Growing Attention*

Yusufcan Masatlioglu† Tri Phu Vu‡

October 31, 2023

Abstract

We model a boundedly rational agent who exhibits limited and heterogeneous attention. The agents' attention can be naturally sorted in an ascending order. We provide a characterization result using two simple behavioral postulates. Unlike earlier models, consideration sets' composition and frequency are uniquely identified. Within our framework, we accommodate unobserved heterogeneity in attention and unobserved preferences. Furthermore, we explore various extensions of our model and provide insights into the comparative statics of attentiveness levels in two probabilistic choices. Finally, we apply our model to a problem of optimal list design where a designer can intentionally manipulate the list.

Keywords: Revealed Preference, Attention, List, Self-Preferencing.

1 Introduction

This paper is based on two well-established observations on decision-making: (i) People do not pay attention to all available alternatives (but rather consider a smaller subset of feasible options), and (ii) people's attention exhibits heterogeneity and often vary during the decision-making process. Regarding (i), the marketing literature has long emphasized that individuals often limit their attention to a small subset of all available alternatives when making choices. This subset is commonly referred to as a

*We thank Filiz-Emel Ozbay and Erkut Ozbay for helpful discussions and comments.

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

‡Department of Economics, University of Maryland. Email: tvuphu@umd.edu.

consideration set and typically comprises options that the decision maker (hereafter DM) perceives as feasible or relevant to their objectives (Wright and Barbour, 1977). For example, when purchasing pain relievers, shampoo, or soft drinks, the median numbers of products people consider are three, four, and five, respectively (Urban, 1975).

Regarding (ii), the formation of consideration sets is an inherently complex process that depends on stimulus, memory, evaluation cost, cognitive ability, and other factors (Hauser and Wernerfelt, 1990; Shapiro et al., 1997). Changes in internal conditions and external environments frequently result in fluctuating levels of attentiveness and consideration sets. As an illustrative example, consider an individual whose attention varies daily based on the previous night's sleep (Durmer and Dinges, 2005). Longer periods of sleep lead to an improved capacity to effectively process and interact with new information that we encounter as part of our daily experiences. Such unobserved heterogeneity of consideration sets, combined with potentially unobserved preferences, presents a substantial challenge for researchers in identifying both *what* the DM considers and *how frequently* a set of options is considered.

In this paper, we study probabilistic choice resulting from both unobserved heterogeneity of consideration sets and possibly unobserved preferences. Our main contribution is developing a framework where it is possible to uniquely identify *both* the contents and frequencies of consideration sets. Such insights carry valuable implications for managerial decisions, potentially leading to more efficient resource allocation, improved product development, and more effective marketing strategies.

Our identification strategy is based on the fact that the more we sleep, the more we pay attention. For example, researchers have shown that if we reduce sleep to 5 hours per night over a week, people pay less attention compared to those allowed to sleep for 8 hours each night (Dorrian et al., 2004). The person's physical state, like tiredness, discomfort, sickness, etc., also plays an important role in how much the person pays attention. Hence, to facilitate these examples in our model, consideration sets satisfy the *expansion* property. Formally, the collection of consideration sets can be ordered such that the consideration sets of higher types (more attention) nest the consideration sets of lower types (less attention).[1]

[1]Without any structure on the collection of consideration sets, the unobserved heterogeneity of

The expansion of consideration sets property is descriptively appealing as individuals often enrich their consideration sets over the decision-making process. On the one hand, the DM can actively search for additional information to ensure a thorough evaluation of options and make informed choices. For instance, in situations involving repeated choices, if previous selections fail to deliver desired outcomes, individuals may change their approach and seek new information to enhance future outcomes. During the information-collecting process, individuals may come across options that were previously unknown or neglected. The discovery of these new alternatives enlarges the DM's awareness, leading to the expansion of their consideration sets. For example, in the sequential experimentation model of Fershtman and Pavan (2023), the DM makes choices in several periods, and the value of each alternative is not necessarily known to the decision maker before exploration. At the beginning of an arbitrary period, the decision maker has a consideration set and can choose to either stick to the initial consideration set and explore the options within the consideration set or search for new alternatives and add these new options to the initial consideration set for exploration in subsequent periods. Consequently, the consideration sets in the subsequent periods nest the preceding ones.

On the other hand, the DM may enrich her consideration sets through interactions with external inputs such as recommendations and advertising content (Eliaz and Spiegler, 2011; Goodman et al., 2013). To exemplify, consider the iterative search model introduced by Masatlioglu and Nakajima (2013). A consumer initiates a search for a product to buy on e-commerce platforms with limited prior knowledge about the available alternatives. As the consumer evaluates some known or heard-of options, the platforms recommend her to look at other related items. The recommendation attracts the consumer's attention and enlarges the consumer's awareness, leading to an evolution of her consideration sets. Also, in the competitive marketing model of Eliaz and Spiegler (2011), firms can use different marketing strategies to manipulate consumer's attention and the consumer's considerate sets evolve through interactions with the advertising content. The list of choice models satisfying the expansion of consideration sets property also includes satisficing (Simon, 1955), rationalization

consideration sets poses a serious challenge to the identification problem. For example, consider the simplest scenario, which is the standard random utility model (RUM). Under RUM, the decision maker evaluates all available alternatives, effectively eliminating concerns about unobserved heterogeneity in consideration sets. However, it is well-known that RUM is not uniquely identified.

(Cherepanov et al., 2013), and rational inattention (Caplin et al., 2019). Details on these examples are given in Section 2.

Our model applies to choices of either a single individual in varying contexts (intrapersonal) or different individuals in the population (interpersonal). In the latter case, it is necessary that the preferences of all individuals are identical. This requirement is justified if individuals' preferences are commonly characterized by specific criteria (e.g., seeking the most affordable housing, purchasing the highest-quality products, choosing the most environmentally friendly cars, etc.). The shared preference assumption is also utilized in the studies by Dardanoni et al. (2020), Cattaneo et al. (2020), and Hagiu et al. (2022). The former paper interprets the single preference as the average utilities within the population.

Besides the expansion of consideration sets property, we investigate choice behaviors under certain assumptions regarding the formation of consideration sets. Our model is based on the limited attention model of Masatlioglu et al. (2012) where the consideration set is unaffected when overlooked alternatives are removed from the feasible set.[2] This is one of the common properties of consideration set formation documented in the psychology literature (Broadbent, 1958). Under the standard choice theory assumption that the decision maker evaluates all available alternatives, this condition is trivially satisfied. The property is also descriptively appealing as it holds when decision-makers use heuristic decision rules to determine their consideration sets (Masatlioglu et al., 2012).

Our first result is a characterization of choice behaviors resulting from such consideration set formation structure and the expansion of consideration set property. We identify two simple and intuitive behavioral postulates: weak monotonicity and independence. Weak monotonicity requires that the likelihood of selecting an option should not increase when a better alternative is added to the choice set. It is similar to but less restrictive than the classic monotonicity axiom. Hence, our model can accommodate monotonicity violations, a feature of a stochastic choice theory (Manzini and Mariotti, 2018). Meanwhile, roughly speaking, independence states that choice frequencies of alternative z must remain unchanged when a higher-ranked alternative

[2]When there is no structure on the formation of consideration sets, any random choice function can be explained in our framework using the expansion property.

x is added to the menu given that there exists an option y ranked between the x and z and chosen with a strictly positive probability (y is attractive). Intuitively, the choice probability of z is independent to the presence of x because the attractiveness of y absorbs any potential shifts.

In our first result, preferences are given exogenously. In certain contexts and applications, the preference might not be observable by the outside analyst and must be inferred from the choice data. Endogenizing the preference also enhances the practical applicability of our model. We discuss how to identify preferences when they are not observable. We show that a regularity violation is sufficient for identifying preferences. We also provide additional conditions to identify preference despite no regularity violation. Generally, the endogenous preference is not unique when choice probabilities satisfy regularity. We argue that such non-uniqueness arises primarily because there are no restrictions on binary choices. We proceed to establish that by imposing a simple property on binary choices, the endogenous preference can be limited to at most two candidates, allowing for unique representation as well.

Next, we discuss how to identify the consideration sets from choice data. Generally, identifying the consideration sets is significantly challenging due to the fact they are typically not observable. The conventional approach in the literature is to use auxiliary data or impose some strong conditions on the formation of consideration sets to narrow it down significantly. Regarding the latter, for example, Caplin et al. (2019) assume that an option is considered at a choice set if and only if it is chosen with a strictly positive probability. Meanwhile, Honka (2014, p. 857) imposes a condition that if an alternative x is considered, so are all better-ranked options. In our framework, both the contents and frequencies of consideration sets can be uniquely identified when the choice data has full support.[3] Even when the choice probabilities do not have full support, the list of candidates for consideration sets can also be significantly narrowed down. The (unique) identification of the composition and frequencies of consideration sets also allows us to perform comparative statics of attentiveness levels at two probabilistic choices. We intuitively show that increased levels of attentiveness result in improved decision-making outcomes. A converse re-

[3] A choice data has full support if every element in a choice set is chosen with a strictly positive probability. Throughout the paper, we will use the full-support random choice function and positive random choice function interchangeably.

lationship also holds as better choices require heightened attentiveness levels.

Additionally, we consider two extensions of our baseline model by introducing a *direction* of consideration sets' expansion. In the first extension, we model agents whose consideration sets follow a ranked list of items. Such situations are ubiquitous in daily life as the list can be offered by a search engine, an online shopping platform, or a voting ballot, for example. In the second extension, we allow for variations of the ranked list in different choice sets. Put differently, the relative positions of two alternatives in the list may depend on the availability of other options. An illustrative instance of this scenario involves a customer conducting searches across multiple platforms (desktop, tablet, mobile) and using the list of search results as her guiding list. In this context, it is well documented that a search engine generates different first-page lists depending on its algorithms for each platform.

Finally, we apply our model to study a problem of optimal list design. In recent years, online platforms such as Amazon, Target, and Apple's App Store are increasingly assuming a dual role, functioning both as marketplaces for third-party sellers and as a seller by offering their own products on their marketplaces (Hagiu et al., 2022; Padilla et al., 2022; Farronato et al., 2023). Empirical studies have indicated that these platforms often manipulate customer search results to promote their own items (Chen and Tsai, 2023; Farronato et al., 2023). This practice, commonly referred to as self-preferencing, has triggered heated policy debates in the United States and European countries due to concerns about antitrust violations and potential negative effects on consumer welfare. Motivated by this self-preferencing phenomenon, we consider a scenario where a designer wants to construct a list to maximize an objective function. Under mild assumption of the designer's objective function, we provide a simple algorithm to identify all optimal lists fully. Our primary finding is that the optimal list may not align perfectly with the designer's priority order. Intuitively, the designer can strategically place alternatives perceived as inferior by customers and of minimal value to her after options that are perceived as superior by customers and hold high value to her. The ordering incentivizes customers to choose the high-value options, resulting in greater benefits for the designer.

The rest of the paper is organized as follows. Section 2 introduces a Growing Attention Model (GAM). Section 3 provides a characterization result. Section 4 states our results for GAM with an endogenous preference. Section 5 identifies the content of consideration sets and conducts a comparative analysis of attentiveness levels within two probabilistic choices. Section 6 extends our baseline models to accommodate a direction of consideration sets' expansion. Section 7 presents an application of our model to list design. Section 8 reviews the literature and compares GAM to other related models of stochastic choice. Finally, Section 9 concludes.

2 Model

Let X be the finite set of alternatives and $\mathcal{X}$ the set of all nonempty subsets of X. We will refer to each element of $\mathcal{X}$ as a menu or a choice set. A consideration set is a map $\Gamma : \mathcal{X} \to \mathcal{X}$ such that $\Gamma(S) \subseteq S$ for all $S \in \mathcal{X}$. Given the consideration set $\Gamma(S)$, the DM selects the best alternative in $\Gamma(S)$ according to a complete and transitive preference $\succ$, which is denoted by $\max(\Gamma(S), \succ)$.[4] Following Masatlioglu et al. (2012), a consideration map $\Gamma : \mathcal{X} \to \mathcal{X}$ is an attention filter if $\Gamma(S) \subseteq S$ and $\Gamma(S) = \Gamma(S \setminus x)$ whenever $x \notin \Gamma(S)$. Let $\mathcal{T}$ be the set of all attention filters. We say a collection of distinct attention filters $\Gamma \subseteq \mathcal{T}$ has a *growing attention* structure if Γ can be sorted $\Gamma = \{\Gamma_1, \Gamma_2, ..., \Gamma_m\}$ such that Γ gradually expands: $\Gamma_1(S) \subseteq \Gamma_2(S) \subseteq ... \subseteq \Gamma_m(S)$ for all $S \in \mathcal{X}$. The expansion property says that the agent becomes gradually more attentive as she considers more and more options. This expansion structure of consideration sets is observed in several models of stochastic choice and consideration set formation. We give some examples below. In all examples, consideration sets are attention filters.

Example 1 (Satisficing, Simon (1955) and Aguiar et al. (2016))**.** Let $u_1, u_2, \ldots, u_m$ be real numbers satisfying $u_1 \geq u_2 \geq \cdots \geq u_m$. Each u_i can be interpreted as a utility threshold. The decision maker considers all alternatives with utilities exceeding the reservation utility u_i. That is, $\Gamma_i(S) = \{x \in S : u(x) \geq u_i\}$, where u is the utility function representing the DM's preference. Clearly, $\Gamma_1(S) \subseteq \Gamma_2(S) \subseteq ... \subseteq \Gamma_m(S)$.

Example 2 (Rationalization, Cherepanov et al. (2013))**.** In the rationalization model, the DM only considers alternatives that can be rationalized by a rationale (a linear order). Let $\mathbb{P} = \{P_1, P_2, \ldots, P_n\}$ be the set of all rationales. The DM's consideration set for a given menu S is $\Gamma(S) = \{x \in S : \exists i \text{ s.t. } xP_iy\ \forall y \neq x, y \in S\}$. Let

[4] Throughout the paper, we use $\max(\Gamma(S), \succ)$ and $\succ$-best element in $\Gamma(S)$ interchangeably.

$R_i \subseteq \mathbb{P}$ be the set rationales used by type i. If higher-order types use more rationales, that is $R_i \subseteq R_j$ when $i < j$, then it follows that consideration sets must expand: $\Gamma_1(S) \subseteq \Gamma_2(S) \subseteq ... \subseteq \Gamma_m(S)$ for all $S \in \mathcal{X}$.

Example 3 (Competitive Marketing, Eliaz and Spiegler (2011))**.** In the model of competitive marketing, the consumer has an initial consideration set coupled with a consideration function that determines whether new products are added for consideration. The consideration function depends on the products and marketing strategies that firms use to advertise the products. Firms can manipulate the customer's attention by employing different marketing techniques. Let X be the set of all products and $\mathcal{M} = \{1, 2, \ldots, M\}$ the set of marketing strategies. Eliaz and Spiegler (2011) interpret elements in $\mathcal{M}$ as advertising intensity with $m_1 > m_2$ implying that the advertisement under strategy m_1 is more intense than under strategy m_2 (section 4.1 in their paper). There is a default option $a^* \in X$ with marketing level $M^{a^*} \in \mathcal{M}$ that is always included in the consideration set. The consideration function is a mapping $\phi : X \times \mathcal{M} \to \{0, 1\}$ with $\phi = 1$ meaning the product is considered. Eliaz and Spiegler (2011) further impose that $\phi(x, M^x) = 1$ if and only if $M^x \geq M^{a^*}$, where M^x is the advertising intensity of product x. Put differently, x is added to the consideration set if and only if it is marketed at least the same level of intensity as the default option. In our model, let M_i^x be the advertising intensity of product x at type i. Suppose higher-order types experience more intensive advertising (higher types are more targeted), and suppose that the advertising intensity of the default option does not vary (it is always considered by the consumer so there is no need to adjust the intensity level). That is, $M_i^{a^*} = M_j^{a^*}$ for all i, j and $M_i^x \geq M_j^x$ for all $x \in X$ whenever $i \geq j$. Then $\phi(x, M_j^x) = 1 \Rightarrow \phi(x, M_i^x) = 1$ if $i \geq j$. Hence, if $x \in \Gamma_j(S)$ then $x \in \Gamma_i(S)$. Put differently, consideration sets gradually expand over the sorted types.

Example 4 (Rational Inattention, Caplin et al. (2019))**.** Let Ω and $\mathcal{X}$ be the state space and action space, respectively. There is a prior distribution over states $\mu \in \Delta(\Omega)$. Utility function is state-dependent $u : \mathcal{X} \times \Omega \to \mathbb{R}$. Given distribution μ, utility u, and a subset of actions $S \in \mathcal{X}$ (a choice set) from which the DM must select, the DM chooses the state-dependent random choice function $\pi : \Omega \to \Delta(A)$, where $\pi(x|\omega)$ is the probability of choosing action $x \in S$ in state $\omega \in \Omega$. The DM's objective is to maximize her expected utility minus the Shannon mutual information costs between actions and states. Put differently, the DM chooses function π to maximize

$$\sum_{\omega\in\Omega}\mu(\omega)\left(\sum_{x\in S}\pi(x|\omega)u(x,\omega)\right)-\lambda\left[\sum_{\omega\in\Omega}\mu(\omega)\left(\sum_{x\in S}\pi(x|\omega)ln\pi(x|\omega)\right)-\sum_{x\in S}\pi(x)ln(x)\right],$$

where $\pi(x)=\sum_{\omega\in\Omega}\mu(\omega)\pi(x|\omega)$ is the unconditional probability that option x is chosen and λ is the marginal cost of information (attention cost). For each state-dependent random choice function π, Caplin et al. (2019) define the consideration set as the set of options that are chosen with strictly positive probabilities: $\Gamma(S)=\{x\in S|\pi(x)>0\}$. In the canonical case (section 3.1 in their paper), the consumer is faced with m options in a set $S=\{x_1,x_2,\ldots,x_m\}$. Among these, there is only one good option, and the remaining ones are bad. The state space is set equal to the action space, implying that option x_i is the good option in state ω_i. The utility function is $u(x_i,\omega_j)=u_G$ if $i=j$ (good option) and $u(x_i,\omega_j)=u_B<u_G$ otherwise (bad option). The prior distribution $\mu(\omega_i)$ is the prior probability that option x_i is the good one, and the states are ordered according to the perceived likelihood: $\mu(\omega_i)\geq\mu(\omega_{i+1})$. Given the marginal cost of Shannon mutual information $\lambda>0$, Caplin et al. (2019) show that the consideration set comprises of the first $K\leq m$ options: $\Gamma_\lambda(S)=\{x_1,x_2,\ldots,x_K\}$ where the unique integer K satisfies

$$\mu(\omega_K)>\frac{\sum_{k=1}^{K}\mu(\omega_k)}{K+\exp\left(\frac{u_G-u_B}{\lambda}\right)-1}\geq\mu(\omega_{K+1}),\ \text{with}\ \mu(\omega_t)=0\ \text{when}\ t>m$$

Clearly, $\Gamma_{\lambda_1}(S)\subseteq\Gamma_{\lambda_2}(S)$ whenever $\lambda_1\geq\lambda_2$. Hence, if higher-index types in our model have a lower marginal cost of information, then the consideration sets gradually expand over the sorted types. Intuitively, the lower attention cost allows the DM to consider more and more options.

A random choice function (RCF) is defined as a mapping $\pi:X\times\mathcal{X}\to[0,1]$ such that $\sum_{x\in S}\pi(x,S)=1$ and $\pi(x,S)=0$ whenever $x\notin S$. Here, $\pi(x,S)$ denotes the probability of choosing option x from menu S. Given the growing attention property, we define our model as follows.

Definition 1. Given a preference $\succ$, a stochastic choice function π has a *growing attention model* (GAM($\succ$)) representation if there exists a collection of growing attention filters $\Gamma = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_m\}$ and a probability measure μ on Γ such that

$$\pi(x, S) = \sum_{i:\, \max(\Gamma_i(S), \succ) = x} \mu(\Gamma_i), \qquad \text{for all } x \in S \text{ and } S \in \mathcal{X} \qquad \text{(GAM } (\succ))$$

3 Behavioral Characterization

Our characterization result uses two simple and intuitive axioms: weak monotonicity (w-MON) and independence. The first axiom is a weakening of the regularity condition.

Axiom 1 (w-MON). $\pi(y, S) \leq \pi(y, S \setminus x)$ if $x \succ y$.

Weak monotonicity states that adding a preferred alternative in the choice set will decrease the choice probability of a dominated option. This axiom is similar to but less restrictive than the usual regularity axiom; the choice probability cannot increase by adding *any* alternative. Our axiom only requires regularity to hold when a better alternative is added to the choice set.

This condition is a necessary condition for our model. Note that w-MON is trivially satisfied when $\pi(y, S) = 0$. When $\pi(y, S) > 0$, there must exist a type Γ_i in the support such that $y = \max(\Gamma_i(S), \succ)$. If $x \in \Gamma_i(S)$ then it is revealed that $y \succ x$, which is a contradiction. Hence, $x \notin \Gamma_i(S)$ and it follows that $\Gamma_i(S) = \Gamma_i(S \setminus x)$ because Γ_i is an attention filter. Therefore, $\max(\Gamma_i(S), \succ) = \max(\Gamma_i(S \setminus x), \succ)$ and it follows $\pi(y, S) \leq \pi(y, S \setminus x)$.

The next axiom states conditions when adding a better alternative does not affect choices. Assume that x is preferred to y and y is preferred to z. Hence, both x and y are better than z. Then, as long as y is chosen with positive probability, adding even a better alternative does not influence the choice probability of z.

Axiom 2 (Independence). Suppose $x \succ y \succ z$ and $S \supseteq \{x, y, z\}$. Suppose $\pi(y, S) > 0$. Then $\pi(z, S \setminus x) = \pi(z, S)$.

Independence is similar to the centrality axiom in Apesteguia et al. (2017).[5] It says

[5] It can be shown that the centrality axiom in Apesteguia et al. (2017) and the usual regularity

that when removing a better alternative from the choice set, the choice probabilities remain unchanged if there exists an alternative that is chosen with a strictly positive probability and ranked between the removed alternative and the considered option.

This condition is also a necessary condition for our model. Since π has a GAM representation and $\pi(y,S)>0$, there must exist a type k such that $y=\max(\Gamma_k(S),\succ)$. Among such k, there is the smallest one, called k^*. Observe that $y=\max(\Gamma_{k^*}(S),\succ)$ and $x\succ y$ imply that $x\notin\Gamma_{k^*}(S)$. Then, it follows that $\Gamma_{k^*}(S)=\Gamma_{k^*}(S\setminus x)$ because Γ_{k^*} is an attention filter. Hence, $y=\max(\Gamma_{k^*}(S),\succ)=\max(\Gamma_{k^*}(S\setminus x),\succ)$. Note that $y=\max(\Gamma_{k^*}(S),\succ)=\max(\Gamma_{k^*}(S\setminus x),\succ)$ implies $\max(\Gamma_t(S),\succ)\neq z\neq\max(\Gamma_t(S\setminus x),\succ)$ for all $t\geq k^*$ because of the expansion property. Also, observe that $\max(\Gamma_t(S),\succ)=\max(\Gamma_t(S\setminus x),\succ)$ for all $t\leq k^*-1$. The reason is that if there exists $t^*\leq k^*-1$ that $\max(\Gamma_{t^*}(S),\succ)\neq\max(\Gamma_{t^*}(S\setminus x),\succ)$ then x must belong to $\Gamma_{t^*}(S)$. However, it could not happen as $y=\max(\Gamma_{k^*}(S),\succ)$ and $x\succ y$ and $\Gamma_{t^*}(S)\subseteq\Gamma_{k^*}(S)$. Finally, $\max(\Gamma_t(S),\succ)=\max(\Gamma_t(S\setminus x),\succ)$ for all $t\leq k^*-1$ and $\max(\Gamma_t(S),\succ)\neq z\neq\max(\Gamma_t(S\setminus x),\succ)$ for all $t\geq k^*$ imply that $\pi(z,S\setminus x)=\pi(z,S)$.

We now state our characterization result for GAM($\succ$).

Theorem 1. RCF π has a GAM($\succ$) representation if and only if π satisfies w-MON and Independence.

Theorem 1 shows that GAM is captured by two simple behavioral postulates that are relatively easy to verify. The proof of Theorem 1 is constructive.[6] First, we use the observed choice frequencies and preferences to identify the $\succ$-best element within every type's consideration set. Such an identification does not require the knowledge about the underlying structure of consideration sets and is possible thanks to a construction technique developed by Filiz-Ozbay and Masatlioglu (2023). Subsequently, we define a consideration set by including its best element that we just identified and all options within the lower contour set of its best element. Finally, we use induction to prove that the consideration sets are attention filters and satisfy the expansion property.

condition imply independence in our model. Hence, SCRUM is a subclass of GAM.

[6]All omitted proofs are given in Appendix A.

4 Endogenous GAM

Our characterization provided in the previous section assumes that the preference order in GAM is exogenous. In certain contexts, the preference might not be observable by the outside analyst. In such instances, it becomes necessary to infer the preference order from the choice data. This section presents the methodology for such identification. We first define an endogenous GAM representation: A random choice function π has an endogenous GAM representation if there exists a preference order $\succ$ such that π has a GAM($\succ$) representation. Given the possibility of multiple endogenous GAM representations, we define a revealed preference as follows.

Definition 2. Suppose π has an endogenous GAM representation. We say that x is revealed to be preferred to y if x is preferred to y in every preference representing π.

The definition of revealed preference is conservative to ensure we avoid making erroneous claims about the decision maker's preferences.[7] Next, we show that a regularity violation is sufficient to uncover the revealed preference between a pair of alternatives. Axiom 1 states that if there exists a regularity violation, it must be that the alternative causing the violation must be the less preferred alternative. In other words, if $\pi(x, S) > \pi(x, S \setminus y)$ for some $S \supseteq \{x, y\}$ then x must be revealed to be preferred to y. The Proposition below states this observation.

Proposition 1 (Revealed Preference-1)**.** Suppose RCF π has an endogenous GAM representation. If $\pi(x, S) > \pi(x, S \setminus y)$ for some $S \supseteq \{x, y\}$ then x is revealed to be preferred to y.

The revealed preference from Proposition 1 is solely based on Axiom 1. Cattaneo et al. (2020) also prove a similar result. In their model, the regularity violations were the only source of revealed preference. Indeed, in the extreme case of no regularity violation in the choice data, there is no revelation. In our model, other observations can reveal the underlying preferences. Specifically, one can identify endogenous preference by checking choices from binary and trinary sets with the help of Axiom 2. Assume we observe $\pi(x, \{x, y, z\}) \neq \pi(x, \{x, y\})$ and $\pi(x, \{x, y\}) < \pi(x, \{x, z\}) <$

[7] Other studies also use the same approach. For example, see Masatlioglu et al. (2012) and Cattaneo et al. (2020), among others.

$\pi(y, \{y, z\})$. By Axiom 2, the former implies that we cannot have $z \succ y \succ x$. Then $\pi(x, \{x, y\}) < \pi(x, \{x, z\})$ rules out $y \succ z \succ x$ following Independence and w-MON as if $y \succ z \succ x$ occurs then $\pi(x, \{x, z\}) = \pi(x, \{x, y, z\}) \leq \pi(x, \{x, y\})$, which is a contradiction. Third, by a similar argument, the inequality $\pi(x, \{x, z\}) < \pi(y, \{y, z\})$ excludes $y \succ x \succ z$. Therefore, all linear orders that rank y above x are effectively ruled out, leading to $x \succ y$. Our next result states this observation and offers a revealed preference relationship that does not require regularity violations.

Proposition 2 (Revealed Preference-2)**.** Suppose RCF π has an endogenous GAM representation. If there exists z such that one of the following occurs

i) $\pi(x, \{x, y, z\}) \neq \pi(x, \{x, y\})$ and $\pi(x, \{x, y\}) < \pi(x, \{x, z\}) < \pi(y, \{y, z\})$; or

ii) $\pi(x, \{x, y, z\}) \neq \pi(x, \{x, z\})$ and $\pi(x, \{x, z\}) < \min\{\pi(y, \{y, z\}), \pi(x, \{x, y\})\}$,

then x is revealed to be preferred to y.

Based on Propositions 1 and 2, we introduce the following binary relation. For $x \neq y$, define:

$$
\begin{array}{lll}
xPy \text{ if} & i) & \exists S \supseteq \{x, y\} \text{ s.t. } \pi(x, S) > \pi(x, S \setminus y) \text{ or;} \\
& ii) & \exists z \text{ s.t. } \pi(x, \{x, y, z\}) \neq \pi(x, \{x, y\}) < \pi(x, \{x, z\}) < \pi(y, \{y, z\}) \text{ or;} \\
& iii) & \exists z \text{ s.t. } \pi(x, \{x, y, z\}) \neq \pi(x, \{x, z\}) < \min\{\pi(y, \{y, z\}), \pi(x, \{x, y\})\}.
\end{array}
$$

By Propositions 1 and 2, if xPy then x is revealed to be preferred to y. Put differently, this is the sufficient condition to identify the preference ranking between x and y. Additionally, since the underlying preference is transitive, we also conclude that the decision maker prefers x to z if xPy and yPz for some y, even when xPz is not directly revealed from the observed choice data. Hence, the underlying preference must include the transitive closure of P.

Using binary relation P, we can further restrict the set of possible candidates for the underlying preferences. Note that Axiom 2 implies that if $x \succ z$ and $y \succ z$ then $\pi(z, \{x, y, z\})$ must be equal to either $\pi(z, \{x, z\})$ or $\pi(z, \{y, z\})$, regardless of the preference ranking between x and y. Hence, xPz and $\pi(z, \{x, y, z\})$ is not equal to neither $\pi(z, \{x, z\})$ nor $\pi(z, \{y, z\})$ imply that z is revealed to be preferred to y. The following Proposition states this observation.

Proposition 3 (Revealed Preference-3)**.** Suppose RCF π has an endogenous GAM representation. Then xPz and $\pi(z, \{x, y, z\}) \notin \{\pi(z, \{x, z\}), \pi(z, \{y, z\})\}$ imply that z is revealed to be preferred to y.

Given Propositions 1-3, our identification strategy of endogenous preferences is as follows. We begin by assuming that the RCF has an endogenous GAM representation, allowing us to apply Propositions 1-3 to limit the possible preference orderings. Let Ψ be the set of linear orders that include the transitive closure of the revealed preferences identified in Propositions 1-3. The set Ψ must contain all endogenous preferences if they exist. We then apply the characterization result in Theorem 1 to Ψ by checking Axioms 1 and 2 for each order in Ψ. If $\Psi = \emptyset$ or $\Psi \neq \emptyset$ but no ordering in Ψ satisfies both axioms, we can conclude that the RCF has no GAM representation. Conversely, if there is one ordering in Ψ satisfying the two axioms, the RCF has at least one GAM representation with the preference ordering we just identified. To illustrate this identification process, we give a simple example below.

Example 5. Table 1 provides a set of parametric probabilistic choices described by π_α, where $\alpha \in (0, 0.5)$.

π_α	$\{x, y, z\}$	$\{x, y\}$	$\{x, z\}$	$\{y, z\}$
x	α	0.30	0.50	$-$
y	$0.60 - \alpha$	0.70	$-$	0.60
z	0.40	$-$	0.50	0.40

Table 1: Probabilistic choice functions.

Suppose π_α has an endogenous GAM representation. When $\alpha \in (0.3, 0.5)$, there is one regularity violation in the data: $\pi_\alpha(x, \{x, y, z\}) > \pi_\alpha(x, \{x, y\})$. Proposition 1 then implies that x is revealed to be preferred to z. Applying part i) of Proposition 2 yields x is revealed to be preferred to y. Given that $\pi_\alpha(y, \{x, y, z\}) \notin \{\pi_\alpha(y, \{x, y\}), \pi_\alpha(y, \{y, z\})\}$ and xPy (by the definition of P), it follows from Proposition 3 that y is revealed to be preferred to z. By transitivity, there is only one candidate for the underlying preferences: $x \succ_1 y \succ_1 z$. Put differently, $\Psi = \{\succ_1\}$. It is routine to show that Axioms 1 and 2 are satisfied with $\succ_1$. Therefore, π_α has a GAM representation, and the underlying preference must be $\succ_1$ (unique identification).

When $\alpha \in (0, 0.3)$, the choice data satisfies regularity, and Proposition 1 is silent on the revealed preferences. Part i) of Proposition 2 still implies that x is revealed to be preferred to y, and Proposition 3 indicates that y is revealed to be preferred to z. Hence, $\Psi = \{\succ_1\}$ and π_α has a GAM($\succ_1$) representation (unique identification). Finally, when $\alpha = 0.3$, Propositions 1-3 identify no reveal preferences and Ψ covers all linear orders. By checking Axioms 1-2, π_α has two GAM representations with $\succ_1$ and $\succ_2$, where $\succ_2$ is the opposite of $\succ_1$: $z \succ_2 y \succ_2 x$.

We conclude this section by commenting on the number of possible preferences representing the data. In Example 5 above, we have identified that there are at most two possible preferences representing data for all $\alpha \in (0, 0.5)$. One might wonder whether this is specific to this example. In general, the smaller set of preferences is desirable for the purpose of welfare analyses since we can reveal more about the underlying preferences. In Appendix B, we show that under two mild properties, there are at most two preferences. We also illustrated that these two properties are implied by the well-known conditions which have been supported by the existing empirical evidence. More importantly, this result is independent of the size of X; see Appendix B for details.

5 Identification of Consideration Sets and Comparative Statics

Identifying consideration sets provides valuable insights into decision-making processes. Such insights are crucial for managerial decisions because they shed light on which items or products the DM actively evaluates. The absence of understanding consideration sets can result in poor-quality managerial choices with highly unfavorable outcomes. A case in point is highlighted by Hauser (2014, p. 1688), who illustrates that substantial investments in aspects like reliability, style, interior design, and quality failed to prevent two American automakers from declaring bankruptcy in 2009. This failure stemmed from consumers not being exposed to the improved products, as these products were never included in their consideration sets in the first place.

In this section, we study how to identify consideration sets within our framework. We begin with a uniqueness result. Theorem 2 below states that the GAM($\succ$)

representation, if it exists, is unique when the choice data has full support.

Theorem 2. Suppose a positive RCF π has a GAM($\succ$) representation with Γ and μ being the collection of distinct growing attention filters and probability measure, respective. Then (Γ, μ) is unique.

The uniqueness of the probability measure μ in Theorem 2 comes from the construction of types outlined in the proof of Theorem 1. The construction provides the exact weights for each type in the support. The uniqueness of the consideration sets results from the full-support assumption of the random choice function. This assumption also allows us to identify the composition of consideration sets uniquely. As mentioned earlier, we can always identify the $\succ$-best element within each type's consideration set from the observed choice data. It turns out that the identification of $\succ$-best elements is sufficient to pin down the consideration set's composition. In the proof of Theorem 2, we show that the consideration set of each type has the following structure: $\Gamma_i(S) = \{x \in S : \max(\Gamma_i(S), \succ) \succsim x\}$ for all i and $S \in \mathcal{X}$. Put differently, each type's consideration set includes its $\succ$-best element and all options within the lower contour set of its best element.

To appreciate the unique identification of both the composition and frequency of consideration sets, note that there are three sources of variation in a GAM. Firstly, we can vary the probability measure μ. Secondly, we can change the *observed* characteristic of the consideration sets, meaning that we can vary the collection of $\succ$-best elements that results from applying preference $\succ$ to the collection of consideration sets. Finally, variations can occur in the *unobserved* characteristic of the consideration sets. This means that even if the collection of $\succ$-best elements remains unchanged, consideration sets can still vary. Theorem 2 consolidates these three types of variation and asserts that no variation is admissible when the choice data has full support.

Comparative Statics of Attentiveness Levels

To provide a practical application of Theorem 2, we illustrate that the (unique) identification of consideration sets allows us to compare the attentiveness levels within two GAM potentially characterized by distinct preferences. Understanding attentiveness levels is important for effective communication as it helps tailor the message appropriately. For instance, in marketing, by understanding how attentive their audience is, firms can create engaging advertisements and promotional campaigns.

We define attentiveness as the ratio between the number of options the decision maker considers and the total number of available alternatives. Put differently, we quantify attentiveness by comparing the size of consideration sets to the size of the menu. For notation simplicity, we write $\pi = (\Gamma, \mu, \succ)$ if π can be written as a probability measure μ over a collection Γ of attention filters given preference $\succ$, where the collection Γ is not necessarily growing. Given two RCFs $\pi = (\Gamma, \mu, \succ)$ and $\pi' = (\Gamma', \mu', \succ')$, we say that π is more attentive than π' if

$$\sum_{i:\, \frac{|\Gamma_i(S)|}{|S|} \geq k} \mu_i \geq \sum_{j:\, \frac{|\Gamma'_j(S)|}{|S|} \geq k} \mu'_j, \text{ for all } S \in \mathcal{X} \text{ and } k \in \left\{ \frac{1}{|S|}, \frac{2}{|S|}, \ldots, \frac{|S|}{|S|} \right\}$$

In other words, π is more attentive than π' if the size of consideration sets under π first-order stochastically dominates that under π'.

To establish a link between observed choice data and attentiveness levels, we need a partial order between two RCFs as follows.

Definition 3 (First-order stochastic dominance, FOSD)**.** Let $\succ$ and $\succ'$ be two preference orders. Let $U_{\succsim}(x, S)$ be the upper contour set of x in S given preference $\succ$ and $\pi(U_{\succsim}(x, S), S)$ the cumulative choice probabilities of all elements in $U_{\succsim}(x, S)$. RCF π first-order stochastically dominates RCF π' (π FOSD π') if $\pi(U_{\succsim}(x, S), S) \geq \pi'(U_{\succsim'}(x', S), S)$ for all $S \in \mathcal{X}$ and $x, x' \in S$ such that $|U_{\succsim}(x, S)| = |U_{\succsim'}(x', S)|$.

Corollary 1 below states the relationship between attentiveness levels and observed choice data. It directly follows from the unique identification of both the composition and frequency of consideration sets in Theorem 2.

Corollary 1. Suppose RCFs π and π' have GAM($\succ$) and GAM($\succ'$) representations, respectively. Suppose both of them have full support. Then π is more attentive than π' if and only if π FOSD π'.

Corollary 1 intuitively says that a higher level of attentiveness leads to improved outcomes. Furthermore, better decision-making similarly requires increased attentiveness levels. Notably, Corollary 1 does not require that the preferences or collections of consideration sets in two RCFs are identical.

The subsequent Corollary partially relaxes the full-support assumption laid out in Corollary 1. Expressly, it only necessitates the presence of one positive random choice function. This relaxation, however, comes at a cost: it no longer guarantees that higher attentiveness levels will invariably lead to improved decision-making outcomes. The Corollary also follows from Theorem 2.

Corollary 2. Suppose RCFs π and π' have GAM($\succ$) and GAM($\succ'$) representations, respectively. Suppose π has full support and FOSD π'. Then π is more attentive than π'.

6 List-based expansion of consideration sets

In our baseline model, we do not impose a particular direction on the expansion of consideration sets. In this section, we introduce such a direction. For example, imagine someone who comes across a ranked list of search results on a search platform. Because people have limited attention spans, it is often not feasible for them to go through all the results exhaustively. Instead, when going through the list, a decision-maker explores various options to create their consideration set, but they may only include a subset of the available alternatives due to their limited attention. We make the assumption that this consideration set follows the list that the individual faces: if an option x is included in the consideration set, then every feasible alternative that appeared before x in the list is also included. In simpler terms, when an alternative is deemed worthy of consideration, all the alternatives listed before are also considered.[8]

Formally, we write $y \lhd x$, or equivalently, $x \rhd y$ if x appears after y in the list, where $\rhd$ is the underlying list order.[9] The consideration sets satisfy the following property: if $x \in \Gamma(S)$ and $x \rhd y$ then $y \in \Gamma(S)$. A stochastic choice function π is said to have a $\text{GAM}_{\rhd}(\succ)$ representation if it has a GAM($\succ$) representation where consideration sets in the support have a list-based structure with respect to $\rhd$.[10] We show that $\text{GAM}_{\rhd}(\succ)$ can be characterized by three axioms: $\rhd$-wMON, $\rhd$-Independence, and Identity (IDE). The first two are reminiscent of w-MON and Independence, while the

[8] This assumption is also used in other studies. For example, see Honka (2014, p. 857), Cattaneo et al. (2021), Yegane (2022), and Manzini et al. (2023), among others.

[9] In Appendix C, we consider the context-dependent lists and provide a characterization result to accommodate such generalization.

[10] Using the proof of Theorem 1, it can be shown that any GAM($\succ$) has a $\text{GAM}_{\succ}(\succ)$ representation.

last axiom captures the list-based structure of consideration sets.

Axiom 3 ($\rhd$-wMON)**.** Suppose $x \trianglerighteq y$ and $y \succ z$. Then $\pi(z, S) \leq \pi(z, S \setminus x)$ for all $S \supseteq \{x, y, z\}$.

To understand $\rhd$-wMON, it is essential to recognize that the condition $x \trianglerighteq y$ indicates that either $x \equiv y$ or x appears after y in the list. This implies that if the agents consider option x, they also consider option y. Therefore, $\rhd$-wMON states that adding an arbitrary alternative (x) to the choice set that, if considered, subsequently leads to the consideration of a preferred option (y) will (weakly) decrease the choice probability of a dominated item (z). This is intuitive as the presence of x in the consideration set ensures the presence of y, and due to the dominance of y over z, the likelihood of selecting z decreases. Notably, $\rhd$-wMON constitutes a generalization of w-MON in Axiom 1. Indeed, setting $x \equiv y$ in Axiom 3 yields Axiom 1. Additionally, in cases where $x \neq y$ but the list order aligns with the preference order on the pair (x, y), $\rhd$-wMON directly follows from w-MON due to the transitivity of the preference.[11]

Axiom 3 is necessary for $\text{GAM}_{\rhd}(\succ)$. To see why, observe that it is trivially satisfied when $\pi(z, S) = 0$. When $\pi(z, S) > 0$, there exists a type i in the support such that $\max(\Gamma_i(S), \succ) = z$. Suppose $x \in \Gamma_i(S)$. Since $\Gamma_i(S)$ has the list-based structure and $x \trianglerighteq y$, it follows $y \in \Gamma_i(S)$. Therefore, $z = \max(\Gamma_i(S), \succ) \succsim y$, which is a contradiction. Hence, $x \notin \Gamma_j(S)$. Then $\Gamma_i(S) = \Gamma_i(S \backslash x)$ since Γ_i is an attention filter. Therefore, $\max(\Gamma_i(S), \succ) = \max(\Gamma_i(S \setminus x), \succ)$ and it follows $\pi(z, S) \leq \pi(z, S \setminus x)$.

Axiom 4 ($\rhd$-Independence)**.** Suppose $x \trianglerighteq y$ and $y \succ z$ and $\pi(z, S) > 0$. Then $\pi(t, S) = \pi(t, S \setminus x)$ for all $S \supseteq \{x, y, z, t\}$ and t such that $z \rhd t$.

To understand $\rhd$-Independence, note that condition $x \trianglerighteq y$ again implies that if the agents consider option x, they also consider option y. Additionally, since z appears after t in the list (because $z \rhd t$) and the agents choose z with a strictly positive probability in S, it must be the case that the agents consider both z and t in S. It follows that z must be preferred to t, hence, $\rhd$-Independence states that introducing an arbitrary alternative (x) to the choice set that, if considered, subsequently results

[11] To see this, suppose $x \trianglerighteq y \Leftrightarrow x \succsim y$. By transitivity, $x \trianglerighteq y$ and $y \succ z$ would imply $x \succsim y \succ z$. Then, w-MON implies $\pi(z, S) \leq \pi(z, S \setminus x)$.

in the consideration of preferred options (y, z) will not influence the choice probability of a dominated item (t). Intuitively, the choice probability of t is independent to the presence of x because the dominance of y and z absorbs any potential changes.

It is worth noting that when the list order corresponds to the preference order, $\rhd$-Independence follows directly from Independence in Axiom 2 because of the transitivity of the preference.[12] To see why $\rhd$-Independence is necessary for a $\text{GAM}_{\rhd}(\succ)$ representation, observe that there exists a type i such that $\max(\Gamma_i(S), \succ) = z$ because $\pi(z, S) > 0$. Among such i, there exists the biggest one i^*. Consider an arbitrary $j \leq i^*$. Suppose $x \in \Gamma_j(S)$. It follows $y \in \Gamma_j(S)$ because $x \unrhd y$. Therefore, $\max(\Gamma_j(S), \succ) \succsim y$. However, this could not happen as $y \succ z = \max(\Gamma_{i^*}(S), \succ) \succsim \max(\Gamma_j(S), \succ) \succsim y$, where $\max(\Gamma_{i^*}(S), \succ) \succsim \max(\Gamma_j(S), \succ)$ comes from the fact that $\Gamma_{i^*}(S)$ nests $\Gamma_j(S)$ since $i^* \geq j$. Hence, $x \notin \Gamma_j(S)$. Since Γ_j is an attention filter, it is the case that $\Gamma_j(S) = \Gamma_j(S \setminus x)$. Therefore, $\max(\Gamma_j(S), \succ) = \max(\Gamma_j(S \setminus x), \succ)$ for all $j \leq i^*$ and it follows $\pi(t, S) = \pi(t, S \setminus x)$ for all $t \in S$ such that $z \rhd t$.

Axiom 5 (Identity - IDE)**.** $x \rhd y$ and $y \succ x$ imply $\pi(x, S) = 0$ for all $S \supseteq \{x, y\}$.

Given a binary pair (x, y), the Identity axiom above says that if the preference and list orders are in conflict, the option appearing later in the list must not be chosen. The intuition behind Axiom is 5 simple. If x is considered, then so is y due to its earlier position in the list. Consequently, x is never selected because the agents prefer y to x. We now state the characterization result for $\text{GAM}_{\rhd}(\succ)$.

Theorem 3. RCF π has a $\text{GAM}_{\rhd}(\succ)$ representation if and only if it satisfies $\rhd$-wMON, $\rhd$-Independence, and IDE.

The proof of Theorem 3 closely parallels that of Theorem 1. Both proofs identify the best element within each type's consideration set. The primary difference between the two lies in the construction of consideration sets. The idea in the proof of Theorem 3 is as follows. Consider a choice set S and suppose a is the $\succ$-best element in the consideration set $\Gamma(S)$ we already identified. We enumerate all elements in S as $x_{|S|} \rhd x_{|S|-1} \rhd \cdots \rhd x_1$, with $|S|$ being the cardinality of S. We then define the

[12] Unlike $\rhd$-wMON that generalizes w-MON, $\rhd$-Independence does not constitute a generalization of Independence. However, $\rhd$-Independence coupled with the Identity in Axiom 5 generalize the Independence axiom.

consideration set as:

$$\Gamma(S) = \{x_1, x_2, \ldots, x_{i^*}\}, \quad \text{where } i^* \text{ is the largest integer such that } a \succsim x_t \, \forall t \leq i^*$$

The construction ensures that $\Gamma(S)$ has the list-based structure with respect to $\rhd$ and a is the $\succ$-best element in $\Gamma(S)$. In the proof of Theorem 3, we demonstrate that this construction is always feasible and results in consideration sets being attention filters and satisfying the expansion property.

Unknown Lists

Our result in Theorem 3 relies on the exogeneity of the list order. Such a situation is justified when the list is observed by an outside analyst (when the list corresponds to a Google search page, for example). In certain contexts, the list may not be discernible. Motivated by this observation, we demonstrate to what extent we can identify the list order from observed choice data. A stochastic choice function π is said to have a GAM($\succ$) representation with unknown lists if there exists a linear order $\rhd$ such that π has a $\text{GAM}_{\rhd}(\succ)$ representation where consideration sets have the list-based structure with respect to $\rhd$. Remark 1 below states our identification result for the unknown lists.

Remark 1. Suppose RCF π has a GAM($\succ$) representation with unknown lists. Then, x must appear after y (denoted as $x\mathbb{L}y$, with $\mathbb{L}$ being a binary relation for the list order) in any list representing π if there exists y such that $x \succ y$ and $\pi(y, S) > 0$ for some $S \supseteq \{x, y\}$.

Remark 1 directly follows from the IDE axiom introduced earlier (Axiom 5). The Identity axiom states that given any pair (x, y), if the list order and preference order contradict each other, the option appearing later in the list has a zero probability of being selected. Remark 1 operationalizes on this idea. In Remark 1, since $x \succ y$ and y is chosen with a strictly positive probability, the list order and preference order must agree on (x, y). Hence, x appears after y in the list.

Whenever the choice data has full support, Remark 1 helps us to uniquely identify the list order. Moreover, in this situation, the list order must coincide with the preference order. When π is not positive, we can partially identify the unknown lists. Any list orders representing π must be included in the transitive closure of the binary relation $\mathbb{L}$ defined in Remark 1.

7 An Application: Optimal List Design

In recent years, there has been a growing body of literature investigating the manipulation of online behaviors. Several studies have documented that online platforms such as Amazon, Walmart, Google Shopping, and Apple App Store can intentionally manipulate their offerings to favor their own products (Hagiu et al., 2022; Padilla et al., 2022; Farronato et al., 2023; Motta, 2023). This manipulation involves tactics such as strategically recommending their own products or prominently featuring their offerings. For example, on Amazon, Farronato et al. (2023) find that Amazon-labeled products consistently receive higher rankings in consumer search results compared to products that are observably comparable.

Motivated by this phenomenon, in our application, we assume that there is a designer who wants to construct a list in order to maximize an objective function.[13] For each item x in the list, let $w(x) \in \mathbb{R}^+$ be the weight that the designer assigns to x. One can interpret $w(x)$ as the designer's utility of x, or in the case of online shopping, the revenue that x brings to the platforms. We assume that the weights are pairwise distinct, that is, $w(x) \neq w(y)$ when $x \neq y$.[14] As before, let $\pi(x, L)$ be the probability that x is selected from a list L. For an arbitrary L, we assume the designer's objective function, denoted as $W(L)$, satisfies three conditions:

1. $W(L)$ depends on the weights of items and their probabilities of being selected.

2. An unchosen option will not impact $W(L)$.

3. For two lists L and L' that only differ in the chosen probabilities of x and y ($\pi(z, L) = \pi(z, L')$ for all $z \neq x, y$)

$$W(L) > W(L') \text{ if } w(x) > w(y) \text{ and } \pi(x, L) > \pi(x, L')$$

The first two conditions above are straightforward. The last condition indicates the

[13] Manzini et al. (2023) study a similar application in a completely different framework where the firm maximizes the approval ratings. In our application, the designer's objective functions and optimal list results are significantly different from theirs.

[14] We do not need this assumption in designing an optimal list. However, it allows us to characterize the set of all optimal lists fully.

monotonicity of the objective function. It simply states that the designer prefers the list where items with greater weights are selected with higher probabilities. The class of functions satisfying these three conditions includes the well-known family of functions with constant elasticity of substitution ($W(L) = (\sum_{x\in L} \pi(x,L)w(x)^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}$, where $\sigma > 1$ is the elasticity of substitution).[15] Before presenting the analysis, we formally define an optimal list.

Definition 4. The list is optimal if it solves $\max_{L\in\mathcal{L}} W(L)$, where $\mathcal{L}$ is the set of all possible lists.

In what follows, we denote a list L as $[x_1, x_2, \ldots, x_{|L|}]$ with x_j being the item in the *jth* position. We write $x_j L x_i$ if x_j appears after x_i in the list, i.e., when $j > i$. Following our GAM framework, suppose there are $|L|$ types of agents. For $i = 1, 2, \ldots, |L|$, agents type i consider the first i options in the list: $\Gamma_i(L) = \{x_1, x_2, \ldots, x_i\}$. All types have the same preference, denoted by $\succ$. The designer observes the preferences of agents, but she is unaware of the probability distribution among them.[16] Furthermore, suppose that the type distribution is positive, i.e., the probability of agents type i is nonzero for all i.[17] Given preference $\succ$ and weights w, we introduce the following simple *L-algorithm* to construct an optimal list.

[15]The expected utility ($W(L) = \sum_{x\in L} w(x)\pi(x,L)$) and Cobb-Douglas objective functions ($W(L) = \prod_{x\in L} w(x)^{\pi(x,L)}$) are special cases of this family when the $\sigma \to \infty$ and $\sigma \to 1$, respectively.

[16]When the preference is not unobserved, we will show that the top position in an optimal list is always identified as the item with the highest weight. Subsequent positions in the optimal list are generally ambiguous.

[17]This assumption, like the pairwise distinction of weights assumption, is unnecessary in designing an optimal list. However, it helps us to identify the set of all optimal lists.

L-algorithm

1. Start the list from the item with the highest weight. That is, $x_1 = \operatorname*{argmax}_{x \in L} w(x)$.

2. Let $L_{\succ}(x_1)$ be the lower contour set of x_1 given preference $\succ$. For the $2nd$, $3rd$, $4th$, $\ldots$, $(k+1)th$ positions in the list, where $k = |L_{\succ}(x_1)|$, randomly pick an element from $L_{\succ}(x_1)$ without replacement until the lower contour set is exhausted. Put differently, $(x_2, x_3, \ldots, x_{k+1})$ constitutes an arbitrary permutation of all elements in $L_{\succ}(x_1)$.

3. Choose $x_{k+2} = \operatorname*{argmax}_{x:\, x \notin L_{\succ}(x_1)} w(x)$, i.e., x_{k+2} is the option with the highest weight among those in the upper contour set of x_1.

4. Follow the same procedure as in step 2 by constructing the lower contour set of x_{k+2} and randomly placing its elements at positions after x_{k+2} until the set is exhausted.

5. Repeat steps 3-4 until every position in the list is occupied.

The *L-algorithm* starts by selecting the item with the highest weight and positions it at the top of the list. Subsequently, it constructs a lower contour set of the initially placed option, from which alternatives are randomly chosen without replacement and gradually added to the list. The process continues by identifying the item with the highest weight among those in the upper contour set of the last positioned item that is chosen with a strictly positive probability and assigning it to the subsequent available position in the list. This step is followed by the construction of the lower contour set for the newly positioned item, and from the set, once again, alternatives are randomly chosen without replacement and added to the list. This iterative process is repeated until all positions in the list are occupied.

Theorem 4 below shows that the set of lists generated by running the *L-algorithm* fully characterizes all optimal lists. This result holds for any positive type distributions and any preferences. Furthermore, it does not necessitate a specific functional form for the objective function and also applies to incomplete lists (when the lists only include a subset of all available options).

Theorem 4. The list is optimal if and only if it results from running the *L-algorithm*.

Theorem 4 implies that the optimal list order does not necessarily agree with the weight ordering. Consequently, the designer might not place their preferred items at the top positions of the list. While the first item is always the one she prefers the most (because of step 1 in the *L-algorithm*), the designer can allocate subsequent positions to options with relatively low weights. These items share common attributes of being unattractive and inferior compared to the top-ranked item. Intuitively, such an order establishes incentives for the agents to choose the top-ranked item, thereby maximizing the designer's utility. To illustrate, suppose the list has six items labeled as $z_1, z_2, \ldots, z_6$. Suppose the weight is $w(z_i) = i$ for all i. The agent's preference is $z_4 \succ z_5 \succ z_1 \succ z_6 \succ z_2 \succ z_3$. The *L-algorithm* identifies two optimal lists: $[z_6, z_2, z_3, z_5, z_1, z_4]$ and $[z_6, z_3, z_2, z_5, z_1, z_4]$. In both lists, items in the $2nd$ and $3rd$ positions have little value to the designer.

The proof of Theorem 4 relies on the following result. Consider an arbitrary list L and a list L' obtained from L by switching the positions of x_i and x_{i+k} while keeping the positions of all other items unchanged. Then the consideration sets of agents type j, with $j \geq i+k$ or $j \leq i-1$, remain the same under L and L'. Hence, the difference in the designer's objective function under L and L' depends entirely on the behaviors of agents type j, with $i \leq j \leq k-1$.

For a given list L, let $\{x_{a_1}, x_{a_2}, \ldots, x_{a_k}\}$ be an ordered set of distinct options that are chosen with strictly positive probabilities under L with $x_{a_k} \succ x_{a_{k-1}} \succ \cdots \succ x_{a_1} = x_1$. This set is important in designing the optimal list because it impacts the designer's utility. The proof of Theorem 4 includes four Observations 1-4 that explore different characteristics of the set $\{x_{a_1}, x_{a_2}, \ldots, x_{a_k}\}$.

Observation 1 below says that when x and y are selected with strictly positive probabilities under the optimal list, x appears before y if and only if the designer's weight of x is higher. This is because if the weight of y is higher and y appears later in the list, the designer is strictly better off by switching the positions of x and y while keeping the positions of all other items unchanged. Observation 1 also implies that the optimal list order must align with the weight ordering on the portion of the list where items are chosen with strictly positive probabilities.

Observation 1. Suppose list L is optimal. Then $w(x_{a_{i+1}}) < w(x_{a_i})$ for all $i = 1, 2, \ldots, k-1$.

Observation 2 below claims that the first item in every optimal list must be the one with the highest weight. This directly verifies step 1 in the *L-algorithm*. Notably, this result holds under every preference. Hence, the top position in the list must have the highest weight regardless of the observability of the agent's preference. The rationale behind Observation 2 is that if the highest-weight item is not placed at the top position, the designer has a strictly higher utility by moving it to the first place while keeping the relative positions of all other items unchanged.

Observation 2. Let $x^* = \arg\max_{x \in L} w(x)$. Then $x_1 = x^*$ in every optimal list.

Observation 3 below establishes that any options in the lower contour set of x_{a_i} must precede $x_{a_{i+1}}$ in the optimal list. It directly verifies step 2 in the *L-algorithm*. The rationale behind Observation 3 is that if there exists an option in the lower contour set of x_{a_i} positioned after $x_{a_{i+1}}$ in the list, the designer is strictly better off moving it to a position before $x_{a_{i+1}}$.

Observation 3. Suppose list L is optimal. Suppose that $x_{a_i} \succ y$ for some $i \leq k-1$. Then $x_{a_{i+1}} L y$.

Observation 4 below states that x_{a_t} must be the one with the highest weight among those in the upper contour set of $x_{a_{t-1}}$. Put differently, any option with a strictly positive probability of being chosen must have the highest weight among those in the upper contour set of the previous option that is selected with a strictly positive probability.

Observation 4. Suppose list L is optimal. Then $x_{a_t} = \underset{x:\, x \in L \text{ s.t. } x \succ x_{a_{t-1}}}{\operatorname{argmax}} w(x)$ for all $t = 2, 3, \ldots, k$.

Together, Observation 3-4 confirm steps 3-5 in the *L-algorithm*. Hence, Observations 1-4 verify the optimality of any list generated from running the *L-algorithm*.

In summary, under the self-preferencing idea where the designer can manipulate the list to maximizer her utility, the optimal list ordering depends on both the agents' preferences and the designer's priority ranking. Moreover, the designer can strategically position items of little value to her near the top of the list. This observation is crucial in understanding the designer's behaviors and identifying list manipulation. It indicates that the presence of low-value options (to the designer) near the top of the list does not automatically eliminate the potential issue of information manipulation.

8 Literature Review

Our paper builds upon the existing literature on consideration sets that has been studied in marketing and economics (Wright and Barbour, 1977; Hauser and Wernerfelt, 1990; Masatlioglu et al., 2012; Hauser, 2014; Caplin et al., 2019). Firstly, our research relates to a class of theoretical papers that studies probabilistic choices with an ordered collection of heterogeneous types. Within the framework of RUM, Apesteguia et al. (2017) introduce the single-crossing random utility model (SCRUM), where the collection of preferences satisfies the well-known single-crossing property. As shown in Section 2, our model nests SCRUM as a particular case. Petri (2023) also studies SCRUM but confines it to binary choices. Apesteguia and Ballester (2023) characterize RUM with ordered menus and constrained domains, accommodating limited data in empirical contexts. Outside the framework of RUM, Filiz-Ozbay and Masatlioglu (2023) introduce the progressive random choice (PRC) model. PRC has an ordered collection of choice functions instead of preferences and hence can accommodate different types of bounded rationality. Our paper differs from these studies as we refrain from directly imposing a structure on the collection of types. Instead, we establish a structure through conditions applied to the underlying consideration sets.

Secondly, our work is also related to the literature addressing the identification of consideration sets. Theoretically, Manzini and Mariotti (2014) demonstrate that the preference relation and alternative-specific consideration probabilities can be uniquely inferred from choice data. Cattaneo et al. (2020) investigate a random attention model and provide partial identification results. Empirically, Abaluck and Adams-Prassl (2021) show that consideration sets can be identified by exploiting the variations in asymmetric demand responses to other product characteristics. Barseghyan et al. (2021) explore decision-making under risk in cases where the collection of preferences satisfies the single-crossing property, demonstrating that identification is feasible in most scenarios. Finally, Dardanoni et al. (2020) develop a framework for identifying the distribution of cognitive characteristics from aggregate choice shares with minimal data, accounting for the heterogeneity of both attention capacities and preferences.

Comparison to related models

As mentioned previously, the SCRUM of Apesteguia et al. (2017) is a specific case within the GAM framework. Since any stochastic choice can be represented using the PRC model proposed by Filiz-Ozbay and Masatlioglu (2023), GAM is nested in PRC. Filiz-Ozbay and Masatlioglu (2023) also characterize a special case of PRC, called less-is-more PRC (L-PRC), where welfare is improved when having fewer options (choice overload phenomenon). L-PRC and GAM are independent. L-PRC requires the monotonicity condition to hold on upper contour sets.[18] Consequently, it requires the choice frequencies of the worst alternative in a set to (weakly) violate the monotonicity condition. GAM, on the other hand, does not allow any strict regularity violations because choice probabilities must adhere to the w-MON axiom.[19]

Cattaneo et al. (2020) propose a random attention model (RAM) of stochastic choice where the randomness of choices comes from the random consideration of the decision-makers. The revealed preference in RAM (Lemma 1 in their paper) relies on regularity violations and is identical to the revealed preference in Proposition 1 in our paper. Since RAM is exclusively characterized by the revealed preference, endogenous GAM is nested in RAM. Having said that, in our model, we can identify preferences even in the absence of regularity violations (Propositions 2-3). Moreover, our model allows for the unique identification of consideration sets. These additional results are not featured in RAM.

Some other well-known models are also included in RAM, including the random utility model (RUM), Manzini and Mariotti (2014) (MM), and Brady and Rehbeck (2016) (BR). First, GAM is distinct from BR because there is a default option in BR, and it is unclear how to remove the default option from the model.[20] Second,

[18] That is, for all $x \in T \subset S \subseteq X$ such that $\pi(x, S) \neq 0$, $\pi(U(x, S), S) \leq \pi(U(x, T), T)$.

[19] It is noteworthy that the intersection of GAM and L-PRC nests a full-support SCRUM, where the choice probabilities of the worst alternative remain unchanged. Specifically, under a SCRUM with full support, if z is the worst alternative in set S, then it holds that $\pi(z, T) = \pi(z, S)$ for all $T \subseteq S$ with $z \in T$.

[20] In their words: "Whether the default option can be removed... remains an open question."

RUM and MM require the classic regularity condition to hold. Hence, GAM is also independent of both RUM and MM since the endogenous GAM can accommodate regularity violations. In this regard, GAM is also distinct from a class of models that satisfies regularity, such as the additive perturbed utility model of Fudenberg et al. (2015), fixed distribution satisficing model of Aguiar et al. (2016), or attribute rule model of Gul et al. (2014).

9 Conclusion

We have introduced the Growing Attention Model to capture the idea that the decision maker's attention can be sorted in an ascending order. Our main theoretical contribution lies in the development of a framework that facilitates the unique identification of both the contents and frequencies of consideration sets, given the unobserved heterogeneity of attention. Even when the preference is not observed, we demonstrate that full identification can still be achieved under certain simple conditions. Such identification of consideration sets holds significant value for managerial decision-making and enables comparisons of attentiveness levels associated with two probabilistic choices. Our work can be extended in several directions. Firstly, exploring alternative structures and properties of consideration sets, such as the dynamics of attention, may yield fresh insights into individual decision-making processes. Secondly, extending our model to empirical settings holds promise for practical applications.

References

Abaluck, J. and A. Adams-Prassl (2021): "What do consumers consider before they choose? Identification from asymmetric demand responses," *The Quarterly Journal of Economics*, 136, 1611–1663.

Aguiar, V. H., M. J. Boccardi, and M. Dean (2016): "Satisficing and Stochastic Choice," *Journal of Economic Theory*, 166, 445–482.

Apesteguia, J. and M. A. Ballester (2023): "Random utility models with ordered types and domains," *Journal of Economic Theory*, 105674.

Apesteguia, J., M. A. Ballester, and J. Lu (2017): "Single Crossing Random Utility Models," *Econometrica*, 85, 661–674.

Barseghyan, L., F. Molinari, and M. Thirkettle (2021): "Discrete choice under risk with limited consideration," *American Economic Review*, 111, 1972–2006.

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

Broadbent, D. E. (1958): *Perception and communication*, Pergamon Press, New York.

Caplin, A., M. Dean, and J. Leahy (2019): "Rational inattention, optimal consideration sets, and stochastic choice," *The Review of Economic Studies*, 86, 1061–1094.

Cattaneo, M. D., P. Cheung, X. Ma, and Y. Masatlioglu (2021): "Attention Overload," .

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

Chen, N. and H.-T. T. Tsai (2023): "Steering via algorithmic recommendations," *RAND Journal of Economics, forthcoming.*

Cherepanov, V., T. Feddersen, and A. Sandroni (2013): "Rationalization," *Theoretical Economics*, 8, 775–800.

Chipman, J. S. (1958): "Stochastic choice and subjective probability," *Econometrica*, 26, 63–63.

Dardanoni, V., P. Manzini, M. Mariotti, and C. J. Tyson (2020): "Inferring cognitive heterogeneity from aggregate choices," *Econometrica*, 88, 1269–1296.

Dorrian, J., N. L. Rogers, and D. F. Dinges (2004): "Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss," *Sleep deprivation*, 39–70.

Durmer, J. S. and D. F. Dinges (2005): "Neurocognitive consequences of sleep deprivation," in *Seminars in neurology*, vol. 25, 117–129.

Eliaz, K. and R. Spiegler (2011): "Consideration sets and competitive marketing," *The Review of Economic Studies*, 78, 235–262.

Farronato, C., A. Fradkin, and A. MacKay (2023): "Self-preferencing at Amazon: evidence from search rankings," *AEA Papers and Proceedings*, 113, 239–43.

Fershtman, D. and A. Pavan (2023): "Searching for arms: Experimentation with endogenous consideration sets," *Working paper.*

Filiz-Ozbay, E. and Y. Masatlioglu (2023): "Progressive random choice," *Journal of Political Economy*, 131, 716–750.

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

Georgescu-Roegen, N. (1958): "Threshold in Choice and the Theory of Demand," *Econometrica: Journal of the Econometric Society*, 157–168.

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

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

Hagiu, A., T.-H. Teh, and J. Wright (2022): "Should platforms be allowed to sell on their own marketplaces?" *The RAND Journal of Economics*, 53, 297–327.

Hauser, J. R. (2014): "Consideration-set heuristics," *Journal of Business Research*, 67, 1688–1699.

Hauser, J. R. and B. Wernerfelt (1990): "An evaluation cost model of consideration sets," *Journal of consumer research*, 16, 393–408.

Honka, E. (2014): "Quantifying search and switching costs in the US auto insurance industry," *The RAND Journal of Economics*, 45, 847–884.

Lea, A. M. and M. J. Ryan (2015): "Irrationality in mate choice revealed by túngara frogs," *Science*, 349, 964–966.

Lindman, H. R. (1971): "Inconsistent preferences among gambles." *Journal of Experimental Psychology*, 89, 390.

Manzini, P. and M. Mariotti (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82, 1153–1176.

——— (2018): "Dual random utility maximisation," *Journal of Economic Theory*, 177, 162–182.

Manzini, P., M. Mariotti, and L. Ulku (2023): "A model of approval with an application to list design," *Working Paper*.

Masatlioglu, Y. and D. Nakajima (2013): "Choice by Iterative Search," *Theoretical Economics*, 8, 701–728.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012): "Revealed Attention," *American Economic Review*, 102, 2183–2205.

Mellers, B. A., S.-j. Chang, M. H. Birnbaum, and L. D. Ordonez (1992): "Preferences, prices, and ratings in risky decision making." *Journal of Experimental Psychology: Human Perception and Performance*, 18, 347.

Motta, M. (2023): "Self-preferencing and foreclosure in digital markets: theories of harm for abuse cases," *International Journal of Industrial Organization*, 102974.

Padilla, J., J. Perkins, and S. Piccolo (2022): "Self-Preferencing in Markets with Vertically Integrated Gatekeeper Platforms," *The Journal of Industrial Economics*, 70, 371–395.

Petri, H. (2023): "Binary single-crossing random utility models," *Games and Economic Behavior*, 138, 311–320.

Rivalan, M., Y. Winter, and V. Nachev (2017): "Principles of economic rationality in mice," *Scientific Reports*, 7, 17441.

Shapiro, S., D. J. MacInnis, and S. E. Heckler (1997): "The effects of incidental ad exposure on the formation of consideration sets," *Journal of consumer research*, 24, 94–104.

Simon, H. A. (1955): "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69, 99–118.

Soltani, A., B. De Martino, and C. Camerer (2012): "A range-normalization model of context-dependent choice: a new model and evidence," *PLoS computational biology*, 8, e1002607.

Tversky, A. and J. E. Russo (1969): "Substitutability and similarity in binary choices," *Journal of Mathematical psychology*, 6, 1–12.

Urban, G. L. (1975): "PERCEPTOR: A model for product positioning," *Management Science*, 21, 858–871.

Wright, P. and F. Barbour (1977): "Phased decision strategies: Sequels to an initial screening," in *Studies in Management Sciences, Multiple Criteria Decision Making*, ed. by M. Starr and M. Zeleny, North-Holland, Amsterdam, 91–109.

Yegane, E. (2022): "Stochastic choice with limited memory," *Journal of Economic Theory*, 205, 105548.

APPENDIX

Appendix A Omitted Proofs

Proof of Theorem 1. The necessity part is shown in the paper. We prove the sufficiency. The proof uses a construction technique in Filiz-Ozbay and Masatlioglu (2023). For a given linear order $\rhd$ and an arbitrary RCF π, Filiz-Ozbay and Masatlioglu (2023) constructively prove that there exist a unique collection of choice functions $\mathbb{C} = \{c_1, c_2, \ldots, c_K\}$ satisfying a *progressive* property with respect to $\rhd$,[21] and a probability distribution μ over $\mathbb{C}$ such that $\pi(x, S) = \sum_{i:\, c_i(S)=x} \mu(c_i)$ for all $x \in S$ and $S \in \mathcal{X}$. We briefly describe the construction here for convenience. Define

$$K = \{\pi(L_{\trianglerighteq}(x), S) | S \subseteq X \text{ and } x \in S\}, \quad \text{where } L_{\trianglerighteq}(x) = \{y : x \trianglerighteq y\}$$

This defines a collection of all cumulative probabilities on lower-contour sets derived from the probabilistic choice. Clearly, K is a finite subset of $[0, 1]$. Next, we sort the strictly positive elements in K from the lowest to the highest, that is, $0 < k_1 < k_2 < .. < k_m = 1$. Define the following lower-contour set operator $L^+_{\rhd}(x, S)$, which yields the set of all alternatives that are $\rhd$-worse than x and chosen with a strictly positive probability in S. That is,

$$L^+_{\rhd}(x, S) = \{y \in S | \pi(y, S) > 0 \text{ and } x \rhd y\}$$

Also, define $L^+_{\trianglerighteq}(x, S) = L^+_{\rhd}(x, S) \cup \{x\}$. For any choice set S, follow the steps below.

STEP 1. Define $c_1(S)$ as the $\rhd$-worst alternative in S with a strictly positive probability of being chosen and $\mu(c_1) = k_1$. That is, $\pi(c_1(S), S) > 0$ and $L^+_{\rhd}(c_1(S), S) = \emptyset$.

STEP i with $2 \leq i \leq m$. Define the *ith* choice function as follows.

$$c_i(S) = \begin{cases} c_{i-1}(S) & \text{if } \pi(L_{\trianglerighteq}(c_{i-1}(S)), S) > k_{i-1}, \\ x & \text{if } \pi(L_{\trianglerighteq}(c_{i-1}(S)), S) = k_{i-1}, \end{cases} \text{ and } \mu(c_i) = k_i - k_{i-1}$$

where $x \in S$ satisfies $\pi(x, S) > 0$ and $L^+_{\rhd}(x, S) \setminus L^+_{\rhd}(c_{i-1}(S), S) = c_{i-1}(S)$. That is, if

[21] c_i is a choice function if $c_i : \mathcal{X} \to X$ and $c_i(S) \in S$ for all $S \in \mathcal{X}$. The collection of choice functions $\mathbb{C} = \{c_1, c_2, \ldots, c_K\}$ satisfies the *progressive* property with respect to $\rhd$ if $c_i(S) \trianglerighteq c_j(S)$ for all $S \in \mathcal{X}$ and $i \geq j$. Here, the linear order $\rhd$ can be arbitrary.

$c_i(S) \neq c_{i-1}(S)$ then $c_i(S)$ is the next alternative in S that is chosen with a strictly positive probability ($c_i(S)$ is the successor of $c_{i-1}(S)$). Filiz-Ozbay and Masatlioglu (2023) call $\mathbb{C} = \{c_1, c_2, \ldots, c_m\}$ a PRC($\triangleright$) representation of π. We first prove the following Lemma.

Lemma 1. Let $\{c_1, c_2, \ldots, c_m\}$ be the PRC($\triangleright$) representation of π and μ the associated probability measure. Fix a choice set S and a type $i \geq 1$. Consider $x \in S$ such that $x \triangleright c_i(S)$. If the two following conditions are satisfied

$(I_1) : \pi(c_i(S), S) \leq \pi(c_i(S), S \setminus x)$ and;

$(I_2) : \pi(t, S) = \pi(t, S \setminus x)$ for all $t \in S$ such that $c_i(S) \triangleright t$,

then $c_j(S) = c_j(S \setminus x)$ for all $j = 1, 2, \ldots, i$.

Proof. Proof by induction. First, we show that $c_1(S) = c_1(S \setminus x)$. Proof by contradiction. Suppose $c_1(S) \neq c_1(S \setminus x)$. Observe that $0 < \pi(c_1(S), S)$ because of the construction of choice functions. Also, $\pi(c_1(S), S) \leq \pi(c_1(S), S \setminus x)$ because of the two conditions on π (note that $c_i(S) \trianglerighteq c_1(S)$ by construction). Hence, $0 < \pi(c_1(S), S \setminus x)$. By definition, $c_1(S \setminus x)$ is the $\triangleright$-worst alternative among those with a strictly positive probability of being chosen when the menu is $S \setminus x$. The inequality $0 < \pi(c_1(S), S \setminus x)$ then implies $c_1(S) \triangleright c_1(S \setminus x)$ (because $c_1(S) \neq c_1(S \setminus x)$). With $c_1(S) \triangleright c_1(S \setminus x)$, it is the case that $c_i(S) \trianglerighteq c_1(S) \triangleright c_1(S \setminus x)$. Hence, condition I_2 is applicable and we have $\pi(c_1(S \setminus x), S) = \pi(c_1(S \setminus x), S \setminus x)$. By definition of $c_1(S \setminus x), \pi(c_1(S \setminus x), S \setminus x) > 0$. This implies $\pi(c_1(S \setminus x), S) > 0$. By definition, $c_1(S \setminus x)$ is the $\triangleright$-worst alternative among those with a strictly positive probability of being chosen when the menu is S. The inequality $0 < \pi(c_1(S \setminus x), S)$ then implies $c_1(S \setminus x) \triangleright c_1(S)$. So, we have $c_1(S \setminus x) \triangleright c_1(S) \triangleright c_1(S \setminus x)$, which is a contradiction. Hence, the initial assumption is wrong and $c_1(S) = c_1(S \setminus x)$.

Second, suppose that $c_t(S) = c_t(S \setminus x)$ for all $t = 1, 2, .., \ldots, j-1$. We will show that $c_j(S) = c_j(S \setminus x)$ (here $j \leq i$). If $c_j(S) = c_{j-1}(S)$ and $c_j(S \setminus x) = c_{j-1}(S \setminus x)$ then we are done. Therefore, it is sufficient to consider the following cases.

Case 1: $c_j(S) \neq c_{j-1}(S)$ **and** $c_j(S \setminus x) = c_{j-1}(S \setminus x)$. By the construction of choice functions, $c_j(S) \neq c_{j-1}(S)$ implies $c_j(S) \triangleright c_{j-1}(S)$. By transitivity, $c_i(S) \trianglerighteq c_j(S)$ (because $i \geq j$) and $c_j(S) \triangleright c_{j-1}(S) = c_{j-1}(S \setminus x)$ imply $c_i(S) \triangleright c_{j-1}(S \setminus x)$. Hence, condition I_2 is applicable and it is the case that $\pi(c_{j-1}(S \setminus x), S) = \pi(c_{j-1}(S \setminus x), S \setminus x)$.

However, it cannot happen as

$$\begin{aligned}
\pi(c_{j-1}(S \setminus x), S \setminus x) \geq \mu(c_j) + \sum_{\substack{k=1 \\ k:\, c_k(S\setminus x)=c_{j-1}(S\setminus x)}}^{j-1} \mu(c_k) &= \mu(c_j) + \sum_{\substack{k=1 \\ k:\, c_k(S)=c_{j-1}(S\setminus x)}}^{j-1} \mu(c_k) \\
&= \mu(c_j) + \pi(c_{j-1}(S \setminus x), S) \\
&> \pi(c_{j-1}(S \setminus x), S)
\end{aligned}$$

The inequality in the first line comes from the fact that $c_{j-1}(S \setminus x) = c_j(S \setminus x)$. The equation in the first line results from our assumption that $c_k(S) = c_k(S \setminus x)$ for all $k = 1, 2, \ldots, j-1$. The equation in the second line holds because $\nexists k \geq j$ such that $c_k(S) = c_{j-1}(S \setminus x)$ as $c_j(S) \rhd c_{j-1}(S \setminus x)$. The last inequality uses $\mu(c_j) > 0$.

<u>**Case 2:** $c_j(S) = c_{j-1}(S)$ **and** $c_j(S \setminus x) \neq c_{j-1}(S \setminus x)$.</u> By the construction of choice functions, $c_j(S \setminus x) \neq c_{j-1}(S \setminus x)$ implies $c_j(S \setminus x) \rhd c_{j-1}(S \setminus x)$. Since $i \geq j$, we have $c_i(S) \trianglerighteq c_j(S) = c_{j-1}(S) = c_{j-1}(S \setminus x)$. By conditions I_1 and I_2, it is the case that $\pi(c_{j-1}(S \setminus x), S) \leq \pi(c_{j-1}(S \setminus x), S \setminus x)$. However, it cannot happen as

$$\begin{aligned}
\pi(c_{j-1}(S \setminus x), S) \geq \mu(c_j) + \sum_{\substack{k=1 \\ k:\, c_k(S)=c_{j-1}(S\setminus x)}}^{j-1} \mu(c_k) &= \mu(c_j) + \sum_{\substack{k=1 \\ k:\, c_k(S\setminus x)=c_{j-1}(S\setminus x)}}^{j-1} \mu(c_k) \\
&= \mu(c_j) + \pi(c_{j-1}(S \setminus x), S \setminus x) \\
&> \pi(c_{j-1}(S \setminus x), S \setminus x)
\end{aligned}$$

The inequality in the first line comes from the fact that $c_{j-1}(S \setminus x) = c_{j-1}(S) = c_j(S)$. The equation in the first line results from our assumption that $c_k(S) = c_k(S \setminus x)$ for all $k = 1, 2, \ldots, j-1$. The equation in the second line holds because $\nexists k \geq j$ such that $c_k(S \setminus x) = c_{j-1}(S \setminus x)$ as $c_j(S \setminus x) \rhd c_{j-1}(S \setminus x)$. The last inequality uses $\mu(c_j) > 0$.

<u>**Case 3:** $c_j(S) \neq c_{j-1}(S)$ **and** $c_j(S \setminus x) \neq c_{j-1}(S \setminus x)$.</u> By construction of the choice functions, $c_j(S) \rhd c_{j-1}(S)$ and $c_j(S \setminus x) \rhd c_{j-1}(S \setminus x)$. Proof by construction. Suppose $c_j(S) \neq c_j(S \setminus x)$.

- Suppose $c_j(S) \rhd c_j(S \setminus x)$. By transitivity, $c_i(S) \trianglerighteq c_j(S)$ (because $j \geq i$) and $c_j(S) \rhd c_j(S \setminus x)$ imply $c_i(S) \rhd c_j(S \setminus x)$. Hence, condition I_2 is applicable and thus $\pi(c_j(S \setminus x), S) = \pi(c_j(S \setminus x), S \setminus x) > 0$. The last inequality comes from the definition of $c_j(S \setminus x)$. Therefore,

$$c_j(S) \rhd c_j(S \setminus x) \rhd c_{j-1}(S \setminus x) = c_{j-1}(S) \text{ and } \pi(c_j(S \setminus x), S) > 0$$

This is a contradiction to the definition of $c_j(S)$ because there is another alternative, $c_j(S \setminus x)$, that is chosen with a strictly positive probability in S but ranked between $c_j(S)$ and $c_{j-1}(S)$.

- Suppose $c_j(S \setminus x) \rhd c_j(S)$. Note that $c_i(S) \trianglerighteq c_j(S)$ so conditions I_1 and I_2 imply that $\pi(c_j(S), S \setminus x) \geq \pi(c_j(S), S) > 0$, where the last inequality comes from the definition of $c_j(S)$. Therefore,

$$c_j(S \setminus x) \rhd c_j(S) \rhd c_{j-1}(S) = c_{j-1}(S \setminus x) \text{ and } \pi(c_j(S), S \setminus x) > 0$$

This is a contradiction to the definition of $c_j(S \setminus x)$ because there is another alternative, $c_j(S)$, that is chosen with a strictly positive probability in $S \setminus x$ but ranked between $c_j(S \setminus x)$ and $c_{j-1}(S \setminus x)$.

Hence, the initial assumption is wrong and it is the case that $c_j(S) = c_j(S \setminus x)$. This completes our proof of the Lemma. □

Next, we prove the Theorem. Suppose $\{c_1, c_2, \ldots, c_m\}$ is the collection of choice functions from the PRC($\succ$) representation of π. Suppose the probability distribution over the choice collection is $\{\mu_1, \mu_2, \ldots, \mu_m\}$. This implies $c_i(S) \succsim c_j(S)$ when $i \geq j$ and $\pi(x, S) = \sum_{i \colon c_i(S) = x} \mu_i$ for all $x \in S$ and $S \in \mathcal{X}$. Define consideration sets as follows

$$\Gamma_i(S) = \{x \in S : c_i(S) \succsim x\} \text{ for all } i \text{ and } S \in \mathcal{X}$$

and the probability measure is $\mu(\Gamma_i) = \mu_i$. First, $\Gamma_i(S) \supseteq \Gamma_j(S)$ when $i \geq j$ because $c_i(S) \succsim c_j(S)$. Hence, the collection of consideration sets $\{\Gamma_1, \Gamma_2, \ldots, \Gamma_m\}$ is growing. Also, $c_i(S) = \max(\Gamma_i(S), \succ)$ by construction. Therefore,

$$\pi(x, S) = \sum_{i \colon c_i(S) = x} \mu_i = \sum_{i \colon \max(\Gamma_i(S), \succ) = x} \mu(\Gamma_i) \quad \forall x \in S \text{ and } S \in \mathcal{X}$$

The last step is to show that Γ_i is an attention filter. Take an arbitrary $x \in S$

but $x \notin \Gamma_i(S)$. It follows $x \succ c_i(S)$ by the construction of Γ_i. We will show that $\Gamma_i(S) = \Gamma_i(S \setminus x)$. The idea is to use Lemma 1 for preference $\succ$. Applying w-MON for $x \succ c_i(S)$, we have $\pi(c_i(S), S) \leq \pi(c_i(S), S \setminus x)$ so condition I_1 in Lemma 1 is satisfied. Consider an arbitrary z such that $c_i(S) \succ z$. Applying Independence for $x \succ c_i(S) \succ z$ and $\pi(c_i(S), S) > 0$, we have $\pi(z, S) = \pi(z, S \setminus x)$ so condition I_2 in Lemma 1 is satisfied. Therefore, Lemma 1 is applicable and it is the case that $c_j(S) = c_j(S \setminus x)$ for all $j = 1, 2, \ldots, i$. By construction, $\Gamma_i(S) = \Gamma_i(S \setminus x)$ and it follows that Γ_i is an attention filter. This completes our proof of the Theorem. ■

Proof of Theorem 2. First, we show that $\Gamma_i(S) = \{x : x \in S \text{ and } \max(\Gamma_i(S), \succ) \succsim x\}$ for all $i \in \{1, 2, .., m\}$ and $S \in \mathcal{X}$. Proof by contradiction. Suppose there exists (x, i, S) such that $x \in S$ and $\max(\Gamma_i(S), \succ) \succsim x$ but $x \notin \Gamma_i(S)$. Clearly, $x \neq \max(\Gamma_i(S), \succ)$ so $\max(\Gamma_i(S), \succ) \succ x$. Because the collection of attention filters is growing, $x \notin \Gamma_i(S)$ implies that $x \notin \Gamma_k(S)$ for all $k = 1, 2, \ldots, i-1$. Therefore, $x \neq \max(\Gamma_k(S), \succ)$ for all $k = 1, 2, \ldots, i-1$. Note that $\max(\Gamma_k(S), \succ) \succsim \max(\Gamma_i(S), \succ)$ for all $k \geq i$ because of the growing attention property. By transitivity, $\max(\Gamma_k(S), \succ) \succsim \max(\Gamma_i(S), \succ)$ and $\max(\Gamma_i(S), \succ) \succ x$ imply $\max(\Gamma_k(S), \succ) \succ x$ for all $k \geq i$. Hence, $x \neq \max(\Gamma_k(S), \succ)$ for all $k = i, i+1, \ldots, m$. It follows that $x \neq \max(\Gamma_k(S), \succ)$ for all k. Hence, x is not chosen at S, which contradicts the assumption that π has full support.

Second, suppose (Γ, μ) represents π with $\Gamma = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_m\}$ containing distinct attention filters. Suppose $\exists i \neq j$ such that $\max(\Gamma_i(S), \succ) = \max(\Gamma_j(S), \succ)$ for all $S \in \mathcal{X}$. It follows from the first part of the proof that $\Gamma_i(S) = \Gamma_j(S)$ for all $S \in \mathcal{X}$. Hence, $\Gamma_i = \Gamma_j$, which is a contradiction. Therefore, for all $i \neq j$, there exists $S \in \mathcal{X}$ such that $\max(\Gamma_i(S), \succ) \neq \max(\Gamma_j(S), \succ)$. Define the following choice functions: $c_i(S) = \max(\Gamma_i(S), \succ)$ for all i and $S \in \mathcal{X}$. Then $c_i(S) \succsim c_j(S)$ for all S and $i \geq j$ because of the expansion property. Also, $c_i(S) \neq c_j(S)$ for some $S \in \mathcal{X}$. Additionally,

$$\pi(x, S) = \sum_{i:\, \max(\Gamma_i(S), \succ)} \mu_i = \sum_{i:\, x = c_i(S)} \mu_i \text{ for all } x \in S \text{ and } S \in \mathcal{X}$$

Therefore, $\{c_1, c_2, \ldots, c_m\}$ is a PRC($\succ$) representation of π. Because the PRC($\succ$) representation is uniquely identified following Filiz-Ozbay and Masatlioglu (2023) (we described the construction in the proof of Theorem 1), it follows that (Γ, μ) is unique and $\max(\Gamma_i(S), \succ)$ is identified. This completes our proof. ■

Proof of Corollary 1. Suppose $\pi = (\Gamma, \mu, \succ)$ and $\pi' = (\Gamma', \mu', \succ')$. Both RCFs are positive, so it follows from the proof of Theorem 2 that $\Gamma_i(S) = \{x \in S : \max(\Gamma_i(S), \succ) \succsim x\}$ and $\Gamma'_j(S) = \{y \in S : \max(\Gamma'_j(S), \succ') \succsim' y\}$. Enumerate elements in S as $a_{|S|} \succ a_{|S|-1} \succ \cdots \succ a_1$ and $b_{|S|} \succ' b_{|S|-1} \succ' \cdots \succ' b_1$. Then, $\pi(U_{\succsim}(a_m, S), S) = \sum_{i:\, \max(\Gamma_i(S), \succ) \succsim a_m} \mu_i = \sum_{i:\, |\Gamma_i(S)| \geq m} \mu_i$, where the second equation comes from the fact that $\Gamma_i(S) = \{x \in S : \max(\Gamma_i(S), \succ) \succsim x\}$. Similarly,

$$\pi'(U_{\succsim'}(b_m, S), S) = \sum_{j:\, \max(\Gamma'_j(S), \succ') \succsim' b_m} \mu'_j = \sum_{j:\, |\Gamma'_j(S)| \geq m} \mu'_j$$

By definition, π FOSD π' if we have $\pi(U_{\succsim}(a_m, S), S) \geq \pi'(U_{\succsim'}(b_m, S), S)$ for all $S \in \mathcal{X}$ and $m \in \{1, 2, \ldots, |S|\}$. This inequality is equivalent to $\sum_{i:\, |\Gamma_i(S)| \geq m} \mu_i \geq \sum_{j:\, |\Gamma'_j(S)| \geq m} \mu'_j$ for all $S \in \mathcal{X}$ and $m \in \{1, 2, \ldots, |S|\}$. Equivalently, π is more attentive than π'. This completes our proof. ■

Proof of Corollary 2. Suppose $\pi = (\Gamma, \mu, \succ)$ and $\pi' = (\Gamma', \mu', \succ')$. Enumerate elements in S as $a_{|S|} \succ a_{|S|-1} \succ \cdots \succ a_1$ and and $b_{|S|} \succ' b_{|S|-1} \succ' \cdots \succ' b_1$. As in the proof of Corollary 1, we have $\pi(U_{\succsim}(a_m, S), S) = \sum_{i:\, \max(\Gamma_i(S), \succ) \succsim a_m} \mu_i = \sum_{i:\, |\Gamma_i(S)| \geq m} \mu_i$ because π has full support. Also, $\pi'(U_{\succsim'}(b_m, S), S) = \sum_{j:\, \max(\Gamma'_j(S), \succ') \succsim' b_m} \mu'_j$. Because π FOSD π', we have $\pi(U_{\succsim}(a_m, S), S) \geq \pi'(U_{\succsim'}(b_m, S), S)$ for all $S \in \mathcal{X}$ and $m \in \{1, 2, \ldots, |S|\}$. Equivalently,

$$\sum_{i:\, |\Gamma_i(S)| \geq m} \mu_i \geq \sum_{j:\, \max(\Gamma'_j(S), \succ') \succsim' b_m} \mu'_j, \forall m \in \{1, 2, \ldots, |S|\} \text{ and } S \in X$$

We will prove that $\sum_{j:\, \max(\Gamma'_j(S), \succ') \succsim' b_m} \mu'_j \geq \sum_{j:\, |\Gamma'_j(S)| \geq m} \mu_j$ so it follows that π is more attentive than π'. It is sufficient to show $\{j : \max(\Gamma'_j(S), \succ') \succsim' b_m\} \supseteq \{j : |\Gamma'_j(S)| \geq m\}$. Take an arbitrary $t \in \{j : |\Gamma'_j(S)| \geq m\}$. It follows $|\Gamma'_t(S)| \geq m$. Hence, it must be the case that $\max(\Gamma'_t(S), \succ') \succsim' b_m$. This is because otherwise we must have $|\Gamma'_t(S)| \leq m - 1$ since $b_{m-1} \succsim' \max(\Gamma'_t(S), \succ')$, which is a contradiction. Observe that $\max(\Gamma'_t(S), \succ') \succsim' b_m$ implies $t \in \{j : \max(\Gamma'_j(S), \succ') \succsim' b_m\}$, which is what we need.

This completes our proof. ■

Proof of Theorem 3. Let $\{c_1, c_2, \ldots, c_m\}$ be the PRC($\rhd$) representation of π. Suppose $\{\mu_1, \mu_2, \ldots, \mu_m\}$ is the associated probability distribution. Enumerate all elements in S as $x_{|S|} \rhd x_{|S|-1} \rhd \cdots \rhd x_1$ with $|S|$ being the cardinality of S. Define the consideration sets as follows

$$\Gamma_i(S) = \{x_1, x_2, \ldots, x_{i^*}\}, \quad \text{where } i^* \text{ is the largest integer s.t. } c_i(S) \succsim x_t \, \forall t \leq i^*$$

Define the probability distribution over $\{\Gamma_1, \Gamma_2, \ldots, \Gamma_m\}$ as $\mu(\Gamma_i) = \mu_i$. Note that $\Gamma_i(S)$ has the list-based structure with respect to $\rhd$ for all i and $S \in \mathcal{X}$. We show that $\Gamma_i(S)$ is well defined by proving that $c_i(S) \succsim x_1$. If $c_i(S) = x_1$ then we are done. If $c_i(S) \neq x_1$ then $c_i(S) \rhd x_1$ because $c_i(S) \in S$. Because $\pi(c_i(S), S) > 0$, by IDE axiom, we have $c_i(S) \succ x_1$. Hence, $c_i(S) \succsim x_1$ in all situations and $\Gamma_i(S)$ is well defined. Observe that it immediately follows from the definition of $\Gamma_i(S)$ that $c_i(S) = \max(\Gamma_i(S), \succ)$. Hence $\pi(x, S) = \sum_{i:\, x = c_i(S)} \mu_i = \sum_{i:\, x = \max(\Gamma_i(S), \succ)} \mu_i$. We show that the collection of consideration sets $\{\Gamma_1, \Gamma_2, \ldots, \Gamma_m\}$ is growing. It is sufficient to show $\Gamma_i(S) \subseteq \Gamma_j(S)$ when $i \leq j$. By construction, this set relationship is equivalent to $i^* \leq j^*$. Proof by contradiction. Suppose $i^* > j^*$. This implies $x_{j^*+1} \in \Gamma_i(S)$. By definition of j^*, it is the case that $x_{j^*+1} \succ c_j(S)$. Hence, $c_i(S) = \max(\Gamma_i(S), \succ) \succsim x_{j^*+1} \succ c_j(S) \succsim c_i(S)$, which is a contradiction. Here, $c_j(S) \succsim c_i(S)$ comes from the IDE axiom as follows. Note that $c_j(S) \trianglerighteq c_i(S)$ since $j \geq i$. Also, both $c_j(S)$ and $c_i(S)$ are chosen with strict probabilities in S. Hence, it follows from the IDE axiom that $c_j(S) \trianglerighteq c_i(S)$ implies $c_j(S) \succsim c_i(S)$.

The last step is to show that Γ_i is an attention filter. Take arbitrary $y \in S$ but $y \notin \Gamma_i(S)$. We will show that $\Gamma_i(S) = \Gamma_i(S \setminus y)$. The idea is to use Lemma 1 for the linear order $\rhd$. Note that $y \in S$ but $y \notin \Gamma_i(S)$ imply $y \trianglerighteq x_{i^*+1}$ with $x_{i^*+1} \in S$. Applying $\rhd$-wMON for $y \trianglerighteq x_{i^*+1}$ and $x_{i^*+1} \succ c_i(S)$, we have $\pi(c_i(S), S) \leq \pi(c_i(S), S \setminus y)$ so condition I_1 in Lemma 1 is satisfied. Applying $\rhd$-Independence for $y \trianglerighteq x_{i^*+1}$ and $x_{i^*+1} \succ c_i(S)$ and $\pi(c_i(S), S) > 0$, we have $\pi(t, S) = \pi(t, S \setminus y)$ for all t such that $c_i(S) \rhd t$ so condition I_2 in Lemma 1 is satisfied. Hence, Lemma 1 is applicable and it follows that $c_j(S) = c_j(S \setminus y)$ for all $j = 1, 2, \ldots, i$. By construction, $\Gamma_i(S) = \Gamma_i(S \setminus y)$. This completes our proof. ■

Proof of Observation 1. Proof by contradiction. Suppose $\exists i$ such that $w(x_{a_{i+1}}) > w(x_{a_i})$ in the optimal list L. Consider another list L' obtained from L by switching the positions of x_{a_i} and $x_{a_{i+1}}$ while keeping the positions of all other items unchanged. Then, $\max(\Gamma_j(L), \succ) = \max(\Gamma_j(L'), \succ)$ for all $j \geq a_{i+1}$ and $j \leq a_i - 1$. Also,

$$\max(\Gamma_j(L), \succ) = x_{a_i} \text{ and } \max(\Gamma_j(L'), \succ) = x_{a_{i+1}}, \quad \forall j \text{ such that } a_i \leq j \leq a_{i+1} - 1$$

Since $w(x_{a_{i+1}}) > w(x_{a_i})$, the designer's utility is strictly higher under list L' (contradiction). Hence, the initial assumption is wrong and we have $w(x_{a_{i+1}}) < w(x_{a_i})$.

Proof of Observation 2. Proof by contradiction. Suppose there exists an optimal list L but $x_1 \neq x^*$. It follows from Observation 1 that x^* is chosen with a zero probability in L (because $w(x^*) > w(x_1)$ by definition of x^*). Consider the following cases.

Case 1: $x^* L x_{a_k}$, i.e., x^* appears after x_{a_k} in the list. Consider a list L' obtained from L by switching x^* and the item in the $(a_k + 1)th$ position in list L (L' can be identical to L if the $(a_k+1)th$ position in L is x^*). The designer's utilities under L and L' are the same because $\max(\Gamma_i(L), \succ) = \max(\Gamma_i(L'), \succ)$ and $\pi(x, L) = \pi(x, L')$ for all $i = 1, 2, \ldots, |L|$ and $x \in L$. Hence, L' is also optimal. Now, consider a list L'' obtained from L' by switching x^* and x_{a_k} in list L'. Then, $\max(\Gamma_i(L'), \succ) = \max(\Gamma_i(L''), \succ)$ for all $i \neq a_k$. Also, $\max(\Gamma_{a_k}(L'), \succ) = x_{a_k}$ and $\max(\Gamma_{a_k}(L''), \succ) \in \{x^*, x_{a_{k-1}}\}$. Note that $w(x^*) > w(x_{a_k})$ by definition of x^*. Additionally, it follows from Observation 1 that $w(x_{a_{k-1}}) > w(x_{a_k})$. Hence, by the monotonicity of the objective function, the designer's utility under list L'' is strictly higher, implying that list L' cannot be optimal (contradiction).

Case 2: $x_{a_{i+1}} L x^* L x_{a_i}$, for some $i \in \{1, 2, \ldots, k\}$, i.e., x^* appears after x_{a_i} but before $x_{a_{i+1}}$ in the list. Again, consider two lists L' and L'' as in case 1. List L' is obtained from L by switching the positions of x^* and $(a_i + 1)th$. By a similar logic as in case 1, list L' is also optimal. List L'' is obtained from L' by switching the positions of x^* and x_{a_i}. Again, being similar to case 1, list L'' gives the designer a higher utility (contradiction).

Proof of Observation 3. Proof by contradiction. Suppose $x_{a_i} \succ y$ but y appears after $x_{a_{i+1}}$ in the optimal list L. Consider the following cases.

Case 1: y appears right after x_{a_j} in the list for some j such that $k \geq j \geq i+1$. Consider a list L' obtained from list L by switching the positions of x_{a_j} and y. Then, $\max(\Gamma_t(L), \succ) = \max(\Gamma_t(L'), \succ)$ for all $t \neq a_j$. Also, $\max(\Gamma_{a_j}(L), \succ) = x_{a_j}$ and $\max(\Gamma_{a_j}(L'), \succ) = x_{a_{j-1}}$ because of the transitivity of preferences: $x_{a_i} \succ y$ and $x_{a_{j-1}} \succsim x_{a_i}$ (because $j-1 \geq i$) imply $x_{a_{j-1}} \succ y$. Since $w(x_{a_{j-1}}) > w(x_{a_j})$ following Observation 1, the designer's utility is strictly higher under list L' (contradiction). It follows that y cannot appear right after x_{a_j} in the list for j such that $k \geq j \geq i+1$.

Case 2: y appears after x_{a_k} in list L. By case 1, $x_{a_k+1} \neq y$. Consider a list L' obtained from list L by switching the positions of y and x_{a_k+1}. List L' is also optimal because $\max(\Gamma_t(L), \succ) = \max(\Gamma_t(L'), \succ)$ and $\pi(x, L) = \pi(x, L')$ for all $t = 1, 2, \ldots, |L|$ and $x \in L$. However, y appears right after x_{a_k} in list L', which cannot happen following case 1.

Case 3: $\exists t \geq 1$ such that $x_{a_{i+t+1}} L y L x_{a_{i+t}}$, i.e., y appears after $x_{a_{i+t}}$ but before $x_{a_{i+t+1}}$, with $i+t+1 \leq k$. First, y cannot appear right after $x_{a_{i+t}}$ because of the result in case 1. Second, consider a list L' obtained from list L by switching the positions of y and $x_{a_{i+t}+1}$. By the same logic as in case 2, list L' is also optimal. However, y appears right after $x_{a_{i+t}}$ in list L', which cannot happen following case 1. This completes our proof. ■

Proof of Observation 4. Proof by induction. Note that $x_1 = x_{a_1} = \max\limits_{x \in L} w(x)$ following Observation 2. First, we show that $x_{a_2} = \operatorname*{argmax}\limits_{x:\, x \succ x_1} w(x)$. Suppose not and $x_{a_2} \neq \operatorname*{argmax}\limits_{x:\, x \succ x_1} w(x)$. Let $\operatorname*{argmax}\limits_{x:\, x \succ x_1} w(x) = x'$. Since $w(x') > w(x_{a_2})$, it follows from Observation 1 that x' is chosen with zero probability under optimal list L. Consider the following cases.

Case 1: $x' L x_{a_k}$, i.e., x' appears after x_{a_k} in the list. Following the same logic as in case 1 in the proof of Observation 2 (by replacing x^* in the proof of Observation 2 by x'), we can show that there is a contradiction.

Case 2: $x_{a_{i+1}} L x' L x_{a_i}$ for some $i \in \{2, \ldots, k\}$, i.e., x' appears after x_{a_i} but before $x_{a_{i+1}}$ in the list. Again, following the same logic as in case 2 in the proof of Observation 2 (by replacing x^* in the proof of Observation 2 by x'), we can reach a contradiction.

Second, suppose $x_{a_j} = \operatorname*{argmax}\limits_{x:\, x \succ x_{a_{j-1}}} w(x)$ holds for $j = 2, 3, \ldots, t$, where $t \leq k-1$.

We will show that $x_{a_{t+1}} = \operatorname*{argmax}_{x:\, x \succ x_{a_t}} w(x)$. Proof by contradiction. Suppose not and $x_{a_{t+1}} \neq \operatorname*{argmax}_{x:\, x \succ x_{a_t}} w(x)$. Let $\operatorname*{argmax}_{x:\, x \succ x_{a_t}} w(x) = x''$. Following Observation 2, x'' is not chosen at list L. Follow the same logic as in cases 1 and 2, we can show that there is a contradiction. This completes our proof. ■

Appendix B Number of Possible Preferences

In order to establish the existence of a maximum of two possible preferences representing the data, we will need two simple properties: strictness and non-additivity. As we will illustrate later, these two conditions are typically satisfied in empirical settings.

First, the strictness condition requires that the RCF must have full support, and the frequencies of selecting the same alternative must vary in different binary menus.

Definition B.1 (Strictness)**.** RCF π is strict if $\pi(x, S) > 0$ for all $S \ni x$ and $\pi(x, \{x, y\}) \neq \pi(x, \{x, z\})$ for all x, y, z pairwise distinct.

The strictness assumption cannot be rejected by any finite data set. Additionally, it is a weakening of a much more restrictive condition usually assumed for estimation purposes: $\pi(x, S) \neq \pi(x, S') > 0$ for any $S \ni x$ and $S' \ni x$; in our strictness condition, we only require that $\pi(x, S) \neq \pi(x, S')$ when S and S' are binary menus.

Meanwhile, the non-additivity condition states that for arbitrary x, y, z, the choice frequency of x in $\{x, z\}$ cannot be decomposed into the choice frequency of x in $\{x, y\}$ plus the choice frequency of y in $\{y, z\}$.

Definition B.2 (Non-Additivity)**.** RCF π satisfies non-additivity if for arbitrary x, y, z pairwise distinct: $\pi(x, \{x, z\}) \neq \pi(x, \{x, y\}) + \pi(y, \{y, z\})$.

The violation of non-additivity can result in three endogenous GAM representations even when $|X| = 3$. While these cases are very rare and typically not encountered in empirical settings, it would be helpful to understand when our identification can yield more than two representations. To illustrate, consider the following example.

Example B.1. Let $X = \{x, y, z\}$ and RCF $\pi_{a,b}$ be given by

$\pi_{a,b}$	$\{x, y, z\}$	$\{x, y\}$	$\{x, z\}$	$\{y, z\}$
x	a	a	$a + b$	$-$
y	b	$1 - a$	$-$	b
z	$1 - a - b$	$-$	$1 - a - b$	$1 - b$

with $0 < a, b$ and $a + b < 1$. There is no regularity violation in the choice data and $\pi_{a,b}$ satisfies strictness. Additionally, $\pi_{a,b}$ violates non-additivity as $\pi_{a,b}(x, \{x, z\}) = \pi_{a,b}(x, \{x, y\}) + \pi_{a,b}(y, \{y, z\})$. $\pi_{a,b}$ has three GAM representations with the following preferences: $y \succ_1 x \succ_1 z$, $x \succ_2 z \succ_2 y$, and $z \succ_3 y \succ_3 x$.

Non-additivity effectively excludes cases as in Example B.1 above. Additionally, non-additivity also implies that for given binary choices among arbitrary x, y, z, there exist at least two random utility models when the grand set is $\{x, y, z\}$. To understand this implication, consider the choice data in Example B.1 above that violates non-additivity. Given binary choices in $\{x, y\}, \{y, z\}$, and $\{x, z\}$, if an arbitrary RCF π' is a RUM, choice probabilities in $\{x, y, z\}$ under π' must satisfy

$$\pi'(x, \{x, y, z\}) \leq a; \pi'(y, \{x, y, z\}) \leq b; \text{ and } \pi'(z, \{x, y, z\}) \leq 1 - a - b,$$

It follows that $1 = \pi'(x, \{x, y, z\}) + \pi'(y, \{x, y, z\}) + \pi'(z, \{x, y, z\}) \leq a + b + 1 - a - b = 1$. Hence, the three inequalities above must hold with equality, and it implies that there is a unique RUM given binary choices. That unique RUM is actually given in Example B.1.

The violation of non-additivity, as the one in Example B.1, is relatively rare. Note that under the full support of the RCF, non-additivity constitutes a generalization of a well-known notion of choice consistency across menus called moderate stochastic transitivity (MST) (Chipman, 1958; Georgescu-Roegen, 1958):

$$\min\{\pi(x, \{x, y\}), \pi(y, \{y, z\})\} \geq \frac{1}{2} \Rightarrow \pi(x, \{x, z\}) \geq \min\{\pi(x, \{x, y\}), \pi(y, \{y, z\})\}$$

MST is one of the common stochastic transitivity properties studied in the literature.[22] Empirically, there is robust evidence suggesting that individual choices satisfy MST. In reviewing experimental data, Mellers et al. (1992, p. 348) note that "moderate stochastic transitivity are often satisfied, although a few exceptions have been noted." Some of the earliest supporting evidence for MST includes perceptual choice data in Tversky and Russo (1969) and gamble choice in Lindman (1971). Recent em-

[22] The two other common stochastic transitivity properties are weak stochastic transitivity (WST): $\min\{\pi(x, \{x, y\}), \pi(y, \{y, z\})\} \geq 1/2 \Rightarrow \pi(x, \{x, z\}) \geq 1/2$, and strong stochastic transitivity (SST): $\min\{\pi(x, \{x, y\}), \pi(y, \{y, z\})\} \geq 1/2 \Rightarrow \pi(x, \{x, z\}) \geq \max\{\pi(x, \{x, y\}), \pi(y, \{y, z\})\}$.

pirical evidence shows that MST is also satisfied in different domains such as choice among lotteries (Soltani et al., 2012) and even in animal studies (Lea and Ryan, 2015; Rivalan et al., 2017). Remark B.1 below states that non-additivity generalizes MST as any positive RCF satisfying the latter also satisfies the former. Hence, whenever the MST is satisfied, then so is non-additivity.

Remark B.1. If a positive RCF π satisfies MST then π satisfies non-additivity.

Proof. Proof by contradiction. Suppose π is positive and satisfies MST but violates non-additivity. Without loss of generality, suppose $\pi(x, \{x, z\}) = \pi(x, \{x, y\}) + \pi(y, \{y, z\})$. Let the binary choices in π be given by

π	$\{x, y\}$	$\{x, z\}$	$\{y, z\}$
x	a	$a + b$	$-$
y	$1 - a$	$-$	b
z	$-$	$1 - a - b$	$1 - b$

where $a, b > 0$ and $a + b < 1$ because π has full support. We show that π violates MST.

Case 1: Suppose $\pi(x, \{x, y\}) = a \geq 1/2$. It follows $\pi(x, \{x, z\}) = a + b > a \geq 1/2$ and $\pi(z, \{y, z\}) = 1 - b > a \geq 1/2$. Hence, $\min\{\pi(x, \{x, z\}), \pi(z, \{y, z\})\} \geq 1/2$. The MST implies that $\pi(x, \{x, y\}) \geq \min\{\pi(x, \{x, z\}), \pi(z, \{y, z\})\}$, or equivalently $a \geq \min\{a + b, 1 - b\}$. This is a contradiction since $0 < a < a + b < 1$.

Case 2: Suppose $\pi(x, \{x, y\}) = a \leq 1/2$. It follows $\pi(y, \{x, y\}) \geq 1/2$.

- If $\pi(x, \{x, z\}) \geq 1/2$ then $\min\{\pi(y, \{x, y\}), \pi(x, \{x, z\})\} \geq 1/2$. The MST implies that $\pi(y, \{y, z\}) \geq \min\{\pi(y, \{x, y\}), \pi(x, \{x, z\})\}$, or equivalently $b \geq \min\{1 - a, a + b\}$. This is a contradiction since $0 < a < a + b < 1$.

- If $\pi(x, \{x, z\}) \leq 1/2$ then $\pi(z, \{x, z\}) \geq 1/2$. It follows $\pi(z, \{y, z\}) \geq 1/2$ (because $1 - b > 1 - a - b$). We have $\min\{\pi(z, \{y, z\}), \pi(y, \{x, y\})\} \geq 1/2$. The MST implies $\pi(z, \{x, z\}) \geq \min\{\pi(z, \{y, z\}), \pi(y, \{x, y\})\}$, or equivalently $1 - a - b \geq \min\{1 - b, 1 - a\}$. This is a contradiction since $0 < a, b$. This completes our proof.[23] ■

[23] It can be further shown that there are some RCFs satisfying non-additivity but do not satisfy

Once strictness and non-additivity properties are satisfied, Theorem B.1 below claims that there are at most two endogenous GAM representations for X of any size.

Theorem B.1 (Number of endogenous GAM representations)**.** Suppose a strict RCF π has an endogenous GAM representation and satisfies non-additivity. There are at most two representations.

Proof. We prove Theorem B.1 by induction based on the number of alternatives in the grand set X. Firstly, we use non-additivity and strictness of RCF to show that for arbitrary x, y, z pairwise distinct, there are at most two preference orders over x, y, z. Subsequently, we generalize this result to X of any size. Before we proceed with the proof, we prove the following Lemmata.

Lemma B.1. Suppose a strict RCF π has an endogenous GAM representation. For arbitrary x, y, and z, if there are two preferences over x, y, z, then no alternative (among x, y, z) is ranked first or last in both preferences.

Proof. Proof by contradiction. Suppose there are two different preferences $\succ_1$ and $\succ_2$ representing π and $\succ_1, \succ_2$ differ over x, y, z for some x, y, z. Suppose one alternative among x, y, z is ranked first or last in both preferences.

Case 1: One alternative is ranked first in the two preferences. Without loss of generality, suppose the alternative is z and suppose $z \succ_1 x \succ_1 y$ and $z \succ_2 y \succ_2 x$. Applying Independence for $z \succ_1 x \succ_1 y$, we have $\pi(y, \{x, y, z\}) = \pi(y, \{x, y\})$. Applying Independence for $z \succ_2 y \succ_2 x$, we have $\pi(x, \{x, y, z\}) = \pi(x, \{x, y\})$. Hence,

$$1 = \pi(x, \{x, y\}) + \pi(y, \{x, y\}) = \pi(x, \{x, y, z\}) + \pi(y, \{x, y, z\})$$

Therefore, $\pi(z, \{x, y, z\}) = 0$ (contradiction because π is a strict probability choice).

MST, so MST is strictly more restrictive. For example, consider the following RCF with full support

π	$\{x, y, z\}$	$\{x, y\}$	$\{x, z\}$	$\{y, z\}$
x	5/12	6/12	8/12	–
y	3/12	6/12	–	2/5
z	4/12	–	4/12	3/5

This RCF has at least one GAM representation with $y \succ x \succ z$ and satisfy non-additivity. However, there is a violation of MST because $\min\{\pi(y, \{x, y\}), \pi(x, \{x, z\})\} \geq 1/2$ but $\pi(y, \{y, z\}) < 1/2$.

Case 2: One alternative is ranked last in the two preferences. Without loss of generality, suppose the alternative is z and suppose $x \succ_1 y \succ_1 z$ and $y \succ_2 x \succ_2 z$. Applying Independence for $x \succ_1 y \succ_1 z$, we have $\pi(z, \{x, y, z\}) = \pi(z, \{y, z\})$. Applying Independence for $y \succ_2 x \succ_2 z$, we have $\pi(z, \{x, y, z\}) = \pi(z, \{x, z\})$. Hence, $\pi(z, \{x, y, z\}) = \pi(z, \{x, z\}) = \pi(z, \{y, z\})$. This is also a contradiction to the fact that π is a strict probability choice. ■

Lemma B.2. Suppose a strict RCF π has an endogenous GAM representation and satisfies non-additivity. There are at most two preference orders over x, y, z for arbitrary x, y, z pairwise distinct.

Proof. Proof by contradiction. Suppose there are three different preferences $\succ_1, \succ_2$ and $\succ_3$ representing π and $\succ_1, \succ_2$ and $\succ_3$ differ over x, y, z for some x, y, z. Since no alternative is ranked first or last in any two preferences following Lemma B.1, the three preferences $\succ_1, \succ_2$ and $\succ_3$ must satisfy: $x \succ_1 y \succ_1 z$, and $y \succ_2 z \succ_2 x$, and $z \succ_3 x \succ_3 y$. We have

$$\pi(z, \{x, y, z\}) = \pi(z, \{y, z\}) \leq \pi(z, \{x, z\}) \text{ (by w-MON and Independence for } \succ_1)$$
$$\pi(x, \{x, y, z\}) = \pi(x, \{x, z\}) \leq \pi(x, \{x, y\}) \text{ (by w-MON and Independence for } \succ_2)$$
$$\pi(y, \{x, y, z\}) = \pi(y, \{x, y\}) \leq \pi(y, \{y, z\}) \text{ (by w-MON and Independence for } \succ_3)$$

Let $\pi(x, \{x, y, z\}) = a > 0$ and $\pi(y, \{x, y, z\}) = b > 0$ with $a + b < 1$. Choices over x, y, z are then given by

π	$\{x, y, z\}$	$\{x, y\}$	$\{x, z\}$	$\{y, z\}$
x	a	$1 - b$	a	$-$
y	b	b	$-$	$a + b$
z	$1 - a - b$	$-$	$1 - a$	$1 - a - b$

However, π violates non-additivity because $\pi(y, \{y, z\}) = \pi(y, \{x, y\}) + \pi(x, \{x, z\})$. Hence, the initial assumption is wrong, and there are at most two preference orders over x, y, z for arbitrary x, y, z pairwise distinct. ■

Next, we will prove the Theorem. Proof by induction based on the number of alternatives in X. First, when $|X| = 3$, there are at most two preferences representing

π as shown in Lemma B.2. Suppose there are at most two preferences representing π when $|X| \leq k-1$. Consider $|X| = k \geq 4$. Proof by contradiction. Suppose there are three different preferences $\succ_1, \succ_2, \succ_3$ representing π. For any $a \in X$ and $i \in \{1,2,3\}$, let $\succ_i^a$ be a preference obtained from $\succ_i$ after removing alternative a. Note that for any $a \in X$, at least two among three $\succ_1^a, \succ_2^a, \succ_3^a$ have to be identical. Otherwise, at $X \setminus a$, there are three different preferences $\succ_1^a, \succ_2^a, \succ_3^a$ representing π. This contradicts our initial assumption as $|X \setminus a| = k-1$.

Consider $\succ_i^a$ for $a \in \{x,y,z,t\} \subseteq X$ and $i \in \{1,2,3\}$. For each a, by the argument above, there exists (i,j) with $i \neq j, i,j \in \{1,2,3\}$ such that $\succ_i^a \equiv \succ_j^a$. There are at most three such pairs but four alternatives a. By Dirichlet's box principle, there exist $i \neq j$ and $a \neq b$ with $a,b \in \{x,y,z,t\}$ such that $\succ_i^a \equiv \succ_j^a$ and $\succ_i^b \equiv \succ_j^b$. Without loss of generality, suppose that $\succ_1^x \equiv \succ_1^x$ and $\succ_1^t \equiv \succ_2^t$. Consider the following cases.

Case 1: x and t do not agree at $\succ_1$ and $\succ_2$. Without loss of generality, suppose $x \succ_1 t$ but $t \succ_2 x$.

Case 1.1: Suppose there exists $y \in X$, $y \neq x, t$ such that $y \succ_1 t$. It follows $y \succ_1^x t$. Note that both y and t appear in $\succ_1^x$ and $\succ_2^x$, and we have $\succ_1^x \equiv \succ_2^x$. Then $y \succ_1^x t$ implies $y \succ_2^x t$ and hence $y \succ_2 t$. By transitivity, $y \succ_2 t$ and $t \succ_2 x$ imply $y \succ_2 x$. Also, note that both y and x appear in $\succ_1^t$ and $\succ_2^t$. It follows that $y \succ_2 x$ implies $y \succ_1 x$. Hence, we have $y \succ_1 x \succ_1 t$ by transitivity. Now, for two preferences $\succ_1$ and $\succ_2$, we have $y \succ_1 x \succ_1 t$ and $y \succ_2 t \succ_2 x$ so y is ranked first among x, y, t in both preferences. This cannot happen because of Lemma B.1.

Case 1.2: Suppose there exists $y \in X$, $y \neq x, t$ such that $y \succ_2 x$. This case is similar to case 1.1. By the similar logic, we have $y \succ_1 x \succ_1 t$ and $y \succ_2 t \succ_2 x$, so y is still ranked first among x, y, t in both preferences. This cannot happen because of Lemma B.1.

Case 1.3: For all $y \in X$, $y \neq x, t$, $t \succ_1 y$ and $x \succ_2 y$ because of results in cases 1.1 and 1.2. By transitivity, we have $x \succ_1 t \succ_1 y$ and $t \succ_2 x \succ_2 y$. Hence, y is ranked last among x, y, t in both preferences. This cannot happen because of Lemma B.1.

Case 2: x and t agree at $\succ_1$ and $\succ_2$. Without loss of generality, suppose $x \succ_1 t$ and $x \succ_2 t$. Take an arbitrary $y \in X$, $y \neq x, t$. Note that $t \succ_1 y \Leftrightarrow t \succ_2 y$ because both t and y appear in $\succ_1^x$ and $\succ_2^x$ and $\succ_1^x \equiv \succ_2^x$. Similarly, $x \succ_1 y \Leftrightarrow x \succ_2 y$ because both x and y appear in $\succ_1^t$ and $\succ_2^t$ and $\succ_1^t \equiv \succ_2^t$. Also, for all $y \neq z$ and y, z different from x, t, $y \succ_1 z \Leftrightarrow y \succ_2 z$. This is because both y and z appear in $\succ_1^t$ and $\succ_2^t$ and $\succ_1^t \equiv \succ_2^t$. Therefore, for all $a, b \in X$, $a \neq b$, $a \succ_1 b$ if and only if $a \succ_2 b$. It follows that $\succ_1$ and $\succ_2$ are identical, which is a contradiction. Hence, the initial assumption is wrong and it follows that there are at most two preferences representing π. This completes our proof. ■

Appendix C Menu-Dependent Lists

To accommodate real-life situations where the relative positions of two options in a list can depend on the availability of other items, we introduce context-dependent lists in this Appendix. Suppose there exists an observed ranked list in each choice set. Mathematically, there is a linear order $\rhd_S$ corresponding to the underlying list in $S \in \mathcal{X}$. The consideration sets still have a list-based structure. That is, if $x \in \Gamma(S)$ and $x \rhd_S y$ then $y \in \Gamma(S)$. A stochastic choice function π is said to have a $\text{GAM}_{\{\rhd_S\}_{S\in\mathcal{X}}}(\succ)$ representation if it has a $\text{GAM}(\succ)$ representation where consideration sets in the support have the menu-dependent list-based structure described as above. Note that $\text{GAM}_{\{\rhd_S\}_{S\in\mathcal{X}}}(\succ)$ generalizes $\text{GAM}_{\rhd}(\succ)$ introduced in section 6. Similar to $\text{GAM}_{\rhd}(\succ)$, $\text{GAM}_{\{\rhd_S\}_{S\in\mathcal{X}}}(\succ)$ is also characterized by three axioms: $\rhd_S$-wMON, $\rhd_S$-Independence, and $\rhd_S$-IDE. All three axioms are obtained from their corresponding ones (in $\text{GAM}_{\rhd}(\succ)$) by replacing the menu-independent list order ($\rhd$) with the menu-dependent list orders ($\rhd_S$) (the only exception is in Axiom C.2 where we replace the condition $z \rhd t$ in Axiom 4 by the condition '$z \rhd_S t$ or $z \rhd_{S\setminus x} t$'). Hence, the rationale behind these axioms and the proof of the characterization result in Theorem C.1 closely align with the case when the list order is menu-independent.

Axiom C.1 ($\rhd_S$-wMON)**.** Suppose $x \trianglerighteq_S y$ and $y \succ z$. Then $\pi(z,S) \le \pi(z, S\setminus x)$ for all $S \supseteq \{x,y,z\}$.

Axiom C.2 ($\rhd_S$-Independence)**.** Suppose $x \trianglerighteq_S y$ and $y \succ z$ and $\pi(z,S) > 0$. Then $\pi(t,S) = \pi(t, S\setminus x)$ for all $S \supseteq \{x,y,z,t\}$ and t such that $z \rhd_S t$ or $z \rhd_{S\setminus x} t$.

Axiom C.3 ($\rhd_S$-IDE)**.** $x \rhd_S y$ and $y \succ x$ imply $\pi(x,S) = 0$ for all $S \supseteq \{x,y\}$.

Theorem C.1. RCF π has a $\text{GAM}_{\{\rhd_S\}_{S\in\mathcal{X}}}(\succ)$ representation if and only if π satisfies $\rhd_S$-wMON, $\rhd_S$-Independence, and $\rhd_S$-IDE.

Proof: We first prove the necessity of the axioms.

$\rhd_S$-wMON: Suppose $x \trianglerighteq_S y$ and $y \succ z$. Note that $\rhd_S$-wMON is trivially satisfied when $\pi(z,S) = 0$. When $\pi(z,S) > 0$, there exists a type i in the support such that $\max(\Gamma_i(S), \succ) = z$. Suppose $x \in \Gamma_i(S)$. Since $x \in \Gamma_i(S)$ and $x \trianglerighteq_S y$, it follows that $y \in \Gamma_i(S)$. Therefore, $z = \max(\Gamma_i(S), \succ) \succsim y$, which is a contradiction. Hence,

$x \notin \Gamma_j(S)$. It follows $\Gamma_i(S) = \Gamma_i(S \setminus x)$ since Γ_i is an attention filter. Therefore, $\max(\Gamma_i(S), \succ) = \max(\Gamma_i(S \setminus x), \succ)$ and it follows $\pi(z, S) \leq \pi(z, S \setminus x)$.

$\rhd_S$-Independence: Suppose $x \unrhd_S y$ and $y \succ z$ and $\pi(z, S) > 0$. There exists a type i in the support such that $\max(\Gamma_i(S), \succ) = z$ because $\pi(z, S) > 0$. Among such i, there exists the biggest one i^*. Consider an arbitrary $j \leq i^*$. Suppose $x \in \Gamma_j(S)$. It follows $y \in \Gamma_j(S)$ because $x \unrhd_S y$. Therefore, $\max(\Gamma_j(S), \succ) \succsim y$. However, this could not happen as $y \succ z = \max(\Gamma_{i^*}(S), \succ) \succsim \max(\Gamma_j(S), \succ) \succsim y$ where $\max(\Gamma_{i^*}(S), \succ) \succsim \max(\Gamma_j(S), \succ)$ comes from the fact that $\Gamma_{i^*}(S)$ nests $\Gamma_j(S)$ since $i^* \geq j$. Hence, $x \notin \Gamma_j(S)$. It follows $\Gamma_j(S) = \Gamma_j(S \setminus x)$ since Γ_i is an attention filter. Therefore, $\max(\Gamma_j(S), \succ) = \max(\Gamma_j(S \setminus x), \succ)$ for all $j \leq i^*$. This means $\pi(t, S) = \pi(t, S \setminus x)$ for all $t \in S$ such that $z \rhd_S t$ or $z \rhd_{S \setminus x} t$.

$\rhd_S$-IDE: Suppose $x \rhd_S y$ and $y \succ x$. If $x \in \Gamma_i(S)$ then $y \in \Gamma_i(S)$ because $x \rhd_S y$. However, since $y \succ x$, it follows that $\max(\Gamma_i(S), \succ) \neq x$. Hence, $\pi(x, S) = 0$.

Now, we prove the sufficiency part. Observe that in the construction of PRC introduced in Theorem 1, a choice set S is initially fixed and the construction only uses the linear order operating within S. Hence, the construction of choice functions in Theorem 1 also applies when the list orders vary in different choice sets. Formally, a collection of choice functions $\{c_1, c_2, \ldots, c_m\}$ is the PRC($\{\rhd_S\}_{S \in \mathcal{X}}$) representation of π with $\{\mu_1, \mu_2, \ldots, \mu_m\}$ being the associated probability distribution if $c_i(S) \unrhd_S c_j(S)$ when $i \geq j$ and $\pi(x, S) = \sum_{i:\, c_i(S)=x} \mu_i$ for all $x \in S$ and $S \in \mathcal{X}$. Enumerate all elements in S as $x_{|S|} \rhd_S x_{|S|-1} \rhd_S \cdots \rhd_S x_1$. Define the consideration sets as follows

$$\Gamma_i(S) = \{x_1, x_2, \ldots, x_{i^*}\}, \quad \text{where } i^* \text{ is the largest integer s.t. } c_i(S) \succsim x_t \, \forall k \leq i^*$$

Define the probability distribution over $\{\Gamma_1, \Gamma_2, \ldots, \Gamma_m\}$ as $\mu(\Gamma_i) = \mu_i$. Following the similar logic in the proof of Theorem 3, $\Gamma_i(S)$ is well defined and

$$\pi(x, S) = \sum_{i:\, x=c_i(S)} \mu_i = \sum_{i:\, x=\max(\Gamma_i(S), \succ)} \mu_i$$

Also, $\Gamma_i(S)$ has the list-based structure with respect to $\rhd_S$ for all i and $S \in \mathcal{X}$. The last step is to show that Γ_i is an attention filter. Take arbitrary $y \in S$ but $y \notin \Gamma_i(S)$. We will show that $\Gamma_i(S) = \Gamma_i(S \setminus y)$. Note that $y \in S$ but $y \notin \Gamma_i(S)$ imply $y \unrhd_S x_{i^*+1}$

with $x_{i^*+1} \in S$. Applying $\rhd_S$-wMON for $y \trianglerighteq_S x_{i^*+1}$ and $x_{i^*+1} \succ c_i(S)$, we have

$$\pi(c_i(S), S) \leq \pi(c_i(S), S \setminus y) \tag{I_1}$$

For any $j \leq i$, applying $\rhd_S$-Independence for $y \trianglerighteq_S x_{i^*+1}$ and $x_{i^*+1} \succ c_j(S)$ and $\pi(c_j(S), S) > 0$, we have

$$\pi(t, S) = \pi(t, S \setminus y) \text{ for all } t \text{ such that } c_j(S) \rhd_S t \text{ or } c_j(S) \rhd_{S \setminus y} t, \text{ for all } j \leq i \tag{I_2}$$

where the binary relationship $x_{i^*+1} \succ c_j(S)$ comes from transitivity as $x_{i^*+1} \succ c_i(S)$ (by definition of $c_i(S)$) and $c_i(S) \succsim c_j(S)$ (since $i \geq j$). The latter follows from $c_i(S) \trianglerighteq_S c_j(S)$ (by construction) and an implication from the $\rhd_S$-IDE axiom stated below

$$\text{for all } S \supseteq \{z, t\} \text{ and } \pi(z, S) \cdot \pi(t, S) > 0 : \quad z \trianglerighteq_S t \Leftrightarrow z \succsim t \tag{I_3}$$

We will show that $c_j(S) = c_j(S \setminus y)$ for all $j = 1, 2, \ldots, i$ by using three conditions I_1, I_2, and I_3. It then immediately follows that $\Gamma_i(S) = \Gamma_i(S \setminus y)$ by construction. Proof by induction.

STEP 1. We will prove $c_1(S) = c_1(S \setminus y)$. Proof by contradiction. Suppose $c_1(S) \neq c_1(S \setminus y)$.

- First, observe that $0 < \pi(c_1(S), S)$ because of the construction of choice functions. Also, $\pi(c_1(S), S) \leq \pi(c_1(S), S \setminus y)$ because of conditions I_1 and I_2 (note that $c_i(S) \trianglerighteq_S c_1(S)$ by construction). Hence, $0 < \pi(c_1(S), S \setminus y)$. By definition, $c_1(S \setminus y)$ is the $\rhd_{S \setminus y}$-worst alternative among those with a strictly positive probability of being chosen when the menu is $S \setminus y$. The inequality $0 < \pi(c_1(S), S \setminus y)$ then implies $c_1(S) \rhd_{S \setminus y} c_1(S \setminus y)$ (because $c_1(S) \neq c_1(S \setminus y)$ by assumption).

- Second, because both $c_1(S)$ and $c_1(S \setminus y)$ are chosen with strictly positive probabilities in $S \setminus y$, $c_1(S) \rhd_{S \setminus y} c_1(S \setminus y)$ implies $c_1(S) \succ c_1(S \setminus y)$ following condition I_3.

- Third, applying condition I_2 for $c_1(S) \rhd_{S \setminus y} c_1(S \setminus y)$, we have $\pi(c_1(S \setminus y), S) = \pi(c_1(S \setminus y), S \setminus y) > 0$, where the last inequality comes from the definition of $c_1(S \setminus y)$. By definition, $c_1(S)$ is the $\rhd_S$-worst alternative among those with a strictly positive probability of being chosen in S. The inequality $\pi(c_1(S\setminus y), S) > 0$ then implies $c_1(S \setminus y) \rhd_S c_1(S)$ (because $c_1(S) \neq c_1(S \setminus y)$ by assumption).

- Fourth, because both $c_1(S)$ and $c_1(S \setminus y)$ are chosen with strictly positive probabilities in S, by condition I_3, $c_1(S \setminus y) \rhd_S c_1(S)$ implies $c_1(S \setminus y) \succ c_1(S)$. However, this is a contradiction to the claim in the second bullet. Therefore, the initial assumption is wrong and $c_1(S) = c_1(S \setminus y)$.

STEP 2. Suppose that $c_t(S) = c_t(S \setminus y)$ for all $t = 1, 2, .., \dots, j-1$. We will show that $c_j(S) = c_j(S \setminus y)$ (here $j \leq i$). If $c_j(S) = c_{j-1}(S)$ and $c_j(S \setminus y) = c_{j-1}(S \setminus y)$ then we are done. Therefore, it is sufficient to consider the following cases.

Case 1: $c_j(S) \neq c_{j-1}(S)$ and $c_j(S \setminus y) = c_{j-1}(S \setminus y)$. By the construction of choice functions, $c_j(S) \neq c_{j-1}(S)$ implies $c_j(S) \rhd_S c_{j-1}(S)$. By transitivity, $c_i(S) \trianglerighteq_S c_j(S)$ (because $i \geq j$) and $c_j(S) \rhd_S c_{j-1}(S) = c_{j-1}(S \setminus y)$ imply $c_i(S) \rhd_S c_{j-1}(S \setminus y)$. Hence, condition I_2 is applicable and it is the case that $\pi(c_{j-1}(S\setminus y), S) = \pi(c_{j-1}(S\setminus y), S\setminus y)$. However, it cannot happen as

$$\begin{aligned}
\pi(c_{j-1}(S \setminus y), S \setminus y) \geq \mu(c_j) + \sum_{\substack{k=1 \\ k:\, c_k(S\setminus y)=c_{j-1}(S\setminus y)}}^{j-1} \mu(c_k) &= \mu(c_j) + \sum_{\substack{k=1 \\ k:\, c_k(S)=c_{j-1}(S\setminus y)}}^{j-1} \mu(c_k) \\
&= \mu(c_j) + \pi(c_{j-1}(S \setminus y), S) \\
&> \pi(c_{j-1}(S \setminus y), S)
\end{aligned}$$

The inequality in the first line comes from the fact that $c_j(S \setminus y) = c_{j-1}(S \setminus y)$. The equation in the first line results from our assumption that $c_k(S) = c_k(S \setminus y)$ for all $k = 1, 2, \dots, j-1$. The equation in the second line holds because $\nexists k \geq j$ such that $c_k(S) = c_{j-1}(S \setminus y)$ as $c_j(S) \rhd c_{j-1}(S \setminus y)$. The last inequality uses $\mu(c_j) > 0$.

Case 2: $c_j(S) = c_{j-1}(S)$ and $c_j(S \setminus y) \neq c_{j-1}(S \setminus y)$. By the construction of choice functions, $c_j(S \setminus y) \neq c_{j-1}(S \setminus y)$ implies $c_j(S \setminus y) \rhd_{S\setminus y} c_{j-1}(S \setminus y)$. Since $i \geq j$, we have $c_i(S) \trianglerighteq_S c_j(S) = c_{j-1}(S) = c_{j-1}(S \setminus y)$. By conditions I_1 and I_2, it is the case that $\pi(c_{j-1}(S \setminus y), S) \leq \pi(c_{j-1}(S \setminus y), S \setminus y)$. However, it cannot happen as

$$\pi(c_{j-1}(S \setminus y), S) \geq \mu(c_j) + \sum_{\substack{k=1 \\ k:\, c_k(S)=c_{j-1}(S\setminus y)}}^{j-1} \mu(c_k) = \mu(c_j) + \sum_{\substack{k=1 \\ k:\, c_k(S\setminus y)=c_{j-1}(S\setminus y)}}^{j-1} \mu(c_k)$$

$$
\begin{aligned}
&= \mu(c_j) + \pi(c_{j-1}(S \setminus y), S \setminus y) \\
&> \pi(c_{j-1}(S \setminus y), S \setminus y)
\end{aligned}
$$

The inequality in the first line comes from the fact that $c_{j-1}(S \setminus y) = c_{j-1}(S) = c_j(S)$. The equation in the first line results from our assumption that $c_k(S) = c_k(S \setminus y)$ for all $k = 1, 2, \ldots, j-1$. The equation in the second line holds because $\nexists k \geq j$ such that $c_k(S \setminus y) = c_{j-1}(S \setminus y)$ as $c_j(S \setminus y) \rhd_{S \setminus y} c_{j-1}(S \setminus y)$. The last inequality uses $\mu(c_j) > 0$.

Case 3: $c_j(S) \neq c_{j-1}(S)$ and $c_j(S \setminus y) \neq c_{j-1}(S \setminus y)$. By construction of the choice functions, $c_j(S) \rhd_S c_{j-1}(S)$ and $c_j(S \setminus y) \rhd_{S \setminus y} c_{j-1}(S \setminus y)$. Proof by construction. Suppose $c_j(S) \neq c_j(S \setminus y)$. By condition I_2, we have $\pi(c_j(S), S \setminus y) \geq \pi(c_j(S), S) > 0$, where the last inequality comes from the construction of choice functions. Hence, $c_j(S)$ is chosen with a strictly positive probability at $S \setminus y$. Consider two following cases:

- ***Case 3a:*** $c_j(S \setminus y) \rhd_{S \setminus y} c_j(S)$. As $c_j(S)$ is chosen with a strictly positive probability at $S \setminus y$, the relationship $c_j(S \setminus y) \rhd_{S \setminus y} c_j(S)$ then implies $\exists k \leq j-1$ such that $c_j(S) = c_k(S \setminus y)$. Using our assumption that $c_t(S \setminus x) = c_t(S)$ for all $t = 1, 2, \ldots, j-1$, it follows that $c_j(S) = c_k(S)$. However, this cannot happen because $c_j(S) \rhd_S c_{j-1}(S) \trianglerighteq_S c_k(S)$ since $k \leq j-1$.

- ***Case 3b:*** $c_j(S) \rhd_{S \setminus y} c_j(S \setminus y)$. As $c_j(S)$ and $c_j(S \setminus y)$ are chosen with strictly positive probabilities at $S \setminus y$, the relationship $c_j(S) \rhd_{S \setminus y} c_j(S \setminus y)$ then implies $c_j(S) \succ c_j(S \setminus y)$ following condition I_3. By condition I_2, we have $\pi(c_j(S \setminus y), S) = \pi(c_j(S \setminus y), S \setminus y) > 0$, where the last inequality comes from the definition of $c_j(S \setminus y)$. Hence, $c_j(S \setminus y)$ (and also $c_j(S)$) is chosen with a strictly positive probability at S. By condition I_3, $c_j(S) \succ c_j(S \setminus y)$ implies $c_j(S) \rhd_S c_j(S \setminus y)$. This means $\exists k \leq j-1$ such that $c_j(S \setminus y) = c_k(S)$. It follows $c_j(S \setminus y) = c_k(S \setminus y)$ because of our assumption that $c_t(S \setminus y) = c_t(S)$ for all $t = 1, 2, \ldots, j-1$. However, this cannot happen since $c_j(S \setminus y) \rhd_{S \setminus y} c_{j-1}(S \setminus y) \trianglerighteq_{S \setminus y} c_k(S \setminus y)$ as $k \leq j-1$. Therefore, the initial assumption is wrong and $c_j(S) = c_j(S \setminus y)$. This completes our proof. ■