S) = \mu(T \cup x, S).$$

Note that Additivity implies Relative Additivity (Axiom 3) and is therefore more restrictive. To elaborate, Additivity implies  $\mu(T, S \setminus x) = \mu(T, S) + \mu(T \cup x, S)$  and  $\mu(T', S) + \mu(T' \cup x, S) = \mu(T', S \setminus x)$ . Multiplying these two equations yields Relative Additivity.

Proposition 9 establishes that Additivity characterizes the RCG $^o$  model.

**Proposition 9 (Characterization of RCG<sup>o</sup>).**  $\mu$  has an RCG<sup>o</sup> representation if and only if it satisfies Additivity.

**Characterization of IC<sup>o</sup>.** With regard to the IC<sup>o</sup> model, Proposition 10 establishes that a full-support  $\mu$  has an IC<sup>o</sup> representation if and only if it satisfies IIS<sup>o</sup> and Additivity. Here, IIS<sup>o</sup> is the variant of IIS (Axiom 1) that allows  $T$  to be empty.<sup>16</sup>

**Proposition 10 (Characterization of IC<sup>o</sup>).** A full-support  $\mu$  has an IC<sup>o</sup> representation if and only if it satisfies IIS<sup>o</sup> and Additivity.

As in the environment where the chosen collection must be non-empty, Proposition 10 implies that the IC<sup>o</sup> model remains equivalent to the intersection of the Logit<sup>o</sup> and RCG<sup>o</sup> models. This result is stated in Corollary 1.

**Corollary 1.** Suppose  $\mu$  has full support. The following are equivalent:

- (i)  $\mu$  has an IC<sup>o</sup> representation
- (ii)  $\mu$  has Logit<sup>o</sup> and RCG<sup>o</sup> representations.

## A.2 Attribute Rule

In a framework related to the original EBA model, Gul et al. (2014) introduce an Attribute Rule model to study choice situations where the decision maker selects a single option. In an Attribute Rule (AR), the DM first picks an attribute according to a logit formula and subsequently chooses an alternative possessing that attribute according to another logit formula. Let  $B(S)$  represent the set of all attributes that are possessed by at least one element in  $S$ . The probability of choosing an option  $x$  from  $S \ni x$  is denoted as  $p_{\text{AR}}(x, S)$  and is given by

$$p_{\text{AR}}(x, S) = \sum_{i \in B(S)} \frac{\theta_i}{\sum_{j \in B(S)} \theta_j} \frac{\eta_x^i}{\sum_{y \in S} \eta_y^i} \quad \text{for all } x \in S.$$

In the formulation of  $p_{\text{AR}}(x, S)$ , function  $\theta$  governs the choice of attributes and maps attributes to  $\mathbb{R}_{++}$ . Meanwhile, function  $\eta$  governs the choice of options and maps a pair of an attribute and an option possessing that attribute to the natural numbers, with  $\eta_x^i = 0$  if option  $x$  does not have attribute  $i$ .

---

<sup>16</sup>Formally, IIS<sup>o</sup> states that  $\mu(T, S)/\mu(T', S) = \mu(T, S')/\mu(T', S')$  for all  $T, T' \subseteq S \cap S'$  whenever the two ratios are well defined. IIS<sup>o</sup> fully characterizes the Logit<sup>o</sup> model when  $\mu$  has full support. Because the characterization of Logit<sup>o</sup> deviates only slightly from that of Logit, we omit it here for brevity.

Let  $B_i$  denote the collection of options possessing attribute  $i$ . As  $\eta_x^i = 0$  if  $x$  does not have attribute  $i$ , it is the case that  $\sum_{y \in S} \eta_y^i = \sum_{y \in (B_i \cap S)} \eta_y^i$ . Equivalently,

$$p_{\text{AR}}(x, S) = \sum_{i \in B(S)} \frac{\theta_i}{\sum_{j \in B(S)} \theta_j} \frac{\eta_x^i}{\sum_{y \in (B_i \cap S)} \eta_y^i} \quad \text{for all } x \in S.$$

This reformulation motivates an alternative interpretation of an Attribute Rule as follows. The decision maker first draws an attribute  $i$  with probability  $\frac{\theta_i}{\sum_{j \in B(S)} \theta_j}$ . The DM then pays attention only to feasible options possessing that attribute, which is the set  $T = B_i \cap S$ . Finally, the DM picks an option  $x$  from  $T$  with probability proportional to its value on attribute  $i$ . Based on this alternative interpretation of the AR model, we can reformulate it as a two-stage decision-making process. In the first stage the DM chooses a collection  $T$  from the choice set  $S$ ; in the second stage she selects an option from  $T$  as the final choice.

Mathematically, let  $I_T = \{i : i \in B(S) \text{ such that } T = B_i \cap S\}$  denote the set of attributes that induce collection  $T$  to be chosen in the first stage. The probability of choosing  $T$  from choice set  $S$  equals the total weight of all attributes  $i$  in  $I_T$ , divided by the total weight of all feasible attributes:

$$\mu_{\text{AR}}(T, S) = \sum_{i \in I_T} \frac{\theta_i}{\sum_{j \in B(S)} \theta_j} \quad \text{for all non-empty } T \subseteq S. \quad (5)$$

In the second stage, the DM selects  $x$  from  $T$  with probability  $\rho_S(x|T)$ , defined as

$$\rho_S(x|T) = \mathbb{E}_{i \in I_T} [\Pr(x|T, i)] = \sum_{i \in I_T} [\Pr(i|T) \cdot \Pr(x|T, i)].$$

Here,  $\mathbb{E}_{i \in I_T}$  denotes expectation over  $i \in I_T$ . Term  $\Pr(x|T, i)$  denotes the probability that an option  $x$  is chosen from  $T$  based on its value on attribute  $i$  and is given by  $\Pr(x|T, i) = \frac{\eta_x^i}{\sum_{y \in T} \eta_y^i}$ . Meanwhile,  $\Pr(i|T)$  represents the probability that attribute  $i$  induced collection  $T$  to be chosen in the first stage. This probability is given by  $\Pr(i|T) = \frac{\theta_i}{\sum_{j \in I_T} \theta_j}$ . Note that  $\rho_S(x|T)$  generally depends on  $S$  as the set  $I_T$  depends on  $S$ ; we therefore make the subscript  $S$  explicit in  $\rho_S(x|T)$ .

It is straightforward to verify that, for each non-empty  $S \subseteq X$  and each non-empty

$T \subseteq S$ ,  $\rho_S(\cdot|T)$  constitutes a probability distribution over  $T$ , and that

$$p_{\text{AR}}(x, S) = \sum_{T: T \subseteq S, T \neq \emptyset} \mu_{\text{AR}}(T, S) \cdot \rho_S(x|T) \quad \text{for all } x \in S, \text{ for all } S \subseteq X.$$

We say that  $\mu$  has an AR representation if it has the form given in equation (5), i.e., if it corresponds to the first stage of the choice procedure in an AR after reformulation. Note that the functional forms of  $\mu_{\text{EBA}}$  in equation (1) and  $\mu_{\text{AR}}$  in equation (5) are identical. To see this, define  $\omega_i = \theta_i$  and  $A_i = B_i$  for all aspects  $i$ . This equates the two numerators in  $\mu_{\text{EBA}}$  and  $\mu_{\text{AR}}$ . Regarding the two denominators, note that  $A_j \cap S \neq \emptyset$  if and only if there is at least one option in  $S$  possessing aspect  $j$ . This is equivalent to saying  $j \in B(S)$ . Hence, the formulations of  $\mu$  in static EBA and AR are identical. This result is stated in Remark 4.

**Remark 4.**  $\mu$  has a static EBA representation if and only if it has an AR representation.

Remark 4 is useful in explaining the differences in the choice probabilities of individual options between the sequential EBA and AR models, as noted by Gul et al. (2014). Remark 4 implies that the disparity comes from distinct choice procedures after the initial collection is selected. Specifically, in an AR, the DM picks an element from the chosen collection according to a logit formula. Meanwhile, in the sequential EBA model, the DM keeps eliminating options until only one remains.

### A.3 Nested Logit Model

In this section, we introduce the formulation of  $\mu$  in the nested logit model and present its connection to other models we investigated. The nested logit is a particular case of NSC model (Section 2.5.3). Given a choice set  $S$ , the probability of choosing a non-empty  $T \subseteq S$  is defined as

$$\mu_{\text{NL}}(T, S) = \frac{\sigma(N_i \cap S)}{\sum_{j=1}^q \sigma(N_j \cap S)} \mathbb{1}(T = N_i \cap S) \quad \text{for all non-empty } T \subseteq S,$$

where  $(N_1, \dots, N_q)$  is a partition of the grand set and  $\sigma$  is the weighting function. Nested logit differs from NSC in that there exist  $\eta_1, \dots, \eta_q \in \mathbb{R}_{++}$  and a function  $v: X \rightarrow \mathbb{R}_{++}$  such that  $\sigma(N_i \cap S) = \left( \sum_{x \in (N_i \cap S)} v(x) \right)^{\eta_i}$  for all nest  $N_i$  and non-empty  $S \subseteq X$ . The function  $v$  represents the DM's preference over alternatives. Meanwhile,

the scalar  $\eta_i$  captures the degree of substitutability or complementarity among elements in nest  $N_i$ . As the value of each collection,  $\sigma(T)$ , is increasing in the total utilities of its elements,  $\sum_{x \in T} v(x)$ , the probability of choosing a collection depends on the DM's underlying preference over alternatives. The stronger the preference for elements within the collection, the higher the probability that the collection is selected.

Corollary 2 describes the relationship between  $\mu$  in nested logit and the other models studied in our paper. It shows that nested logit is highly distinct from those models. The corollary follows from Propositions 7 and 8.

**Corollary 2.** The following are equivalent:

- (i)  $\mu$  has a singleton representation
- (ii)  $\mu$  has nested logit and RRM representations
- (iii)  $\mu$  has nested logit and endogenous EBA (or RCG) representations.

## Appendix B Omitted Proofs

**Proof of Proposition 1:** The necessity is straightforward. We show the sufficiency. Define  $\pi(T) = \mu(T, X)$  for all non-empty  $T \subseteq X$ . As  $\mu$  has full support,  $\pi(T) > 0$  for all non-empty  $T \subseteq X$ . By definition,  $\frac{\pi(T)}{\pi(T')} = \frac{\mu(T, X)}{\mu(T', X)}$  for all non-empty  $T, T' \subseteq X$ . It immediately follows that  $\frac{\pi(T)}{\sum_{T': \emptyset \neq T' \subseteq X} \pi(T')} = \frac{\mu(T, X)}{\sum_{T': \emptyset \neq T' \subseteq X} \mu(T', X)} = \mu(T, X)$  for all non-empty  $T \subseteq X$ , where the second equation uses the fact that  $\sum_{T': \emptyset \neq T' \subseteq X} \mu(T', X) = 1$ . Now, consider an arbitrary choice set  $S \subset X$  such that  $S \neq \emptyset$ . We have

$$\frac{\mu(T, S)}{\mu(T', S)} = \frac{\mu(T, X)}{\mu(T', X)} = \frac{\pi(T)}{\pi(T')} \quad \text{for all non-empty } T, T' \subseteq S.$$

The first equation holds because of the IIS axiom. The second equation holds because of the definition of  $\pi$ . It follows from the equation above that  $\mu(T, S) = \frac{\mu(T, S)}{\sum_{T': \emptyset \neq T' \subseteq S} \mu(T', S)} = \frac{\pi(T)}{\sum_{T': \emptyset \neq T' \subseteq S} \pi(T')}$ . Hence,  $\mu$  has a Logit representation. This completes our proof of Proposition 1. ■

**Proof of Proposition 2:** We first state and prove the following Lemma.

**Lemma 1.** Suppose  $\mu$  satisfies Positivity-1 and Relative Additivity. Then for all  $S \ni x$ ,  $|S| \geq 2$ , and all non-empty  $T \subseteq S \setminus x$

$$(i) \quad \mu(T, S \setminus x) = 0 \text{ implies } \mu(T, S) + \mu(T \cup x, S) = 0$$

$$(ii) \quad \mu(T, S \setminus x) > 0 \text{ implies } \mu(T, S) + \mu(T \cup x, S) > 0.$$

*Proof.* For part (i), suppose  $\mu(T^1, S \setminus x) = 0$  with  $T^1 \subseteq S \setminus x$  and  $x \in S$  and  $T^1 \neq \emptyset$ . This implies there must exist a non-empty  $T' \subseteq S \setminus x$  such that  $\mu(T', S \setminus x) > 0$ . Plugging such  $(T^1, T')$  into Relative Additivity (Axiom 3):

$$\mu(T^1, S \setminus x)[\mu(T', S) + \mu(T' \cup x, S)] = \mu(T', S \setminus x)[\mu(T^1, S) + \mu(T^1 \cup x, S)].$$

The LHS of the equality above is equal to 0 because  $\mu(T^1, S \setminus x) = 0$  by the selection of  $T^1$ . Hence, the RHS of the equality above is also 0. As  $\mu(T', S \setminus x) > 0$  by the selection of  $T'$ , it follows  $\mu(T^1, S) + \mu(T^1 \cup x, S) = 0$ .

For part (ii), suppose  $\mu(T^2, S \setminus x) > 0$  with  $T^2 \subseteq S \setminus x$  and  $x \in S$  and  $T^2 \neq \emptyset$ . We want to show  $\mu(T^2, S) + \mu(T^2 \cup x, S) > 0$ . Proof is by a contradiction. Suppose

$\mu(T^2, S) + \mu(T^2 \cup x, S) = 0$ . Take an arbitrary non-empty  $T'' \subseteq S \setminus x$ . Plugging such  $(T^2, T'')$  into Relative Additivity (Axiom 3), we have

$$\mu(T^2, S \setminus x)[\mu(T'', S) + \mu(T'' \cup x, S)] = \mu(T'', S \setminus x)[\mu(T^2, S) + \mu(T^2 \cup x, S)].$$

The RHS of the equality above is zero because  $\mu(T^2, S) + \mu(T^2 \cup x, S) = 0$  by assumption. Hence, the LHS of the equality above must be 0. As  $\mu(T^2, S \setminus x) > 0$ , this implies that  $\mu(T'', S) + \mu(T'' \cup x, S) = 0$ . As  $T''$  is chosen arbitrarily, this implies

$$0 = \sum_{T'': \emptyset \neq T'' \subseteq S \setminus x} [\mu(T'', S) + \mu(T'' \cup x, S)] = 1 - \mu(\{x\}, S).$$

The second equation comes from the fact that  $\sum_{T: \emptyset \neq T \subseteq S} \mu(T, S) = 1$  by the definition of  $\mu$ . Note that  $1 - \mu(\{x\}, S) = 0$  implies  $\mu(\{x\}, S) = 1$ . This contradicts with Positivity-1 (Axiom 2). To elaborate, consider an arbitrary  $y \in S \setminus x$ . Such  $y$  exists because  $S \setminus x$  is non-empty. Then there does not exist a non-empty  $T \subseteq S$  with  $y \in T$  such that  $\mu(T, S) > 0$  because  $\mu(\{x\}, S) = 1$  (contradiction). This completes our proof of Lemma 1.  $\square$

Back to the main proof of Proposition 2. Suppose  $\mu$  satisfies Positivity-1 and Relative Additivity. This implies that for all  $x \in S \subseteq X$  and  $T, T' \subseteq S \setminus x$  such that  $T, T' \neq \emptyset$

$$\mu(T, S \setminus x)[\mu(T', S) + \mu(T' \cup x, S)] = \mu(T', S \setminus x)[\mu(T, S) + \mu(T \cup x, S)].$$

For each category  $C \in \mathcal{X}$ , define  $m(C) = \mu(C, X)$ . By Positivity-1 (Axiom 2), for every  $x \in X$ , there exists  $T \ni x$  such that  $\mu(T, X) > 0$ . It follows that  $x$  belongs to at least one category  $C$  with  $m(C) > 0$ ; the category  $C$  can be chosen to be identical to  $T$ . Now, for all non-empty  $T \subseteq S$ , we show

$$\mu(T, S) = \sum_C \frac{m(C) \mathbb{1}(T = S \cap C)}{\sum_{C: S \cap C \neq \emptyset} m(C)} = \frac{1}{\sum_{C: S \cap C \neq \emptyset} m(C)} \sum_{C: C = T \cup A \text{ with } A \subseteq X \setminus S} m(C) \quad (6)$$

by induction based on the number of alternatives in  $S$  by “stepping down.”

Step 1: When  $S = X$ , by definition  $\mu(T, X) = m(T) = \sum_{C \in \mathcal{X}} m(C) \mathbb{1}(T = X \cap C)$ . Additionally, note that  $\sum_{C: X \cap C \neq \emptyset} m(C) = \sum_{C: C \subseteq X, C \neq \emptyset} m(C) = \sum_{C: C \subseteq X, C \neq \emptyset} \mu(C, X) = 1$ .

Hence,  $\mu$  has the representation at  $X$ .

Step 2: Suppose equation (6) holds for all  $(T, S)$  such that  $T \subseteq S$ ,  $T \neq \emptyset$ , and  $|S| = k + 1, k + 2, \dots, |X|$ . Consider a choice set  $S$  with  $x \in S$  and  $|S| = k + 1$ . We show equation (6) holds at  $S \setminus x$ . For notational simplicity, let  $\beta_S = \sum_{C: S \cap C \neq \emptyset} m(C)$ .

Case 2.1: Take non-empty  $T \subseteq S \setminus x$  such that  $\mu(T, S \setminus x) = 0$ . It follows from Lemma 1 that  $\mu(T, S) + \mu(T \cup x, S) = 0$ . Therefore, we can write  $\mu(T, S \setminus x) = 0 = \mu(T, S) + \mu(T \cup x, S)$ . This is equivalent to

$$\begin{aligned}
& \underbrace{\frac{1}{\beta_S} \sum_{C: C=T \cup A \text{ with } A \subseteq X \setminus S} m(C)}_{\mu(T, S)} + \underbrace{\frac{1}{\beta_S} \sum_{C: C=(T \cup x) \cup A \text{ with } A \subseteq X \setminus S} m(C)}_{\mu(T \cup x, S)} \\
&= \frac{1}{\beta_S} \left[ \sum_{C: C=T \cup A', A' \subseteq ((X \setminus S) \cup x) \text{ and } x \notin A'} m(C) + \sum_{C: C=T \cup A', A' \subseteq ((X \setminus S) \cup x) \text{ and } x \in A'} m(C) \right] \\
&= \frac{1}{\beta_S} \left[ \sum_{C: C=T \cup A', A' \subseteq X \setminus (S \setminus x) \text{ and } x \notin A'} m(C) + \sum_{C: C=T \cup A', A' \subseteq X \setminus (S \setminus x) \text{ and } x \in A'} m(C) \right] \\
&= \frac{1}{\beta_S} \left[ \sum_{C: C=T \cup A', A' \subseteq X \setminus (S \setminus x)} m(C) \right] = \frac{1}{\beta_{S \setminus x}} \left[ \sum_{C: C=T \cup A', A' \subseteq X \setminus (S \setminus x)} m(C) \right].
\end{aligned}$$

Note that we replace  $\beta_S$  by  $\beta_{S \setminus x}$  in the last equation. This does not impact the equality as the expression inside the big square bracket equals 0 because  $\mu(T, S) + \mu(T \cup x, S) = \mu(T, S \setminus x) = 0$  by the selection of  $T$ . Hence, equation (6) holds for all non-empty  $T \subseteq S \setminus x$  such that  $\mu(T, S \setminus x) = 0$ .

Case 2.2: Fix a non-empty  $T \subseteq S \setminus x$  such that  $\mu(T, S \setminus x) > 0$ . Lemma 1 implies that  $\mu(T, S) + \mu(T \cup x, S) > 0$ . Applying Relative Additivity with  $(T, T')$ , where non-empty  $T' \subseteq S \setminus x$  can be chosen arbitrarily, we have

$$\frac{\mu(T', S \setminus x)}{\mu(T, S \setminus x)} = \frac{\mu(T', S) + \mu(T' \cup x, S)}{\mu(T, S) + \mu(T \cup x, S)} = \frac{\sum_{C: C=T' \cup A \text{ with } A \subseteq X \setminus (S \setminus x)} m(C)}{\sum_{C: C=T \cup A \text{ with } A \subseteq X \setminus (S \setminus x)} m(C)},$$

where the second equation follows a logic similar to that of case 2.1. Take a summation over all non-empty  $T' \subseteq S \setminus x$

$$\sum_{T': T' \subseteq S \setminus x, T' \neq \emptyset} \frac{\mu(T', S \setminus x)}{\mu(T, S \setminus x)} = \sum_{T': T' \subseteq S \setminus x, T' \neq \emptyset} \frac{\sum_{C: C=T' \cup A \text{ with } A \subseteq X \setminus (S \setminus x)} m(C)}{\sum_{C: C=T \cup A \text{ with } A \subseteq X \setminus (S \setminus x)} m(C)}$$

$$\begin{aligned}
\Leftrightarrow \frac{1}{\mu(T, S \setminus x)} &= \frac{\beta_{S \setminus x}}{\sum_{C: C=T \cup A \text{ with } A \subseteq X \setminus (S \setminus x)} m(C)} \\
\Leftrightarrow \mu(T, S \setminus x) &= \frac{1}{\beta_{S \setminus x}} \sum_{C: C=T \cup A \text{ with } A \subseteq X \setminus (S \setminus x)} m(C),
\end{aligned}$$

where the second equation uses  $\sum_{T': T' \subseteq S \setminus x, T' \neq \emptyset} \mu(T', S \setminus x) = 1$  and the definition of  $\beta_{S \setminus x}$

$$\beta_{S \setminus x} = \sum_{C: (S \setminus x) \cap C \neq \emptyset} m(C) = \sum_{T': T' \subseteq S \setminus x, T' \neq \emptyset} \left[ \sum_{C: C=T' \cup A \text{ with } A \subseteq X \setminus (S \setminus x)} m(C) \right]$$

Hence, equation (6) holds for all non-empty  $T \subseteq S \setminus x$  such that  $\mu(T, S \setminus x) > 0$ . This completes our proof of Proposition 2. ■

**Proof of Proposition 3:** The necessity can be easily verified. We show the sufficiency. As  $\mu$  has full support and satisfies IIS, by Proposition 1, we can write  $\mu(T, S) = \frac{\pi(T)}{\sum_{T': \emptyset \neq T' \subseteq S} \pi(T')}$  for all non-empty  $T \subseteq S$ , where  $\pi(T) = \mu(T, X)$  for all non-empty  $T \subseteq X$ . Applying Relative Additivity (Axiom 3) to this functional form of  $\mu$  and note that  $\mu$  has full support, we have

$$\frac{\pi(T)}{\pi(T')} = \frac{\pi(T) + \pi(T \cup x)}{\pi(T') + \pi(T' \cup x)} \Rightarrow \frac{\pi(T \cup x)}{\pi(T)} = \frac{\pi(T' \cup x)}{\pi(T')} \quad (7)$$

for all non-empty  $T, T' \subseteq X$  such that  $x \notin T, T'$ . Define a function  $\gamma: X \rightarrow (0, 1)$  such that  $\gamma(x) = \frac{\pi(X)}{\pi(X) + \pi(X \setminus x)}$  for all  $x \in X$ . Equivalently,  $\frac{\gamma(x)}{1 - \gamma(x)} = \frac{\pi(X)}{\pi(X \setminus x)}$ . We show  $\frac{\pi(X)}{\pi(X \setminus T)} = \prod_{x \in T} \frac{\gamma(x)}{1 - \gamma(x)}$  for all non-empty  $T \subseteq X$  such that  $T \neq X$ . Consider some  $T = \{x_1, x_2, \dots, x_k\}$ , where  $x_i \neq x_j$  when  $i \neq j$  such that  $T \neq X$ . Let  $T_0 = \emptyset$  and  $T_i = \{x_1, \dots, x_i\}$  for  $i = 1, \dots, k$ . By definition  $T = T_k$ . Equation (7) implies that  $\frac{\pi(X \setminus T_i)}{\pi(X \setminus T_{i+1})} = \frac{\pi(X)}{\pi(X \setminus x_{i+1})} = \frac{\gamma(x_{i+1})}{1 - \gamma(x_{i+1})}$ , where the second equation comes from the definition of  $\gamma$  function. Note that  $\frac{\pi(X)}{\pi(X \setminus T)} = \prod_{i=0}^{k-1} \frac{\pi(X \setminus T_i)}{\pi(X \setminus T_{i+1})}$  as  $T_0 = \emptyset$  and  $T_k = T$ . Hence,  $\frac{\pi(X)}{\pi(X \setminus T)} = \prod_{i=0}^{k-1} \frac{\gamma(x_{i+1})}{1 - \gamma(x_{i+1})} = \prod_{x \in T} \frac{\gamma(x)}{1 - \gamma(x)}$  for all non-empty  $T \subseteq X$  such that  $T \neq X$ .

Next, we express  $\pi(X)$  in terms of  $\gamma$ . Note that,

$$\sum_{T: T \subseteq X, T \neq \emptyset} \pi(T) = \sum_{T: T \subseteq X, T \neq \emptyset} \mu(T, X) = 1.$$

Therefore,

$$\sum_{T:T \subseteq X, T \neq \emptyset} \frac{\pi(X \setminus T)}{\pi(X)} = \frac{1 - \pi(X)}{\pi(X)} \Rightarrow \sum_{T:T \subseteq X, T \neq \emptyset} \prod_{x \in T} \frac{1 - \gamma(x)}{\gamma(x)} = \frac{1 - \pi(X)}{\pi(X)}.$$

This yields  $\pi(X) = \frac{\prod_{x \in X} \gamma(x)}{1 - \prod_{x \in X} (1 - \gamma(x))}$  using the binomial-style expansion. Since  $\frac{\pi(X)}{\pi(X \setminus T)} = \prod_{x \in T} \frac{\gamma(x)}{1 - \gamma(x)}$ , it follows  $\pi(T) = \frac{\prod_{x \in T} \gamma(x) \prod_{y \in X \setminus T} (1 - \gamma(y))}{1 - \prod_{x \in X} (1 - \gamma(x))}$  for all non-empty  $T \subseteq X$ . Plugging this  $\pi(T)$  into the logit representation of  $\mu$  gives the IC representation. This completes our proof of Proposition 3. ■

**Proof of Remark 2:** First, suppose  $\mu$  has an RCG representation. We define *artificial aspects* based on  $\mu$ . At the grand set, enumerate all nonempty  $T$  with positive probabilities of being chosen as  $T_1, T_2, \dots, T_K$ , where  $T_i \neq T_j$  when  $i \neq j$ . Define  $K$  artificial aspects by calling them aspect 1, aspect 2, ..., aspect  $K$ . Let the collection of options possessing artificial aspect  $i$ ,  $A_i$ , be  $A_i = T_i$ . Let option  $x$  possess attribute  $i$  if and only if  $x \in A_i$ . By definition of the artificial aspects and the RCG representation of  $\mu$ ,  $\mu$  satisfies Positivity-2 (Axiom 4). Additionally,  $\mu$  satisfies Relative Additivity because it has an RCG representation. It follows from Proposition 4 that  $\mu$  has an EBA representation with the aspects we defined, which implies that it has an endogenous EBA representation.

Now, suppose  $\mu$  has an endogenous EBA representation. Let  $\{i_1, \dots, i_M\}$  be the set of all (artificial) attributes. Let the set of options possessing attribute  $i_t$  be  $A_{i_t}$ . Let each category  $C$  be a set  $A_{i_t}$ . We show that  $\mu$  has an RCG representation with these categories. First, we show that  $\mu$  satisfies Positivity-1. By the definition of endogenous EBA, for an arbitrary option  $x$ , there exists attribute  $i_j$  such that  $x$  possesses attribute  $i_j$ , i.e.,  $x \in A_{i_j}$ . Hence, for each option  $x$  the category  $C = A_{i_j}$  includes  $x$ . This observation, together with the fact that  $\mu$  satisfies Positivity-2 (because it has an endogenous EBA representation), implies that  $\mu$  satisfies Positivity-1. Second, by Proposition 4,  $\mu$  also satisfies Relative Additivity since it has an endogenous EBA representation. Therefore  $\mu$  satisfies both Positivity-1 and Relative Additivity, and it follows from Proposition 2 that  $\mu$  has an RCG representation. This completes our proof of Remark 2. ■

**Proof of Proposition 5:** First, we show the necessity of Axiom 7. The necessity of Axiom 8 is similar to we omit a formal proof.

Necessity of Axiom 7: Suppose  $\mu$  has an RRM representation with  $\{\mathcal{Q}(x)\}_x$  and  $\{s_x\}_x$  being the constraint sets and salience parameters. Take  $x \in S$  and  $T \subseteq S \setminus x$  such that  $T \neq \emptyset$ . Let

$$\begin{aligned} C_1 &= \{y : y \in S \setminus x \text{ such that } T = \mathcal{Q}(y) \cap (S \setminus x)\} \\ C_2 &= \{y : y \in S \text{ such that } T = \mathcal{Q}(y) \cap S\} \\ C_3 &= \{y : y \in S \text{ such that } T \cup x = \mathcal{Q}(y) \cap S\}. \end{aligned}$$

Clearly,  $C_2$  and  $C_3$  are disjoint. Using definitions of  $C_1, C_2, C_3$  and the RRM representation of  $\mu$ , we can write

$$\begin{aligned} \mu(T, S \setminus x) &= \left( \sum_{z \in S \setminus x} s_z \right)^{-1} \sum_y s_y \mathbb{1}(y \in C_1) \\ \mu(T, S) &= \left( \sum_{z \in S} s_z \right)^{-1} \sum_y s_y \mathbb{1}(y \in C_2) \\ \mu(T \cup x, S) &= \left( \sum_{z \in S} s_z \right)^{-1} \sum_y s_y \mathbb{1}(y \in C_3). \end{aligned}$$

To show that Axiom 7 is necessary, it is sufficient to show  $C_1 = C_2 \cup C_3$  when  $T \neq \mathcal{Q}(x) \cap (S \setminus x)$  and  $T$  is non-empty. We do that in three steps.

*Step 1:* Take an arbitrary  $y_1 \in C_1$ . It follows that  $y_1 \in S$ . If  $x \notin \mathcal{Q}(y_1)$  then  $T = \mathcal{Q}(y_1) \cap (S \setminus x) = \mathcal{Q}(y_1) \cap S$ . Consequently,  $y_1 \in C_2$ . If  $x \in \mathcal{Q}(y_1)$  then  $\mathcal{Q}(y_1) \cap S = T \cup x$ . It follows that  $y_1 \in C_3$ . Therefore, for each  $y_1 \in C_1$ , either  $y_1 \in C_2$  or  $y_1 \in C_3$ . It follows  $C_1 \subseteq C_2 \cup C_3$ .

*Step 2:* Take an arbitrary  $y_2 \in C_2$ . It follows that  $y_2 \in S$ . First, note that  $x \notin \mathcal{Q}(y_2)$ . This is because if  $x \in \mathcal{Q}(y_2)$  then  $x \in \mathcal{Q}(y_2) \cap S$  because  $x \in S$ . But then it follows that  $x \in T = \mathcal{Q}(y_2) \cap S$ , which contradicts with  $T \subseteq S \setminus x$ . Second,  $x \notin \mathcal{Q}(y_2)$  implies that  $T = \mathcal{Q}(y_2) \cap S = \mathcal{Q}(y_2) \cap (S \setminus x)$ . Additionally,  $x \notin \mathcal{Q}(y_2)$  implies  $y_2 \neq x$  because  $y_2 \in \mathcal{Q}(y_2)$  by assumption. Hence,  $y_2 \in C_1$ .

*Step 3:* Take an arbitrary  $y_3 \in C_3$ . It follows that  $y_3 \in S$ . By definition of  $C_3$ , the facts that  $y_3 \in C_3$  and  $T \subseteq S \setminus x$  imply that  $T = \mathcal{Q}(y_3) \cap (S \setminus x)$ . Additionally, it must be the case that  $y_3 \neq x$ . Suppose not and  $y_3 = x$ . Then it implies that  $T = \mathcal{Q}(x) \cap (S \setminus x)$ , which contradicts the initial assumption that  $T \neq \mathcal{Q}(x) \cap (S \setminus x)$ . With  $y_3 \neq x$ ,  $y_3 \in S$ , and  $T = \mathcal{Q}(y_3) \cap (S \setminus x)$ , it follows that  $y_3 \in C_1$ . Combining

the three steps, we have  $C_1 = C_2 \cup C_3$ .

*The sufficiency part:* Suppose  $\mu$  satisfies Axioms 5-8. We show  $\mu$  has an RRM representation. Define the nest  $\mathcal{Q}(x)$  as the revealed nest  $\mathcal{Q}^R(x)$ . For each  $x \in X$ , define the salience parameter  $s_x = \mu(\mathcal{Q}^R(x), X)$ . Note that  $s_x > 0$  (because of Axiom 6) and  $\sum_{x \in X} s_x = 1$  because  $\sum_{T: T \subseteq X, T \neq \emptyset} \mu(T, X) = 1$ . We first state and prove the following Lemma.

**Lemma 2.** Axiom 6 implies that for all  $x, T, S$  with  $x \in S$ ,  $T \subseteq S \setminus x$ , and  $T \neq \emptyset$ , if  $\mu(T, S \setminus x) > 0$  then  $\mu(T, S) + \mu(T \cup x, S) > 0$ .

*Proof.* Suppose  $\mu(T, S \setminus x) > 0$ . Then Axiom 6 implies that there exists  $y \in S \setminus x$  such that  $T = \mathcal{Q}^R(y) \cap (S \setminus x)$ . If  $x \in \mathcal{Q}^R(y)$  then  $T \cup x = \mathcal{Q}^R(y) \cap S$ . Therefore, Axiom 6 implies that  $\mu(T \cup x, S) > 0$ . Otherwise, if  $x \notin \mathcal{Q}^R(y)$  then  $T = \mathcal{Q}^R(y) \cap S$ . Axiom 6 then implies that  $\mu(T, S) > 0$ . In either case,  $\mu(T, S) + \mu(T \cup x, S) > 0$ . This completes our proof of Lemma 2.  $\square$

Back to the main proof of Proposition 5. We show the sufficiency by induction based on the number of alternatives in  $S$  by “stepping down.” Suppose  $S = X$ . First, consider a non-empty  $T \subseteq X$  such that  $\mu(T, X) = 0$ . Axiom 6 implies that  $T \neq \mathcal{Q}^R(x) \cap X$  for all  $x \in X$ . Hence, we can write,

$$\mu(T, X) = 0 = \sum_{y \in X} \frac{s_y}{\sum_{z \in X} s_z} \mathbb{1}(T = \mathcal{Q}^R(y) \cap X)$$

Second, consider a non-empty  $T \subseteq X$  such that  $\mu(T, X) > 0$ . Axiom 6 implies that there is at least one  $y \in X$  such that  $T = \mathcal{Q}^R(y) \cap X = \mathcal{Q}^R(y)$ . Axiom 5 claims that such  $y$  is unique. Hence,  $\mu(T, X) = \mu(\mathcal{Q}^R(y), X) = s_y = \sum_{z \in X} s_z \mathbb{1}(T = \mathcal{Q}^R(z) \cap X)$ . The first two equations come from the definition of  $y$  and parameters  $\{s_z\}_{z \in X}$ . The last equation comes from the uniqueness of  $y$ .

Now, suppose that  $\mu$  has an RRM representation at all choice sets  $S$  with  $|S| = k + 1, k + 2, \dots, |X|$  with  $k \geq 2$ . Take an arbitrary  $S$  such that  $|S| = k + 1$ . Let  $x$  be an arbitrary element in  $S$ . We show that  $\mu$  also has an RRM representation at  $S \setminus x$ . First, take a non-empty  $T \subseteq S \setminus x$  such that  $\mu(T, S \setminus x) = 0$ . It follows from Axiom 6 that  $\nexists y \in S \setminus x$  such that  $T = \mathcal{Q}^R(y) \cap (S \setminus x)$ . Hence, we can write

$$\mu(T, S \setminus x) = 0 = \frac{\sum_{y \in S \setminus x} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x))}{\sum_{z \in S \setminus x} s_z}.$$

Second, take a non-empty  $T \subseteq S \setminus x$  such that  $\mu(T, S \setminus x) > 0$ . Consider the following cases.

Case 1: Suppose  $\mathcal{Q}^R(x) \cap (S \setminus x) = \emptyset$ . In this case, Axiom 7 applies to all  $T, T' \subseteq S \setminus x$  such that  $T, T' \neq \emptyset$ . As  $\mu(T, S \setminus x) > 0$ , Lemma 2 implies that  $\mu(T, S) + \mu(T \cup x, S) > 0$ . Hence, for all non-empty  $T' \subseteq S \setminus x$

$$\begin{aligned}
\frac{\mu(T', S \setminus x)}{\mu(T, S \setminus x)} &= \frac{\mu(T', S) + \mu(T' \cup x, S)}{\mu(T, S) + \mu(T \cup x, S)} \\
&= \frac{\sum_{y \in S} s_y \mathbb{1}(T' = \mathcal{Q}^R(y) \cap S) + \sum_{y \in S} s_y \mathbb{1}(T' \cup x = \mathcal{Q}^R(y) \cap S)}{\sum_{y \in S} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap S) + \sum_{y \in S} s_y \mathbb{1}(T \cup x = \mathcal{Q}^R(y) \cap S)} \\
&= \dots (\text{following the same logic as in the necessary part in Axiom 7}) \\
&= \frac{\sum_{y \in S \setminus x} s_y \mathbb{1}(T' = \mathcal{Q}^R(y) \cap (S \setminus x))}{\sum_{y \in S \setminus x} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x))}.
\end{aligned}$$

Take the summation over all non-empty  $T' \subseteq S \setminus x$ , we have

$$\begin{aligned}
\frac{\sum_{T': \emptyset \neq T' \subseteq S \setminus x} \mu(T', S \setminus x)}{\mu(T, S \setminus x)} &= \frac{\sum_{T': \emptyset \neq T' \subseteq S \setminus x} \sum_{y \in S \setminus x} s_y \mathbb{1}(T' = \mathcal{Q}^R(y) \cap (S \setminus x))}{\sum_{y \in S \setminus x} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x))} \\
&\Leftrightarrow \frac{1}{\mu(T, S \setminus x)} = \frac{\sum_{y \in S \setminus x} s_y}{\sum_{y \in S \setminus x} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x))} \\
&\Leftrightarrow \mu(T, S \setminus x) = \frac{\sum_{y \in S \setminus x} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x))}{\sum_{y \in S \setminus x} s_y}.
\end{aligned}$$

In the second equation above, the numerator of the ratio on the LHS is equal to 1 by definition of the  $\mu$  function. The numerator of the ratio on the RHS is equal to  $\sum_{y \in S \setminus x} s_y$  because for each  $y \in S \setminus x$ , there exists a unique non-empty  $T' \subseteq S \setminus x$  such that  $T' = \mathcal{Q}^R(y) \cap (S \setminus x)$ . The non-emptiness of such  $T'$  comes from the fact that  $y \in \mathcal{Q}^R(y)$  (by definition of the  $\mathcal{Q}^R$  function) and  $y \in S \setminus x$  (by the selection of  $y$ ). The last equation implies that  $\mu(T, S \setminus x)$  has an RRM representation at  $S \setminus x$ .

Case 2: Suppose  $\mathcal{Q}^R(x) \cap (S \setminus x) = T^*$  and  $T^* \neq \emptyset$ . In this case, Axiom 7 is only applicable to all non-empty  $T, T' \subseteq S \setminus x$  such that  $T, T' \neq T^*$ .

Case 2.1: Suppose  $\mu(T^*, S \setminus x) = 1$ . This implies  $\mu(T, S) = 0$  for all non-empty  $T \subseteq S \setminus x$  and  $T \neq T^*$ . Axiom 6 implies that for all non-empty  $T \subseteq S \setminus x$  such that  $T \neq T^*$ ,  $\nexists y \in S \setminus x$  such that  $T = \mathcal{Q}^R(y) \cap (S \setminus x)$ . Additionally, for all  $z \in S \setminus x$ , it must be the case that  $\mathcal{Q}^R(z) \cap (S \setminus x) = T^*$ . Otherwise, there exists another set

$T^{**} = \mathcal{Q}^R(z) \cap (S \setminus x)$  that is non-empty (because  $z \in S \setminus x$  and  $z \in \mathcal{Q}^R(z)$ ) such that  $\mu(T^{**}, S) > 0$  and  $T^{**} \neq T^*$ , which is a contradiction. Hence, we can write

$$\begin{cases} \mu(T^*, S \setminus x) &= \frac{\sum_{y \in S \setminus x} s_y \mathbb{1}(T^* = \mathcal{Q}^R(y) \cap (S \setminus x))}{\sum_{z \in S \setminus x} s_z} \\ \mu(T, S \setminus x) &= 0 = \frac{\sum_{y \in S \setminus x} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x))}{\sum_{z \in S \setminus x} s_z} \end{cases} \text{ for all } T \subseteq S \setminus x \text{ such that } T \neq \emptyset, T^*.$$

The first equation comes from the fact that  $\mathcal{Q}^R(z) \cap (S \setminus x) = T^*$  for all  $z \in S \setminus x$ . The second equation results from the fact that  $\nexists y \in S \setminus x$  such that  $\mathcal{Q}^R(y) \cap (S \setminus x) = T$  for all  $T \subseteq S \setminus x$  such that  $T \neq \emptyset$  and  $T \neq T^*$ .

*Case 2.2:* Suppose  $\mu(T^*, S \setminus x) < 1$ . Take non-empty  $T \subseteq S \setminus x$  such that  $\mu(T, S \setminus x) > 0$  and  $T \neq T^*$ . Such  $T$  must exist because the choice probabilities sum to 1. As  $\mu(T, S \setminus x) > 0$ , it follows from Lemma 2 that  $\mu(T, S) + \mu(T \cup x, S) > 0$ . Hence, for all  $T' \subseteq S \setminus x$  such that  $T' \neq \emptyset$  and  $T' \neq T^*$

$$\begin{aligned} \frac{\mu(T', S \setminus x)}{\mu(T, S \setminus x)} &= \frac{\mu(T', S) + \mu(T' \cup x, S)}{\mu(T, S) + \mu(T \cup x, S)} \\ &= \frac{\sum_{y \in S} s_y \mathbb{1}(T' = \mathcal{Q}^R(y) \cap S) + \sum_{y \in S} s_y \mathbb{1}(T' \cup x = \mathcal{Q}^R(y) \cap S)}{\sum_{y \in S} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap S) + \sum_{y \in S} s_y \mathbb{1}(T \cup x = \mathcal{Q}^R(y) \cap S)} \\ &= \dots (\text{following the same logic as in the necessary part in Axiom 7}) \\ &= \frac{\sum_{y \in S \setminus x} s_y \mathbb{1}(T' = \mathcal{Q}^R(y) \cap (S \setminus x))}{\sum_{y \in S \setminus x} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x))}. \end{aligned}$$

Take the summation over all non-empty  $T' \subseteq S \setminus x$  such that  $T' \neq T^*$ , we have

$$\frac{\sum_{T': \emptyset, T^* \neq T' \subseteq S \setminus x} \mu(T', S \setminus x)}{\mu(T, S \setminus x)} = \frac{\sum_{T': \emptyset, T^* \neq T' \subseteq S \setminus x} \sum_{y \in S \setminus x} s_y \mathbb{1}(T' = \mathcal{Q}^R(y) \cap (S \setminus x))}{\sum_{y \in S \setminus x} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x))},$$

which can be simplified to

$$\frac{1 - \mu(T^*, S \setminus x)}{\mu(T, S \setminus x)} = \frac{\sum_{y \in S \setminus x} s_y - \sum_{z \in S \setminus x} s_z \mathbb{1}(T^* = \mathcal{Q}^R(z) \cap (S \setminus x))}{\sum_{y \in S \setminus x} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x))}. \quad (8)$$

Here, the numerator of the ratio on the LHS is equal to  $1 - \mu(T^*, S \setminus x)$  because we leave out  $T^*$  when taking the summation. The same intuition applies to the numerator of the ratio on the RHS. Now, applying Axiom 8 for  $(T^*, T)$  with  $\mu(T, S) > 0$  (by

selection of  $T$ ), we have

$$\frac{\mu(T^*, S \setminus x)}{\mu(T, S \setminus x)} = \frac{\mu(T^*, S) + \mu(T^* \cup x, S) - \frac{\mu(\mathcal{Q}^R(x), X)}{\sum_{y \in S} \mu(\mathcal{Q}^R(y), X)}}{\mu(T, S) + \mu(T \cup x, S)}.$$

By definition of parameters  $\{s_x\}_{x \in X}$ , the ratio  $\frac{\mu(\mathcal{Q}^R(x), X)}{\sum_{y \in S} \mu(\mathcal{Q}^R(y), X)}$  is indeed  $\frac{s_x}{\sum_{y \in S} s_y}$ . Using the fact that  $\mu$  has an RRM representation at choice set  $S$ , we have

$$\begin{aligned} & \frac{\mu(T^*, S) + \mu(T^* \cup x, S) - \frac{\mu(\mathcal{Q}^R(x), X)}{\sum_{y \in S} \mu(\mathcal{Q}^R(y), X)}}{\mu(T, S) + \mu(T \cup x, S)} \\ &= \frac{\sum_{y \in S} s_y \mathbb{1}(T^* = \mathcal{Q}^R(y) \cap S) + \sum_{y \in S} s_y \mathbb{1}(T^* \cup x = \mathcal{Q}^R(y) \cap S) - s_x}{\sum_{y \in S} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap S) + \sum_{y \in S} s_y \mathbb{1}(T \cup x = \mathcal{Q}^R(y) \cap S)} \\ &= \frac{\sum_{y \in S \setminus x} s_y \mathbb{1}(T^* = \mathcal{Q}^R(y) \cap (S \setminus x))}{\sum_{y \in S \setminus x} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x))}. \end{aligned}$$

Here, the second equation follows by the same reasoning as the necessity of Axiom 7, together with the fact that  $T^* = \mathcal{Q}^R(x) \cap (S \setminus x)$ . Therefore,

$$\frac{\mu(T^*, S \setminus x)}{\mu(T, S \setminus x)} = \frac{\sum_{y \in (S \setminus x)} s_y \mathbb{1}(T^* = \mathcal{Q}^R(y) \cap (S \setminus x))}{\sum_{y \in S \setminus x} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x))}. \quad (9)$$

Combine equations (8) and (9), we get

$$\frac{1}{\mu(T, S \setminus x)} = \frac{\sum_{y \in S \setminus x} s_y}{\sum_{y \in S \setminus x} s_y \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x))} \Leftrightarrow \mu(T, S \setminus x) = \sum_{y \in S \setminus x} \frac{s_y}{\sum_{z \in S \setminus x} s_z} \mathbb{1}(T = \mathcal{Q}^R(y) \cap (S \setminus x)).$$

Therefore,  $\mu$  has an RRM representation at choice set  $S \setminus x$  under all cases. By induction,  $\mu$  has an RRM representation. This completes our proof of Proposition 5.  $\blacksquare$

**Proof of Proposition 6:** The necessity is straightforward. We prove the sufficiency. Suppose the revealed nests are  $N_1^R, N_2^R, \dots, N_q^R$ . Define the nests as the revealed nests. If  $q = 1$  then  $N_1^R = X$ . It follows from Axiom 11 that  $\mu(S, S) = 1$  for all non-empty  $S \subseteq X$  and  $\mu$  has an NSC representation for any weighting function  $\sigma$ . Suppose  $q \geq 2$ . Fix  $x_1^* \in N_1^R$  and  $x_2^* \in N_2^R$ . Take an arbitrary non-empty  $T_i \subseteq N_i^R$  for all  $i = 1, 2, \dots, q$ . Define the  $\sigma$  function as follows. Assign  $\sigma(\{x_1^*\}) = 1$ . Let  $\sigma(T_i) = \frac{\mu(T_i, T_i \cup x_1^*)}{\mu(\{x_1^*\}, T_i \cup x_1^*)}$  for  $i = 2, 3, \dots, q$  and  $\sigma(T_1) = \frac{\mu(T_1, T_1 \cup x_2^*)}{\mu(\{x_2^*\}, T_1 \cup x_2^*)} \frac{\mu(\{x_2^*\}, \{x_1^*, x_2^*\})}{\mu(\{x_1^*\}, \{x_1^*, x_2^*\})}$ . The  $\sigma$  function is well-defined following Axiom 11.

Take arbitrary non-empty  $T, T' \subseteq S$  such that  $\mu(T, S) > 0$  and  $\mu(T', S) > 0$ . We want to show  $\frac{\mu(T, S)}{\mu(T', S)} = \frac{\sigma(T)}{\sigma(T')}$ . As  $\mu(T, S) > 0$  and  $\mu(T', S) > 0$ , Axiom 11 implies that  $T = S \cap N_i^R$  and  $T' = S \cap N_j^R$  for some values of  $i$  and  $j$ . In general,  $i$  and  $j$  are not necessarily distinct. We consider the following cases.

*Case 1:* Suppose  $i, j \neq 1$ . It follows  $T \cap N_1^R = T' \cap N_1^R = \emptyset$ . Let  $T'' = \{x_1^*\}$ ,  $S' = T \cup x_1^*$ , and  $S'' = T' \cup x_1^*$ . Axiom 11 implies that  $\mu(T, S')$ ,  $\mu(T'', S')$ ,  $\mu(T'', S'')$  and  $\mu(T', S'')$  are all positive. By applying Axiom 9, we have

$$\frac{\mu(T, S)}{\mu(T', S)} = \frac{\mu(T, S')}{\mu(T'', S')} \frac{\mu(T'', S'')}{\mu(T', S'')} = \frac{\mu(T, T \cup x_1^*)}{\mu(\{x_1^*\}, T \cup x_1^*)} \frac{\mu(\{x_1^*\}, T' \cup x_1^*)}{\mu(T', T' \cup x_1^*)} = \frac{\sigma(T)}{\sigma(T')},$$

where the second equation comes from the definitions of  $T'', S', S''$  and the third equation uses the definition of  $\sigma$  function.

*Case 2:* Suppose  $i = j = 1$ . It follows  $T, T' \subseteq N_1^R$ . Let  $T'' = \{x_2^*\}$ ,  $S' = T \cup x_2^*$ , and  $S'' = T' \cup x_2^*$ . Again, Axiom 11 implies that  $\mu(T, S')$ ,  $\mu(T'', S')$ ,  $\mu(T'', S'')$  and  $\mu(T', S'')$  are all positive. By applying Axiom 9, we have

$$\frac{\mu(T, S)}{\mu(T', S)} = \frac{\mu(T, S')}{\mu(T'', S')} \frac{\mu(T'', S'')}{\mu(T', S'')} = \frac{\mu(T, T \cup x_2^*)}{\mu(\{x_2^*\}, T \cup x_2^*)} \frac{\mu(\{x_2^*\}, T' \cup x_2^*)}{\mu(T', T' \cup x_2^*)} = \frac{\sigma(T) \frac{\mu(\{x_1^*\}, \{x_1^*, x_2^*\})}{\mu(\{x_2^*\}, \{x_1^*, x_2^*\})}}{\sigma(T') \frac{\mu(\{x_1^*\}, \{x_1^*, x_2^*\})}{\mu(\{x_2^*\}, \{x_1^*, x_2^*\})}} = \frac{\sigma(T)}{\sigma(T')}.$$

The first equation above uses Axiom 9. The second equation comes from definitions of  $T'', S',$  and  $S''$ . The third equation uses the definition of the  $\sigma$  function.

*Case 3:* Suppose  $i = 1 \neq j$ . Then  $T \subseteq N_1^R$  and  $T' \cap N_1^R = \emptyset$ . Applying Axiom 9, for given  $T$  and  $T'$ , we have

$$\begin{aligned} & \frac{\mu(T, T \cup x_2^*)}{\mu(\{x_2^*\}, T \cup x_2^*)} \frac{\mu(\{x_2^*\}, \{x_1^*, x_2^*\})}{\mu(\{x_1^*\}, \{x_1^*, x_2^*\})} = \frac{\mu(T, S)}{\mu(T', S)} \frac{\mu(T', T' \cup x_1^*)}{\mu(\{x_1^*\}, T' \cup x_1^*)} \\ \Leftrightarrow & \underbrace{\frac{\mu(T, T \cup x_2^*)}{\mu(\{x_2^*\}, T \cup x_2^*)} \frac{\mu(\{x_2^*\}, \{x_1^*, x_2^*\})}{\mu(\{x_1^*\}, \{x_1^*, x_2^*\})}}_{\sigma(T)} \underbrace{\frac{\mu(\{x_1^*\}, T' \cup x_1^*)}{\mu(T', T' \cup x_1^*)}}_{1/\sigma(T')} = \frac{\mu(T, S)}{\mu(T', S)} \\ & \Leftrightarrow \frac{\sigma(T)}{\sigma(T')} = \frac{\mu(T, S)}{\mu(T', S)}, \end{aligned}$$

where the second equation comes from the definition of the  $\sigma$  function.

Now, as  $\frac{\mu(T, S)}{\mu(T', S)} = \frac{\sigma(T)}{\sigma(T')}$  for all  $T, T', S$  such that  $\mu(T, S) > 0$  and  $\mu(T', S) > 0$ , we have  $\mu(T, S) = \frac{\mu(T, S)}{\sum_{T': \mu(T', S) > 0} \mu(T', S)} = \frac{\sigma(T)}{\sum_{T': \mu(T', S) > 0} \sigma(T')} = \sum_i \frac{\sigma(T)}{\sum_{j=1}^q \sigma(N_j \cap S)} \mathbb{1}(T = N_i \cap S)$ . Here, the last equation uses the fact that we define the nests as the revealed nests,

and the revealed nests are pairwise disjoint. This completes our proof. ■

**Proof of Proposition 7:** We sequentially show that part (i) is equivalent to part (ii) and part (iii).

*Equivalence of Part (i) and Part (ii):* We show that  $\mu$  has RRM and endogenous EBA representations if and only if it has an RRM representation with  $\mathcal{Q}(x) = x$ . The if part is straightforward. If  $\mu$  has an RRM representation with  $\mathcal{Q}(x) = x$  then  $\mu$  is given by

$$\begin{cases} \mu(\{x\}, S) = \frac{s_x}{\sum_{y \in S} s_y} & \text{for all } x \in S \subseteq X \\ \mu(T, S) = 0 & \text{for all } T \subseteq S \subseteq X \text{ and } |T| \geq 2. \end{cases}$$

This  $\mu$  has a singleton representation. It is routine to check that this  $\mu$  satisfies Relative Additivity and Positivity-1. Hence, it follows from Proposition 2 and Remark 2 that  $\mu$  has an endogenous EBA representation. For the only if part, the proof is by a contradiction. Suppose  $\mu$  has RRM and endogenous EBA representations. Suppose there exists  $x^*$  such that  $\mathcal{Q}^R(x^*) \neq \{x^*\}$ . Take an arbitrary  $S$  such that  $x^* \in S$  and  $(S \setminus x^*) \cap \mathcal{Q}^R(x^*) \neq \emptyset$ . Note that  $\mu$  having an endogenous EBA representation implies that it satisfies Relative Additivity (Axiom 3). That is, for all non-empty  $T, T' \subseteq S \setminus x^*$  and  $x^* \in S$ , we have

$$\mu(T, S \setminus x^*)[\mu(T', S) + \mu(T' \cup x^*, S)] = \mu(T', S \setminus x^*)[\mu(T, S) + \mu(T \cup x^*, S)]. \quad (10)$$

Meanwhile,  $\mu$  having an RRM representation implies that it satisfies Axiom 8, which says that when  $T = \mathcal{Q}^R(x^*) \cap (S \setminus x^*)$  and  $T \neq \emptyset$  then for all non-empty  $T' \subseteq S \setminus x^*$  such that  $T' \neq T$

$$\mu(T, S \setminus x^*)[\mu(T', S) + \mu(T' \cup x^*, S)] = \mu(T', S \setminus x^*) \left[ \mu(T, S) + \mu(T \cup x^*, S) - \frac{\mu(\mathcal{Q}^R(x^*), X)}{\sum_{y \in S} \mu(\mathcal{Q}^R(y), X)} \right]. \quad (11)$$

Because of our assumption that  $\mathcal{Q}^R(x^*) \neq \{x^*\}$  and by selection of  $S$ , such  $T$  in equation (11) always exists. Note that  $\frac{\mu(\mathcal{Q}^R(x^*), X)}{\sum_{y \in S} \mu(\mathcal{Q}^R(y), X)} > 0$  because  $\mu$  has an RRM representation (see Axiom 6). Hence, the two equations (10) and (11) can occur simultaneously if and only if  $\mu(T', S \setminus x^*) = 0$  for all non-empty  $T' \subseteq S \setminus x^*$  such that  $T' \neq T$ . This implies that  $\mu(T, S \setminus x^*) = \mu(\mathcal{Q}^R(x^*) \cap (S \setminus x^*), S \setminus x^*) = 1$ . Consequently, for all  $y \in S \setminus x^*$  we have  $\mathcal{Q}^R(y) \cap (S \setminus x^*) = \mathcal{Q}^R(x^*) \cap (S \setminus x^*)$ . Otherwise, the decision

maker chooses  $\mathcal{Q}^R(y) \cap (S \setminus x^*)$  with a strictly positive probability under choice set  $S \setminus x^*$ , contradicting the fact that  $\mu(\mathcal{Q}^R(x^*) \cap (S \setminus x^*), S \setminus x^*) = 1$ .

Now take  $S = X$ . Using the argument above, for all  $y \in X \setminus x^*$ ,

$$\mathcal{Q}^R(y) \cap (X \setminus x^*) = \begin{cases} \mathcal{Q}^R(y) \setminus x^* & \text{if } x^* \in \mathcal{Q}^R(y) \\ \mathcal{Q}^R(y) & \text{if } x^* \notin \mathcal{Q}^R(y). \end{cases}$$

Note that  $\mathcal{Q}^R(x^*) \cap (X \setminus x^*) = \mathcal{Q}^R(x^*) \setminus x^*$ . Therefore, using  $\mathcal{Q}^R(y) \cap (X \setminus x^*) = \mathcal{Q}^R(x^*) \cap (X \setminus x^*)$  for all  $y \in X \setminus x^*$  (note that  $S = X$  here) we have

$$\mathcal{Q}^R(x^*) \setminus x^* = \begin{cases} \mathcal{Q}^R(y) \setminus x^* & \text{if } x^* \in \mathcal{Q}^R(y) \\ \mathcal{Q}^R(y) & \text{if } x^* \notin \mathcal{Q}^R(y). \end{cases} \quad (12)$$

When  $|X| \geq 3$ , there are at least two alternatives  $y, z \in X \setminus x^*$  such that  $y \neq z$ . If  $x^* \in \mathcal{Q}^R(y)$  then (12) implies that  $\mathcal{Q}^R(x^*) \setminus x^* = \mathcal{Q}^R(y) \setminus x^*$ , which then implies  $\mathcal{Q}^R(y) = \mathcal{Q}^R(x^*)$ . This contradicts the assumption that constraint sets associated with two different options are distinct. Now, suppose  $x^* \notin \mathcal{Q}^R(y)$  and  $x^* \notin \mathcal{Q}^R(z)$ . Then it follows from (12) that  $\mathcal{Q}^R(x^*) \setminus x^* = \mathcal{Q}^R(y)$  and  $\mathcal{Q}^R(x^*) \setminus x^* = \mathcal{Q}^R(z)$ . Consequently,  $\mathcal{Q}^R(y) = \mathcal{Q}^R(z)$ , which is also a contradiction. Hence, the initial assumption is wrong and  $\mathcal{Q}^R(x) = x$  for all  $x \in X$ .

Now, the equivalence of parts (i) and (ii) immediately follows from the fact that  $\mu$  has a singleton representation if and only if it has an RRM representation with  $\mathcal{Q}(x) = x$ .

Equivalence of Part (i) and Part (iii): First, suppose  $\mu$  has a singleton representation. To obtain an NSC representation of  $\mu$ , let each nest contain only one alternative, so there are  $|X|$  nests in total. To obtain an RRM representation of  $\mu$ , let the constraint associated with option  $x$  be  $\mathcal{Q}(x) = x$ . Define the weighting function  $\sigma$  in NSC and salience parameter  $s_x$  in RRM as  $\sigma(\{x\}) = s_x = \mu(\{x\}, X)$  for all  $x \in X$ . It is routine to check that with the collection of nests  $\{\{x\}, \{y\}, \{z\}, \dots\}$  and weighting function  $\sigma$ ,  $\mu$  has an NSC representation. Additionally, with the constraint sets  $\{\mathcal{Q}(x)\}_x$  and salience parameters  $\{s_x\}_x$  defined as above,  $\mu$  has an RRM representation.

Now, suppose  $\mu$  has RRM and NSC representations. Consider choice probabilities at the grand set. Note that Axiom 6 in RRM states that  $\mu(T, X) > 0$  if and only

if  $T = \mathcal{Q}^R(x)$  for some value of  $x$ , where  $\mathcal{Q}^R()$  is the revealed constraint function. Meanwhile, Axiom 11 in NSC states that  $\mu(T, X) > 0$  if and only if  $T = N_i^R$  for some value of  $i$ , where  $N_i^R$  is a *revealed nest*. It follows from these two axioms that each revealed constraint set  $\mathcal{Q}^R(x)$  identically maps to one revealed nest  $N_i^R$ . Note that for each  $x \in X$  there is one  $\mathcal{Q}^R(x)$ . Also, the constraint sets associated with two different options are assumed to be distinct. Furthermore, the revealed nests are pairwise disjoint and must form a partition of the grand set (Axiom 10 in NSC). This implies that there are  $|X|$  distinct revealed nests  $N_1^R, N_2^R, \dots, N_{|X|}^R$  with  $N_i^R = \{x\}$  for some  $x \in X$ . Using this result, it follows from Axiom 11 (or Axiom 6) that at any choice set  $S$ ,  $\mu(\{x\}, S) > 0$  for all  $x \in S$  and  $\mu(T, S) = 0$  when  $|T| \geq 2$ . Additionally, it follows from the Path-Independence of Irrelevant Sets (Axiom 9 in NSC) that  $\frac{\mu(\{x\}, S)}{\mu(\{y\}, S)} = \frac{\mu(\{x\}, S')}{\mu(\{y\}, S')}$  for all  $S, S' \supseteq \{x, y\}$ . Hence,  $\mu$  satisfies both conditions in Definition 3, which implies that it has a singleton representation. This completes our proof of Proposition 7. ■

**Proof of Proposition 8:** We sequentially show that part (i) is equivalent to part (ii) and part (iii).

*Equivalence of Part (i) and Part (ii):* First, suppose  $\mu$  has a nest-invariant representation. Clearly  $\mu$  has an NSC representation. To show that  $\mu$  has an endogenous EBA representation, it is sufficient to show that  $\mu$  satisfies Positivity-1 (Axiom 2) and Relative Additivity (Axiom 3). It is straightforward to see that  $\mu$  satisfies Positivity-1. For Relative Additivity, we want to show that for all  $x, T, T', S$  such that  $x \in S$  and  $T, T' \subseteq S \setminus x$  and  $T, T' \neq \emptyset$

$$\mu(T, S \setminus x)[\mu(T', S) + \mu(T' \cup x, S)] = \mu(T', S \setminus x)[\mu(T, S) + \mu(T \cup x, S)]. \quad (13)$$

If  $\mu(T, S \setminus x) = 0$ , then the NSC representation of  $\mu$  implies that there does not exist nest  $N_i$  such that  $T = N_i \cap (S \setminus x)$ . This observation, together with the fact that  $x \in S \setminus T$ , implies that there does not exist nest  $N_i$  such that  $T = N_i \cap S$  or  $T \cup x = N_i \cap S$ . It follows  $\mu(T, S) = \mu(T \cup x, S) = 0$  so  $0 = \mu(T, S \setminus x) = \mu(T, S) + \mu(T \cup x, S)$ . In this case, (13) holds for any non-empty  $T' \subseteq S \setminus x$ .

Now, suppose  $\mu(T, S \setminus x) > 0$  and  $\mu(T', S \setminus x) > 0$ . The NSC representation of  $\mu$  implies that there exist nests  $N_i$  and  $N_j$  (not necessarily distinct) such that  $T = N_i \cap (S \setminus x)$  and  $T' = N_j \cap (S \setminus x)$ . Consider the following cases.

*Case 1:* Suppose  $x \in N_i$  and  $x \in N_j$ . It follows that  $N_i$  and  $N_j$  are identical. In this case,  $T$  and  $T'$  are identical and (13) holds trivially.

*Case 2:* Suppose  $x \in N_i$  and  $x \notin N_j$  (the case  $x \notin N_i$  and  $x \in N_j$  is similar). It follows that  $T \cup x = N_i \cap S$  and  $T' = N_j \cap S$ . Consequently,  $\mu(T, S) = \mu(T' \cup x, S) = 0$ ,  $\mu(T \cup x, S) > 0$  and  $\mu(T', S) > 0$ . Equation (13) is then equivalent to

$$\frac{\mu(T, S \setminus x)}{\mu(T', S \setminus x)} = \frac{\mu(T \cup x, S)}{\mu(T', S)} \Leftrightarrow \frac{\sigma(T)}{\sigma(T')} = \frac{\sigma(T \cup x)}{\sigma(T')},$$

which is true because  $\sigma(T) = \sigma(T \cup x)$  as both  $T$  and  $T \cup x$  belong to the nest  $N_i$ .

*Case 3:* Suppose  $x \notin N_i$  and  $x \notin N_j$ . It follows that  $T = N_i \cap S$  and  $T' = N_j \cap S$ . Consequently,  $\mu(T \cup x, S) = \mu(T' \cup x, S) = 0$ ,  $\mu(T, S) > 0$  and  $\mu(T', S) > 0$ . Equation (13) is then equivalent to

$$\frac{\mu(T, S \setminus x)}{\mu(T', S \setminus x)} = \frac{\mu(T, S)}{\mu(T', S)} \Leftrightarrow \frac{\sigma(T)}{\sigma(T')} = \frac{\sigma(T)}{\sigma(T')},$$

which is true.

For the only if part, suppose  $\mu$  has NSC and endogenous EBA representations. Suppose the nests in NSC representation are  $N_1, N_2, \dots, N_q$ . If  $q = 1$  then there is only one nest being identical to the grand set:  $N_1 = X$ . It follows that  $\mu(S, S) = 1$  for all non-empty  $S \subseteq X$  regardless of the weighting function  $\sigma$ . As the weighting function  $\sigma$  can be chosen arbitrarily, we can select  $\sigma$  such that  $\sigma(T) = \sigma(X)$  for all non-empty  $T \subseteq X$ . It follows that  $\mu$  has a nest-invariant representation.

Now, suppose  $q \geq 2$  so there are at least two distinct nests. We show that  $\sigma(T) = \sigma(N_i)$  for all non-empty  $T \subseteq N_i$ . Note that this trivially holds when nest  $N_i$  consists of a single option. Suppose there are at least two elements in  $N_i$ . Take an arbitrary non-empty subset  $T$  of  $N_i$  such that  $T \neq N_i$ . As  $T$  is a strict subset of  $N_i$ , there exists  $x \in N_i$  such that  $T \subseteq N_i \setminus x$ . Take an arbitrary non-empty  $T' \subseteq N_j$  for  $j \neq i$ . We can always choose such  $T'$  as there are at least two nests. By selection,  $T$  and  $T'$  are in different nests. Consider choice set  $S = T \cup \{x\} \cup T'$ . Note that as  $\mu$  has an NSC representation, it follows from Axiom 11 that all four terms  $\mu(T, S \setminus x)$ ,  $\mu(T', S \setminus x)$ ,  $\mu(T \cup x, S)$ , and  $\mu(T', S)$  are strictly positive and  $\mu(T' \cup x, S) = \mu(T, S) = 0$ . As  $\mu$  has an endogenous EBA representation, it satisfies

Relative Additivity (Axiom 3). This implies

$$\begin{aligned}
\mu(T, S \setminus x)[\mu(T', S) + \mu(T' \cup x, S)] &= \mu(T', S \setminus x)[\mu(T, S) + \mu(T \cup x, S)] \\
\Rightarrow \mu(T, S \setminus x)\mu(T', S) &= \mu(T', S \setminus x)\mu(T \cup x, S) \\
\Leftrightarrow \frac{\mu(T, S \setminus x)}{\mu(T', S \setminus x)} &= \frac{\mu(T \cup x, S)}{\mu(T', S)} \\
\Leftrightarrow \sigma(T)/\sigma(T') &= \sigma(T \cup x)/\sigma(T'),
\end{aligned}$$

where the second equation uses  $\mu(T' \cup x, S) = \mu(T, S) = 0$ , the third equation uses  $\mu(T', S \setminus x) > 0$ ,  $\mu(T', S) > 0$ , and the last equation uses the NSC representation of  $\mu$ . Note that the last equation implies that  $\sigma(T) = \sigma(T \cup x)$ . As we choose  $T$  and  $x$  arbitrarily, by applying this process iteratively, we have  $\sigma(T) = \sigma(N_i)$  for all non-empty  $T \subseteq N_i$ .

Equivalence of Part (i) and Part (iii): First, suppose  $\mu$  has a nest-invariant representation. This implies that  $\sigma(T) = \sigma(T')$  if  $T$  and  $T'$  belong to the same nest and  $T, T' \neq \emptyset$ . Take arbitrary  $x, T, S$  such that  $x \in S \setminus T$  and  $T \subseteq S \setminus x$  and  $T \neq \emptyset$ . We want to show that  $\mu$  satisfies the probabilistic attention filter condition; that is,  $\mu(T, S) = \mu(T, S \setminus x)$  when  $\mu(T, S) > 0$ ,  $\mu(T, S \setminus x) > 0$ , and  $\mu(\{x\}, S) = 0$ . Note that  $\mu(T, S) > 0$  implies that there exists a nest  $N_i$  with  $T = N_i \cap S$ . Obviously,  $x \notin N_i$  because otherwise  $x \in N_i \cap S$  and this implies  $x \in T$  (contradiction). It follows that  $N_i \cap (S \setminus x) = T$ . Using the NSC representation of  $\mu$ , we have

$$\mu(T, S) = \frac{\sigma(N_i \cap S)}{\sum_j \sigma(N_j \cap S)} \quad \text{and} \quad \mu(T, S \setminus x) = \frac{\sigma(N_i \cap (S \setminus x))}{\sum_j \sigma(N_j \cap (S \setminus x))}.$$

As  $N_i \cap S = T = N_i \cap (S \setminus x)$ , we have  $\sigma(N_i \cap S) = \sigma(N_i \cap (S \setminus x))$ . Hence, to show that  $\mu(T, S) = \mu(T, S \setminus x)$ , it is sufficient to prove  $\sum_j \sigma(N_j \cap S) = \sum_j \sigma(N_j \cap (S \setminus x))$ . We show  $\sigma(N_j \cap S) = \sigma(N_j \cap (S \setminus x))$  for each nest  $N_j$ . Consider the following cases.

*Case 1:* Suppose  $N_j \cap S = N_j \cap (S \setminus x) = \emptyset$ . Then  $\sigma(N_j \cap S) = \sigma(N_j \cap (S \setminus x))$  trivially holds.

*Case 2:* Suppose  $N_j \cap S \neq \emptyset$  and  $N_j \cap (S \setminus x) \neq \emptyset$ . Then both  $N_j \cap S$  and  $N_j \cap (S \setminus x)$  are non-empty subsets of  $N_j$ . Using the fact that  $\sigma$  is invariant across non-empty subsets of the same nest, we have  $\sigma(N_j \cap S) = \sigma(N_j \cap (S \setminus x))$ .

*Case 3:* Suppose  $N_j \cap S \neq \emptyset$  but  $N_j \cap (S \setminus x) = \emptyset$ . Note that  $N_j \cap S \neq \emptyset$

but  $N_j \cap (S \setminus x) = \emptyset$  implies  $N_j = E \cup x$  for some  $E \subseteq X \setminus S$ . Hence, it follows that  $N_j \cap S = \{x\}$  and  $\mu(\{x\}, S) > 0$ . This contradicts the initial assumption that  $\mu(\{x\}, S) = 0$ . Therefore, the case in which  $N_j \cap S \neq \emptyset$  but  $N_j \cap (S \setminus x) = \emptyset$  cannot happen.

*Case 4:* Suppose  $N_j \cap S = \emptyset$  but  $N_j \cap (S \setminus x) \neq \emptyset$ . This case cannot happen because  $N_j \cap (S \setminus x)$  is a subset of  $N_j \cap S$ .

Now, suppose  $\mu$  has an NSC representation and satisfies the probabilistic attention filter condition. We want to show  $\sigma(T) = \sigma(T')$  whenever  $T, T' \neq \emptyset$  and  $T, T'$  belong to the same nest. If there is just one nest  $N_1 = X$  then the NSC representation of  $\mu$  implies that  $\mu(T, S) = 0$  when  $T \neq S$  and  $\mu(S, S) = 1$  for all non-empty  $S \subseteq X$ . In this case, the probabilistic attention filter condition trivially holds. Now, suppose that there exist at least two distinct nests,  $N_i$  and  $N_j$ . We show that  $\sigma(T)$  is the same for all non-empty  $T \subseteq N_i$ . This trivially holds if  $N_i$  contains a single option. Suppose  $N_i$  has at least two elements. Consider  $x \in N_i$  and non-empty  $T, T'$  such that  $T \subseteq N_i \setminus x$  and  $T' \subseteq N_j$ . Let  $S = T \cup \{x\} \cup T'$ . We have  $\mu(T', S) > 0$  and  $\mu(T', S \setminus x) > 0$  because  $T' = N_j \cap S = N_j \cap (S \setminus x)$ . Additionally,  $\mu(\{x\}, S) = 0$  because  $x \in N_i$  but  $N_i \cap S = T \cup x \neq \{x\}$ . Then the probabilistic attention filter condition is applicable and implies that  $\mu(T', S) = \mu(T', S \setminus x)$ . Using the NSC representation of  $\mu$  and the definition of  $(T, T', S)$ ,

$$\mu(T', S) = \mu(T', S \setminus x) \quad \text{implies} \quad \frac{\sigma(T')}{\sigma(T') + \sigma(N_i \cap S)} = \frac{\sigma(T')}{\sigma(T') + \sigma(N_i \cap (S \setminus x))}.$$

Note that  $\sigma(T') > 0$  as  $T' \neq \emptyset$ . Consequently, the equation above implies  $\sigma(N_i \cap S) = \sigma(N_i \cap (S \setminus x))$ . Equivalently,  $\sigma(T \cup x) = \sigma(T)$ . As  $x$  and  $T$  are chosen arbitrarily, applying a similar logic iteratively implies that  $\sigma(T)$  must be the same for all non-empty  $T \subseteq N_i$ . This completes our proof of Proposition 8. ■

**Proof of Proposition 9:** The necessity part is straightforward. We prove the sufficiency. For all category  $C \subseteq X$ , define  $m(C) = \mu(C, X)$ . For all  $T \subseteq S$ , we show

$$\mu(T, S) = \sum_{C: C \subseteq X} m(C) \mathbb{1}(T = S \cap C) = \sum_{C: C = T \cup A \text{ with } A \subseteq X \setminus S} m(C) \quad (14)$$

by induction based on the number of alternatives in  $S$  by “stepping down”.

Step 1: When  $S = X$ , by definition  $\mu(T, X) = m(T) = \sum_{C \subseteq X} m(C) \mathbb{1}(T = X \cap C)$  because  $X \cap C = C$  for all categories  $C$ .

Step 2: Suppose equation (14) holds for all  $(T, S)$  with  $T \subseteq S$  and  $|S| = k + 1, k + 2, \dots, |X|$ , where  $k \geq 2$ . Consider some  $S$  with  $|S| = k + 1$ . Take  $x \in S$ . By Axiom 12, for all  $T \subseteq S \setminus x$ , we have

$$\begin{aligned}
\mu(T, S \setminus x) &= \mu(T, S) + \mu(T \cup x, S) \\
&= \underbrace{\sum_{C: C=T \cup A \text{ with } A \subseteq X \setminus S} m(C)}_{\mu(T, S)} + \underbrace{\sum_{C: C=(T \cup x) \cup A, A \subseteq X \setminus S} m(C)}_{\mu(T \cup x, S)} \\
&= \sum_{C: C=T \cup A', A' \subseteq ((X \setminus S) \cup x), x \notin A'} m(C) + \sum_{C: C=T \cup A', A' \subseteq ((X \setminus S) \cup x), x \in A'} m(C) \\
&= \sum_{C: C=T \cup A', A' \subseteq X \setminus (S \setminus x), x \notin A'} m(C) + \sum_{C: C=T \cup A', A' \subseteq X \setminus (S \setminus x), x \in A'} m(C) \\
&= \sum_{C: C=T \cup A', A' \subseteq X \setminus (S \setminus x)} m(C).
\end{aligned}$$

Hence, equation (14) holds at choice set  $S \setminus x$ . By induction,  $\mu$  has a representation at any non-empty choice set. This completes our proof of Proposition 9. ■

**Proof of Proposition 10:** The necessity can be easily verified. We show the sufficiency. As  $\mu$  has full support and satisfies IIS, by a similar logic as in the proof of Proposition 1, we can write  $\mu(T, S) = \frac{\pi(T)}{\sum_{T' \subseteq S} \pi(T')}$  for all  $T \subseteq S \subseteq X$ , where  $\pi(T) = \mu(T, X) > 0$  because  $\mu$  has full support. Apply Additivity (Axiom 12) to this functional form of  $\mu$ , for all  $T, T' \subseteq X$  such that  $x \notin T, T'$ , we have

$$\begin{aligned}
\frac{\mu(T, X \setminus x)}{\mu(T', X \setminus x)} &= \frac{\mu(T, X) + \mu(T \cup x, X)}{\mu(T', X) + \mu(T' \cup x, X)} \Leftrightarrow \frac{\pi(T)}{\pi(T')} = \frac{\pi(T) + \pi(T \cup x)}{\pi(T') + \pi(T' \cup x)} \\
&\Rightarrow \frac{\pi(T \cup x)}{\pi(T)} = \frac{\pi(T' \cup x)}{\pi(T')}.
\end{aligned}$$

The last equation is equation (7) in the proof of Proposition 3. From here one follows the same logic used in the proof of Proposition 3 to obtain the representation for the IC<sup>o</sup> model. This completes our proof. ■

**Proof of Corollary 2:** Suppose  $\mu$  has a nested logit representation. If each nest in the nested logit representation of  $\mu$  consists of a single item, then  $\mu$  has a singleton representation following the proof of Proposition 8. Suppose there exists a nest  $N_i$

containing at least two elements. Take a proper subset  $T$  of  $N_i$  with  $T \neq \emptyset$ . As the utility of an option is strictly positive, it follows that  $\sigma(T) \neq \sigma(N_i)$ . Hence,  $\mu$  does not have a nest-invariant representation. The remaining results in the Corollary directly follow from Propositions 7 and 8. ■