

# Attention Overload\*

Matias D. Cattaneo<sup>†</sup>   Paul H.Y. Cheung<sup>‡</sup>   Xinwei Ma<sup>§</sup>  
Yusufcan Masatlioglu<sup>¶</sup>

May 28, 2023

## Abstract

We introduce an Attention Overload Model that captures the idea that alternatives compete for the decision maker’s attention, and hence the attention that each alternative receives decreases as the choice problem becomes larger. Using this non-parametric restriction on the random attention formation, we show that a fruitful revealed preference theory can be developed and provide testable implications on the observed choice behavior that can be used to (point or partially) identify the decision maker’s preference and attention frequency. We then enhance our attention overload model to accommodate heterogeneous preferences. Due to the nonparametric nature of our identifying assumption, we must discipline the amount of heterogeneity in the choice model: we propose the idea of List-based Attention Overload, where alternatives are presented to the decision makers as a list that correlates with both heterogeneous preferences and random attention. We show that preference and attention frequencies are (point or partially) identifiable under nonparametric assumptions on the list and attention formation mechanisms, even when the true underlying list is unknown to the researcher. Building on our identification results, for both preference and attention frequencies, we develop econometric methods for estimation and inference that are valid in settings with a large number of alternatives and choice problems, a distinctive feature of the economic environment we consider. We also provide a software package in **R** implementing our empirical methods, and illustrate them in a simulation study.

Keywords: attention frequency, limited and random attention, revealed preference, partial identification, high-dimensional inference.

---

\*We thank Bart Lipman and three anonymous reviewers for their constructive feedback on a previous version of this paper and, in particular, for encouraging us to develop our new heterogeneous preference attention overload model (Section 3). We also thank Whitney Newey, and seminar and conference participants at MIT, Cornell, UC Davis, Rice, Pittsburgh, Rutgers, UCLA, the 2020 Risk, Uncertainty & Decision Conference, the 2020 Southern Economic Association Annual Meeting, and the 2022 Cowles Foundation Econometrics Conference for their comments and suggestions. Cattaneo gratefully acknowledges financial support from the National Science Foundation (SES-2241575).

<sup>†</sup>Department of Operations Research and Financial Engineering, Princeton University.

<sup>‡</sup>Jindal School of Management, University of Texas at Dallas.

<sup>§</sup>Department of Economics, UC San Diego.

<sup>¶</sup>Department of Economics, University of Maryland.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Choice under Attention Overload</b>	<b>5</b>
2.1	Behavioral Implications and Revealed Preferences . . . . .	7
2.2	Revealed Attention . . . . .	11
2.3	Comparison to Other Random Attention Models . . . . .	12
2.4	Econometric Methods . . . . .	14
<b>3</b>	<b>Heterogeneous Preferences over a List</b>	<b>19</b>
3.1	List-Based Attention Overload . . . . .	19
3.2	Characterization with Known List . . . . .	21
3.3	Relationship between AOM and HAOM . . . . .	24
3.4	Characterization with Unknown List . . . . .	25
3.5	Econometric Methods . . . . .	26
<b>A</b>	<b>Appendix: Proofs</b>	<b>27</b>
A.1	Proof of Theorems 1 and 3 . . . . .	27
A.2	Proof of Lemma A.2 . . . . .	29
A.3	Proof of Lemma A.3 . . . . .	30
A.4	Proof of Theorem 4 . . . . .	32
A.5	Proof of Theorem 5 . . . . .	35
A.6	Proof of Theorem 6 . . . . .	36
A.7	Proof of Theorem 7 . . . . .	40
A.8	Proof of Theorem 8 . . . . .	40
A.9	Proof of Theorem 9 . . . . .	41
A.10	Proof of Theorems 10 and 11 . . . . .	42

# 1 Introduction

Decision-making is becoming a bustling task for consumers due to the abundance of options. For example, Amazon US sells more than 606 million products (87 million products in Home & Kitchen and 62 million in Books). This phenomenon is also witnessed in other domains such as healthcare plans, car insurance, or financial services. The proliferation of so many options forces products to compete with each other for consumer attention.<sup>1</sup> Consumers cannot pay attention to all products: some are going to be more appealing than the others, while many are often unnoticed.

It is now well established in economics, marketing, and other social and behavioral sciences that attention is both limited and stochastic (Hauser and Wernerfelt, 1990; Shocker, Ben-Akiva, Boccara, and Nedungadi, 1991). While being a scarce resource, attention may be even more fragmented and scattered due to fierce advertising competition. According to Statista, the US spent over \$253 billion dollars in advertising in 2019. The burden on consumers becomes excessive with the bombardment of advertisements, cognitive overload, and an abundance of alternatives, resulting in them not paying attention to some of the available products.<sup>2</sup> As a consequence, the larger the number of options available, the less likely consumers can pay attention to more of them.

This paper studies decision making in settings where the decision makers confront an abundance of options, their consideration sets are random, and their attention span is limited. We assume that the attention any alternative receives will (weakly) decrease as the number of rivals increases, a nonparametric restriction on the attention rule of decision makers, which we called *Attention Overload*. If attention was deterministic, our proposed behavioral assumption would simply say that if a product grabs the consumer’s consideration in a large supermarket, then it will grab her attention in a small convenience store as there are fewer alternatives (Reutskaja and Hogarth, 2009; Visschers, Hess, and Siegrist, 2010; Reutskaja, Nagel, Camerer, and Rangel, 2011; Geng, 2016).

Our baseline choice model has two components: a random attention rule and a homogeneous preference ordering, but we later enhance our model to allow for heterogeneous preferences. The random attention rule is the probability distribution on all possible consideration sets. To introduce our attention overload assumption formally, we define the amount of attention a product receives as the frequency it enters the consideration set, termed *Attention Frequency*. Attention overload then implies that the attention frequency should not increase as the choice set expands. For preferences, we assume that the decision makers have a complete and transitive (first homogeneous, later heterogeneous) preferences over the alternatives, and that they pick the best alternative in their consideration sets. In this gen-

---

<sup>1</sup>The limited attention phenomenon has been illustrated in different markets: investment decisions (Huberman and Regev, 2001), school choice (Laroche, Rosenblatt, and Sinclair, 1984; Rosen, Curran, and Greenlee, 1998), job search (Sheridan, Richards, and Slocum, 1975), household grocery consumption (Demuyne and Seel, 2018), PC purchases (Goeree, 2008), and airport choice (Başar and Bhat, 2004), just to mention a few examples.

<sup>2</sup>For example, recent Ipsos eye-tracking research suggests that the majority of TV advertising time (55%) is not paid attention to due to multitasking, switching channels, and fast-forwarding. It is also noticed in the literature that consumers consider fewer items (Reutskaja and Hogarth, 2009; Reutskaja, Nagel, Camerer, and Rangel, 2011) or choose the outside option more often (Iyengar and Lepper, 2000) as choice sets expand.

eral setting where attention is random and limited, and products compete for attention, we aim to elicit compatible preference orderings and attention frequencies solely from observed choices.

Existing random attention models cannot capture, or are incompatible with, attention overload. For example, [Manzini and Mariotti \(2014\)](#) consider a parametric attention model with independent consideration where each alternative has a constant attention frequency even when there are more alternatives, and therefore their model does not allow decision makers to be more attentive in smaller decision problems. [Aguiar \(2017\)](#) also share the same feature of constant attention frequency. On the other hand, recent research has tried to incorporate menu-dependent attention frequency ([Demirkan and Kimya 2020](#)) under the framework of independent consideration. This model is so general that it allows for the opposite of attention overload—being more attentive in larger choice sets. The recent models of [Brady and Rehbeck \(2016\)](#) and [Cattaneo, Ma, Masatlioglu, and Suleymanov \(2020\)](#) also allow for the possibility that an alternative receives less attention even when the choice set gets smaller. See [Section 2.3](#) and the supplemental appendix ([Section SA-1](#)) for more discussion on related literature.

We contribute to the decision theory literature by introducing the attention overload non-parametric restriction on the attention frequency, which is the key building block to achieve both preference ordering and attention frequency (point or partial) identification from observed choice data. Our results do not require the attention rule to be observed, nor to satisfy other restrictions beyond those implicitly imposed by attention overload. Since our revealed preference and attention elicitation results are derived from nonparametric restrictions on the consideration set formation, without committing to any particular parametric attention rule, they are more robust to misspecification biases ([Matzkin, 2007, 2013](#); [Molinari, 2020](#)).

The fact that attention is not observed poses unique challenges to both identification of and statistical inference on the decision maker’s preference. This is because one can only identify (and consistently estimate) the choice probabilities from typical choice data, while our main restriction is imposed on the attention rule. Furthermore, as our attention overload assumption does not require a parametric model of consideration set formation, the set of compatible attention rules is usually quite large. In other words, the attention rule is almost never uniquely identified in our model. We nonetheless show that our attention overload assumption, despite being very general, still delivers nontrivial empirical content: we prove in [Section 2](#) that a preference ordering is compatible with our attention overload model if and only if the choice probability satisfies a system of inequality constraints, which corresponds to a form of regularity violation. Furthermore, to improve computation and practical implementation, we discuss how to leverage binary choice problems for identification, estimation, and inference.

Besides revealed preference, information about attention frequency is also an object of interest. For example, it enables marketers to gauge the effectiveness of their marketing strategies, or policy-makers to assess whether consumers allocate their attention to better products. Despite the fact that the underlying attention rule may not be identifiable, we show in [Section 2](#) that our nonparametric attention overload behavioral assumption allows for (point or partial) identification of the attention frequency using standard choice data. This result appears to be the first nonparametric identification result of a relevant feature

of an attention rule in the random limited attention literature: revealed attention analysis has not been possible under nonparametric identifying restrictions in prior work.

In Section 3, we enhance our attention overload model to accommodate multiple decision makers with heterogeneous preference orderings. Due to the nonparametric nature of our identifying assumption, we must discipline the amount of heterogeneity in the model: we propose the idea of *List-Based Attention Overload*, where alternatives are presented to the customers as a list that correlates with both heterogeneous preferences and random limited attention. Many real-life situations involve consumers encountering alternatives in the form of a list (Simon, 1955; Rubinstein and Salant, 2006). The model is motivated by the observation that an item’s placement on a list has a profound impact on its recollection and evaluation by subjects (Ellison and Ellison, 2009; Augenblick and Nicholson, 2016; Biswas, Grewal, and Roggeveen, 2010; Levav, Heitmann, Herrmann, and Iyengar, 2010). For example, a ranked list of search results provided by a web platform can affect both the search behavior and the perception of individuals about the quality of products (Reutskaja, Nagel, Camerer, and Rangel, 2011).

Our underlying idea is that a common set of characteristics among the decision makers is taken into account to construct the list. For example, a list emerging from search results can be a good proxy for their preferences if decision makers perceive that the search result reflects the true quality of the listed items (Westerwick, 2013).<sup>3</sup> Indeed, many commercial websites collect individual consumers’ behavioral data and try to match each consumer with specific products. The list can be thought of as the outcome of personalized recommendations, and therefore individuals facing the same list would share similar tastes. However, individuals might tend to favor their status quo and assign a relatively higher rank to their reference point compared to the rest of the items in the original list. Thus, the existence of the list allows for both heterogeneous preferences and random attention, but restricts the total number of potential preference orderings and attention rules allowed in the choice model.

Attention overload implies that it is often impractical for decision makers to conduct exhaustive searches when many products are on the list. We thus assume that a decision maker investigates alternatives to construct her limited attention consideration set through the list: she might consider only a subset of the alternatives available. Our list-based attention overload model imposes three basic behavioral restrictions on the consideration set formation for a given list: (i) whenever an alternative is considered, all alternatives in the list before it are also taken into account; (ii) if an alternative is not recognized in a smaller set, then it cannot be recognized in a larger set; and (iii) in binary problems, both options are always considered. These assumptions are, for example, supported by eye-tracking studies showing that people tend to scan search engine results in order of appearance, and then fixate on the top-ranked results even if lower-ranked results are more relevant (Pernice, Whinton, Nielsen, et al., 2018). To capture heterogeneity in cognitive ability, the model accommodates individuals with different consideration sets as long as they satisfy the above behavioral restrictions, thereby allowing for list-based heterogeneity in random limited attention.

---

<sup>3</sup>Pan, Hembrooke, Joachims, Lorigo, Gay, and Granka (2007) and Purcell, Rainie, and Brenner (2012) showed that decision makers who have high confidence in search engines think that search engines rank the options best suited to their needs higher. They also showed that people use the search engine’s rankings as a criterion for the quality of products.

To discipline the amount of preference heterogeneity, we also introduce three behavioral axioms characterizing our proposed heterogeneous (preference) attention overload model for a given list. The first axiom captures a restricted form of regularity violation: removing alternatives will not decrease the choice probabilities of a product as long as there is another product listed before it in both decision problems. The second axiom states that binary choice probabilities decrease as the opponent is ranked higher in the list. The last axioms requires that the total binary choice probabilities against the immediate predecessor in the list must be less than or equal to one. We then show that preference and attention frequencies are (point or partially) identifiable under nonparametric assumptions on the list and attention formation mechanisms, even when the true underlying list is unknown to the researcher.

Based on our identification results, we develop econometric methods for revealed preference and attention analysis in both homogeneous and heterogeneous preference settings, which are directly applicable to standard choice data. We only assume that a random sample of choice problems and choices selections is observed, and then provide methods for estimation of and inference on the preference ordering (homogeneous case), or the preferences frequency (heterogeneous case), and the attention frequency of the decision makers. For example, our methods allow for (i) test whether a specific preference ordering is compatible with our attention overload model, (ii) construct (asymptotically) valid confidence sets, (iii) conduct overall model specification testing, and (iv) estimate preference frequencies in heterogeneous settings. For revealed attention, we obtain (point or partial) identification estimates for attention frequencies in both homogeneous and heterogeneous attention overload models. To establish the validity of our econometric methods, we employ the latest results on high-dimensional normal approximation ([Chernozhukov, Chetverikov, Kato, and Koike, 2022](#)). This is crucial because the number of inequality constraints involved in our statistical inference procedures may not be small relative to the sample size. While allowing the dimension (complexity) of the inference problems to be much larger than the sample size, we explicitly characterize the error from a normal approximation for the estimated choice probabilities, thereby shedding light on the finite-sample performance of our proposed econometric methods.

Econometric methods based on revealed preference theory have a long tradition in economics and many other social and behavioral sciences. See [Matzkin \(2007, 2013\)](#), [Molinari \(2020\)](#), and references therein. There is only a handful of recent studies bridging decision theory and econometric methods by connecting discrete choice and limited consideration. Contributions to this new research area include [Abaluck and Adams \(2021\)](#), [Barseghyan, Coughlin, Molinari, and Teitelbaum \(2021\)](#), [Barseghyan, Molinari, and Thirkettle \(2021\)](#), [Cattaneo, Ma, Masatlioglu, and Suleymanov \(2020\)](#), and [Dardanoni, Manzini, Mariotti, and Tyson \(2020\)](#), among others. Each of these papers imposes different (parametric) identification assumptions on the random consideration and the preferences, producing different levels of identification of preference orderings and consideration set rules. Our paper contributes to this emerging literature by providing new nonparametric results on the identification of and inference for the preference and attention frequencies when decision makers only pay attention to a subset of possibly too many alternatives at random.

The rest of the paper proceeds as follows. Section 2 introduces the setup and our key attention overload assumption under homogeneous preferences, and then proves our main

characterization result. That section also presents computationally attractive identification results based on binary comparisons, discusses partial identification of the attention frequency, and outlines valid econometric methods for revealed preference and revealed attention analyses with homogeneous preferences. Section 3 introduces the idea of list-based attention overload to allow for heterogeneous preferences, and presents (point or partial) identification of preference and attention frequencies. That section considers settings where the underlying true list may or may not be known, and also discusses principled econometric methods for estimation and inference using only observed choice data. The appendix contains the proofs of our results, while the online supplemental appendix collects complementary material: (i) in-depth related literature discussion, (ii) simulation evidence, and (iii) omitted technical lemmas and other methodological discussions. We also provide a software package and replication code in R implementing our empirical methods.

## 2 Choice under Attention Overload

The theoretical analysis in this section revolves around the assumption that attention frequency is monotonic and preferences are homogeneous. Then, in Section 3, we enhance our choice model to allow for heterogeneous preferences.

We denote the grand alternative set as  $X$ , and its cardinality by  $|X|$ . A typical element of  $X$  is denoted by  $a$ . We let  $\mathcal{D}$  be a collection of non-empty subsets of  $X$  representing the collection of choice problems. In this section, we allow incomplete data where  $\mathcal{D}$  is a strict subset of all non-empty subsets of  $X$ , which makes the model still applicable when there is missing data. A *choice rule* is a map  $\pi : X \times \mathcal{D} \rightarrow [0, 1]$  such that  $\pi(a|S) = 0$  for all  $a \notin S$  and  $\sum_{a \in S} \pi(a|S) = 1$  for all  $S \in \mathcal{D}$ .  $\pi(a|S)$  represents the probability that the decision maker chooses alternative  $a$  from the choice problem  $S$ . We assume that the choice rule is identifiable from data; that is, it is known or estimable for the purpose of learning about features of the underlying data generating process in econometrics language (Section 2.4).

An important feature of our model is that consideration sets can be random. An *attention rule* is a map  $\mu : 2^X \times \mathcal{D} \rightarrow [0, 1]$  such that  $\mu(T|S) = 0$  for all  $T \not\subseteq S$  and  $\sum_{T \subseteq S} \mu(T|S) = 1$  for all  $S \in \mathcal{D}$ .  $\mu(T|S)$  represents the probability of paying attention to the consideration set  $T \subseteq S$  when the choice problem is  $S$ . This formulation also allows for deterministic attention rules (e.g.,  $\mu(S|S) = 1$  represents full attention).

The choice rule and attention rule are standard features of (rational) choice models with random attention. In this paper, we consider a novel feature of these models that is related to the amount of attention each alternative captures for a given  $\mu$ . We can extract this information from the attention rule by simply summing up the frequencies of consideration sets containing the alternative.

**Definition 1 (Attention Frequency).** Given  $\mu$ , the *attention frequency* map  $\phi_\mu : X \times \mathcal{D} \rightarrow [0, 1]$  is

$$\phi_\mu(a|S) := \sum_{T \subseteq S: a \in T} \mu(T|S).$$

$\phi_\mu(a|S)$  represents the total probability that  $a$  attracts attention in  $S$ . Whenever  $\mu$  is clear

from the context, we will omit the subscript  $\mu$  to reduce notation. In deterministic attention models, the attention that one alternative receives is either zero or one (i.e., whether it is being considered or not). However, in stochastic environments, attention is probabilistic: this means that the attention one alternative receives may not be binary.

When decision makers are overwhelmed by an abundance of options, every choice alternative competes for attention. This implies that as the number of alternatives increases, the competition gets more fierce: the attention frequency to a product should decrease weakly when the set of available alternatives is expanded by adding more options. We call this property *Attention Overload*, the novel nonparametric identifying restriction in this paper.

**Assumption 1 (Attention Overload).** For any  $a \in T \subseteq S$ ,  $\phi_\mu(a|S) \leq \phi_\mu(a|T)$ .

If we allow the consideration set to be empty, then we should also require that the frequency of paying attention to nothing increases when the choice set expands. This is related to the choice overload behavioral phenomenon. At this point, we exclude the possibility of paying attention to nothing for simplicity (i.e.,  $\mu(\emptyset|S) = 0$  for all  $S \in \mathcal{D}$ ). An attention rule  $\mu$  satisfies attention overload if its corresponding attention frequency is monotonic in the sense of Assumption 1. Section 2.3 compares and contrasts Assumption 1 with other choice models in the literature (see also Section SA-I in the supplemental appendix).

Given the nonparametric attention overload restriction in Assumption 1, the choice rule can be defined accordingly. A (rational) decision maker who follows the attention overload choice model maximizes her utility according to a preference ordering  $\succ$  under each realized consideration set.

**Definition 2 (Attention Overload Representation).** A choice rule  $\pi$  has an attention overload representation if there exists a preference ordering  $\succ$  over  $X$  and an attention rule  $\mu$  satisfying attention overload (Assumption 1) such that

$$\pi(a|S) = \sum_{T \subseteq S} \mathbb{1}(a \text{ is } \succ\text{-best in } T) \cdot \mu(T|S)$$

for all  $a \in S$  and  $S \in \mathcal{D}$ . In this case, we say that  $\pi$  is represented by  $(\succ, \mu)$  or, alternatively, that  $\pi$  is an *Attention Overload Model* (AOM). We also say that  $\succ$  represents  $\pi$  if there exists an attention rule  $\mu$  satisfying attention overload such that  $\pi$  is represented by  $(\succ, \mu)$ .

To summarize, the unknown model primitives are the attention rule  $\mu$  and the preference ordering  $\succ$ . We only assume that the choice rule  $\pi$  is observable (i.e., point identifiable and estimable from data), and we do not require additional information beyond standard observable choice data. We next investigate the behavioral implications of AOM. Section 2.1 shows that it is possible to (point or partially) identify the underlying preference ordering by exploiting the attention overload Assumption 1, and Section 2.2 presents (point or partial) identification results for the attention frequency. Section 2.4 builds on those identification results and develops feasible econometric methods. Section 3 further demonstrates how to incorporate and study heterogeneous preferences in the presence of attention overload.



## 2.1 Behavioral Implications and Revealed Preferences

We first investigate characterization and preference elicitation, as they have a close relationship. We aim to determine whether a data generating process possesses an AOM representation, and if it is feasible to identify preference orderings from observed choice data. To accomplish this, we investigate whether a specific preference ordering can accurately represent the data. This is a challenging task due to several potential issues. First, the data may not have an AOM representation at all. Second, even if an AOM representation exists, the actual preference may differ from the proposed preference ordering. Third, even if the proposed preference aligns with the underlying preference ordering, it is still necessary to construct an attention rule that satisfies attention overload and accurately represents the data. Our first main result addresses these challenges by providing a tight representation without the requirement of constructing an attention rule, which can be a laborious task when there are many alternatives.

AOM has several behavioral implications. Assume that  $(\succ, \mu)$  represents  $\pi$ . Since attention is a requirement for a choice, any choice probability is always bounded above by attention frequency, i.e.,  $\phi(a|S) \geq \pi(a|S)$ . Then, by attention overload, we must have  $\phi(a|T) \geq \phi(a|S) \geq \pi(a|S)$  for  $T \subseteq S$ . In addition, the difference  $\phi(a|T) - \pi(a|T)$  captures the probability that  $a$  receives attention but is not chosen in  $T$ . As a consequence, in these cases, a better option must be chosen in  $T$ , which implies  $\phi(a|T) - \pi(a|T) \leq \pi(U_{\succ}(a)|T)$ , where  $U_{\succ}(a)$  denotes the strict upper contour set of  $a$  with respect to  $\succ$ . (With a slight abuse of notation, we set  $\pi(U_{\succ}(a)|T) = \pi(U_{\succ}(a) \cap T|T) = \sum_{b \in T: b \succ a} \pi(b|T)$ .) Combining these observations, we get  $\pi(a|S) \leq \phi(a|S) \leq \phi(a|T) \leq \pi(U_{\succeq}(a)|T)$ , where  $U_{\succeq}(a)$  denotes the upper contour set of  $a$  with respect to  $\succ$ . It follows that

$$\pi(a|S) \leq \pi(U_{\succeq}(a)|T)$$

whenever  $\succ$  represents the data. This condition only refers to preferences, not to the attention rule. Therefore, the following axiom must be satisfied whenever  $\succ$  represents the data.

**Axiom 1 ( $\succ$ -Regularity).** For all  $a \in T \subseteq S$ ,  $\pi(U_{\succeq}(a)|T) \geq \pi(a|S)$ .

Axiom 1 applies to the choice rule  $\pi$ , which is point identifiable and estimable from standard choice data and is stated in terms of a preference ordering  $\succ$ , a key unobservable primitive of our model. Given  $\succ$ , it is routine to check whether  $\pi$  satisfies  $\succ$ -Regularity. This axiom is closely related to, but different from, the classical regularity condition. Axiom 1 trivially implies the regularity condition for the best alternative  $a^*$  in  $T$ , as  $U_{\succeq}(a^*) = \{a^*\}$  and  $\pi(U_{\succeq}(a^*)|T) = \pi(a^*|T) \geq \pi(a^*|S)$ . Hence, the full power of regularity is assumed. For other alternatives, the regularity condition is partially relaxed. At the other extreme,  $\succ$ -Regularity does not restrict the choice probabilities for the worst alternative,  $a_*$ , since  $U_{\succeq}(a_*) = X$ , and hence  $\pi(U_{\succeq}(a_*)|T) = 1 \geq \pi(a_*|S)$  for all  $T$ , implying that  $\succ$ -Regularity holds trivially.

The following result shows that  $\succ$ -Regularity is not only necessary but also a sufficient condition for  $\succ$  to represent the data.

**Theorem 1 (Characterization).**  $\pi$  has an AOM representation with  $\succ$  if and only if  $\pi$  satisfies  $\succ$ -Regularity.

An immediate corollary is that  $\pi$  is AOM if and only if there exists  $\succ$  such that  $\succ$ -Regularity is satisfied. We provided above the proof of the necessity of  $\succ$ -Regularity. The proof of sufficiency, which relies on Farkas’s Lemma, is given in the appendix.  $\succ$ -Regularity informs us whether  $\pi$  has an AOM representation with  $\succ$ . Of course, it is possible that  $\succ$ -Regularity can be violated for  $\succ$  but is satisfied for another preference  $\succ'$ . Hence,  $\succ$ -Regularity allows us to identify all possible preference orderings without constructing the underlying attention rule  $\mu$ .

We now turn to the discussion of revealed preference. We say that  $b$  is revealed to be preferred to  $a$  if  $b \succ a$  for all  $\succ$  representing  $\pi$ . Theorem 1 suggests that  $\succ$ -Regularity could be a handy method to identify the underlying preference. We first define a binary relation based on  $\succ$ -Regularity property:

$$bP_{\pi}a \quad \text{if } b \succ a \text{ for all preference orderings such that } \pi \text{ satisfies } \succ\text{-Regularity.} \quad (1)$$

Again,  $P_{\pi}$  does not require the construction of all AOM representations. Given a candidate preference, finding the corresponding attention rule satisfying attention overload could be a daunting task. On the other hand, checking whether  $\pi$  satisfying  $\succ$ -Regularity is straightforward because  $\succ$ -Regularity does not require finding the underlying attention rule. Indeed, we utilize this fact in Section 2.4 to develop econometric methods. The next result states the revealed preference of this model.

**Corollary 1 (Revealed Preference).** Let  $\pi$  be an AOM. Then,  $b$  is revealed preferred to  $a$  if and only if  $bP_{\pi}a$ .

Although Theorem 1 and Corollary 1 bypass the need of constructing the underlying attention rule for revealed preference analysis, checking all possible  $|X|!$  preference orderings can be computationally expensive. In addition, if the analyst is interested in learning the preference between two alternatives,  $a$  and  $b$ , then some constraints suggested by  $\succ$ -Regularity may provide little to no relevant information. Fortunately, a key observation is that *regularity violations* at binary choice problems can reveal the decision maker’s preference. Although this result does not exhaust the nonparametric identification power of Assumption 1, it can be handy and computationally more attractive.

More specifically, if  $a, b \in S$  and  $\pi(a|S) > \pi(a|\{a, b\})$ , then any  $(\succ, \mu)$  representing  $\pi$  must rank  $b$  above  $a$ , hence  $b$  must be preferred to  $a$ . To reach such a conclusion, assume the contrary: there exists  $(\succ, \mu)$  representing  $\pi$  such that  $a \succ b$  and  $\mu$  satisfying attention overload. First, the attention frequency is always greater (or equal) than the choice probability for any alternative and in any choice set:  $\pi(a|S) \leq \phi(a|S)$ . In addition, they are equal for the best alternative in any choice set:  $a$  is  $\succ$ -best in  $S$  implies  $\pi(a|S) = \phi(a|S)$ . Given  $a \succ b$ , we have  $\phi(a|\{a, b\}) = \pi(a|\{a, b\}) < \pi(a|S) \leq \phi(a|S)$ . This contradicts our attention overload assumption. The next proposition formalizes this observation.

**Proposition 1 (Regularity Violation at Binary Comparisons).** Let  $\pi$  be an AOM with  $(\mu, \succ)$  and  $a, b \in S$ . If  $\pi(a|S) > \pi(a|\{a, b\})$ , then  $b \succ a$ .

Proposition 1 provides a guideline to easy-to-implement revealed preference analysis without knowledge about each particular representation. A natural question is whether we can generalize the implication of Proposition 1 for an arbitrary set  $T \subseteq S$  instead of only for binary sets. The answer is not straightforward because identification from regularity violation may not be as sharp when there are more than two alternatives in the smaller set. From Proposition 1, we are able to claim revealed preference between two alternatives, but when the smaller set contains more than two alternatives, we only know there are some alternatives better than  $a$  in the smaller set. To see this, suppose not, and  $a$  is the best alternative in the smaller set. We must have  $\pi(a|T) = \phi(a|T)$ . Given that  $\phi(a|T) = \pi(a|T) < \pi(a|S) \leq \phi(a, S)$ , it contradicts the attention overload assumption. We put this observation in the following proposition.

**Proposition 2 (Regularity Violation at Bigger Choice Problems).** Let  $\pi$  be an AOM with  $(\mu, \succ)$  and  $a \in T \subset S$ . If  $\pi(a|S) > \pi(a|T)$ , then there exists  $b \in T$  such that  $b \succ a$ .

Both Proposition 1 and 2 are based on regularity violations, and Proposition 2 implies Proposition 1 when  $T$  is a binary set. The following example demonstrates how these propositions can be used to limit possible preferences orderings first in order to then apply Theorem 1: regularity violation alone does not exhaust the nonparametric identification power of attention overload in this example.

**Example 1.** Consider the following choice data:

$\pi(\cdot S)$	$a$	$b$	$c$	$d$
$\{a, b, c, d\}$	0.05	0.1	0.1	0.75
$\{a, b, c\}$	0.8	0.2	0	–
$\{b, c, d\}$	–	0.7	0.3	0
$\{a, b\}$	0.9	0.1	–	–

where there are  $4! = 24$  candidate preference orderings  $\succ$ . First, by applying Proposition 1 from  $\{a, b, c\}$  to  $\{a, b\}$ , we know that any  $\succ$  with  $b \succ a$  would not be represented by the model. Therefore, there are only 12 compatible preferences with AOM. Second, two other regularity violations involve non-binary sets. From  $\{a, b, c, d\}$  to  $\{a, b, c\}$ , since  $c$  violate regularity, by Proposition 2, we know that  $c \succ a$  and  $c \succ b$  would not hold simultaneously, which eliminates 4 additional preference candidates. Analogously, from  $\{a, b, c, d\}$  to  $\{b, c, d\}$ , since  $d$  violates regularity, it is impossible that  $d$  is preferred to both  $b$  and  $c$ , which further eliminate 4 preferences. Only four preference orderings remain:  $a \succ b \succ c \succ d$ ,  $a \succ b \succ d \succ c$ ,  $a \succ c \succ b \succ d$ , and  $a \succ c \succ d \succ b$ . Finally, we can apply Theorem 1: by checking Axiom 1, we can see that  $\succ$  must include  $b \succ d$  and  $c \succ d$ . This eliminates two preferences. Therefore, the only two possible candidates are  $a \succ b \succ c \succ d$  and  $a \succ c \succ b \succ d$ .  $\square$

In the above example, we show that (i) the data has multiple AOM representations, (ii)  $a \succ_1 b \succ c \succ_1 d$  and  $a \succ_2 c \succ_2 b \succ_2 d$  represent the data, and (iii) since only  $\succ_1$  and  $\succ_2$  satisfy Axiom 1, the revealed preference  $P_\pi$  is only missing information on  $b$  and  $c$  (otherwise it is complete). While we had 24 possible candidates, Proposition 1 implied only 12 of them

were viable candidates. Then, Proposition 2 eliminated 8 of the remaining ones. Finally, only two satisfied Axiom 1. Hence, Example 1 demonstrates how our main results can be used constructively to identify the set of plausible preferences, while also substantially reducing the computational burden.

## Attentive at Binaries

We established how revealed preferences analysis can be done with our nonparametric attention overload assumption. Due to the nature of attention overload, one might suspect that the decision makers are more likely to pay full attention when there are only two alternatives. We now assume that extreme limited attention (i.e., considering only a single option) at binaries cannot exceed a preset probability level, and investigate its implications for revealed preference.

Assume that for any  $\eta \geq 0.5$ ,

$$\eta \geq \max \left\{ \mu(\{a\}|\{a, b\}), \mu(\{b\}|\{a, b\}) \right\} \quad \text{for all } a, b \in X. \quad (2)$$

This condition puts an upper limit on the magnitude of extreme limited attention. As  $\eta$  increases, the probability of limited attention can increase:  $\eta = 1$  imposes no constraint on attention behavior. This assumption does not impose a lower bound for full attention. Even in the extreme case, where  $\eta$  is equal to 0.5, it is still possible that there is no full attention ( $\mu(\{a, b\}|\{a, b\}) = 0$ ). Hence, it is not a demanding condition when compared to full attention at binaries.

Condition (2) can generate additional revealed preferences:  $aP_B b$  if  $\pi(a|\{a, b\}) > \eta$ , for any  $\eta \geq 0.5$ . Whenever we observe  $\pi(a|\{a, b\}) > \eta$ , the choice probability of  $a$  cannot be entirely attributed to the attention on the singleton set,  $\mu(\{a\}|\{a, b\}) \leq \eta$ . Then, we must have  $\mu(\{a, b\}|\{a, b\}) > 0$  and the decision maker chooses  $a$  over  $b$  when she pays attention to both alternatives. Hence, it implies that  $a$  must be better than  $b$ .

We could also interpret the parameter  $\eta$  as a measure of how cautious the policy-maker is when making a welfare judgment. If  $\eta = 1$ , the policy-maker would not draw any conclusion from binary comparisons only ( $P_B = \emptyset$ ). The choice  $\eta = 0.5$  is commonly used in the literature (Marschak, 1959; Fishburn, 1998), which would refer to the largest  $P_B$  in our setup—almost uniquely identified.

Condition (2) provides additional revealed preference information if the data on binary comparisons are available. Under (2), the revealed preferences of our model must include  $P_B$ . We can then extend our characterization theorem: if  $\pi$  satisfies  $\succ$ -Regularity where  $\succ$  includes  $P_B$ , then the data has an AOM representation with  $\mu$  satisfying (2). More importantly,  $P_B$  improves the result of Proposition 2 (and hence Proposition 1) by restricting the set of plausible preferences. We revisit Example 1 to illustrate this point.

**Example 1. (continued)** Assume that we observe additional data at  $\{b, c\}$ :  $\pi(b|\{b, c\}) = 0.7$ . Then, by Condition (2), we conclude that the only possible preference consistent with the observed choice data is  $a \succ b \succ c \succ d$  whenever  $\eta \in [0.5, 0.7)$ , thereby achieving point identification of the preference ordering.

## 2.2 Revealed Attention

Our attention overload model builds on the simple nonparametric requirement that each alternative gets weakly less attention in bigger choice problems, which is captured by monotonicity in attention frequency (Assumption 1). Given a dataset, one might want to learn how the attention frequency changes across different alternatives and choice problems. For example, marketers might want to gauge the effectiveness of their marketing strategies, or policy makers could be interested in assessing whether consumers allocate their attention to better products. Since we do not put any restriction on the attention rule, the attention frequency can vary depending on the actual attention rule that the decision maker has. This section shows that it is possible to develop upper and lower bounds for the attention frequency and thus achieve partial identification of  $\phi$ .

Consider bounding  $\phi$  from below first. For any superset  $R \supseteq S$ , the attention overload assumption implies that  $\pi(a|R) \leq \phi(a|R) \leq \phi(a|S)$ . Therefore, for any  $S$ ,  $\phi(a|S) \geq \max_{R \supseteq S} \pi(a|R)$ . This lower bound on the attention frequency only uses information from the choice rule, which is estimable from standard choice data. Importantly, this lower bound does not require a particular AOM representation, that is, it does not require knowledge of the underlying attention rule. It is also possible to derive an upper bound for  $\phi$ , although in this case the bound will depend on the preference ordering. Consider a preference  $\succ$  and an attention rule  $\mu$  satisfying attention overload, so that  $\pi$  is an AOM with  $(\succ, \mu)$ . Then, for any subset  $T \subseteq S$ ,  $\phi(a|S) \leq \phi(a|T) \leq \pi(U_{\succ}(a)|T)$ , which implies that  $\phi(a|S) \leq \min_{T \subseteq S} \pi(U_{\succ}(a)|T)$ . These observations give the following theorem.

**Theorem 2 (Revealed Attention).** Let  $\pi$  be an AOM and  $(\mu, \succ)$  represent  $\pi$ . Then, for every  $a$  and  $S$  such that  $a \in S$ ,

$$\max_{R \supseteq S} \pi(a|R) \leq \phi_{\mu}(a|S) \leq \min_{T \subseteq S: a \in T} \pi(U_{\succ}(a)|T).$$

We now consider three extreme cases of Theorem 2. If both the lower bound and the upper bound are 1, we say  $a$  attracts full attention at  $S$  (Revealed Full Attention). If both bounds are zero, then we say  $a$  does not attract any attention at  $S$  (Revealed Inattention). The third case happens when the lower bound is zero and the upper bound is one (No Revealed Attention). Indeed, these three cases are the only possibilities when the data is deterministic, which was studied by Lleras, Masatlioglu, Nakajima, and Ozbay (2017). However, they did not provide any characterization result for revealed attention. Theorem 2 provides such characterization not only for stochastic choice but also for its deterministic counterpart, and hence our theorem is also a novel contribution in the competing attention framework for deterministic choice theory.

Since the stochastic data is richer, Theorem 2 covers another interesting case, which we call Partial Revealed Attention: the upper bound is strictly below one and/or the lower bound is strictly above zero. To illustrate revealed attention, we revisit Example 1.

**Example 1. (continued)** Given the two possible preferences, we can have bounds on attention frequency. Here, we focus on the choice set  $\{a, b, c, d\}$ . By applying Theorem 2,  $\phi(a|\{a, b, c, d\})$  must be 0.05,  $\phi(b|\{a, b, c, d\})$  and  $\phi(c|\{a, b, c, d\})$  must be between 0.1 and

0.25 and  $\phi(d|\{a, b, c, d\})$  must be between 0.75 and 1.

In some cases, the attention frequency will be uniquely identified for certain alternatives. For instance, in addition to Example 1, we have  $\pi(c|R) = 0.25$  for some  $R \supseteq \{a, b, c, d\}$ . Then, the lower bound for  $\phi(c|\{a, b, c, d\})$ , while being free of the underlying preference, must be 0.25. Hence, the attention frequency is point-identified to be 0.25 since the upper bound is also 0.25.

Theorem 2 is useful in real world applications to inform a firm/government how much attention each product/policy receives among other options. While the lower bound can be interpreted as the pessimistic evaluation for attention, the upper bound captures optimistic evaluation. The question is whether these local pessimistic (optimistic) evaluations hold globally, that is, we ask whether there is an underlying attention rule  $\mu$  satisfying attention overload such that the attention frequencies agree with the pessimistic (optimistic) evaluations for every set. Due to the richness in attention rule allowed by our Assumption 1, it turns out that the answer is affirmative.

**Theorem 3 (Pessimistic Evaluation for Attention).** Let  $\pi$  be an AOM and  $(\succ, \mu)$  represent  $\pi$ . Then there exists a pessimistic attention rule  $\mu^*$  such that  $(\succ, \mu^*)$  is also an AOM representation of  $\pi$ . That is, for all  $S$ ,  $\phi_{\mu^*}(a|S) = \max_{R \supseteq S} \pi(a|R)$ .

This theorem concerns the pessimistic evaluation case, but an analogous result can be established for the optimistic evaluation. As a consequence, in econometrics language, Theorem 2 delivers the sharp identified set for  $\phi$ .

## 2.3 Comparison to Other Random Attention Models

We compare AOM to other existing (random) attention models: Tversky (1972), Manzini and Mariotti (2014), Brady and Rehbeck (2016), Aguiar (2017), Cattaneo, Ma, Masatlioglu, and Suleymanov (2020), and Demirkan and Kimya (2020). With the exception of Cattaneo, Ma, Masatlioglu, and Suleymanov (2020), which imposes a nonparametric restriction, all the other models introduce the idea of random limited attention with a parametric restriction on the attention rule. We show that none of these models can capture the attention overload assumption by comparing their underlying attention rules. Section SA-1 in the supplemental appendix provides further comparisons between AOM and the related literature.

Consider two individuals, Ann and Ben. In a larger decision problem,  $S$ , Ann pays attention to all alternatives with probability one (full attention,  $\mu_{\text{Ann}}(S|S) = 1$ ), while Ben experiences attention overload and focuses only on a single alternative  $a$  in  $S$  while ignoring the rest (limited attention,  $\mu_{\text{Ben}}(\{a\}|S) = 1$ ). We chose these two extreme cases to make our point clear. Existing evidence suggests that, as the size of the available options decreases, the phenomenon of choice overload becomes less evident, leading decision makers to overlook less alternatives. Hence, assuming  $|T| < |S|$ , Ann continues to exhibit full attention in  $T$  but Ben considers more alternatives in  $T$ . Table 1 summarizes the comparisons with the literature using these two decision makers.

The Independence Attention Model (IAM) of Manzini and Mariotti (2014) and “Elimination by Aspects” Model (EAM) (Tversky, 1972; Aguiar, 2017) make the same predictions

	Ann	Ben
	Larger set $S$	
	Full attention $\mu_{\text{Ann}}(S S) = 1$	Limited attention $\mu_{\text{Ben}}(\{a\} S) = 1$
	Predictions for a smaller set $a \in T \subset S$	
Manzini and Mariotti (2014)	$\mu_{\text{Ann}}(T T) = 1$	$\mu_{\text{Ben}}(\{a\} T) = 1$
Tversky (1972); Aguiar (2017)	$\mu_{\text{Ann}}(T T) = 1$	$\mu_{\text{Ben}}(\{a\} T) = 1$
Brady and Rehbeck (2016)	No restriction	$\mu_{\text{Ben}}(\{a\} T) = 1$
Cattaneo, Ma, Masatlioglu, and Suleymanov (2020)	No restriction	$\mu_{\text{Ben}}(\{a\} T) = 1$
Demirkan and Kimya (2020)	No restriction	No restriction
Attention Overload Model (AOM)	$\mu_{\text{Ann}}(T T) = 1$	Increasing Attention

Table 1: Predictions of each model for a smaller set  $T$  for Ann and Ben, given that they exhibit full attention and extremely limited attention in the larger set, respectively.

for Ann and Ben. [Manzini and Mariotti \(2014\)](#) considers a parametric model of limited attention where each alternative has a constant attention frequency. Full attention on the larger selection implies that the attention frequency for each alternative in  $S$  is one. (Some of the parametric models require the attention parameters be strictly between zero and one, but we can capture these examples either by allowing the parameters to be equal to zero and one, or by taking a limit.) “Elimination by Aspects” attention is an adaptation of [Tversky \(1972\)](#) into limited attention, where alternatives are exogeneously bundled into categories, and the decision maker considers each of these categories with certain probabilities. [Aguiar \(2017\)](#) characterizes a special case of this model with the default option where each category includes the default option. For Ann, both models make the same prediction, which is consistent with attention overload: Ann should pay full attention in  $T$ . However, these models do not allow Ben to be more attentive for smaller sets: Ben must pay attention to the singleton  $\{a\}$  with probability one. In IAM, this is because the attention frequencies for other alternatives in  $S$  are zero, while in EAM,  $a$  is the only alternative belonging to the most popular category. Hence, [Manzini and Mariotti \(2014\)](#) and ([Tversky, 1972](#); [Aguiar, 2017](#)) are too restrictive to accommodate attention overload. Furthermore, [Demirkan and Kimya \(2020\)](#) drops the menu-independence assumption in IAM, which leads to no restriction on the attention rule for  $T$ , and hence cannot accommodate attention overload either.

The Logit Attention Model (LAM) of [Brady and Rehbeck \(2016\)](#) and the Random Attention Model (RAM) of [Cattaneo, Ma, Masatlioglu, and Suleymanov \(2020\)](#) make the same prediction for our individuals. LAM could be interpreted as a parametric limited attention model where each subset could be the consideration set with some probability. Since Ann exhibits full attention in the larger set,  $S$  must be the most probable consideration set, and since  $S$  is not a subset of  $T$ , Ann is allowed to focus on any subset of  $T$ , including a single alternative. For Ben, this model implies Ben must continue to pay attention only to

*a.* RAM imposes a (nonparametric) monotonicity assumption on the attention rule, so that  $\mu(T|S) \leq \mu(T|S \setminus b)$  for every  $b \notin T \subseteq S$ . In RAM, the monotonic attention does not impose any restriction on Ann’s behavior (for example, it could be  $\mu_{\text{Ann}}(\{a\}|T) = 1$ ), but Ben must pay attention to the same single alternative ( $\mu_{\text{Ben}}(\{a\}|T) = 1$ ). In other words, Ann can exhibit the opposite of choice overload, while Ben must exhibit limited attention even in a smaller problem. Therefore, these two models stand in contrast to the concept of attention overload.

In contrast to the existing literature, our AOM introduces a novel nonparametric assumption on attention frequency, which measures how much attention each alternative (rather than each consideration set) receives. That is,  $\phi(b|S) \leq \phi(b|S \setminus c)$  for every  $b \in S$ . If Ann behaves according to AOM,  $\mu_{\text{Ann}}(S|S) = 1$  implies  $\phi_{\text{Ann}}(b|S) = 1$  for all  $b$ , which dictates  $\phi_{\text{Ann}}(b|T) = 1$  for all  $b \in T$  (full attention). For Ben, AOM imposes that he must consider  $a$  for sure ( $\phi_{\text{Ben}}(a|T) = 1$ ), but attention frequency for other alternatives can increase. In particular, AOM allows that  $\phi_{\text{Ben}}(b|T) = 1$  for all  $b \in T$  (full attention). This discussion makes it clear that our attention overload property is distinct from all other (parametric or nonparametric) random limited attention models. As a consequence, our AOM captures novel empirical findings and describes novel attention allocation behaviors compared to the existing models in the literature.

## 2.4 Econometric Methods

We obtained several testable implications and related results for the AOM: Theorem 1, Propositions 1 and 2, and Theorem 2. Our next goal is to develop econometric methods to implement these findings using real data, which can help elicit preferences, conduct empirical testing of our AOM, and provide confidence sets for attention frequencies. To this end, we rely on a random sample of observations consisting of choice data for  $n$  units indexed by  $i = 1, 2, \dots, n$ . Each unit faces a choice problem  $Y_i$ , and her choice is denoted by  $y_i \in Y_i$ . This is formally stated in the assumption below. We recall that  $\mathcal{D} \subseteq 2^X \setminus \emptyset$  is a collection of choice problems.

**Assumption 2 (Choice Data).** The data consists of a random sample of choice problems and observed choices  $\{(Y_i, y_i) : 1 \leq i \leq n\}$  with  $Y_i \in \mathcal{D}$  and  $\mathbb{P}[y_i = a|Y_i = S] = \pi(a|S)$ .

A given preference ordering  $\succ$  is compatible with our AOM if and only if  $\succ$ -Regularity holds. In other words, each preference ordering corresponds to a collection of inequality constraints, which we collect in the following hypothesis:

$$\mathbf{H}_0 : \max_{\substack{a \in T \subseteq S \\ T, S \in \mathcal{D}}} D(a|S, T) \leq 0, \quad \text{where } D(a|S, T) = \pi(a|S) - \pi(U_{\succeq}(a)|T). \quad (3)$$

To construct a test statistic for testing the hypotheses in (3), we can replace the unknown choice probabilities by their estimates,  $\hat{\pi}(a|S)$  and  $\hat{\pi}(U_{\succeq}(a)|T)$ , leading to  $\hat{D}(a|S, T) =$



$\widehat{\pi}(a|S) - \widehat{\pi}(U_{\succeq}(a)|T)$ . For example,

$$\widehat{\pi}(a|S) = \frac{1}{N_S} \sum_{i=1}^n \mathbb{1}(Y_i = S, y_i = a), \quad N_S = \sum_{i=1}^n \mathbb{1}(Y_i = S).$$

We define the following test statistic

$$\mathbb{T}(\succ) = \max \left\{ \max_{\substack{a \in T \subseteq S \\ T, S \in \mathcal{D}}} \frac{\widehat{D}(a|S, T)}{\widehat{\sigma}(a|S, T)}, 0 \right\},$$

where  $\widehat{\sigma}(a|S, T)$  is the standard error of  $\widehat{D}(a|S, T)$  and takes the form

$$\widehat{\sigma}(a|S, T) = \sqrt{\frac{1}{N_S} \widehat{\pi}(a|S)(1 - \widehat{\pi}(a|S)) + \frac{1}{N_T} \widehat{\pi}(U_{\succeq}(a)|T)(1 - \widehat{\pi}(U_{\succeq}(a)|T))}.$$

The outer max operation in  $\mathbb{T}(\succ)$  guarantees that we will never reject the null hypothesis if none of the estimated differences  $\widehat{D}(a|S, T)$  are strictly positive. In other words, a preference is not ruled out by our analysis if  $\succ$ -Regularity holds in the sample. The statistic depends on a specific preference ordering which we would like to test against for: such dependence is explicitly reflected by the notation  $\mathbb{T}(\succ)$ .

We investigate the statistical properties of the test statistic in order to construct valid inference procedures, building on the recent work of [Chernozhukov, Chetverikov, and Kato \(2019\)](#) and [Chernozhukov, Chetverikov, Kato, and Koike \(2022\)](#) for many moment inequality testing. (See also [Molinari \(2020\)](#) for an overview and further references.) We seek for a critical value, denoted by  $\text{cv}(\alpha, \succ)$ , such that under the null hypothesis (i.e., when the preference is compatible with our AOM),

$$\mathbb{P}[\mathbb{T}(\succ) > \text{cv}(\alpha, \succ)] \leq \alpha + \mathfrak{r}_{\succ}, \quad (4)$$

where  $\alpha \in (0, 1)$  denotes the desired significance level of the test, and  $\mathfrak{r}_{\succ}$  denotes a quantifiable error of approximation (which should vanish in large samples with possibly many inequalities).

To provide some intuition on the critical value construction, the Studentized test statistic is approximately normally distributed in large samples as follows:

$$\frac{\widehat{D}(a|S, T)}{\widehat{\sigma}(a|S, T)} \underset{a}{\sim} \text{Normal} \left( \frac{D(a|S, T)}{\sigma(a|S, T)}, 1 \right).$$

Since  $D(a|S, T) \leq 0$  under the null hypothesis, the above normal distribution will be first-order stochastically dominated by the standard normal distribution. Letting  $\widehat{\mathbf{D}}$  be the column vector collecting all  $\widehat{D}(a|S, T)$ , and  $\mathbf{\Omega}$  be its correlation matrix, then our test statistic  $\mathbb{T}(\succ)$  will be dominated by the maximum of a normal vector with a zero mean and a variance of  $\mathbf{\Omega}$ , up to the error from normal approximation. Using properties of Bernoulli random variables, an estimate of  $\mathbf{\Omega}$  can be constructed with the estimated choice probabilities and

the effective sample sizes. We denote the estimated correlation matrix by  $\widehat{\Omega}$ . The critical value is then the  $(1 - \alpha)$ -quantile of the maximum of a Gaussian vector, and is precisely defined as

$$cv(\alpha, \succ) = \inf \left\{ t : \mathbb{P}[\mathbb{T}^g(\succ) \leq t | \text{Data}] \geq 1 - \alpha \right\}, \quad \mathbb{T}^g(\succ) = \max \left\{ \max(\widehat{\Omega}^{1/2} \mathbf{z}), 0 \right\},$$

where the inner max operation computes the maximum over the elements of  $\widehat{\Omega}^{1/2} \mathbf{z}$ , and  $\mathbf{z}$  denotes a standard normal random vector of suitable dimension. Precise definitions and omitted details are given in the supplemental appendix to conserve space. The theorem below offers formal statistical guarantee on the validity of our proposed test.

**Theorem 4 (Preference Elicitation with Theorem 1).** Assume Assumption 2 holds. Let  $\mathbf{c}_1$  be the number of comparisons (i.e., inequalities) in (3), and  $\mathbf{c}_2 = \left( \min_{S \in \mathcal{D}} N_S \right) \cdot \left( \min_{\substack{a \in T_{CS} \\ T, S \in \mathcal{D}}} \sigma(a|S, T) \right)$ . Then, under the null hypothesis  $H_0$  in (3), the error in rejection probability in (4) holds with  $\mathbf{r}_\succ = C \cdot \left( \log^5(n\mathbf{c}_1)/\mathbf{c}_2^2 \right)^{1/4}$ , where  $C$  denotes an absolute constant.

This theorem shows that the error in distributional approximation,  $\mathbf{r}_\succ$ , only depends on the dimension of the problem (i.e.,  $\mathbf{c}_1$ ) logarithmically, and therefore our estimation and inference procedures remain valid even if the test statistic  $\mathbb{T}(\succ)$  involves comparing “many” pairs of estimated choice probabilities. By providing a statistical guarantee as in (4), our procedures can accommodate situations where both the number of alternatives and the number of choice problems are large, and hence they are expected to perform well in finite samples, leading to more robust welfare analysis results and policy recommendations.

It is routine to incorporate condition (2) into our econometric implementation. Specifically, the test statistic and the critical value we introduced above are based on moment inequality testing. To accommodate the new assumption on attentive at binaries, one only needs to include additional probability comparisons corresponding to the  $\eta$ -constrained revealed preference.

Given the testing procedures we developed, it is easy to construct valid confidence sets by test inversion. To be precise, a dual asymptotically valid  $100(1 - \alpha)\%$  level confidence set is

$$\text{CS}(1 - \alpha) = \left\{ \succ : \mathbb{T}(\succ) \leq cv(\alpha, \succ) \right\}.$$

Therefore, for any preference  $\succ$  that is compatible with our AOM, we have the statistical guarantee on coverage:  $\mathbb{P}[\succ \in \text{CS}(1 - \alpha)] \geq 1 - \alpha - \mathbf{r}_\succ$ .

Revealed preference between two alternatives,  $a$  and  $b$ , can be analyzed based on Theorem 1 by checking, say, if  $a \succ b$  for all identified preferences  $\succ$  in  $\text{CS}(1 - \alpha)$ . This approach to revealed preference has the advantage that it exhausts the identification power of our AOM. On the other hand, it comes with a nontrivial computational cost as discussed in Section 2.1. Therefore, we also discuss how Proposition 1 can be implemented in practice. The econometric methods stemming from Theorem 1 and Proposition 1 are complementary: while the former provides a more systematic framework for preference revelation, the latter can be handy if binary comparisons are available in the data or if the analyst is particularly

interested in inferring preference ordering among pairs of alternatives. (Also see Example 1, which demonstrates that applying Proposition 1 first to a choice data may greatly reduce the number of preference orderings to be tested against  $\succ$ -Regularity.)

We fix two alternatives, say  $a$  and  $b$ , and let  $\mathcal{D}_{ab}$  be the collection of choice problems containing both  $a$  and  $b$ , excluding the binary comparison; that is,  $\mathcal{D}_{ab} = \{S \in \mathcal{D} : S \not\supseteq \{a, b\}\}$ . We also recall the simplified notation  $\pi(a|b) = \pi(a|\{a, b\})$  for binary comparisons. Then, we may deduce the preference ordering  $b \succ a$  if we are able to reject the following null hypothesis:

$$H_0 : \max_{S \in \mathcal{D}_{ab}} D(a|S, b) \leq 0, \quad \text{where } D(a|S, b) = \pi(a|S) - \pi(a|b). \quad (5)$$

Constructing a test statistic is straightforward:

$$T(ab) = \max \left\{ \max_{S \in \mathcal{D}_{ab}} \frac{\widehat{D}(a|S, b)}{\widehat{\sigma}(a|S, b)}, 0 \right\},$$

where  $\widehat{D}(a|S, b) = \widehat{\pi}(a|S) - \widehat{\pi}(a|b)$ , and  $\widehat{\sigma}(a|S, b)$  is its standard error. We employ the same technique to construct a critical value. Letting  $\widehat{\mathbf{D}}$  be the column vector collecting all  $\widehat{D}(a|S, b)$ , and  $\widehat{\mathbf{\Omega}}$  be its estimated correlation matrix. The critical value is then defined as

$$cv(\alpha, ab) = \inf \left\{ t : \mathbb{P}[T^g(ab) \leq t | \text{Data}] \geq 1 - \alpha \right\}, \quad T^g(ab) = \max \left\{ \max(\widehat{\mathbf{\Omega}}^{1/2} \mathbf{z}), 0 \right\}.$$

The following result is a corollary of Theorem 4.

**Corollary 2 (Preference Elicitation with Proposition 1).** Assume Assumption 2 holds. Let  $\mathbf{c} = \left( \min \{ \min_{S \in \mathcal{D}_{ab}} N_S, N_{\{a, b\}} \} \right) \cdot \left( \min_{S \in \mathcal{D}_{ab}} \sigma(a|S, b) \right)$ . Then, under the null hypothesis  $H_0$  in (5), the error in rejection probability  $\mathbb{P}[T(ab) > cv(\alpha, ab)] \leq \alpha + \mathbf{r}_{ab}$  holds with  $\mathbf{r}_{ab} = C \cdot (\log^5(n|\mathcal{D}_{ab}|)/\mathbf{c}^2)^{1/4}$ , where  $C$  denotes an absolute constant.

One of our main econometric contributions is a careful study of the properties of the estimated correlation matrix,  $\widehat{\mathbf{\Omega}}$ , where we provide an explicit bound on the supremum of the entry-wise estimation error  $\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}\|_\infty$ . This is further combined with the results in Chernozhukov, Chetverikov, Kato, and Koike (2022) to establish a normal approximation for the centered and normalized inequality constraints, say  $(\widehat{D}(a|S, T) - D(a|S, T))/\widehat{\sigma}(a|S, T)$ . Cattaneo, Ma, Masatlioglu, and Suleymanov (2020) considered preference elicitation under a monotonic attention assumption, and proposed estimation and inference procedures based on pairwise comparison of choice probabilities as in (3). However, their econometric analysis assumed “fixed dimension,” and hence did not allow the complexity of the problem to be “large” relative to the sample size, which is required in this paper.

We now discuss how to operationalize the partial identification result in Theorem 2 on attention frequency. We will illustrate with the lower bound,  $\phi(a|S) \geq \max_{R \supseteq S} \pi(a|R)$ , since the upper bound follows analogously. A naïve implementation would replace the unknown choice probabilities by their estimates. Unfortunately, the uncertainty in the estimated choice probabilities will be amplified by the maximum operator, leading to over-estimated

lower bounds. Our aim is to provide a construction of the lower bound, denoted as  $\widehat{\phi}(a|S)$  such that

$$\mathbb{P} \left[ \phi(a|S) \geq \widehat{\phi}(a|S) \right] \geq 1 - \alpha + \mathbf{r}_{\phi(a|S)} \quad (6)$$

with  $\alpha \in (0, 1)$  denoting the desired significance level, and  $\mathbf{r}_{\phi(a|S)}$  denoting the error in approximation, which ideally should become smaller as the sample size increases.

Our construction is based on computing the maximum over a collection of adjusted empirical choice probabilities. To be very precise, we define

$$\widehat{\phi}(a|S) = \max_{R \supseteq S, R \in \mathcal{D}} \left\{ \widehat{\pi}(a|R) - \text{cv}(\alpha, \phi(a|S)) \cdot \widehat{\sigma}(a|R) \right\},$$

where  $\widehat{\sigma}(a|R) = \sqrt{\widehat{\pi}(a|R)(1 - \widehat{\pi}(a|R))/N_R}$  is the standard error of the estimated probability  $\widehat{\pi}(a|R)$ , and

$$\text{cv}(\alpha, \phi(a|S)) = \inf \left\{ t : \mathbb{P} \left[ \max(\mathbf{z}) \leq t \right] \geq 1 - \alpha \right\},$$

with  $\mathbf{z}$  is a standard normal random vector of dimension  $|\{R \in \mathcal{D} : R \supseteq S\}|$ , the number of supersets of  $S$ .

To provide some intuition for the construction, we begin with the normal approximation:  $\widehat{\pi}(a|R) \overset{a}{\sim} \text{Normal}(\pi(a|R), \sigma(a|R))$ . Then, the estimated choice probabilities are mutually independent since they are constructed from different subsamples, and

$$\begin{aligned} & \mathbb{P} \left[ \widehat{\pi}(a|R) - \text{cv}(\alpha, \phi(a|S)) \cdot \widehat{\sigma}(a|R) \leq \pi(a|R), \forall R \supseteq S, R \in \mathcal{D} \right] \\ &= \mathbb{P} \left[ \max_{R \supseteq S, R \in \mathcal{D}} \frac{\widehat{\pi}(a|R) - \pi(a|R)}{\widehat{\sigma}(a|R)} \leq \text{cv}(\alpha, \phi(a|S)) \right] \\ &\approx \mathbb{P} \left[ \max(\mathbf{z}) \leq \text{cv}(\alpha, \phi(a|S)) \right] = 1 - \alpha. \end{aligned}$$

where  $\approx$  denotes an approximation in large samples. Heuristically, the above demonstrates that with high probability (approximately  $1 - \alpha$ ) the true choice probabilities,  $\pi(a|R)$ , are bounded from below by  $\widehat{\pi}(a|R) - \text{cv}(\alpha, \phi(a|S)) \cdot \widehat{\sigma}(a|R)$ . This is made possible by the adjustment term we added to the estimated choices probabilities. This idea is formalized in the following theorem, which offers precise probability guarantees.

**Theorem 5 (Attention Frequency Elicitation with Theorem 2).** Let  $\pi$  be an AOM, and assume Assumption 2 holds. Define  $\mathbf{c}_1 = |\{R \in \mathcal{D} : R \supseteq S\}|$  to be the number of supersets of  $S$ , and  $\mathbf{c}_2 = \left( \min_{R \supseteq S, R \in \mathcal{D}} N_R \right) \cdot \left( \min_{R \supseteq S, R \in \mathcal{D}} \sigma(a|R) \right)$ . Then (6) holds with  $\mathbf{r}_{\phi(a|S)} = C \cdot \left( \log^5(n\mathbf{c}_1)/\mathbf{c}_2^2 \right)^{1/4}$ , where  $C$  denotes an absolute constant.

Due to space limitations, the supplemental appendix reports simulation evidence showcasing the empirical performance of our theoretical and methodological econometric results. To be more precise, we test against four preference orderings using simulated data, and in this process, we vary the number of choice problems available ( $\mathcal{D}$ ) in the data, and the effective sample size of each choice problem ( $N_S$ ). Overall, our procedure performs well: it is

able to reject preference orderings that are not compatible with  $\succ$ -Regularity with nontrivial power, while at the same time maintaining size.

### 3 Heterogeneous Preferences over a List

The analysis so far considered homogeneous preferences, and hence our choice model assumed that every decision maker has the same taste but different levels of attentiveness. This section builds on our previous results and presents a model that describes the choice behavior at both the individual and population levels allowing for preference heterogeneity in the underlying data generating process. Our heterogeneous preference choice model can be used empirically with aggregate data on a group of distinct decision makers where each of them may differ not only on what they pay attention to but also on what they prefer. The model also allows heterogeneous preferences that may correlate with attention, full independence being a special case.

Limited attention presents particular difficulties for both identifying and drawing econometric conclusions about the decision maker’s preferences. Furthermore, it is well-known that, even under full attention, the classical Random Utility Model (RUM) suffers from non-uniqueness due to varying tastes (Fishburn, 1998; Turansick, 2022). Therefore, simply adding heterogeneous preferences to our AOM renders the identification exercise hopeless without additional structure. To address this challenge, we impose a nonparametric restriction on possible preference types and consideration sets to discipline our proposed choice model with heterogeneous preferences: our key idea is that alternatives are presented to the decision maker as a list that correlates with both heterogeneous preferences and random (limited) attention. Our modeling strategy that individuals come across various options presented as a list is motivated by real-world scenarios. For example, customers often browse Amazon’s ordered search results or receive a ranked list of advertisements, not to mention decision makers employing web search engines.

We first assume that the list is observable. This assumption may be reasonable in situations where we can observe Amazon’s product list for each product category, a ballot for a specific election, Google’s search results for a keyword, etc. In Section 3.4, we relax this assumption and endogenize the list, allowing for identification of heterogeneous preferences when the true underlying list is unknown to the researcher.<sup>4</sup>

#### 3.1 List-Based Attention Overload

Consider individuals who encounter a ranked list of results provided by a search platform, where each individual in the population faces the same list. Formally, we assume there is

---

<sup>4</sup>The supplementary appendix (Section SA-1) gives a review of the related literature on limited attention and heterogeneous preferences (e.g., Aguiar, Boccardi, Kashaev, and Kim, 2023; Kashaev and Aguiar, 2022; Dardanoni, Manzini, Mariotti, and Tyson, 2020; Gibbard, 2021; Dardanoni, Manzini, Mariotti, Petri, and Tyson, 2023; Barseghyan, Molinari, and Thirkettle, 2021; Abaluck and Adams, 2021; Barseghyan and Molinari, 2023), and choice over list (e.g., Rubinstein and Salant, 2006; Guney, 2014; Yildiz, 2016; Ishii, Kovach, and Ülkü, 2021).

a list of items represented by the linear order  $\triangleright$ . Let  $\langle a_1, a_2, \dots, a_{|X|} \rangle$  be the enumeration of the elements in  $X$  with respect to  $\triangleright$ , where  $a_j$  denotes the item in the  $j$ th position, and  $|X|$  is the size of the grand set  $X$ . In other words,  $j < k$  implies that  $a_j$  appears earlier in the list than  $a_k$ , which is equivalent to saying  $a_j \triangleright a_k$ . We will use both notation,  $j < k$  and  $a_j \triangleright a_k$ , interchangeably. For a choice problem  $S \subseteq X$ , we also enumerate its elements as  $\langle a_{s_1}, a_{s_2}, \dots, a_{s_{|S|}} \rangle$ .

It is often impractical for consumers to conduct exhaustive searches because their attention is limited. Through the list, a decision maker investigates alternatives to construct her consideration set but she might consider only a subset of the alternatives available to her due to limited attention. Our proposed model will impose three basic behavioral restrictions on the consideration set formation for a given list. Let  $\Gamma(S)$  be the consideration set when the choice problem is  $S$ . First, we assume that the consideration set obeys the underlying order: if  $a_k$  belongs to the consideration set, so does every feasible alternative that appeared before  $a_k$ , i.e.,  $a_k \in \Gamma(S)$  and  $a_j \in S$  such that  $j < k$  imply  $a_j \in \Gamma(S)$ . In words, whenever an alternative is considered, all alternatives in the list before it are also taken into account. Second, we assume that alternatives are in competition for consumers' attention: if  $a$  belongs to the consideration set of a larger choice problem, then  $a$  will be considered for smaller choice problems, i.e., if  $a \in \Gamma(S)$  and  $a \in T \subset S$ , then  $a \in \Gamma(T)$ . Intuitively, if an alternative is not recognized in a smaller set, then it cannot be recognized in a larger set. Finally, we assume that each individual always considers both items in binary problems, i.e.,  $\Gamma(S) = S$  whenever  $|S| = 2$ . We collect these conditions in the following definition.

**Definition 3 (List-based Attention Overload).** A deterministic consideration set mapping  $\Gamma$  satisfies list-based attention overload on  $\triangleright$  (i.e.,  $\langle a_1, a_2, \dots, a_{|X|} \rangle$ ) if (i) for every  $T \subseteq S$ ,  $a_k \in \Gamma(S)$  and  $a_j \in T$ ,  $j \leq k$  implies  $a_j \in \Gamma(T)$ ; (ii)  $\Gamma(S) = S$  whenever  $|S| = 2$ .

Any  $\Gamma$  satisfying list-based attention overload is an attention rule satisfying our attention overload assumption. We denote  $\mathcal{AO}_\triangleright$  as the set of all consideration set rules satisfying list-based attention overload with respect to  $\triangleright$ .

Individuals are also heterogeneous in terms of their preferences. Unlike RUM, we assume that the set of preferences is related to the underlying list. First, our model recognizes that some individuals perceive search results as reflecting the true quality of listed items. Indeed, many commercial websites collect individual consumers' behavioral data and try to match each consumer with personally relevant products. The list can be thought of as the outcome of personalized recommendations. Individuals facing the same list share similar tastes. However, our model also captures the idea that individuals might favor their status quo, meaning that they assign a relatively higher rank to their reference point compared to other items in the original list. This assumption restricts potential preferences exhibited in the model. For all  $j < k$ , define  $\succ_{kj}$  as a linear order where the  $k$ th alternative in  $\triangleright$  is moved to the  $j$ th position. To give some examples,  $\succ_{21}$  corresponds to the ordering  $\langle a_2, a_1, a_3, a_4, \dots, a_{|X|} \rangle$ , and  $\succ_{42}$  is  $\langle a_1, a_4, a_2, a_3, \dots, a_{|X|} \rangle$ . We call  $\succ_{kj}$  a single improvement of  $\triangleright$ , and we use  $\mathcal{P}_\triangleright$  to denote the set of all single improvements of  $\triangleright$  including  $\triangleright$  itself. For notational convenience, we let  $\succ_{kk} = \triangleright$ . If  $\succ \in \mathcal{P}_\triangleright$ , this implies that there exists a unique alternative  $a_k$  such that its relative ranking improved with respect to  $\triangleright$ . Orderings involving multiple changes to the original list order  $\triangleright$ , such as  $\langle a_2, a_1, a_4, a_3, \dots, a_{|X|} \rangle$ , are not allowed.

Each pair  $(\Gamma, \succ) \in \mathcal{AO}_\triangleright \times \mathcal{P}_\triangleright$  induces a deterministic choice function, which corresponds to a particular type in our model. Let  $\tau$  be a probability distribution on  $\mathcal{AO}_\triangleright \times \mathcal{P}_\triangleright$ , so  $\tau(\Gamma, \succ)$  is the probability of  $(\Gamma, \succ)$  being the choice type. Each  $\tau$  naturally induces a probabilistic choice function:

$$\pi(a|S) = \tau\left(\left\{(\Gamma, \succ) \in \mathcal{AO}_\triangleright \times \mathcal{P}_\triangleright : a \text{ is } \succ\text{-best in } \Gamma(S)\right\}\right). \quad (7)$$

The probabilistic choice function induced by  $\tau$  sets the probability of an alternative  $a$  being chosen from an alternative set  $S$  as the sum of probabilities of choice types that select  $a$  from  $S$ .

Our goal is to tackle the “inverse problem”: given a probabilistic choice function  $\pi$  (which is identified from a choice data), whether it can be represented with our list-based attention overload model. We first state the precise definition.

**Definition 4 (Heterogeneous Preference Attention Overload).** We say that a probabilistic choice function  $\pi$  has a *Heterogeneous Preference Attention Overload* representation with respect to  $\triangleright$  (HAOM $_\triangleright$ ) if there exists  $\tau$  on  $\mathcal{AO}_\triangleright \times \mathcal{P}_\triangleright$  such that (7) holds.

HAOM $_\triangleright$  introduces heterogeneity both in terms of preferences and attention. This feature makes this model independent of RUM. (If the support of  $\tau$  consists of only choice types with  $\Gamma(S) = S$ , then the model becomes a special case of RUM.) Due to limited attention, HAOM $_\triangleright$  allows choice types outside of the preference maximization paradigm, and therefore it captures behaviors outside of RUM. On the other hand, since RUM allows more preference types, some choice behaviors can be only captured by RUM but not by HAOM $_\triangleright$ . Having said that, HAOM $_\triangleright$  still encompasses more choice types than RUM, even though we restrict the set of possible preferences. This is because HAOM $_\triangleright$  considers two types of heterogeneity (attention and preferences), while RUM only allows for heterogeneity in preference ordering. Our AOM can be regarded as another extreme of HAOM $_\triangleright$ , as it requires that every choice type has the same preferences but it imposes a mild nonparametric restriction on attention. See Section 3.3 for more discussion.

## 3.2 Characterization with Known List

We now provide a list of behavioral postulates describing the implications of HAOM $_\triangleright$  for a given list  $\triangleright$ . Since the model is more involved in terms of types, the following results require that  $\mathcal{D}$  includes all non-empty subsets of  $X$ , which is a common assumption on choice models that involve random utility.

We first highlight that HAOM $_\triangleright$  allows violations of the regularity condition, while regularity always holds in RUM. However, HAOM $_\triangleright$  limits the types of regularity violations that are permissible. For example, there will be no regularity violation as long as the first alternative in the list is always present. This is because only choice types  $\{\Gamma : a_k \in \Gamma(S)\} \times \{\succ_{k1}\}$  will pick  $a_k$  in the presence of  $a_1$  (assuming  $k > 1$ ), and they will continue choosing  $a_k$  in a smaller choice problem  $T \subset S$ . Indeed, we can generalize this intuition: removing alternatives will not decrease the choice probabilities of a product as long as there is another product listed before it in both decision problems. Consider two products  $a_k$  and  $a_j$  such that

$j < k$ , then the choice probability of  $a_k$  obeys regularity conditions, i.e.,  $\pi(a_k|T) \geq \pi(a_k|S)$  for  $T \subset S$  provided that  $a_j \in T$ . This is our first behavioral postulate for  $\text{HAOM}_\triangleright$ .

**Axiom 2 (List-Regularity).** For all  $a_j, a_k \in T \subset S$  with  $j < k$ ,  $\pi(a_k|T) \geq \pi(a_k|S)$ .

The following property imposes a structure on binary choices. It simply says everything else equal, being listed earlier increases choice probabilities on binary comparisons. For example, compared to the third product in the list, the first product is chosen more frequently than the second one. In other words, binary choice probabilities decrease as the opponent is ranked higher in the list. For example,  $a_3$  is going to be chosen against  $a_2$  more often than against  $a_1$ . This is because the individual considers both alternatives in every binary comparison. Hence, being listed earlier is reflected in choice probabilities. We adopt the following notation for binary comparisons:  $\pi(a_k|a_\ell) := \pi(a_k|\{a_k, a_\ell\})$ .

**Axiom 3 (List-Monotonicity).** For all  $a_j, a_k, a_\ell$  such that  $j < k < \ell$ ,  $\pi(a_\ell|a_k) \geq \pi(a_\ell|a_j)$ .

Again consider binary comparisons. Assume we have  $\pi(a_2|a_1) = 0.6$ . This implies that the frequency of  $\succ_{21}$  must be 0.6, and those types must prefer  $a_2$  over  $a_3$  as well (our model only allows preferences that are single improvements over the original list order). Hence,  $\pi(a_3|a_2)$  must be smaller than  $0.4 = 1 - 0.6$ . The next axiom generalizes this intuition: the total binary choice probabilities against the immediate predecessor in the list must be less than or equal to 1.

**Axiom 4 (List-Boundedness).**  $\sum_{j=2}^{|X|} \pi(a_j|a_{j-1}) \leq 1$ .

$\text{HAOM}_\triangleright$  introduces some compatibility among all the preference types in the population because each preference type is a single improvement of  $\triangleright$ . However, it allows for a significant level of heterogeneity in attention. Our next theorem indicates that  $\text{HAOM}_\triangleright$  can still make predictions because it states that the three axioms completely characterize  $\text{HAOM}_\triangleright$ .

**Theorem 6 (Characterization).** Given  $\triangleright$ , a choice rule  $\pi$  satisfies Axioms 2–4 if and only if  $\pi$  has an  $\text{HAOM}_\triangleright$  representation.

Theorem 6 establishes both a necessary and sufficient condition for  $\text{HAOM}_\triangleright$ . The importance of this theorem lies in its applicability to any dataset, even when RUM is not applicable. This makes it a powerful tool for studying choice behaviors beyond utility maximization. In contrast to RUM, our choice model enjoys strong identification for preferences in  $\mathcal{P}_\triangleright$ : the frequency of each preference type in  $\mathcal{P}_\triangleright$  is uniquely (point) identified. To state it formally, we define the preference type frequency for each  $\succ$ ,  $\tau(\succ) := \tau(\{(\Gamma, \succ) : \Gamma \in \mathcal{AO}_\triangleright\})$ ; that is,  $\tau(\succ)$  represents the total probability of  $\succ$  being the underlying preferences. If both  $\tau_1$  and  $\tau_2$  are  $\text{HAOM}_\triangleright$  representations of  $\pi$ , then  $\tau_1(\succ) = \tau_2(\succ)$ . The uniqueness of  $\text{HAOM}_\triangleright$  is in sharp contrast to the non-uniqueness of RUM.

Given the uniqueness result, we now ask whether we can reveal the frequency of specific preference types. Our revealed preference serves (at least) two purposes. First, it identifies the specific form of heterogeneity in the population in terms of preferences. Second, it provides a unique weight for each preference type within this heterogeneity. For example, by



using these weights, a policymaker can evaluate how a particular policy affects each agent in the heterogeneous population and then combine these effects with precisely identified weights. Furthermore, when we interpret  $\succ_{kj}$  as the status quo bias preferences,  $\tau(\succ_{kj})$  is the frequency of people whose default option is  $a_k$  and the strength of bias is  $j - k$  (the difference between the original and final ranking of the default  $a_k$ ). Hence, our identification result can be used to measure the status quo bias in the data. For the theorem below, we recall the adopted convention that  $\succ_{11} := \triangleright$ , which corresponds to choice types whose preference coincides with the original list order.

**Theorem 7 (Revealed Preference Types).** Let  $\tau$  be a  $\text{HAOM}_\triangleright$  representation of  $\pi$ . Then,

$$\tau(\succ_{kj}) = \begin{cases} \pi(a_k|a_j) - \pi(a_k|a_{j-1}) & \text{if } k > j > 1 \\ \pi(a_k|a_j) & \text{if } k > j = 1 \end{cases},$$

and  $\tau(\succ_{11}) = 1 - \sum_{k=2}^{|X|} \pi(a_k|a_{k-1})$ .

To provide some intuition, recall from Definition 3 that decision makers pay full attention in binary comparisons. This implies that for  $j < k$ , the choice probability  $\pi(a_k|a_j)$  is just the frequency of choice types who rank  $a_k$  at or higher than the  $j$ th position; that is,  $\pi(a_k|a_j) = \sum_{\ell \leq j} \tau(\succ_{k\ell})$ . This observation justifies the first part of the theorem. Now take  $j = k - 1$ . Then  $\pi(a_k|a_{k-1})$  is the frequency of choice types who do not agree with  $\triangleright$  on the ranking of  $a_k$ , which leads to the second conclusion of the theorem.

The identification result provided by Theorem 7 is based on binary choices. While it provides point identification, it can only be used when all binary comparisons  $\{a_k, a_j\}$  such that  $k > j$  are available in the data. When the data is incomplete, in the sense that not all possible comparisons are observed (or identifiable and estimable in econometrics language), we can provide bounds for the frequency of preference types: if a non-top alternative is chosen, it must be attributed to the types who like that alternative better than the top alternative. In this sense, the choice probability is the lower bound of those types since some of these decision maker types might not have looked far enough down the list (i.e., random and limited attention).

**Proposition 3 (Bounds for Preference Types).** Let  $\tau$  be a  $\text{HAOM}_\triangleright$  representation of  $\pi$ . Fix  $S$  and let  $a_{s_1}$  be its top-listed item. Then, for  $k > s_1$ ,  $\pi(a_k|S) \leq \tau(\{(\Gamma, \succ) : \Gamma \in \mathcal{AO}_\triangleright \text{ and } a_k \succ a_{s_1}\})$ .

To close this subsection, we demonstrate how to perform revealed attention analysis on our choice model with heterogeneous preference and random attention; c.f., Section 2.2. Given  $\tau$ , we define the attention frequency for an alternative  $\phi_\tau(a|S) := \tau(\{(\Gamma, \succ) : a \in \Gamma(S)\})$ ; that is,  $\phi_\tau(a|S)$  represents the total probability of  $a$  being considered in the population when the choice set is  $S$ . Under RUM, this is assumed to be equal to one for all alternatives in  $S$  (full attention). In our model, while full attention holds for the first alternative on the list, it might not hold for other alternatives. Even though we cannot point identify the attention frequency for other alternatives, we provide upper and lower bounds. For notation convenience, we will drop the subscript and use  $\phi(a|S) = \phi_\tau(a|S)$ .

**Theorem 8 (Revealed Attention).** Let  $\tau$  be a  $\text{HAOM}_\triangleright$  representation of  $\pi$ . Fix  $S$  and let  $a_{s_1}$  be its top-listed item. Then, (i)  $\phi(a_{s_1}|S) = 1$ ; (ii) for  $a_k \in S$  and  $k > s_1$ ,

$$\max_{R \supseteq S} \sum_{\ell \geq k} \pi(a_\ell | R) \leq \phi(a_k | S) \leq 1 - \sum_{s_1 < j \leq k: a_j \in S} \left( \max_{R \supseteq \{a_{s_1}, a_j\}} \pi(a_j | R) - \min_{\{a_{s_1}, a_j\} \subseteq T \subseteq S} \pi(a_j | T) \right).$$

### 3.3 Relationship between AOM and HAOM

We compare AOM,  $\text{HAOM}_\triangleright$ , and RUM given our characterization results. Since AOM assumes a fixed preference ordering while  $\text{HAOM}_\triangleright$  allows for heterogeneous preferences, we assume the underlying preference in AOM is  $\triangleright$  and thus use the notation  $\text{AOM}_\triangleright$ . To start, RUM is characterized by the positivity of Block-Marschak conditions (Falmagne, 1978), which implies regularity, and thus RUM satisfies List-regularity. However, List-regularity does not imply the positivity of Block-Marschak conditions, and therefore RUM is independent of  $\text{HAOM}_\triangleright$ .

We give a simple example below to illustrate the difference between  $\text{AOM}_\triangleright$ ,  $\text{HAOM}_\triangleright$  and RUM. Table 2 provides a set of parametric probabilistic choices described by  $\pi_{(\alpha, \beta)}$ , where  $\alpha \in [0, 0.8]$  and  $\beta \in [0, 1]$ . Let  $\triangleright$  be  $\langle a_1, a_2, a_3 \rangle$ . When  $\alpha > \beta$ , we have  $\pi_{(\alpha, \beta)}(a_1 | \{a_1, a_2, a_3\}) > \pi_{(\alpha, \beta)}(a_1 | \{a_1, a_2\})$ . Hence,  $\pi_{(\alpha, \beta)}$  cannot be represented by RUM (due to regularity violation) or  $\text{AOM}_\triangleright$  (due to the revelation  $a_3 \triangleright a_1$ ) when  $\alpha > \beta$ . On the other hand,  $\pi_{(\alpha, \beta)}$  cannot be represented by  $\text{HAOM}_\triangleright$  when  $\beta < 0.40$  (Axiom 4 is violated). Therefore, if  $\alpha < \beta < 0.40$ ,  $\pi_{(\alpha, \beta)}$  can be represented by RUM and  $\text{AOM}_\triangleright$  but not  $\text{HAOM}_\triangleright$ . Similarly, if  $\alpha > \beta > 0.40$ ,  $\pi_{(\alpha, \beta)}$  can be represented by  $\text{HAOM}_\triangleright$  but not RUM nor  $\text{AOM}_\triangleright$ . Figure 1 gives a full graphical representation.

$\pi_{(\alpha, \beta)}$	$\{a_1, a_2, a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$
$a_1$	$\alpha$	$\beta$	0.70	—
$a_2$	$0.80 - \alpha$	$1 - \beta$	—	0.60
$a_3$	0.20	—	0.30	0.40

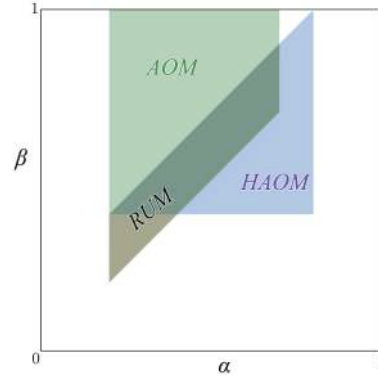


Table 2: Probabilistic choice functions.

Figure 1: Areas covered by each model.

We now consider a particular choice:  $\pi_{(0.75, 0.50)}$ , which can be represented by  $\text{HAOM}_\triangleright$  but not  $\text{AOM}_\triangleright$ . In other words, there is no attention rule satisfying attention overload property such that  $(\triangleright, \mu)$  represents  $\pi_{(0.75, 0.50)}$ . On the other hand,  $\text{HAOM}_\triangleright$  can accommodate this behavior since it allows multiple preferences. Indeed, we use Theorem 7 to identify uniquely each preference type. Then, we must have  $\tau(\succ_{11} = \triangleright) = 0.10$ ,  $\tau(\succ_{21}) = 0.50$ ,  $\tau(\succ_{31}) = 0.30$ , and  $\tau(\succ_{32}) = 0.10$ . Similarly, we can provide bounds for revealed attention for each alternative. Since  $a_1$  is the first item in the list, we must have  $\phi(a_1 | \{a_1, a_2, a_3\}) = 1$ . For other alternatives, Theorem 8 dictates that  $0.25 \leq \phi(a_2 | \{a_1, a_2, a_3\}) \leq 0.55$ , and  $0.2 \leq$

$$\phi(a_3|\{a_1, a_2, a_3\}) \leq 0.45.$$

### 3.4 Characterization with Unknown List

Our results so far have operated under the assumption that the list is exogenously given, which may be appropriate in settings where the true list is discernible to the researcher (e.g., Amazon search results, a voting ballot, or a Google search). Axioms 2–4 were expressed in relation to the list  $\triangleright$ , which is a primitive of our choice model with heterogeneous preferences. Thus, they should be interpreted as an identification mechanism that takes in both a probabilistic choice function and a list. In this section, we relax the assumption that the list  $\triangleright$  is known, and show how to obtain preference elicitation in our choice model.

We describe to what extent one can identify the list of a given probabilistic choice function for a  $\text{HAOM}_{\triangleright}$  representation. First, we introduce the following definition.

**Definition 5 (Strict Choice Rule).** A probabilistic choice function  $\pi$  is *strict* if  $\pi(a|b) \neq \pi(a|c)$  for all distinct  $a, b, c$ .

For the revelation of the list, we assume that  $\pi$  is strict and that the model  $\text{HAOM}_{\triangleright}$  is correctly specified, and we investigate whether the underlying list is identifiable from choice data. There are at least two ways to elicit  $\triangleright$  in our choice model. First, if we have  $\pi(a|S) < \pi(a|T)$  for some  $a \in S \subseteq T$ , then we know that  $a$  must appear before all alternatives in  $T$  by Axiom 2. Second, we can use the information coming from binary menus. Suppose we have  $\pi(c|b) > \pi(c|a)$ , we can immediately conclude that it cannot be  $b \triangleright a \triangleright c$  by Axiom 3. In other words, if  $c$  is after  $a$  and  $a$  is after  $b$ ,  $c$  should be chosen more often with the one ranked closer. Since we observe the opposite situation, then it must not be that  $c$  is after  $a$ , and  $a$  is after  $b$ . Note that  $\pi(c|b) > \pi(c|a)$  also implies  $\pi(c|b) + \pi(a|c) > 1$ , which would violate the boundedness imposed by Axiom 4 if it is  $b \triangleright c \triangleright a$ . On the other hand, by analogy, observing  $\pi(b|c) > \pi(b|a)$  will rule out the other two possibilities, i.e.  $c \triangleright a \triangleright b$  and  $c \triangleright b \triangleright a$ . Therefore, the remaining possibilities are  $a \triangleright b \triangleright c$  and  $a \triangleright c \triangleright b$ . In either case, we know that  $a$  appears before  $b$  and  $c$ . We define the following  $L_{\pi}$  binary relation:

$$aL_{\pi}b \text{ if } (i) \text{ there exists } \{a, b\} \subseteq S \subseteq T \text{ such that } \pi(a|S) < \pi(a|T), \text{ or} \\ (ii) \text{ there exists } c \in a \text{ such that } \pi(c|b) > \pi(c|a) \text{ and } \pi(b|c) > \pi(b|a).$$

Our discussion shows that if any  $\pi$  has a  $\text{HAOM}_{\triangleright}$  representation, then  $L_{\pi}$  must be a subset of  $\triangleright$ . While this is an important observation, the application could be limited if  $L_{\pi}$  is incomplete. The next theorem illustrates that  $L_{\pi}$  is “almost” complete. That is,  $L_{\pi}$  includes all binary comparisons except the binary comparison of the last two alternatives in the list. In other words,  $L_{\pi}$  identifies the list up to the last two elements. It is possible that some data also reveals the position of the last two elements.

**Theorem 9.** If a strict  $\pi$  has a  $\text{HAOM}_{\triangleright}$  representation, the list is uniquely identified up to the last two elements by  $L_{\pi}$ .

Theorem 9 guarantees that the list is almost point identified:  $L_{\pi}$  is missing one binary comparison. Hence, there are two completions of  $L_{\pi}$ . It is routine to check whether Axioms

2–4 are satisfied by (at least) one of those completions. We can state the following corollary as the characterization result for the unknown list environment.

**Corollary 3 (Characterization).** A strict  $\pi$  has a  $\text{HAOM}_{\triangleright}$  representation if and only if (i)  $L_{\pi}$  ranks everything except the last two alternatives and (ii)  $\pi$  satisfies Axioms 2–4 according to one possible completion of  $L_{\pi}$ .

For necessity, the first requirement is given by Theorem 9, and the second one is given by the fact that  $\pi$  has a  $\text{HAOM}_{\triangleright}$  representation in  $\triangleright$ , which is a completion of  $L_{\pi}$  according to Theorem 9. The sufficiency is given by Theorem 6.

### 3.5 Econometric Methods

We first discuss how to empirically bound the frequency of different preference types. To implement the partial identification result in Proposition 3, we fix two positions,  $1 \leq j < k$ . We are interested in bounding the fraction of decision maker who rank  $a_k$  at or above the  $j$ th position. In other words, we consider the following parameter of interest:

$$\theta_{kj} = \sum_{\ell \leq j} \tau(\succ_{k\ell}).$$

We have  $\theta_{kj} \geq \theta_{ks_1} \geq \pi(a_k|S)$  for any choice problem  $\langle a_{s_1}, a_{s_2}, \dots, a_{s_{|S|}} \rangle$  satisfying (i)  $a_k \in S$ , and (ii)  $s_1 \leq j$ , where the first inequality follows from the definition of  $\theta_{kj}$ , and the second inequality follows by Proposition 3.

We construct a lower bound for  $\theta_{kj}$  by computing the maximum over a collection of adjusted empirical choice probabilities. (The same idea has been used to empirically bound the attention frequency in Section 2.4 and Theorem 5.) Let  $\mathcal{D}_{kj} = \{S : a_k \in S, s_1 \leq j\}$ . We define

$$\hat{\theta}_{kj} = \max_{S \in \mathcal{D}_{kj}} \left\{ \hat{\pi}(a_k|S) - \text{cv}(\alpha, \underline{\theta}_{kj}) \cdot \hat{\sigma}(a_k|S) \right\}.$$

As before,  $\hat{\sigma}(a_k|S) = \sqrt{\hat{\pi}(a_k|S)(1 - \hat{\pi}(a_k|S))/N_S}$  is the standard error of the estimated probability  $\hat{\pi}(a_k|S)$ , and the critical value is given by

$$\text{cv}(\alpha, \underline{\theta}_{kj}) = \inf \left\{ t : \mathbb{P}[\max(\mathbf{z}) \leq t] \geq 1 - \alpha \right\},$$

where  $\mathbf{z}$  is a standard normal random vector of dimension  $|\mathcal{D}_{kj}|$ . The statistical validity of the proposed lower bound is guaranteed by the following theorem.

**Theorem 10 (Revealed Preference Distribution with Proposition 3).** Let  $\pi$  be an  $\text{HAOM}_{\triangleright}$ , and assume Assumption 2 holds. Define  $\mathbf{c} = (\min_{S \in \mathcal{D}_{kj}} N_S) \cdot (\min_{S \in \mathcal{D}_{kj}} \sigma(a_k|S))$ . Then,  $\mathbb{P}[\theta_{kj} \geq \hat{\theta}_{kj}] \geq 1 - \alpha + \mathbf{r}_{\underline{\theta}_{kj}}$ , where  $\mathbf{r}_{\underline{\theta}_{kj}} = C \cdot (\log^5(n|\mathcal{D}_{kj}|)/\mathbf{c}^2)^{1/4}$  with  $C$  an absolute constant.

We now discuss how to bound the attention frequency using the results of Theorem 8. We will illustrate with the lower bound since the upper bound follows analogously. Fix some

choice problem  $S$ , and some option  $a_k \in S$  which differs from the top-listed one. As before,  $\mathcal{D}$  is the collection of choice problems available in the data. We form the lower bound as

$$\widehat{\phi}(a_k|S) = \max_{R \supseteq S, R \in \mathcal{D}} \left\{ 1 - \widehat{\pi}(U_{\triangleright}(a_k)|R) - \text{cv}(\alpha, \underline{\phi}(a_k|S)) \cdot \widehat{\sigma}(U_{\triangleright}(a_k)|R) \right\},$$

where  $\widehat{\pi}(U_{\triangleright}(a_k)|R)$  is the empirical probability of choosing an option listed before  $a_k$  in  $R$ , and  $\widehat{\sigma}(U_{\triangleright}(a_k)|R) = \sqrt{\widehat{\pi}(U_{\triangleright}(a_k)|R)(1 - \widehat{\pi}(U_{\triangleright}(a_k)|R))/N_R}$  is its standard error. The critical value is constructed as

$$\text{cv}(\alpha, \underline{\phi}(a_k|S)) = \inf \left\{ t : \mathbb{P}[\max(\mathbf{z}) \leq t] \geq 1 - \alpha \right\},$$

where  $\mathbf{z}$  is a standard normal random vector of dimension  $|\{R \in \mathcal{D} : R \supseteq S\}|$ , the number of supersets of  $S$ .

**Theorem 11 (Attention Frequency Elicitation with Theorem 8).** Let  $\pi$  be an HAOM $_{\triangleright}$ , and assume Assumption 2 holds. Define  $\mathbf{c}_1 = |\{R \in \mathcal{D} : R \supseteq S\}|$  to be the number of supersets of  $S$ , and  $\mathbf{c}_2 = (\min_{R \supseteq S, R \in \mathcal{D}} N_R) \cdot (\min_{R \supseteq S, R \in \mathcal{D}} \sigma(B_{\triangleright}(a_k)|R))$ . Then,  $\mathbb{P}[\phi(a_k|S) \geq \widehat{\phi}(a_k|S)] \geq 1 - \alpha + \mathbf{r}_{\underline{\phi}(a_k|S)}$ , where  $\mathbf{r}_{\underline{\phi}(a_k|S)} = C \cdot (\log^5(n\mathbf{c}_1)/\mathbf{c}_2^2)^{1/4}$ , and  $C$  denotes an absolute constant.

To showcase the performance of our proposed econometric methods, we report the results of a simulation study on preference elicitation in the supplemental appendix. Our numerical findings illustrate how the lower bound on preference distribution (i.e.,  $\theta_{kj}$  above) changes as we vary the effective sample size and the number of choice problems available in the data.

## A Appendix: Proofs

This appendix contains the proofs of our results. Proofs of the technical lemmas used herein are available in the supplemental appendix.

### A.1 Proof of Theorems 1 and 3

Corollary 1 is implied by Theorem 1, that is,  $\succ$ -Regularity captures all of the empirical content that our AOM delivers for revealed preference. In addition, we already showed the necessity of  $\succ$ -Regularity, and hence we only need to prove sufficiency for Theorem 1. On the other hand, Theorem 3 will be shown in the following proof because we are also proving the existence of an attention rule that uses pessimistic evaluation. For optimistic evaluation, see footnote 5 below.

The proof is divided in two parts. The first part sets up a system of linear equations that pins down the attention rule satisfying the desired property. Some algebraic operations are devoted into lining up the system in preparation for the second part of the proof, which utilizes Farkas's Lemma to prove the existence of a solution to the system of equations for any parameter value which satisfies  $\succ$ -Regularity.

Assume  $(\pi, \succ)$  satisfies property  $\succ$ -Regularity. For every  $S$  and  $x \in S$ , a compatible attention rule should explain the data, i.e.,  $\sum_{\substack{x \in T \subseteq S \\ x \text{ is } \succ\text{-best in } T}} \mu(T|S) = \pi(x|S)$ . In addition,

we would like to set the attention rule such that it gives the pessimistic evaluation, i.e.,  $\phi(x|S) = \max_{R \supseteq S} \pi(x|R)$ .<sup>5</sup> If the above is feasible, then the resulting attention rule will satisfy the attention overload assumption. It remains to show that there exists a solution to the system of linear equations. Let  $x_1 \succ x_2 \succ \dots \succ x_n$ . Then, we have for  $i = 1, \dots, n$

$$\sum_{\substack{x_i \in T \subseteq S \\ x_i \text{ is } \succ\text{-best in } T}} \mu(T|S) = \pi(x_i|S) \quad (\text{denoted by } \mathcal{P}_i)$$

$$\phi(x_i|S) = \max_{R \supseteq S} \pi(x_i|R). \quad (\text{denoted by } \mathcal{M}_i)$$

For  $x_1$ ,  $\succ$ -Regularity requires that, for any  $R \supseteq S$ ,  $\phi(x_1|S) \geq \pi(x_1|S) \geq \pi(x_1|R)$ , which implies  $\max_{R \supseteq S} \pi(x_1|R) = \phi(x_1|S) = \pi(x_1|S)$ . In addition, we also have  $\sum_{\substack{x_1 \in T \subseteq S \\ x_1 \text{ is } \succ\text{-best in } T}} \mu(T|S) = \sum_{x_1 \in T \subseteq S} \mu(T|S) = \phi(x_1|S) = \pi(x_1|S)$ . It gives us  $\mathcal{P}_1 = \mathcal{M}_1$ ; the probability that the best alternative is chosen is the attention it received. On the other hand,  $\mathcal{P}_n$  is  $\pi(x_n|S) = \mu(\{x_n\}|S)$ , which immediately gives the solution to the “unknown”  $\mu(\{x_n\}|S)$ . Hence, we are left with  $\mathcal{P}_i$  for  $i = 1, \dots, n-1$  and  $\mathcal{M}_i$  for  $i = 2, \dots, n$ . Then, we create  $\mathcal{M}'_i \equiv \sum_{j \leq i} \mathcal{P}_j - \mathcal{M}_i$  for every  $i = 2, \dots, n$ , that is,

$$\sum_{j < i} \sum_{\substack{x_j \in T \subseteq S \\ x_j \text{ is } \succ\text{-best in } T}} \mu(T|S) = \sum_{j \leq i} \pi(x_j|S) - \max_{R \supseteq S} \pi(x_i|R). \quad (\text{denoted by } \mathcal{M}'_i)$$

The above makes sense because  $\sum_{j \leq i} \pi(x_j|S) - \max_{R \supseteq S} \pi(x_i|R) \geq 0$  for  $i = 2, \dots, n$ , which is required by  $\succ$ -Regularity. Lastly, we define  $\mathcal{P}'_1 \equiv \mathcal{P}_1 - \sum_{j > 1} \mathcal{M}_j$ . We are left with  $\mathcal{P}'_i$ ,  $\mathcal{P}_i$  for  $i = 2, \dots, n-1$ , and  $\mathcal{M}'_i$  for  $i = 2, \dots, n$ . We utilize Farkas’s Lemma to prove the existence of solution to the above system of linear equations. The system is straightforward when  $n \leq 2$ , so we focus on  $n \geq 3$ .

**Lemma A.1 (Farkas’ Lemma).** Let  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{b} \in \mathbb{R}^m$ . Then exactly one of the following is true:

1. There exists an  $\mathbf{x} \in \mathbb{R}^n$  such that  $\mathbf{A}\mathbf{x} = \mathbf{b}$  and  $\mathbf{x} \geq 0$ .
2. There exists a  $\mathbf{y} \in \mathbb{R}^m$  such that  $\mathbf{y}\mathbf{A} \geq 0$  and  $\mathbf{y}\mathbf{b} < 0$ .

We let  $\mathbf{A}$  be the matrix and  $\mathbf{b}$  be the vector such that the above system of linear equations is represented by  $\mathbf{A}\boldsymbol{\mu} = \mathbf{b}$ . Specifically,  $\mathbf{A} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{2n-2})^\top$ , and  $\mathbf{b} = (b_1, b_2, \dots, b_{2n-2})^\top$ , where  $\mathbf{r}_j$ ’s are column vectors. In particular, we let  $\mathbf{r}_1$  and  $b_1$  correspond to the LHS and RHS of  $\mathcal{P}'_1$  respectively;  $\mathbf{r}_j$  and  $b_j$  correspond to the LHS and RHS of  $\mathcal{M}'_{n+2-j}$  respectively for  $j = 2, \dots, n$ ;  $\mathbf{r}_j$  and  $b_j$  correspond to the LHS and RHS of  $\mathcal{P}_{-n+1+j}$  respectively for  $j = n+1, \dots, 2n-2$ . To save notation, let  $m_i := \max_{R \supseteq S} \pi(x_i|R)$ ,  $\pi_i := \pi(x_i|S)$  and  $k_i := \sum_{j \leq i} \pi(x_j|S) - \max_{R \supseteq S} \pi(x_i|R) = \sum_{j \leq i} \pi_j - m_i$ , for all  $i$ . Let  $\mathcal{B}$  be

<sup>5</sup> We set  $\phi(x|S) = \max_{R \supseteq S} \pi(x|R)$ , which is the pessimistic evaluation. An alternative proof can use the optimistic evaluation, i.e.,  $\phi_\mu(a|S) = \min_{T \subseteq S} \pi(U_\geq(a)|T)$ , and same proof strategy goes through. In fact, for any attention frequency between these bounds, our proof remains valid if we choose  $\phi$  such that it satisfies attention overload.

the collection of  $\mathbf{b}$  subject to the condition  $\succ$ -Regularity:

$$\mathcal{B} = \left\{ \mathbf{b} \in \mathbb{R}^{2n-2} : b_1 = \pi_1 - k_n - k_{n-1} - \cdots - k_2, \quad b_i = k_{n-i+2} \text{ for } i = 2, \dots, n, \right. \\ \left. b_i = \pi_{i+1-n} \text{ for } i = n+1, \dots, 2n-2, \text{ where } \pi(\cdot|\cdot) \text{ satisfies } \succ\text{-Regularity} \right\}.$$

We show that there does not exist  $\mathbf{y} = (y_1, y_2, y_3, \dots, y_{2n-2}) \in \mathbb{R}^{2n-2}$  such that  $\mathbf{y}\mathbf{A} \geq 0$  and  $\mathbf{y}\mathbf{b} < 0$  for any  $\mathbf{b} \in \mathcal{B}$ . We define the set  $\mathcal{Y}(\mathbf{A})$  as the set of  $\mathbf{y}$  which satisfies  $\mathbf{y}\mathbf{A} \geq 0$ . Hence, it suffices to show that for all  $\mathbf{b} \in \mathcal{B}$ ,  $\min_{\mathbf{y} \in \mathcal{Y}(\mathbf{A})} \mathbf{y}\mathbf{b} \geq 0$ . Except for  $b_1$ , all the other  $b_j$  are positive for all possible  $\pi(\cdot|\cdot)$  as long as the choice rule satisfies  $\succ$ -Regularity. The key insight in the following proof is to show how we can guarantee  $\mathbf{y}\mathbf{b} \geq 0$  despite the fact the possibility of  $b_1$  being negative.

By construction,  $\mathbf{A}$  admits a reduced row-echelon form. Since  $\mathbf{A}$  admits a reduced row-echelon form, the leading entry is 1 and the leading entry in each row is the only non-zero entry in its column. Then, we know that it gives  $y_j \geq 0$  for all  $j$  (i.e., for all  $\mathbf{y} \in \mathcal{Y}(\mathbf{A})$ ,  $\mathbf{y} \geq 0$ ). With this observation, we can see that if  $b_1 \geq 0$ , then the proof becomes trivial. Therefore, we assume  $b_1 < 0$ . We then further explore the restriction of  $y$  under the requirement that  $\mathbf{y}\mathbf{A} \geq 0$ .

**Lemma A.2.** For all  $\mathbf{y} \in \mathcal{Y}(\mathbf{A})$ , non-empty  $P \subseteq \{2, 3, \dots, n-1\}$ , and  $j = \bar{P} + 1, \bar{P} + 2, \dots, 2n - \bar{P}$ , we have  $\sum_{i \in P} y_i + y_j \geq |P| \cdot y_1$ , where  $|P|$  is the cardinality of  $P$ , and  $\bar{P}$  is the largest element in  $P$ .

We need to show an auxiliary minimization problem to complete the proof. Let  $\mathbf{c}_n$  and  $\mathbf{z}_n$  be two vectors. To be consistent with the above in notation, both vectors start with subscript 2 and end with  $2n-2$ . i.e.  $\mathbf{c}_n = (c_2, c_3, \dots, c_{2n-2})^\top$ .

**Lemma A.3.** For all  $n \geq 3$ ,  $\min_{\mathbf{c}_n \in \mathbf{C}_n, \mathbf{z}_n \in \mathbf{Z}_n} \mathbf{c}_n \cdot \mathbf{z}_n \geq 1$ , where  $\mathbf{C}_n = \{ \mathbf{c}_n \in \mathbb{R}_+^{2n-3} \mid \sum_{i=2}^{n+1-j} c_i + \sum_{i=n+3-j}^n c_i + \sum_{i=n+1}^{n-2+j} c_i \geq 1, j = 2, 3, \dots, n \}$ ,  $\mathbf{Z}_n = \{ \mathbf{z}_n \in \mathbb{R}_+^{2n-3} \mid \sum_{i \in P} z_i + z_j \geq |P|, \forall \text{ non-empty } P \subseteq \{2, 3, \dots, n-1\}, j = \bar{P} + 1, \bar{P} + 2, \dots, 2n - \bar{P} \}$  and  $\bar{P}$  denotes the largest element in  $P$ .

Since  $b_1 < 0$ , we can apply Lemma A.3 by setting  $c_i = -\frac{b_i}{b_1}$  and  $z_i = \frac{y_i}{y_1}$  for  $i = 2, \dots, 2n-2$ . Firstly, Lemma A.2 implies that the set of constraints in  $\mathbf{Z}_n$  is fulfilled in  $\mathcal{Y}(\mathbf{A})$  after we plug in  $z_i = \frac{y_i}{y_1}$ . Secondly, all the constraint in the set  $\mathbf{C}_n$  is fulfilled after we plugged in  $c_i = -\frac{b_i}{b_1}$ , due to the way  $\mathcal{B}$  is constructed. Therefore, the statement that all  $\mathbf{b} \in \mathcal{B}$ ,  $\min_{\mathbf{y} \in \mathcal{Y}(\mathbf{A})} \mathbf{y}\mathbf{b} \geq 0$  is implied by the statement that  $\min_{\mathbf{c}_n \in \mathbf{C}_n, \mathbf{z}_n \in \mathbf{Z}_n} \mathbf{c}_n \cdot \mathbf{z}_n \geq 1$ . It remains to prove Lemmas Lemma A.2 and A.3.

## A.2 Proof of Lemma A.2

For any set  $P$ , we get  $\sum_{i \in P} y_i + y_j \geq |P| \cdot y_1$  from the column of  $\mu(S - \cup_{i \in P \cup \{j\}} \{x_{i+n-2}\} | S)$  for any  $j \in \{\bar{P} + 1, \bar{P} + 2, \dots, \bar{P} + n\}$ . For the LHS: it is because for any  $i \in P$ , the vector  $r_{n-i+2}$  has the coefficient of 1 in the column of  $\mu(S - \cup_{i \in P \cup \{j\}} \{x_{i+n-2}\} | S)$  by construction. For the RHS: it is because the vector  $r_1$  has the coefficient of  $-|P|$  in the same column by construction. Also, we get  $\sum_{i \in P} y_i + y_j \geq |P| \cdot y_1$  from the column of  $\mu(S - \cup_{i \in P} \{x_{i+n-2}\} -$

$\cup_{i < j-n} \{x_i\} | S)$  for any  $j \in \{n+1, n+2, \dots, 2n-\bar{P}\}$ . For the LHS: it is because for any  $i \in P$ , the vector  $r_{n-i+2}$  has the coefficient of 1 in the column of  $\mu(S - \cup_{i \in P} \{x_{i+n-2}\} - \cup_{i < j-n} \{x_i\} | S)$  by construction. For the RHS: it is because the vector  $r_1$  has the coefficient of  $-|P|$  in the same column by construction. Hence, we have covered any  $j$  in  $\{\bar{P} + 1, \bar{P} + 2, \dots, 2n - \bar{P}\}$ , which concludes the proof.

### A.3 Proof of Lemma A.3

We prove by induction. Consider  $n = 3$ . We have  $\mathbf{C}_3 = \{\mathbf{c}_3 \in \mathbb{R}_+^3 | c_2 \geq 1, c_4 + c_3 \geq 1\}$  and  $\mathbf{Z}_3 = \{\mathbf{z}_3 \in \mathbb{R}_+^3 | z_2 + z_3 \geq 1, z_2 + z_4 \geq 1\}$ . It is straight-forward to see that, if  $z_2 \geq 1$ ,  $\mathbf{c}_3 \cdot \mathbf{z}_3 \geq c_2 z_2 \geq 1$ . Therefore, we consider the case that  $z_2 < 1$ . Then, we have, by putting in all the constraints, we get  $\mathbf{c}_3 \cdot \mathbf{z}_3 = z_2 c_2 + z_3 c_3 + z_4 c_4 \geq z_2(1) + c_3(1 - z_2) + c_4(1 - z_2) = z_2 + (c_3 + c_4)(1 - z_2) \geq 1$ . Hence, it is true for  $n = 3$ . Therefore, suppose the claim holds for  $n = k - 1$ , and consider  $n = k$ . We set up the Lagrangian minimization problem and assigns Lagrangian multipliers  $\lambda_i$  to the constraints in  $\mathbf{C}_n$ . For notational convenience, we label each multiplier by all the subscripts involved in the corresponding constraint. Take  $n = 3$  as an example, then we would have multipliers  $\lambda_2$  for  $c_2 \geq 1$  and  $\lambda_{3,4}$  for  $c_4 + c_3 \geq 1$ . It is simple to check that each constraint has its own unique subscript. We collect all possible subscript labels into the set  $\Lambda_k$ . (The Lagrangian multiplier for the constraints in  $\mathbf{Z}_n$  is not much used in the proof.) We then get the first order condition of the Lagrangian equation with the complementary slackness conditions:

$$\frac{\partial L}{\partial c_i} = z_i - \sum_{i \in I \in \Lambda_k} \lambda_I \geq 0, \quad \left( z_i - \sum_{i \in I \in \Lambda_k} \lambda_I \right) c_i = 0, \quad i = 2, 3, \dots, 2k - 2.$$

By plugging in first order condition, we can have

$$\mathbf{c}_k \cdot \mathbf{z}_k \geq \sum_{i=2}^{2k-2} c_i \left( \sum_{i \in I \in \Lambda_k} \lambda_I \right) = \sum_{I \in \Lambda_k} \lambda_I \left( \sum_{i \in I} c_i \right) \geq \sum_{I \in \Lambda_k} \lambda_I,$$

where the last inequality applies the inequality constraint in  $\mathbf{C}_k$ . If  $c_i \neq 0$  for all  $i$ , then  $\sum_{I \in \Lambda_k} \lambda_I \geq 1$ . For example, if  $c_{2k-2} \neq 0$  and  $c_2 \neq 0$ , we can get binding constraint due to complementary slackness such that  $z_2 = \sum_{2 \in I \in \Lambda_k} \lambda_I$  and  $z_{2k-2} = \sum_{2k-2 \in I \in \Lambda_k} \lambda_I$ . Then, apply the respective constraint from  $\mathbf{Z}_k$ ,

$$z_2 + z_{2k-2} \geq 1 \Rightarrow \sum_{2 \in I \in \Lambda_k} \lambda_I + \sum_{2k-2 \in I \in \Lambda_k} \lambda_I \geq 1 \Rightarrow \sum_{I \in \Lambda_k} \lambda_I \geq 1$$

In fact, it is straight-forward to check that as long as

$$(c_2 \neq 0 \text{ and } c_{2k-2} \neq 0) \text{ where we use } \sum_{i \in \{2\}} z_i + z_{2k-2} \geq |\{2\}| \text{ or}$$

$$(c_2 \neq 0, c_3 \neq 0 \text{ and } c_{2k-3} \neq 0) \text{ where we use } \sum_{i \in \{2,3\}} z_i + z_{2k-3} \geq |\{2,3\}| \text{ or}$$



⋮

$(c_2 \neq 0, c_3 \neq 0, \dots, c_{k-1} \neq 0 \text{ and } c_{k+1} \neq 0)$  where we use  $\sum_{i \in \{2, 3, \dots, k-1\}} z_i + z_{k+1} \geq |\{2, 3, \dots, k-1\}|$  or

$(c_2 \neq 0, c_3 \neq 0, \dots, c_{k-1} \neq 0 \text{ and } c_k \neq 0)$  where we use  $\sum_{i \in \{2, 3, \dots, k-1\}} z_i + z_k \geq |\{2, 3, \dots, k-1\}|$

Then  $\sum_{I \in \Lambda_k} \lambda_I \geq 1$ . For cases outside the above, we check sequentially and apply induction hypothesis in each scenario.

**Case 1:**  $c_2 = 0$ . Notice that under the specification of  $c_2 = 0$ , the set of permissible choice of  $\mathbf{c}_k$  is smaller. Then, we re-number some of the variables. In particular, we write  $z'_i = z_{i+1}$  and  $c'_i = c_{i+1}$  for  $i = 2, 3, \dots, 2(k-1) - 2$ . We name this set of relabeled constraint as  $\mathbf{C}_k|_{\text{Case 1}}$  where both  $c'_i$  and  $c_j$  for some  $i, j$  co-exist. In particular, we have now  $\mathbf{c}_k = (c_2, c_3, \dots, c_{2k-2}) := (0, c'_2, c'_3, \dots, c'_{2(k-1)-2}, c_{2k-2})$ . We perform the same procedure on  $\mathbf{Z}_k$ . Then, by restricting attention to  $c'_i$  and  $z'_i$ , we can see that  $\mathbf{C}_k|_{\text{Case 1}} \subset \mathbf{C}'_{k-1}$  and  $\mathbf{Z}_k|_{\text{Case 1}} \subset \mathbf{Z}'_{k-1}$ , where  $\mathbf{C}'_{k-1}$  is the same set as  $\mathbf{C}_{k-1}$  by just relabeling  $\mathbf{c}$  to  $\mathbf{c}'$ . Hence, in this case, by induction hypothesis,

$$\min_{\mathbf{c}_k \in \mathbf{C}_k|_{\text{Case 1}}, \mathbf{z}_k \in \mathbf{Z}_k|_{\text{Case 1}}} \mathbf{c}_k \cdot \mathbf{z}_k \geq \min_{\mathbf{c}_{k-1} \in \mathbf{C}'_{k-1}, \mathbf{z}_{k-1} \in \mathbf{Z}'_{k-1}} \mathbf{c}_{k-1} \cdot \mathbf{z}_{k-1} \geq 1$$

Therefore, if  $c_2 = 0$ , the proof is done. Let  $c_2 \neq 0$ . As shown above, if  $c_{2k-2} \neq 0$ , the proof is done. Then, we look into cases where  $c_{2k-2} = 0$ . It comes Case 2, where we first assume  $c_3 = 0$ .

**Case 2:**  $c_2 \neq 0, c_3 = 0$  and  $c_{2k-2} = 0$ . We re-label the variable, in particular, write  $c'_i = c_i$  for  $i = 2$ ,  $c'_i = c_{i+1}$  for  $i = 3, \dots, 2(k-1) - 2$ . Then, we have  $\mathbf{c}_k = (c_2, c_3, \dots, c_{2k-2}) := (c'_2, 0, c'_3, \dots, c'_{2(k-1)-3}, c'_{2(k-1)-2}, 0)$ , where  $0$ 's are guaranteed by the case supposition. Analogously, we do the same for  $\mathbf{z}$ . By a similar argument. We show can  $\min_{\mathbf{c}_k \in \mathbf{C}_k|_{\text{Case 2}}, \mathbf{z}_k \in \mathbf{Z}_k|_{\text{Case 2}}} \mathbf{c}_k \cdot \mathbf{z}_k \geq 1$ . Therefore, if  $c_3 = 0$  and  $c_{2k-2} = 0$ , the proof is done. Let  $c_3 \neq 0$ . As shown above, since  $c_2 \neq 0$ , if  $c_{2k-3} \neq 0$ , the proof is done. Thus, on top of the assumption that  $c_{2k-2} = 0$ , we look into cases where  $c_{2k-3} = 0$ . It comes Case 3, where we assume  $c_4 = 0$ .

**Case 3:**  $c_2, c_3 \neq 0, c_4 = 0$  and  $c_{2k-3} = c_{2k-2} = 0$ . We re-label the variable, in particular, write  $c'_i = c_i$  for  $i = 2, 3$ ,  $c'_i = c_{i+1}$  for  $i = 4, \dots, 2(k-1) - 3$ , and  $c'_i = c_{i+2}$  for  $i = 2(k-1) - 2$ . Then, we have  $\mathbf{c}_k = (c_2, c_3, \dots, c_{2k-2}) := (c'_2, c'_3, 0, c'_5, \dots, c'_{2(k-1)-3}, 0, c'_{2(k-1)-2})$ , where  $0$ 's are guaranteed by the case supposition (we do not list all of the  $0$ 's in the specification so that one can see easier where the  $c'$  are defined). Analogously, we relabel  $\mathbf{z}$ . By a similar argument. We can show that  $\min_{\mathbf{c}_k \in \mathbf{C}_k|_{\text{Case 3}}, \mathbf{z}_k \in \mathbf{Z}_k|_{\text{Case 3}}} \mathbf{c}_k \cdot \mathbf{z}_k \geq 1$ . Continuing this argument, we skip to Case  $k-2$ .

**Case  $k-2$ :**  $c_2, c_3, c_{k-2} \neq 0, c_{k-1} = 0$  and  $c_{k+2} = \dots = c_{2k-3} = c_{2k-2} = 0$ . Write  $c'_i = c_i$  for  $i = 2, \dots, k-2$ ,  $c'_i = c_{i+1}$  for  $i = k-1, k$ , and  $c'_i = c_{i+2}$  for  $i = k+1, \dots, 2(k-1) - 2$ . Analogously, we do the same for  $\mathbf{z}$ . By a similar argument, we can show  $\min_{\mathbf{c}_k \in \mathbf{C}_k|_{\text{Case } k-2}, \mathbf{z}_k \in \mathbf{Z}_k|_{\text{Case } k-2}} \mathbf{c}_k \cdot \mathbf{z}_k \geq 1$ .

**Case  $k-1$  (Last Case):**  $c_2, c_3, c_{k-1} \neq 0, c_k = c_{k+1} = \dots = c_{2k-2} = 0$ . Write  $c'_i = c_i$  for

$i = 2, \dots, k-1$  and  $c'_i = c_{i+2}$  for  $i = k, k+1, \dots, 2(k-1) - 2$ . Analogously, we do the same for  $\mathbf{z}$ . By a similar argument, we can show  $\min_{\mathbf{c}_k \in \mathbf{C}_k | \text{Case } k-1, \mathbf{z}_k \in \mathbf{Z}_k | \text{Case } k-1} \mathbf{c}_k \cdot \mathbf{z}_k \geq 1$ .

Therefore, the above covers every possible case and it shows that in every case the minimum is greater than or equal to 1 for  $n = k$ . By induction, the minimum of the objective function is greater than or equal to 1 for all  $n \geq 3$ .

## A.4 Proof of Theorem 4

We denote the choice probabilities by the vector  $\boldsymbol{\pi}$ , and the constraints implied by  $\succ$ -Regularity will be collected in the matrix  $\mathbf{R}_\succ$ .  $\boldsymbol{\sigma}$  contains the standard deviations, that is,  $\boldsymbol{\sigma}^2$  are the diagonal elements in the covariance matrix  $\mathbf{R}_\succ \mathbb{V}[\hat{\boldsymbol{\pi}}] \mathbf{R}_\succ^\top$ . As before,  $\boldsymbol{\Omega}$  and  $\hat{\boldsymbol{\Omega}}$  are the true and estimated correlation matrices of  $\mathbf{R}_\succ \hat{\boldsymbol{\pi}}$ . The operation  $\max(\cdot)$  computes the largest element in a vector/matrix.  $\odot$  denotes Hadamard (element-wise) division. The supremum norm of a vector/matrix is  $\|\cdot\|_\infty$ . Finally, we use  $c$  to denote some constant, whose value may differ depending on the context.

The next lemma provides a Gaussian approximation to the infeasible centered and scaled sum.

**Lemma A.4.** Let  $\check{\mathbf{z}}$  be a mean-zero Gaussian random vector with the covariance matrix  $\boldsymbol{\Omega}$ . Then,

$$\varrho_1 = \sup_{\substack{A \subseteq \mathbb{R}^{\mathbf{c}_1} \\ A \text{ rectangular}}} \left| \mathbb{P} \left[ (\mathbf{R}_\succ \hat{\boldsymbol{\pi}} - \mathbf{R}_\succ \boldsymbol{\pi}) \odot \boldsymbol{\sigma} \in A \right] - \mathbb{P} \left[ \check{\mathbf{z}} \odot \boldsymbol{\sigma} \in A \right] \right| \leq c \left( \frac{\log^5(n\mathbf{c}_1)}{\mathbf{c}_2^2} \right)^{\frac{1}{4}}.$$

The next step is to replace the infeasible standard error.

**Lemma A.5.** Let  $\xi_1, \xi_2 > 0$  with  $\xi_2 \rightarrow 0$ . Then,

$$\begin{aligned} & \mathbb{P} \left[ \left\| (\mathbf{R}_\succ \hat{\boldsymbol{\pi}} - \mathbf{R}_\succ \boldsymbol{\pi}) \odot \boldsymbol{\sigma}_\succ - (\mathbf{R}_\succ \hat{\boldsymbol{\pi}} - \mathbf{R}_\succ \boldsymbol{\pi}) \odot \hat{\boldsymbol{\sigma}}_\succ \right\|_\infty \geq \xi_1 \xi_2 \right] \\ & \leq c \xi_1^{-1} \sqrt{\log \mathbf{c}_1} + c \left( \frac{\log^5(n\mathbf{c}_1)}{\mathbf{c}_2^2} \right)^{\frac{1}{4}} + c \exp \left\{ -\frac{1}{c} \mathbf{c}_2^2 \xi_2^2 + \log \mathbf{c}_1 \right\}. \end{aligned}$$

As a result,

$$\begin{aligned} & \mathbb{P} \left[ \left| \max \left( (\mathbf{R}_\succ \hat{\boldsymbol{\pi}} - \mathbf{R}_\succ \boldsymbol{\pi}) \odot \boldsymbol{\sigma}_\succ \right) - \max \left( (\mathbf{R}_\succ \hat{\boldsymbol{\pi}} - \mathbf{R}_\succ \boldsymbol{\pi}) \odot \hat{\boldsymbol{\sigma}}_\succ \right) \right| \geq \xi_1 \xi_2 \right] \\ & \leq c \xi_1^{-1} \sqrt{\log \mathbf{c}_1} + c \left( \frac{\log^5(n\mathbf{c}_1)}{\mathbf{c}_2^2} \right)^{\frac{1}{4}} + c \exp \left\{ -\frac{1}{c} \mathbf{c}_2^2 \xi_2^2 + \log \mathbf{c}_1 \right\}. \end{aligned}$$

As we will be employing a simulation-based approach to construct the critical value, it is necessary to provide an error bound for the estimated covariance/correlation matrix. We first state a lemma on the error bound of variance estimation.

**Lemma A.6.** Let  $\xi_2 > 0$  with  $\xi_2 \rightarrow 0$ . Then,

$$\mathbb{P} \left[ \left\| (\hat{\boldsymbol{\sigma}}_{\succ} - \boldsymbol{\sigma}_{\succ}) \oslash \boldsymbol{\sigma}_{\succ} \right\|_{\infty} \geq \xi_2 \right] \leq c \exp \left\{ -\frac{1}{c} \mathbf{c}_2^2 \xi_2^2 + \log \mathbf{c}_1 \right\}.$$

We now state the properties of the estimated correlation matrix.

**Lemma A.7.** Let  $\xi_2 > 0$  with  $\xi_2 \rightarrow 0$ . Then,

$$\mathbb{P} \left[ \left\| \hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega} \right\|_{\infty} \geq \xi_2 \right] \leq c \exp \left\{ -\frac{1}{c} \mathbf{c}_2^2 \xi_2^2 + 2 \log \mathbf{c}_1 \right\}.$$

Now we are ready to state the following result, which provides a feasible Gaussian approximation to  $\check{\mathbf{z}}$  (defined in Lemma A.4).

**Lemma A.8.** Let  $\mathbf{z}$  be a mean-zero Gaussian random vector with a covariance matrix  $\hat{\boldsymbol{\Omega}}$ . Take  $\xi_3 > 0$  such that  $\xi_3 \rightarrow 0$ . Then,

$$\mathbb{P} \left[ \hat{\rho}_2 \leq c \xi_3^{\frac{1}{2}} \log \mathbf{c}_1 \right] \geq 1 - c \exp \left\{ -\frac{1}{c} \mathbf{c}_2^2 \xi_3^2 + 2 \log \mathbf{c}_1 \right\}, \quad \hat{\rho}_2 = \sup_{\substack{A \subseteq \mathbb{R}^{\mathbf{c}_1} \\ A \text{ rectangular}}} \left| \mathbb{P}[\check{\mathbf{z}} \in A] - \mathbb{P}[\mathbf{z} \in A | \text{Data}] \right|.$$

The proof considers a sequence of approximations, which will rely on the following statistics:

$$\begin{aligned} \mathbb{T}(\succ) &= \max \left\{ (\mathbf{R}_{\succ} \hat{\boldsymbol{\pi}}) \oslash \hat{\boldsymbol{\sigma}}_{\succ}, 0 \right\}, \\ \mathbb{T}^{\circ}(\succ) &= \max \left\{ (\mathbf{R}_{\succ} (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})) \oslash \hat{\boldsymbol{\sigma}}_{\succ}, 0 \right\}, \quad \check{\mathbb{T}}^{\circ}(\succ) = \max \left\{ (\mathbf{R}_{\succ} (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})) \oslash \boldsymbol{\sigma}_{\succ}, 0 \right\}, \\ \mathbb{T}^{\mathbf{g}}(\succ) &= \max \left\{ \mathbf{z}, 0 \right\}, \quad \check{\mathbb{T}}^{\mathbf{g}}(\succ) = \max \left\{ \check{\mathbf{z}}, 0 \right\}. \end{aligned}$$

We will also define the following quantiles/critical values.

$$\begin{aligned} \text{cv}(\alpha, \succ) &= \inf \left\{ t \geq 0 : \mathbb{P} \left[ \mathbb{T}^{\mathbf{g}}(\succ) \leq t \mid \text{Data} \right] \geq 1 - \alpha \right\}, \\ \check{\text{cv}}(\alpha, \succ) &= \inf \left\{ t \geq 0 : \mathbb{P} \left[ \check{\mathbb{T}}^{\mathbf{g}}(\succ) \leq t \right] \geq 1 - \alpha \right\}, \\ \tilde{\text{cv}}(\alpha, \succ) &= \inf \left\{ t \geq 0 : \mathbb{P} \left[ \max(\check{\mathbf{z}}) \leq t \right] \geq 1 - \alpha \right\}. \end{aligned}$$

To show the validity of the above critical value, we first need an error bound on the two critical values,  $\text{cv}(\alpha, \succ)$  and  $\check{\text{cv}}(\alpha, \succ)$ . The following lemma will be useful.

**Lemma A.9.** The critical values,  $\text{cv}(\alpha, \succ)$  and  $\check{\text{cv}}(\alpha, \succ)$ , satisfy

$$\mathbb{P} \left[ \check{\text{cv}} \left( \alpha + c \xi_3^{\frac{1}{2}} \log \mathbf{c}_2, \succ \right) \leq \text{cv}(\alpha, \succ) \leq \check{\text{cv}} \left( \alpha - c \xi_3^{\frac{1}{2}} \log \mathbf{c}_2, \succ \right) \right] \geq 1 - c \exp \left\{ -\frac{1}{c} \mathbf{c}_2^2 \xi_3^2 + 2 \log \mathbf{c}_1 \right\}.$$

To wrap up the proof of Theorem 4, we rely on a sequence of error bounds. First notice

that, under  $H_0$ ,

$$\begin{aligned} \mathbb{P}[\mathbb{T}(\gamma) > c\nu(\alpha, \gamma)] &\leq \mathbb{P}\left[\mathbb{T}(\gamma) > \check{c}\check{\nu}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1, \gamma\right)\right] + c \exp\left\{-\frac{1}{c}\mathbf{c}_2^2\xi_3^2 + 2 \log \mathbf{c}_1\right\} \\ &\leq \mathbb{P}\left[\mathbb{T}(\gamma)^\circ > \check{c}\check{\nu}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1, \gamma\right)\right] + c \exp\left\{-\frac{1}{c}\mathbf{c}_2^2\xi_3^2 + 2 \log \mathbf{c}_1\right\}. \end{aligned} \quad (8)$$

Next, we apply Lemma A.5 and obtain that

$$\begin{aligned} \mathbb{P}\left[\mathbb{T}^\circ(\gamma) > \check{c}\check{\nu}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1, \gamma\right)\right] &\leq \mathbb{P}\left[\tilde{\mathbb{T}}^\circ(\gamma) > \check{c}\check{\nu}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1, \gamma\right) - \xi_1\xi_2\right] \\ &\quad + c\xi_1^{-1}\sqrt{\log \mathbf{c}_1} + c\left(\frac{\log^5(n\mathbf{c}_1)}{\mathbf{c}_2^2}\right)^{\frac{1}{4}} + c \exp\left\{-\frac{1}{c}\mathbf{c}_2^2\xi_2^2 + \log \mathbf{c}_1\right\}. \end{aligned}$$

The last error bound in our analysis is due to Lemma A.4, which gives

$$\mathbb{P}\left[\tilde{\mathbb{T}}^\circ(\gamma) > \check{c}\check{\nu}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1, \gamma\right) - \xi_1\xi_2\right] \leq \mathbb{P}\left[\tilde{\mathbb{T}}^{\mathbf{G}}(\gamma) > \check{c}\check{\nu}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1, \gamma\right) - \xi_1\xi_2\right] + c\left(\frac{\log^5(n\mathbf{c}_1)}{\mathbf{c}_2^2}\right)^{\frac{1}{4}}.$$

Collecting all pieces, we have

$$\begin{aligned} \mathbb{P}[\mathbb{T}(\gamma) > c\nu(\alpha, \gamma)] &\leq \mathbb{P}\left[\tilde{\mathbb{T}}^{\mathbf{G}}(\gamma) > \check{c}\check{\nu}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1, \gamma\right) - \xi_1\xi_2\right] \\ &\quad + c\left(\frac{\log^5(n\mathbf{c}_1)}{\mathbf{c}_2^2}\right)^{\frac{1}{4}} + c\xi_1^{-1}\sqrt{\log \mathbf{c}_1} + c \exp\left\{-\frac{1}{c}\mathbf{c}_2^2(\xi_2 \wedge \xi_3)^2 + 2 \log \mathbf{c}_1\right\}. \end{aligned}$$

To proceed, we employ the following anti-concentration result of normal random vectors (Lemma D.4 in the Online Supplement to Chernozhukov, Chetverikov, and Kato 2019).

**Lemma A.10.** Let  $\check{\mathbf{z}} \in \mathbb{R}^{\mathbf{c}_1}$  be a mean-zero normal random vector such that  $\mathbb{V}[\check{z}_\ell] = 1$  for all  $1 \leq \ell \leq \mathbf{c}_1$ . Then for any  $t \in \mathbb{R}$  and any  $\epsilon > 0$ , one has

$$\mathbb{P}[|\max(\mathbf{z}) - t| \leq \epsilon] \leq 4\epsilon(\sqrt{2 \log \mathbf{c}_1} + 1).$$

First assume  $\check{c}\check{\nu}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1, \gamma\right) > 0$ , then it will be true that  $\check{c}\check{\nu}(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1, \gamma) = \check{c}\check{\nu}(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1, \gamma)$ . Then by applying Lemma A.10, we have

$$\check{c}\check{\nu}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1 + 4\xi_1\xi_2(\sqrt{2 \log \mathbf{c}_1} + 1), \gamma\right) \leq \check{c}\check{\nu}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1, \gamma\right) - \xi_1\xi_2,$$

which further implies that

$$\begin{aligned} \mathbb{P}\left[\tilde{\mathbb{T}}^{\mathbf{G}}(\gamma) > \check{c}\check{\nu}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1, \gamma\right) - \xi_1\xi_2\right] &\leq \mathbb{P}\left[\tilde{\mathbb{T}}^{\mathbf{G}}(\gamma) > \check{c}\check{\nu}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1 + 4\xi_1\xi_2(\sqrt{2 \log \mathbf{c}_1} + 1), \gamma\right)\right] \\ &\leq \alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1 + 4\xi_1\xi_2(\sqrt{2 \log \mathbf{c}_1} + 1). \end{aligned}$$

Finally we have that

$$\begin{aligned} \mathbb{P}[\mathbf{T}(\succ) > \text{cv}(\alpha, \succ)] &\leq \alpha + c\xi_3^{\frac{1}{2}} \log \mathbf{c}_1 + 4\xi_1\xi_2(\sqrt{2\log \mathbf{c}_1} + 1) \\ &\quad + c \left( \frac{\log^5(n\mathbf{c}_1)}{\mathbf{c}_2^2} \right)^{\frac{1}{4}} + c\xi_1^{-1} \sqrt{\log \mathbf{c}_1} + c \exp \left\{ -\frac{1}{c} \mathbf{c}_2^2 (\xi_2 \wedge \xi_3)^2 + 2 \log \mathbf{c}_1 \right\}. \end{aligned}$$

To control the above error, we need to verify a few side conditions we used in the derivation. Consider

$$\xi_1^{-2} = \xi_2 = \xi_3 = \frac{\sqrt{2c \log \mathbf{c}_1 + \frac{c}{2} \log \mathbf{c}_2}}{\mathbf{c}_2}.$$

Then the last term in the above becomes

$$c \exp \left\{ -\frac{1}{c} \mathbf{c}_2^2 (\xi_2 \wedge \xi_3)^2 + 2 \log \mathbf{c}_1 \right\} = \frac{c}{\sqrt{\mathbf{c}_2}}.$$

In addition, the requirement that  $\xi_2 \rightarrow 0$  will follow from the assumption that  $\log(\mathbf{c}_1)/\mathbf{c}_2^2 \rightarrow 0$ . The other terms in the error bound can be shown to be bounded by  $c \left( \frac{\log^5(n\mathbf{c}_1)}{\mathbf{c}_2^2} \right)^{\frac{1}{4}}$  as well.

## A.5 Proof of Theorem 5

The proof again relies on bounding the errors in the normal approximation and variance estimation. Let  $\mathbf{z}$  be a standard normal random vector of suitable dimension. Then

$$\begin{aligned} &\mathbb{P} \left[ \hat{\pi}(a|R) - \text{cv}(\alpha, \underline{\phi}(a|S)) \cdot \hat{\sigma}(a|R) \leq \pi(a|R), \forall R \supseteq S, R \in \mathcal{D} \right] \\ &\leq \mathbb{P} \left[ \max(\mathbf{z}) \leq \text{cv}(\alpha, \underline{\phi}(a|S)) \right] \\ &+ \left| \mathbb{P} \left[ \frac{\hat{\pi}(a|R) - \pi(a|R)}{\hat{\sigma}(a|R)} \leq \text{cv}(\alpha, \underline{\phi}(a|S)), \forall R \supseteq S, R \in \mathcal{D} \right] - \mathbb{P} \left[ \max(\mathbf{z}) \leq \text{cv}(\alpha, \underline{\phi}(a|S)) \right] \right|. \end{aligned}$$

By the construction of the critical value,  $\text{cv}(\alpha, \underline{\phi}(a|S))$ , the first term is exactly  $1 - \alpha$ . As a result, the error term in the theorem can be taken as

$$\mathbf{r}_{\underline{\phi}(a|S)} = \left| \mathbb{P} \left[ \frac{\hat{\pi}(a|R) - \pi(a|R)}{\hat{\sigma}(a|R)} \leq \text{cv}(\alpha, \underline{\phi}(a|S)), \forall R \supseteq S, R \in \mathcal{D} \right] - \mathbb{P} \left[ \max(\mathbf{z}) \leq \text{cv}(\alpha, \underline{\phi}(a|S)) \right] \right|,$$

or any further bound thereof. In the following, we first provide a lemma on normal approximation. Define  $\mathbf{R}_{\underline{\phi}(a|S)}$  as the matrix extracting the relevant choice probabilities for constructing the lower bound in the theorem. We use  $\boldsymbol{\sigma}_{\underline{\phi}(a|S)}$  to collect the standard deviations of  $\mathbf{R}_{\underline{\phi}(a|S)} \hat{\boldsymbol{\pi}}$ , and its estimate is represented by  $\hat{\boldsymbol{\sigma}}_{\underline{\phi}(a|S)}$ .

**Lemma A.11.** The following normal approximation holds

$$\varrho_1 = \sup_{\substack{A \subseteq \mathbb{R}^{\mathfrak{c}_1} \\ A \text{ rectangular}}} \left| \mathbb{P} \left[ \left( \mathbf{R}_{\underline{\phi}(a|S)} \hat{\boldsymbol{\pi}} - \mathbf{R}_{\underline{\phi}(a|S)} \boldsymbol{\pi} \right) \otimes \boldsymbol{\sigma}_{\underline{\phi}(a|S)} \in A \right] - \mathbb{P}[\mathbf{z} \in A] \right| \leq c \left( \frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2} \right)^{\frac{1}{4}}.$$

The next step is to replace the infeasible standard errors by its estimate. The following lemma provides an error bound which arises as we “take the hat off.”

**Lemma A.12.** Let  $\xi_1, \xi_2 > 0$  with  $\xi_2 \rightarrow 0$ . Then

$$\begin{aligned} & \mathbb{P} \left[ \left\| \left( \mathbf{R}_{\underline{\phi}(a|S)} \hat{\boldsymbol{\pi}} - \mathbf{R}_{\underline{\phi}(a|S)} \boldsymbol{\pi} \right) \otimes \boldsymbol{\sigma}_{\underline{\phi}(a|S)} - \left( \mathbf{R}_{\underline{\phi}(a|S)} \hat{\boldsymbol{\pi}} - \mathbf{R}_{\underline{\phi}(a|S)} \boldsymbol{\pi} \right) \otimes \hat{\boldsymbol{\sigma}}_{\underline{\phi}(a|S)} \right\|_{\infty} \geq \xi_1 \xi_2 \right] \\ & \leq c \xi_1^{-1} \sqrt{\log \mathfrak{c}_1} + c \left( \frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2} \right)^{\frac{1}{4}} + c \exp \left\{ -\frac{1}{c} \mathfrak{c}_2^2 \xi_2^2 + \log \mathfrak{c}_1 \right\}. \end{aligned}$$

As a result,

$$\begin{aligned} & \mathbb{P} \left[ \left| \max \left( \left( \mathbf{R}_{\underline{\phi}(a|S)} \hat{\boldsymbol{\pi}} - \mathbf{R}_{\underline{\phi}(a|S)} \boldsymbol{\pi} \right) \otimes \boldsymbol{\sigma}_{\underline{\phi}(a|S)} \right) - \max \left( \left( \mathbf{R}_{\underline{\phi}(a|S)} \hat{\boldsymbol{\pi}} - \mathbf{R}_{\underline{\phi}(a|S)} \boldsymbol{\pi} \right) \otimes \hat{\boldsymbol{\sigma}}_{\underline{\phi}(a|S)} \right) \right| \geq \xi_1 \xi_2 \right] \\ & \leq c \xi_1^{-1} \sqrt{\log \mathfrak{c}_1} + c \left( \frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2} \right)^{\frac{1}{4}} + c \exp \left\{ -\frac{1}{c} \mathfrak{c}_2^2 \xi_2^2 + \log \mathfrak{c}_1 \right\}. \end{aligned}$$

To close the proof of the theorem, we provide the further bound that

$$\begin{aligned} & \left| \mathbb{P} \left[ \frac{\hat{\pi}(a|R) - \pi(a|R)}{\hat{\sigma}(a|R)} \leq \text{cv}(\alpha, \underline{\phi}(a|S)), \forall R \supseteq S, R \in \mathcal{D} \right] - \mathbb{P} \left[ \max(\mathbf{z}) \leq \text{cv}(\alpha, \underline{\phi}(a|S)) \right] \right| \\ & \leq c \xi_1^{-1} \sqrt{\log \mathfrak{c}_1} + c \left( \frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2} \right)^{\frac{1}{4}} + c \exp \left\{ -\frac{1}{c} \mathfrak{c}_2^2 \xi_2^2 + \log \mathfrak{c}_1 \right\} + \mathbb{P} \left[ \left| \max(\mathbf{z}) - \text{cv}(\alpha, \underline{\phi}(a|S)) \right| \leq \xi_1 \xi_2 \right] \\ & \leq c \xi_1^{-1} \sqrt{\log \mathfrak{c}_1} + c \left( \frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2} \right)^{\frac{1}{4}} + c \exp \left\{ -\frac{1}{c} \mathfrak{c}_2^2 \xi_2^2 + \log \mathfrak{c}_1 \right\} + c \xi_1 \xi_2 \sqrt{\log \mathfrak{c}_1}, \end{aligned}$$

where the second term follows from Lemma A.10. Finally, we set

$$\xi_1^{-2} = \xi_2 = \frac{\sqrt{2c \log \mathfrak{c}_1 + \frac{c}{2} \log \mathfrak{c}_2}}{\mathfrak{c}_2}.$$

## A.6 Proof of Theorem 6

*Proof.* Given a linear order  $\triangleright$ , we consider a partition of  $\mathcal{P}_{\triangleright}$ . We denote  $\mathcal{P}_{\triangleright}(x)$  as the subset of single improvement of  $\triangleright$  where each  $\succ$  in  $\mathcal{P}_{\triangleright}(x)$  captures each improvement of  $x$  in  $\triangleright$ . Also,  $\bigcup_{x \in X} \mathcal{P}_{\triangleright}(x) = \mathcal{P}_{\triangleright}$  and  $\mathcal{P}_{\triangleright}(x) \cap \mathcal{P}_{\triangleright}(y) = \emptyset$  for  $x \neq y$ . For each type  $(\Gamma, \succ)$ , we denote  $c_{(\Gamma, \succ)}(\cdot)$  as the corresponding choice function. Multiple types might exhibit the same choice behavior. Moreover, we also define induced choice data given  $\tau$ . To do that, we first define  $\pi_z^\tau$  given each  $z$  where  $\pi_z^\tau(x, S) := \tau(\{(\Gamma, \succ) \in \mathcal{AO}_{\triangleright} \times \mathcal{P}_{\triangleright}(z) : c_{(\Gamma, \succ)}(S) = x\})$ . It is easy to

see that  $\sum_{z \in X} \pi_z^\tau(x, S) = \pi^\tau(x, S)$ . We first prove necessity. To proceed, we denote  $m_{S, \triangleright}$  as the top element in  $S$  according to  $\triangleright$ . Also, we denote  $m := m_{X, \triangleright}$ . For Axiom 1, we proceed by proving two claims. The first one shows that each type chooses either the first item in the decision problem or their reference points.

**Claim 1.** For any  $(\Gamma, \succ) \in \mathcal{CF}_\triangleright \times \mathcal{P}_\triangleright(x)$  and all  $S$ , we must have either  $c_{(\Gamma, \succ)}(S) = m_{S, \triangleright}$  or  $c_{(\Gamma, \succ)}(S) = x$ . Therefore, if  $x \notin S$ , we must have  $c_{(\Gamma, \succ)}(S) = m_{S, \triangleright}$ .

*Proof.* Suppose not. i.e.  $c_{(\Gamma, \succ)}(S) = y$  and  $y \notin \{m_{S, \triangleright}, x\}$ . Since  $y \neq m_{S, \triangleright}$ , we must have  $m_{S, \triangleright} \triangleright y$ . Since each linear order  $\succ$  in  $\mathcal{P}_\triangleright(x)$  agrees with  $\triangleright$  except for  $x$ , we have  $m_{S, \triangleright} \succ y$ . However, by the definition of choice function and the property of competition filter on list, it must be that  $y \succ m_{S, \triangleright}$ , which implies that  $y \succ m_{S, \triangleright} \succ y$ , a contradiction.  $\square$

Then the second claim is that when  $x$  is not the first item in the list, then  $\pi_z^\tau(x, \cdot)$  decreasing in the second component when  $z = x$ , and is constant otherwise.

**Claim 2.** For all  $S \subseteq T$  and  $x, z \in S$  with  $x \neq z$  such that there exists  $y \triangleright x$  and  $y \in S$ , we have

i)  $\pi_x^\tau(x, S) \geq \pi_x^\tau(x, T)$ ; ii)  $\pi_z^\tau(x, S) = \pi_z^\tau(x, T) = 0$ .

*Proof.* Since there exists  $y \triangleright x$  and  $y \in S \subseteq T$ , we know that  $x \neq m_{S, \triangleright}$  and  $x \neq m_{T, \triangleright}$ . Then, by Claim 1, for  $z \neq x$ , it must be that  $c_{(\Gamma, \succ)}(S) \neq x$  for any  $(\Gamma, \succ) \in \mathcal{CF}_\triangleright \times \mathcal{P}_\triangleright(z)$  and for any  $S \subseteq X$ . Therefore,  $\pi_z^\tau(x, S) = \pi_z^\tau(x, T) = 0$ . (ii) is proven.

For (i), it suffices to show that for each  $\succ \in \mathcal{P}_\triangleright(x)$  and any  $\Gamma$ , if  $c_{(\Gamma, \succ)}(T) = x$ , then it must be that  $c_{(\Gamma, \succ)}(S) = x$ . To see this, suppose not. Then, by Claim 1, we must have  $c_{(\Gamma, \succ)}(S) = m_{S, \triangleright}$ . However, since  $c_{(\Gamma, \succ)}(T) = x$ , we know that it must be that  $x \succ m_{T, \triangleright}$ . Since  $x \in \Gamma(T)$ , by competition filter, we must have  $x \in \Gamma(S)$  so that  $m_{S, \triangleright} \succ x$ . If  $m_{T, \triangleright} = m_{S, \triangleright}$ , then  $x \succ m_{T, \triangleright} = m_{S, \triangleright} \succ x$ , which is a contradiction. On the other hand, if  $m_{T, \triangleright} \neq m_{S, \triangleright}$ , then it must be that  $m_{T, \triangleright} \triangleright m_{S, \triangleright}$ . Since  $\succ \in \mathcal{P}_\triangleright(x)$ , it agrees with  $\triangleright$  over binary relation on  $X \setminus x$ . We have  $m_{T, \triangleright} \succ m_{S, \triangleright}$ . However, we can then deduce that  $x \succ m_{T, \triangleright} \succ m_{S, \triangleright} \succ x$ , a contradiction.  $\square$

Using the above claim, we can see that for all  $S \subseteq T$  and  $x \in S$  such that there exists  $y \triangleright x$  and  $y \in S$ , we have  $\pi(x, S) = \sum_{z \in X} \pi_z^\tau(x, S) \geq \sum_{z \in X} \pi_z^\tau(x, T) = \pi^\tau(x, T)$ . Axiom 1 is proven. For Axiom 2, due to Claim 2, it suffices to show that  $\pi_x^\tau(x, \{x, y\}) \geq \pi_x^\tau(x, \{x, z\})$  for all  $x, y, z$  such that  $z \triangleright y \triangleright x$ . Since  $\Gamma$  has full attention over binary sets, for any  $(\Gamma, \succ) \in \mathcal{CF}_\triangleright \times \mathcal{P}_\triangleright(x)$ , if  $c_{(\Gamma, \succ)}(\{x, z\}) = x$ , then it must be that  $c_{(\Gamma, \succ)}(\{x, y\}) = x$ . Note that  $c_{(\Gamma, \succ)}(\{x, z\}) = x$  implies that  $x \succ z$ . Also, since  $\succ$  agrees with  $\triangleright$  over binary relation on  $X \setminus x$ , we know that  $z \succ y$ . It must be that  $x \succ y$ . Due to full attention over binary sets, it must be that  $c_{(\Gamma, \succ)}(\{x, y\}) = x$ . Hence, it is proven. For Axiom 3, it suffices to show that for  $x \neq m$ , we have

$$\tau(\{(\Gamma, \succ) \in \mathcal{AO}_\triangleright \times \mathcal{P}_\triangleright(x)\}) = \pi^\tau(x, \{b_x, x\})$$

It is because if it holds, then  $1 - \sum_{x \in X \setminus m} \pi^\tau(x, \{b_x, x\}) = \tau(\{(\Gamma, \succ) \in \mathcal{AO}_\triangleright \times \mathcal{P}_\triangleright(m)\}) \geq 0$ . To see why it holds, note that for any  $(\Gamma, \succ) \in \mathcal{CF}_\triangleright \times \mathcal{P}_\triangleright(x)$ , it must be that  $c_{(\Gamma, \succ)}(\{b_x, x\}) = x$ , since  $\succ$  is a single improvement of  $\triangleright$  and  $\Gamma$  has full attention over binary sets. Necessity is complete.

We then prove sufficiency. We put the focus on the choice functions generated from  $(\Gamma, \succ) \in \mathcal{CF}_\triangleright \times \mathcal{P}_\triangleright(x)$ . We first state a characterization of this type of choice function

**Claim 3.** Let  $x \in X$ . A choice function  $c$  is represented by some  $(\Gamma, \succ) \in \mathcal{CF}_\triangleright \times \mathcal{P}_\triangleright(x)$  if and only if i) whenever  $c(S) \neq m_{S,\triangleright}$ , we have  $c(S) = x$ ; ii)  $c(T) = x$  implies  $c(S) = x$  for  $x \in S \subseteq T$ ; iii)  $c(\{y, x\}) = x$  for some  $y \in X$  implies  $c(\{z, x\}) = x$  for all  $z \triangleleft y$ ; iv)  $c(\{z, x\}) = x$  for all  $z \triangleleft x$ .

*Proof.* We first prove the only-if part. Note that i) is given by Claim 1. For ii), suppose not. Then, by (i), it must be that  $c(S) = m_{S,\triangleright} \triangleright x$ . By competition filter on list, we know that i)  $m_{S,\triangleright}, x \in \Gamma(S) \subseteq \Gamma(T)$ . However,  $c(S)$  implies  $x \triangleright m_{S,\triangleright}$  but  $c(S) = m_{S,\triangleright}$  implies  $m_{S,\triangleright} \triangleright x$ . A contradiction. For iii), since  $\Gamma$  assign full attention over binary sets, if  $c(\{y, x\}) = x$ , by  $\succ \in \mathcal{P}_\triangleright(x)$ , we have  $x \succ z$  for all  $z \triangleleft y$ . For iv), it is clear that  $\Gamma(\{z, x\}) = \{z, x\}$  and  $x \succ z$  as  $x \triangleright z$ . Only-if part is complete.

For the if-part, we can construct  $(\Gamma_c, \succ_c)$  as follows: Firstly, we consider  $\Gamma_c$ . For  $|S| > 2$ , We let  $\Gamma_c(S) = \{m_{S,\triangleright}\}$  for every  $S$  such that  $c(S) = m_{S,\triangleright}$  and we let  $\Gamma_c(S) = U_\triangleright(x) \cap S$  for every  $S$  such that  $c(S) = x \neq m_{S,\triangleright}$ . For  $|S| = 2$ , we let  $\Gamma_c(S) = S$ . Secondly, for  $\succ_c$ , we construct the binary relation as follows. For  $y, z \neq x$ , we set  $y \succ_c z$  if  $y \triangleright z$  for every  $y, z \neq x$ . For binary relations involving  $x$ , we set  $x \succ_c y$  if  $c(\{x, y\}) = x$ ,  $y \succ_c x$  if  $c(\{x, y\}) = y$ .

To check that  $(\Gamma_c, \succ_c) \in \mathcal{CF}_\triangleright \times \mathcal{P}_\triangleright(x)$ , we first check preference. To check that it is complete, for  $y, z \neq x$ , it follows from  $\triangleright$ . For  $x, y$ , we must have either  $x \succ_c y$  or  $y \succ_c x$  since  $c(\{x, y\})$  is non-empty. To show that it is transitive. For  $w, y, z \neq x$ , if  $w \succ y$  and  $y \succ z$ , we must have  $w \succ z$  since it follows from  $\triangleright$ . Otherwise, firstly, consider that  $w \succ_c x$  and  $x \succ_c z$ . Therefore, we know that  $c(\{w, x\}) = w$  and  $c(\{x, z\}) = x$ . Hence, we know that  $w \triangleright x$  and  $x \triangleright z$ . Hence, it must be that  $w \triangleright z$ . Hence, by i) it must be that  $c(\{w, z\}) = w$ . Secondly, consider that  $x \succ_c w$  and  $w \succ_c z$ . Then, we know that  $c(\{w, x\}) = x$  and  $w \triangleright z$ . Then, by iii), we know that  $c(\{z, x\}) = x$ . Lastly, consider that  $w \succ_c z$  and  $z \succ_c x$ , then we know  $w \triangleright z$  and  $c(\{z, x\}) = z$ . It must be that  $c(\{w, x\}) = w$ . Suppose not, i.e.  $c(\{w, x\}) = x$ . Then, by iii), we must have  $c(\{z, x\}) = x$ . A contradiction. Transitivity is complete. By construction, it is clear that  $\succ_c \in \mathcal{P}_\triangleright(x)$ .

Then, we check the consideration set mapping  $\Gamma_c$ . Firstly, it satisfies full attention over binary set by constructions. Secondly, to see that it is a competition filter on list  $\triangleright$ , we let  $S \subseteq T$  and  $y \in \Gamma_c(T)$ . Let  $z \in U_\triangleright(y) \cap S$ , we aim to show  $z \in \Gamma_c(S)$ . By construction, it must be either  $\Gamma_c(T) = m_{T,\triangleright}$  or  $\Gamma_c(T) = U_\triangleright(x) \cap T$ . If  $y = m_{T,\triangleright}$ , then it must be that  $U_\triangleright(y) \cap S = \{y\}$  and  $z = y$ . We then have  $z \in \Gamma_c(S)$  since it is the top element on the list. If  $y \neq m_{T,\triangleright}$ , then it must be that  $\Gamma_c(T) = U_\triangleright(x) \cap T$  so that  $c(T) = x$  and  $y \triangleright x$ . By ii), we have  $c(S) = x$  so that  $\Gamma_c(S) = U_\triangleright(x) \cap S$ . Since  $z \in U_\triangleright(y) \cap S$ ,  $z \triangleright y \triangleright x$ , we have  $z \in \Gamma_c(S)$ .

Lastly, we show that it explains the choice function. For  $|S| = 2$ , it is immediate; for  $|S| > 2$ , notice that  $c_{(\Gamma_c, \succ_c)}(S) = m_{\triangleright, S}$  if and only if  $\Gamma(S) = \{m_{\triangleright, S}\}$  if and only if  $c(S) = m_{\triangleright, S}$ . If  $c(S) = x \neq m_{\triangleright, S}$ , we have  $\Gamma_c(S) = U_\triangleright(x) \cap S$ . By ii), we must have  $c(\{x, z\}) = x$  for every  $z \in S$ . Hence,  $c_{(\Gamma_c, \succ_c)}(S) = x$ . The proof is complete.  $\square$

The idea of the rest of the proof goes as follows. We denote the set of all choice functions generated from  $(\Gamma, \succ) \in \mathcal{CF}_\triangleright \times \mathcal{P}_\triangleright(x)$  as  $\mathbb{C}_{\triangleright, x}$ . We will construct a sequence of choice



functions  $\{c_1, c_2, \dots, c_n\}$  where  $c_1(S) = x$  for all  $S$  containing  $x$  and whenever  $c_i(S) \neq x$ , we have  $c_i(S) = m_{S, \triangleright}$  for all  $i$ . We denote the set of all such sequences as  $\mathbb{Q}_{\triangleright, x}$  with typical element  $q_x$ . Then, given the choice rule  $\pi$ , for each  $x \neq m$ , we select a sequence  $q_x \in \mathbb{Q}_{\triangleright, x}$ . We will show that the  $q_x$  that we choose is a subset of  $\mathbb{C}_{\triangleright, x}$ . We will then assign weights to them and show that they jointly explain the data. To abuse notation, we write  $\tau(\cdot)$  as also the probability measure over choice functions  $\bigcup_{x \in X} \mathbb{C}_{\triangleright, x}$ . Firstly, for each  $x \in X$  and  $x \neq m$ , we gather all the conditions from Axiom 2 and 3 which are related to  $x$ , which are  $\pi(x, S) \geq \pi(x, T)$  for  $x \neq m_{S, \triangleright}$  and  $S \subseteq T$ ;  $\pi(x, \{x, y\}) \geq \pi(x, \{x, z\})$  for  $z \triangleright y \triangleright x$ .

We denote the sets containing  $x$  and appearing in the inequalities as  $\mathcal{S}_x$ . We let  $\mathcal{B}_0 \subseteq \mathcal{S}_x$  be the sets in  $\mathcal{S}_x$  that is *non-dominating*: If there does not exist  $S' \in \mathcal{S}_x$  and  $S' \neq S$  such that  $\pi(x, S) \geq \pi(x, S')$ , then  $S \in \mathcal{B}_0$ . Therefore,  $\mathcal{B}_0 = \{X\}$ . Then, we construct  $\mathcal{B}_i$  for  $i = 1, \dots, |\mathcal{S}_x|$  and  $B_i$  for  $i = 1, \dots, |\mathcal{S}_x|$  as follows.

$$B_i = \arg \min_{S \in \mathcal{B}_{i-1}} \rho(x, S) \text{ for } i = 1, \dots, |\mathcal{S}_x|$$

$$\mathcal{B}_i = \text{the sets in } \mathcal{S}_x \setminus \{B_0, \dots, B_i\} \text{ that are non-dominating, for } i = 1, \dots, |\mathcal{S}_x| - 1$$

Here, we assume that the minimizer  $B_i$  is unique for simplicity. Hence, since  $\mathcal{B}_0 = \{X\}$ , we have  $B_1 = X$ . Also,  $\mathcal{B}_1$  equals the sets of size  $N - 1$  containing  $x$ . Lastly, since minimizer is unique and there are  $|\mathcal{S}_x|$  sets to begin with, we have  $B_{|\mathcal{S}_x|} = \{b_x, x\}$ , where  $b_x$  is the immediate predecessor of  $x$  in  $X$  according to  $\triangleright$ , as  $p(x, \{b_x, x\}) \geq p(x, S)$  for all  $S \in \mathcal{S}_x$  by Axiom 2 and Axiom 3. Then, we pick  $q_x = \{c_1, c_2, \dots, c_{|\mathcal{S}_x|}\} \in \mathbb{Q}_{\triangleright, x}$  such that, for  $i = 2, 3, \dots, |\mathcal{S}_x|$

$$c_i(S) = \begin{cases} x & \text{if } S \notin \{B_1, \dots, B_{i-1}\} \\ m_{S, \triangleright} & \text{if } S \in \{B_1, \dots, B_{i-1}\} \end{cases}$$

We first verify that each of these choice functions  $c_i \in q_x$  belongs to  $\mathbb{C}_{\triangleright, x}$ . We check each condition in Claim 3. For *i*) it is immediate. For *ii*), assume that  $c_i(T) = x$  and let  $S \subset T$ . Since  $c_i(T) = x$ , it must be that  $T \notin \{B_1, \dots, B_{i-1}\}$ . Then, it must be that  $S \notin \{B_1, \dots, B_{i-1}\}$ . To see this, suppose not, i.e.  $S \in \{B_1, \dots, B_{i-1}\}$  and let  $S = B_j$  and  $j \leq i - 1$ . Since  $S = B_j := \arg \min_{S' \in \mathcal{B}_{j-1}} \rho(x, S')$ , we know  $S \in \mathcal{B}_{j-1}$ .  $\mathcal{B}_{j-1}$  are the sets in  $S_x \setminus \{B_0, \dots, B_{j-1}\}$  that are non-dominating. Yet, since  $T \notin \{B_1, \dots, B_{i-1}\}$ , we have  $T \in S_x \setminus \{B_0, \dots, B_j\}$ . Therefore,  $S$  is dominating in  $S_x \setminus \{B_0, \dots, B_j\}$  since  $\rho(x, S) \geq \rho(x, T)$  by Axiom 2. A contradiction. For *iii*), let  $c_i(\{y, x\}) = x$ , we need to show  $c_i(\{z, x\}) = x$  for all  $z \triangleright y$ . One can prove it using the same argument as above by using Axiom 3. For *iv*), for all  $z \triangleright x$ , the sets  $\{z, x\}$  do not appear in Axiom 2 or 3. Hence,  $\{z, x\} \notin \mathcal{S}_x$  and  $c_i(\{z, x\}) = x$  for all  $i$ .

Then, for  $x \neq m$ , we assign weights to  $q_x$ . In particular, we let  $\tau_\pi(\{c_1\}) = \pi(x, B_1) \geq 0$  and  $\tau_\pi(\{c_i\}) = \pi(x, B_i) - \pi(x, B_{i-1}) \geq 0$  for  $i = 2, \dots, |\mathcal{S}_x|$ . These are non-negative by construction. Also, It is easy to see that  $\tau_\pi(q_x) = \pi(x, \{b_x, x\})$  since  $B_{|\mathcal{S}_x|} = \{b_x, x\}$ . Lastly, we endow the (unique) choice function  $c^*$  in  $\mathbb{C}_{\triangleright, m}$  with weight  $1 - \sum_{x \in X \setminus m} \pi(x, \{b_x, x\})$ , i.e.  $\tau_\pi = 1 - \sum_{x \in X \setminus m} \pi(x, \{b_x, x\})$ , which is non-negative by Axiom 4. Hence, we get  $\tau_\pi(\bigcup_{x \in X} \mathbb{C}_{\triangleright, x}) = \sum_{x \in X} \tau_\pi(\mathbb{C}_{\triangleright, x}) = \tau_\pi(\{c^*\}) + \sum_{x \neq m} \tau_\pi(q_x) = 1 - \sum_{x \in X \setminus m} \pi(x, \{b_x, x\}) +$

$\sum_{x \in X \setminus m} \pi(x, \{b_x, x\}) = 1$ . Hence,  $\tau_\pi$  is a probability measure. Moreover, to see that  $\tau_\pi$  explains the choice data, consider  $x \neq m_{S, \triangleright}$ , for every  $S$ ,  $\pi_{\tau_\pi}(x, S) = \tau_\pi(\{c \in \mathbb{C}_{\triangleright, x} : c(S) = x\}) = \tau_\pi(\{c_1, c_2, \dots, c_n : B_n = S\}) = \pi(x, B_1) + \sum_{i=2}^n \pi(x, B_i) - \pi(x, B_{i-1}) = \pi(x, B_n) = \pi(x, S)$ . Lastly, for every  $S$ ,  $\pi_{\tau_\pi}(m_{S, \triangleright}, S) = 1 - \sum_{x \neq m_{S, \triangleright}} \pi_{\tau_\pi}(x, S) = 1 - \sum_{x \neq m_{S, \triangleright}} \pi(x, S) = \pi(m_{S, \triangleright}, S)$ . Hence, it explains the choice rule. On the other hand, when the minimizers  $B_i$ 's are not unique, one can set assign zero weight for some choice functions in  $q_x$  and the proof is basically the same. Hence, the proof is complete.  $\square$

## A.7 Proof of Theorem 7

*Proof.* Since  $\Gamma$  assigns full attention at  $S$  when  $|S| = 2$ , we focus on the preference types and skip  $\Gamma$  when denoting the choice function for each preference type for menus of two alternatives. Consider choice probability at  $S = \{a_i, a_j\}$  and  $i > j$ . Notice that for types  $\succ_{kl}$  where  $k \neq j$ ,  $c_{\succ_{kl}}(\{a_i, a_j\}) = a_j$ , since the list  $\triangleright$  agree with  $\succ_{kl}$  over alternatives other than  $a_k$ ; for types  $\succ_{kl}$  where  $k = j$  and  $i < l$ ,  $c_{\succ_{kl}}(\{a_i, a_j\}) = a_j$ , since  $\succ_{kl}$  still rank  $a_i$  higher than  $a_j$  even though  $a_j$  is moved higher; lastly, for types  $\succ_{kl}$  where  $k = j$  and  $i \geq l$ ,  $c_{\succ_{kl}}(\{a_i, a_j\}) = a_i$ , since  $\succ_{kl}$  moves  $a_i$  higher than  $a_k$ . Therefore, we know that, for  $i > j$   $\pi(a_i, \{a_i, a_j\}) = \tau\{(\Gamma, \succ_{il}) : l \leq j\}$ . Hence, for  $i \neq j \neq 1$ , we have  $\pi(a_i, \{a_i, a_j\}) - \pi(a_i, \{a_i, a_{j-1}\}) = \tau\{(\Gamma, \succ_{il}) : l \leq j\} - \{(\Gamma, \succ_{il}) : l \leq j-1\} = \tau(\succ_{ij})$ . Also, for  $i \neq j = 1$ ,  $\pi(a_i, \{a_i, a_j\}) = \tau\{(\Gamma, \succ_{il}) : l \leq 1\} = \tau(\succ_{i1})$ . Lastly, given a linear order  $\triangleright$ , we consider a partition of  $\mathcal{P}_\triangleright$ . We denote  $\mathcal{P}_\triangleright(x)$  as the subset of single improvement of  $\triangleright$  where each  $\succ$  in  $\mathcal{P}_\triangleright(x)$  only disagrees with  $\triangleright$  over  $x$ . By definition,  $\tau(\mathcal{P}_\triangleright(a_k)) = \tau\{(\Gamma, \succ_{kl}) : l \leq k-1\} = \pi(a_k, \{a_{k-1}, a_k\})$  for  $k \neq 1$ . Hence, for  $\triangleright$ , which is denoted as  $\succ_{ii}$  by an abuse of notation, we have  $\tau(\triangleright) = 1 - \sum_{k=2}^n \tau(\mathcal{P}_\triangleright(a_k)) = 1 - \sum_{k=2}^n \pi(a_k, \{a_{k-1}, a_k\})$ .  $\square$

## A.8 Proof of Theorem 8

*Proof.* We first show the lower bound. Suppose that the maximum is achieved at  $R \supseteq S$ . We enumerate the alternative in  $R$  by  $\langle a_{r_1}, \dots, a_{r_{|R|}} \rangle$  so that  $a_k = a_{r_\ell}$  and  $1 < \ell$ . For some  $a_{r_i}$  where  $i \geq \ell$  to be chosen, it must be that  $a_{r_i}$  is considered by certain choice types. Therefore, by List-based Attention Overload, in set  $R$ , these choice types must also have considered everything before  $a_{r_i}$ , including  $a_{r_\ell}$ . Hence, we have  $\phi(a_k | R) \geq \sum_{\ell \geq k} \pi(a_\ell | R)$ . Lastly, since List-based Attention Overload satisfies Attention Overload, it must be that  $\phi(a_k | S) \geq \phi(a_k | R)$ .

In the following, we let  $U_{\triangleright}(a_k)$  be the alternatives in  $X$  which are listed before  $a_k$  (including  $a_k$ ); that is, it is the weak upper contour set of  $a_k$  according to the list order. Therefore, for the upper bound, it suffices to show the bound  $\phi(a_k | S) \leq 1 - \sum_{b \in U_{\triangleright}(a_k) \cap S \setminus a_{s_1}} (\pi(b | a_{s_1}) - \pi(b | S))$ . It is because, given by Axiom 2,  $\max_{R \supseteq \{a_{s_1}, b\}} \pi(b | R) = \pi(b | a_{s_1})$  and  $\min_{a_{s_1}, b \in T \subseteq S} \pi(b | T) = \pi(b | S)$ . Firstly, fix an  $b \in U_{\triangleright}(a_k) \cap S \setminus a_{s_1}$ . For  $b$  to be chosen is  $S$ , it must have been considered by the preference types which rank  $b$  before  $a_{s_1}$ . i.e.  $(\Gamma, \succ)$  where  $\succ \in \{\succ_{ba_{s_1}}, \succ_{ba_{s_1-1}}, \dots, \succ_{ba_1}\}$ . If all of these types have paid attention to  $b$  in  $S$ , then  $\pi(b | a_1) - \pi(b | S) = 0$ , since full attention is assumed at binary sets. Therefore, the difference  $\pi(b | a_1) - \pi(b | S)$

captures the types which have not noticed  $b$  but would have chosen  $b$  if otherwise it is (counterfactually) discovered. Also, by List-based Attention Overload, these types must not have considered  $a_k$ , since  $a_k$  is after  $b$ . Since the types are independent, i.e. for two different  $b, b'$ ,  $(\Gamma, \succ)$  where  $\succ \in \{\succ_{ba_{s_1}}, \succ_{ba_{s_1-1}}, \dots, \succ_{ba_1}\}$  and  $(\Gamma, \succ)$  where  $\succ \in \{\succ_{b'a_{s_1}}, \succ_{b'a_{s_1-1}}, \dots, \succ_{b'a_1}\}$  are independent,  $\sum_{b \in U_{\succeq}(a_k) \cap S \setminus a_{s_1}} (\pi(b|a_{s_1}) - \pi(b|S))$  reveals the types who must not have paid attention to  $a_k$ . Therefore,  $1 - \sum_{b \in U_{\succeq}(a_k) \cap S \setminus a_{s_1}} (\pi(b|a_{s_1}) - \pi(b|S))$  is an upper bound for  $\phi(a_k|S)$ .

Lastly, we show that the upper bound is greater than the lower bound. Again, suppose that the maximum for the lower bound is achieved at  $R \supseteq S$ . We have  $1 - \sum_{b \in U_{\succeq}(a_k) \cap S \setminus a_{s_1}} (\pi(b|a_{s_1}) - \pi(b|S)) - \sum_{j \geq k} \pi(a_j|R) \geq (1 - \sum_{a_{s_j} \in U_{\succeq}(a_k) \cap S \setminus a_{s_1}} \pi(a_{s_j}|a_{s_1})) - \sum_{j > k} \pi(a_j|R) \geq (1 - \sum_{a_{s_j} \in U_{\succeq}(a_k) \cap S \setminus a_{s_1}} \pi(a_{s_j}|a_{s_j-1}) - \sum_{j > k} \pi(a_j|R) \geq \sum_{j > k} \pi(a_j|a_{j-1}) - \sum_{j > k} \pi(a_j|R) \geq \sum_{j > k} \pi(a_j|a_{s_1}) - \sum_{j > k} \pi(a_j|R) \geq 0$ . The second, third, fourth, and fifth inequalities are given by Axiom 3, Axiom 4, Axiom 3, and Axiom 2, respectively. Therefore, the proof is complete.  $\square$

## A.9 Proof of Theorem 9

*Proof.* Suppose a strict  $\pi$  has a HAOM $_{\triangleright}$  representation in  $\triangleright$  where  $a_1 \triangleright a_2 \dots \triangleright a_n$ . We will show that it must be that  $L_{\pi} = \triangleright \setminus \{(a_{n-1}, a_n)\}$ .

We will first prove  $\subseteq$ . Suppose there exists  $\{x, y\} \subseteq S \subseteq T$  such that  $\pi(x, S) < \pi(x, T)$ . By Axiom 2, it must be that  $x$  is the  $\triangleright$ -most in  $S$  and hence  $x \triangleright y$ . On other hand, suppose there exist  $z \in X$  such that  $\pi(z, \{y, z\}) > \pi(z, \{x, z\})$  and  $\pi(y, \{y, z\}) > \pi(y, \{x, y\})$ . By  $\pi(z, \{y, z\}) > \pi(z, \{x, z\})$  and Axiom 3, we know that it must not be the case that  $y \triangleright x \triangleright z$ . Also, by a rearrangement,  $\pi(z, \{y, z\}) > \pi(z, \{x, z\})$  implies  $\pi(z, \{y, z\}) + \pi(x, \{x, z\}) > 1$ . It must not be the case that  $y \triangleright z \triangleright x$ . To see this, suppose  $y \triangleright z \triangleright x$ . Suppose the immediate predecessor of  $z$  in  $X$  is  $z_X$ , and the immediate predecessor of  $x$  in  $X$  is  $x_X$ . Axiom 4 implies that  $\rho(z, \{z_X, z\}) + \rho(x, \{x, x_X\}) \leq 1$ . Also, by Axiom 4, we know that  $\rho(z, \{z_X, z\}) \geq \rho(z, \{y, z\})$  and  $\rho(x, \{x, x_X\}) \geq \rho(x, \{x, z\})$ . Therefore, we have  $\rho(z, \{y, z\}) + \rho(x, \{x, z\}) \leq 1$ . A contradiction. Therefore, it cannot be  $y \triangleright z \triangleright x$ . Analogously,  $\pi(y, \{y, z\}) > \pi(y, \{x, y\})$  imply that it cannot be either  $z \triangleright x \triangleright y$  or  $z \triangleright y \triangleright x$ . Therefore, it must be either  $x \triangleright y \triangleright z$  or  $x \triangleright z \triangleright y$ . In either case, we have  $(x, y) \in \triangleright$ .

For  $\supseteq$ , suppose  $(a_k, a_l) \in \triangleright$  where  $k < l$  and  $(a_k, a_l) \neq (a_{n-1}, a_n)$ . Therefore, there exists  $a_h$  such that  $k < l < h$ . Also, since  $\pi$  is strict, by Axiom 3, it must be that  $\pi(a_h, \{a_h, a_l\}) > \pi(a_h, \{a_h, a_k\})$ . On the other hand, it must also be that  $\pi(a_l, \{a_l, a_h\}) > \pi(a_l, \{a_l, a_k\})$ . Suppose instead  $\pi(a_l, \{a_l, a_h\}) < \pi(a_l, \{a_l, a_k\})$ . Therefore, we have  $1 < \pi(a_l, \{a_l, a_k\}) + \pi(a_h, \{a_l, a_h\})$ . Axiom 4 implies that  $\rho(a_l, \{a_l, a_{l-1}\}) + \rho(a_h, \{a_h, a_{h-1}\}) \leq 1$ . Then, by Axiom 3, we have  $\pi(a_l, \{a_l, a_k\}) + \pi(a_h, \{a_l, a_h\}) \leq 1$ . A contradiction. Therefore, we have both  $\pi(a_h, \{a_h, a_l\}) > \pi(a_h, \{a_h, a_k\})$  and  $\pi(a_l, \{a_l, a_h\}) > \pi(a_l, \{a_l, a_k\})$ . Then, we have  $(a_k, a_l) \in L_{\pi}$ . The proof is done.  $\square$

## A.10 Proof of Theorems 10 and 11

*Proof.* The lower bound constructions rely on the same idea used in Section 2.4. See the proof of Theorem 5.  $\square$

## References

- ABALUCK, J., AND A. ADAMS (2021): “What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses,” *Quarterly Journal of Economics*, 136(3), 1611–1663.
- AGUIAR, V. H. (2017): “Random Categorization and Bounded Rationality,” *Economics Letters*, 159, 46–52.
- AGUIAR, V. H., M. J. BOCCARDI, N. KASHAEV, AND J. KIM (2023): “Random utility and limited consideration,” *Quantitative Economics*, 14(1), 71–116.
- AUGENBLICK, N., AND S. NICHOLSON (2016): “Ballot position, choice fatigue, and voter behaviour,” *The Review of Economic Studies*, 83(2), 460–480.
- BARSEGHYAN, L., M. COUGHLIN, F. MOLINARI, AND J. C. TEITELBAUM (2021): “Heterogeneous Choice Sets and Preferences,” *Econometrica*, 89(5), 2015–2048.
- BARSEGHYAN, L., AND F. MOLINARI (2023): “Risk Preference Types, Limited Consideration, and Welfare,” *Journal of Business & Economic Statistics*, forthcoming.
- BARSEGHYAN, L., F. MOLINARI, AND M. THIRKETTLE (2021): “Discrete Choice under Risk with Limited Consideration,” *American Economic Review*, 111(6), 1972–2006.
- BAŞAR, G., AND C. BHAT (2004): “A Parameterized Consideration Set Model for Airport Choice: An Application to the San Francisco Bay Area,” *Transportation Research Part B: Methodological*, 38(10), 889–904.
- BISWAS, D., D. GREWAL, AND A. ROGGEVEEN (2010): “How the order of sampled experiential products affects choice,” *Journal of Marketing Research*, 47(3), 508–519.
- BRADY, R. L., AND J. REHBECK (2016): “Menu-Dependent Stochastic Feasibility,” *Econometrica*, 84(3), 1203–1223.
- CATTANEO, M. D., X. MA, Y. MASATLIOGLU, AND E. SULEYMANOV (2020): “A Random Attention Model,” *Journal of Political Economy*, 128(7), 2796–2836.
- CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2019): “Inference on Causal and Structural Parameters using Many Moment Inequalities,” *Review of Economic Studies*, 86(5), 1867–1900.
- CHERNOZHUKOV, V., D. CHETVERIKOV, K. KATO, AND Y. KOIKE (2022): “Improved Central Limit Theorem and Bootstrap Approximations in High Dimensions,” *Annals of Statistics*, 50(5), 2562–2586.

- DARDANONI, V., P. MANZINI, M. MARIOTTI, H. PETRI, AND C. J. TYSON (2023): “Mixture Choice Data: Revealing Preferences and Cognition,” *Journal of Political Economy*, 131(3), 687–715.
- DARDANONI, V., P. MANZINI, M. MARIOTTI, AND C. J. TYSON (2020): “Inferring Cognitive Heterogeneity From Aggregate Choices,” *Econometrica*, 88(3), 1269–1296.
- DEMIRKAN, Y., AND M. KIMYA (2020): “Hazard Rate, Stochastic Choice and Consideration Sets,” *Journal of Mathematical Economics*, 87, 142–150.
- DEMUYNCK, T., AND C. SEEL (2018): “Revealed Preference with Limited Consideration,” *American Economic Journal: Microeconomics*, 10(1), 102–131.
- ELLISON, G., AND S. F. ELLISON (2009): “Search, obfuscation, and price elasticities on the internet,” *Econometrica*, 77(2), 427–452.
- FALMAGNE, J.-C. (1978): “A representation theorem for finite random scale systems,” *Journal of Mathematical Psychology*, 18(1), 52–72.
- FISHBURN, P. C. (1998): “Stochastic Utility,” in *Handbook of Utility Theory*, ed. by S. Barbera, P. Hammond, and C. Seidl, pp. 273–318. Kluwer Dordrecht.
- GENG, S. (2016): “Decision Time, Consideration Time, and Status Quo Bias,” *Economic Inquiry*, 54(1), 433–449.
- GIBBARD, P. (2021): “Disentangling preferences and limited attention: Random-utility models with consideration sets,” *Journal of Mathematical Economics*, 94, 102468.
- GOEREE, M. S. (2008): “Limited Information and Advertising in the US Personal Computer Industry,” *Econometrica*, 76(5), 1017–1074.
- GUNEY, B. (2014): “A theory of iterative choice in lists,” *Journal of Mathematical Economics*, 53, 26–32, Special Section: Economic Theory of Bubbles (I).
- HAUSER, J. R., AND B. WERNERFELT (1990): “An Evaluation Cost Model of Consideration Sets,” *Journal of Consumer Research*, 16(4), 393–408.
- HUBERMAN, G., AND T. REGEV (2001): “Contagious Speculation and a Cure for Cancer: A Nonevent That Made Stock Prices Soar,” *Journal of Finance*, 56(1), 387–396.
- ISHII, Y., M. KOVACH, AND L. ÜLKÜ (2021): “A model of stochastic choice from lists,” *Journal of Mathematical Economics*, 96, 102509.
- IYENGAR, S. S., AND M. R. LEPPER (2000): “When choice is demotivating: Can one desire too much of a good thing?,” *Journal of Personality and Social Psychology*, 79(6), 995–1006.
- KASHAEV, N., AND V. H. AGUIAR (2022): “A random attention and utility model,” *Journal of Economic Theory*, 204, 105487.

- LAROCHE, M., J. ROSENBLATT, AND I. SINCLAIR (1984): “Brand Categorization Strategies in an Extensive Problem Solving Situation: A Study of University Choice,” *ACR North American Advances*, NA-11.
- LEVAV, J., M. HEITMANN, A. HERRMANN, AND S. IYENGAR (2010): “Order in Product Customization Decisions: Evidence from Field Experiments,” *Journal of Political Economy*, 118(2), 274–299.
- LLERAS, J. S., Y. MASATLIOGLU, D. NAKAJIMA, AND E. Y. OZBAY (2017): “When More Is Less: Limited Consideration,” *Journal of Economic Theory*, 170, 70–85.
- MANZINI, P., AND M. MARIOTTI (2014): “Stochastic Choice and Consideration Sets,” *Econometrica*, 82(3), 1153–1176.
- MARSCHAK, J. (1959): “Binary choice constraints on random utility indicators,” .
- MATZKIN, R. L. (2007): “Nonparametric Identification,” in *Handbook of Econometrics, Volume VIB*, ed. by J. Heckman, and E. Leamer, pp. 5307–5368. Elsevier Science B.V.
- (2013): “Nonparametric Identification in Structural Economic Models,” *Annual Review of Economics*, 5, 457–486.
- MOLINARI, F. (2020): “Microeconometrics with Partial Identification,” in *Handbook of Econometrics, Volume VIIA*, ed. by S. Durlauf, L. Hansen, J. Heckman, and R. Matzkin, pp. 355–486. Elsevier Science B.V.
- PAN, B., H. HEMBROOKE, T. JOACHIMS, L. LORIGO, G. GAY, AND L. GRANKA (2007): “In Google we trust: Users’ decisions on rank, position, and relevance,” *Journal of computer-mediated communication*, 12(3), 801–823.
- PERNICE, K., K. WHITENTON, J. NIELSEN, ET AL. (2018): “How People Read Online: The Eyetracking Evidence,” .
- PURCELL, K., L. RAINIE, AND J. BRENNER (2012): “Search engine use 2012,” .
- REUTSKAJA, E., AND R. M. HOGARTH (2009): “Satisfaction in Choice as a Function of the Number of Alternatives: When ”Goods Siate”,” *Psychology and Marketing*, 26(3), 197–203.
- REUTSKAJA, E., R. NAGEL, C. F. CAMERER, AND A. RANGEL (2011): “Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study,” *American Economic Review*, 101(2), 900–926.
- ROSEN, D. E., J. M. CURRAN, AND T. B. GREENLEE (1998): “College Choice in a Brand Elimination Framework: The High School Students Perspective,” *Journal of Marketing for Higher Education*, 8(3), 73–92.
- RUBINSTEIN, A., AND Y. SALANT (2006): “A model of choice from lists,” *Theoretical Economics*, 1(1), 3–17.

- SHERIDAN, J. E., M. D. RICHARDS, AND J. W. SLOCUM (1975): “Comparative Analysis of Expectancy and Heuristic Models of Decision Behavior,” *Journal of Applied Psychology*, 60(3), 361–368.
- SHOCKER, A. D., M. BEN-AKIVA, B. BOCCARA, AND P. NEDUNGADI (1991): “Consideration Set Influences on Consumer Decision-making and Choice: Issues, Models, and Suggestions,” *Marketing Letters*, 2(3), 181–197.
- SIMON, H. A. (1955): “A behavioral model of rational choice,” *The quarterly journal of economics*, pp. 99–118.
- TURANSICK, C. (2022): “Identification in the Random Utility Model,” *Journal of Economic Theory*, p. 105489.
- TVERSKY, A. (1972): “Elimination by aspects: A theory of choice.,” *Psychological review*, 79(4), 281.
- VISSCHERS, V. H., R. HESS, AND M. SIEGRIST (2010): “Health Motivation and Product Design Determine Consumers Visual Attention to Nutrition Information on Food Products,” *Public Health Nutrition*, 13(7), 1099–1106.
- WESTERWICK, A. (2013): “Effects of sponsorship, web site design, and Google ranking on the credibility of online information,” *Journal of Computer-Mediated Communication*, 18(2), 194–211.
- YILDIZ, K. (2016): “List-rationalizable choice,” *Theoretical Economics*, 11(2), 587–599.