Attention Overload

Matias D. Cattaneo[*] Paul Cheung[†] Xinwei Ma[‡] Yusufcan Masatlioglu[§]

October 19, 2021

Abstract

We introduce an Attention Overload Model that captures the idea that alternatives compete for the decision maker's attention, and hence the attention frequency each alternative receives decreases as the choice problem becomes larger. Using this nonparametric 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 partially identify the decision maker's preference. Furthermore, we provide novel partial identification results on the underlying attention frequency, thereby offering the first nonparametric identification result of (a feature of) the random attention formation mechanism in the literature. Building on our partial identification results, for both preferences and attention frequency, we develop econometric methods for estimation and inference. Importantly, our econometric procedures remain valid even in settings with large number of alternatives and choice problems, an important 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.

[*]Department of Operations Research and Financial Engineering, Princeton University.
[†]Department of Economics, University of Maryland.
[‡]Department of Economics, UC San Diego.
[§]Department of Economics, University of Maryland.

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.[1] It is without doubt that consumers cannot pay attention to all products: some are going to be more appealing than the others, while other are completely unnoticed.

It is now well established in economics, marketing, and other 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 fragmentized and scattered due to fierce advertising competition. According to Statista, the US has spent over $253 billion dollars in advertisement in 2019. The burden on consumers becomes excessive with the bombardment of advertisement, cognitive overload, and abundance of alternative, resulting in them not paying attention to some of the available products.[2] 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. More precisely, 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). In line with the literature, our model has two components: a random attention rule and a preference ordering. 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. Following the traditional insight from economics, we assume that decision makers have a complete and transitive preference over the alternatives, and that they pick the best alternative in their consideration sets. In this general setting where attention is random and

[1]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 (Demuynck and Seel, 2018), PC purchases (Goeree, 2008), and airport choice (Başar and Bhat, 2004), just to mention a few examples.

[2]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 less item (Reutskaja and Hogarth, 2009; Reutskaja, Nagel, Camerer, and Rangel, 2011) or choose outside option more often (Iyengar and Lepper, 2000), as choice sets expands.

limited, and products compete for attention, our goal is twofold: (i) to uncover preferences and (ii) to uncover the attention frequency (and hence to learn about a feature of the attention rule), in both cases solely from observed choices.

Using our proposed (attention overload) nonparametric restriction on the attention frequency, we achieve both preference ordering and attention frequency elicitation from observed data, which in the econometrics language translates into (point or partial) identification of those model primitives. Neither our identification results nor the companion econometric methods we present in this paper 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 (Molinari, 2020).

Interestingly, many existing (random) attention models cannot capture, or are incompatible with, attention overload. Manzini and Mariotti (2014) consider an attention model with independent consideration where each alternative has a constant attention frequency even when there are more alternatives. Hence, their model cannot provide insights on consumer behavior when attention frequency is strictly monotonic. Aguiar (2017), Horan (2019) and Gibbard (2021) 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. Notice that as long as the assumption of attention overload is satisfied, those models will become special cases of our model because we do not assume a particular (parametric) form of attention rule. Lastly, the recent models of Brady and Rehbeck (2016) and Cattaneo, Ma, Masatlioglu, and Suleymanov (2020) fail to satisfy the attention overload assumption. Hence, it is possible that an alternative is getting less attention even when the choice set gets smaller. In Section 5 we provide an in-depth discussion of related literature leveraging our notation and results introduced in the upcoming sections.

Nonetheless, there are a number of special cases of our attention overload model. First, the idea of competition filter in Lleras, Masatlioglu, Nakajima, and Ozbay (2017) is a special case of our model in the deterministic environment, which says that any item winning the consumer's attention would also prevail in the smaller set. Second, notions of rationalization (e.g. Cherepanov, Feddersen, and Sandroni 2013) and categorization (e.g. Manzini and Mariotti 2012) are nested by our model. It means that any revealed preference obtained under our nonparametric attention overload restriction would also hold in their (more specialized) environments. See Section 5 for more details.

The fact that attention is not observed by the analyst 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 a 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. Nevertheless, we show that our attention overload assumption, despite being very general, still delivers nontrivial empirical content. To be more precise, we first 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 we call *Attention Compensation*. We will also demonstrate in Section 4 how our baseline model can be augmented with additional restrictions on the attention rule to strengthen identification.

Besides revealed preference, information about the attention frequency also has practical application and is an object of interest. For example, it enables marketers to gauge the effectiveness of their marketing strategies, and policy-markers to assess whether consumers allocate their attention on the better products. Despite the underlying attention rule may not be identifiable, we also show in Section 2 that our nonparametric behavioral restriction (attention overload) allows us to partially identify the attention frequency using standard choice data, where sometimes the identified set is a singleton, thereby giving point identification. More specifically, we show that lower and upper bounds for the attention frequency can be constructed from the estimable choice rule. This appears to be the first such nonparametric identification result of a relevant feature of an attention rule in the random limited attention literature. In other words, until this paper, revealed attention analysis has not been possible under nonparametric identifying restrictions.

Based on our main elicitation results, we provide econometric methods for revealed preference and attention analysis in Section 3, which can be applied to standard choice data. That is, we only assume that a random sample of choice problems and choices selections is observed, and then develop methods for estimation of and inference on the preference ordering and the attention frequency of the decision makers. To be precise, for revealed preference, our econometric inference technique compares the estimated choice probabilities to a critical value, where the latter can be found by simulation from (possibly high-dimensional) multivariate normal distributions. This allows us to (i) test whether a specific preference ordering is compatible with our attention overload model, (ii) construct (asymptotically) valid confidence sets, and (iii) conduct overall model specification testing. For revealed attention, we operationalize our partial identification results by first adjusting the estimated choice probabilities before forming nonparametric bounds. The adjustment term, which also relies on a critical value simulated from normal distributions, accounts for estimation uncertainty.

To establish the validity of our econometric methods, we employ the latest results on high-dimensional normal approximation (Chernozhukov, Chetverikov, Kato, and Koike, 2021). This is crucial because the number of inequality constraints involved in our statistical inference procedure may not be small relative to the sample size. While allowing the dimension (complexity) of the problem to be potentially much larger than the sample size, we explicitly characterize the error from a normal approximation to the estimated choice probabilities, which sheds light on the finite-sample performance of our proposed econometric methods. To be more precise, we prove that the false rejection frequency (i.e., the probability of rejecting a preference ordering which is actually compatible with our attention overload model) is bounded by α, the desired significance level, plus a normal approximation error. This approximation error depends on the sample size, the standard error of the estimated choice probabilities, and the number of inequality constraints. Importantly, the dependence on the number of inequality constraints is logarithmic, meaning that our econometric methods remain valid even if "many" inequality constraints are involved. A similarly result holds for the estimated upper and lower bounds of the attention frequency: we show that the confidence region has the desired coverage, $1-\alpha$, plus an error term from normal approximation that we precisely quantify.

Econometric methods based on revealed preference theory are of crucial importance, and have a long tradition, in economics and many other social and behavioral sciences. See Matzkin (2007, 2013), Molinari (2020), and references therein, for review of econometric methods for partial elicitation of preferences based on revealed preference theory. There is only a handful of recent studies bridging behavioral 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 identification assumptions on the random consideration and the preference, producing different levels of identification of preference and consideration sets. See Section 5 for more details. Our paper contributes to this emerging literature by providing new results on the identification of and inference on the preference ordering and the underlying attention frequency when decisions makers only pay attention to a subset of possibly too many alternatives at random.

After presenting our main indetification, estimation and inference results under attention overload in the canonical setup, we consider two main generalizations in Section 4. First, as we mentioned above, one of the behavioral consequences of having too many alternatives in the choice set is the likelihood of deferring choice. That is, consumers tend to choose a default option more often as the number of options increases (Iyengar and Lepper, 2000). To capture this phenomenon, we consider an extension of our baseline attention overload model and introduce a default option. We show that our model explain the phenomenon that the default option is chosen more often when the choice set expands. The intuition is simple: the competition between alternatives gets more fierce as the decision problems get bigger, and therefore the decision maker tends not to consider any alternative and choose the default option. Interestingly, existing models of random attention rule predict the other way.[3]

Due to the generality of our model, preference ordering over some alternatives may not be identifiable in some datasets, that is, in some applications the observe data may be uninformative.

[3] That is, the outside option is chosen more often in the smaller set. In Manzini and Mariotti (2014) and Brady and Rehbeck (2016), this is borne out in their attention rule. For Cattaneo, Ma, Masatlioglu, and Suleymanov (2020), the restriction is implied by their monotonic attention assumption.

In other words, the revealed preference needs not be complete. This should not be surprising as we only impose one nonparametric restriction on the attention rule. However, we also show in Section 4 that our baseline attention overload model can be augmented with additional restrictions, which can help to bring additional empirical content and hence lead to additional preference revelation. Among the many possibilities, we investigate nonparametric restrictions over binary comparisons. We show that it aids in revealed preference and we provide a joint characterization result. Another nonparametric restriction on the attention rule that complements ours is the monotonic attention assumption introduced by Cattaneo, Ma, Masatlioglu, and Suleymanov (2020). As we explained above, their monotonic attention assumption does not nest our model, nor does our model nest theirs. If the choice data can be explained both by our attention overload model and by their random attention model, then it is possible to combine the two restrictions to facilitate preference revelation. We further discuss this point in Section 5, and provide numerical evidence in Section 6.

The remaining of the paper proceeds as follows. In the next section (Section 2), we introduce our setup and key attention overload assumption. We also prove the main characterization result that suggests that a specific preference ordering is compatible with our attention overload model if and only if the choice behavior satisfies a system of inequality constraints (i.e., the attention compensation property). That section also presents the partial identification of attention frequency, and show that sharp upper and lower bounds can be formed from the choice rule. Building on these theoretical results, we provide econometric methods in Section 3 for revealed preference and revealed attention analyses. Section 4 contains important extensions to our baseline attention overload setting, including how a default option can be accommodated, and how other restrictions on the attention rule can be incorporated to bring addition identification power. To streamline the presentation, an in-depth related literature discussion is given in Section 5, where we show precisely how our attention overload model distinguishes itself from existing models featuring (random) limited attention as well as other models of rational and behavioral choice. Section 6 presents simulation evidence showcasing the empirical performance of our theoretical and methodological results for preference and attention elicitation. Section 7 concludes. The appendix contains the main proofs of our results, and the online supplemental appendix collects complementary material (additional discussions and proofs of the technical lemmas). Last but not least, we also provide a software package 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. To introduce formally this idea, we first define the primitives in the environment. We denote the grand alternative set as X, and its cardinality by $|X|$. A typical element of X is denoted by a. We let $\mathcal{X}$ be the collection of all non-empty subsets of X, so that each member of $\mathcal{X}$ defines a choice problem. We then define the choice rule (i.e., choice probabilities).

Definition 1 (Choice Rule)**.** A choice rule is a map $\pi : X \times \mathcal{X} \to [0,1]$ such that for all $S \in \mathcal{X}$, $\pi(a|S) \geq 0$ for all $a \in S$, $\pi(a|S) = 0$ for all $a \notin S$, and $\sum_{a\in S} \pi(a|S) = 1$.

Therefore, $\pi(a|S)$ represents the probability that the decision maker chooses alternative a from the choice problem S. Of course, this formulation allows for deterministic choice rules, where $\pi(a|S)$ will be either 0 or 1. The key ingredient of our model is that consideration sets can be random. Given a choice problem S, each non-empty subset of S could be a consideration set with certain probability, that is, each consideration set has a frequency between 0 and 1, where the frequencies of consideration sets add up to 1. Formally,

Definition 2 (Attention Rule)**.** An attention rule is a map $\mu : \mathcal{X} \times \mathcal{X} \to [0,1]$ such that for all $S \in \mathcal{X}$, $\mu(T|S) \geq 0$ for all $T \subseteq S$, $\mu(T|S) = 0$ for all $T \not\subseteq S$, and $\sum_{T\subseteq S} \mu(T|S) = 1$.

Thus, $\mu(T|S)$ 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. For example, $\mu(S|S) = 1$ represents full attention.

Another economically important quantity, which captures one of the main innovations in our paper, is the amount of attention each *alternative* captures for a given μ. We can extract this information from μ by simply summing up the frequencies of consideration sets containing the alternative. That is, for a given μ, the probability that a attracts attention in S is defined as

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

When μ is clearly defined, we omit the subscript μ for convenience. 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 necessary be a zero-or-one dichotomy. We regard it as the *attention frequency* of an alternative.

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

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

Note that 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, which we further investigate in Section 4.1. At this point we exclude the possibility of paying attention to nothing for simplicity.

We say that an attention rule μ satisfies attention overload if its corresponding attention frequency is monotonic in the sense of Assumption 1. Section 5 lists several choice models satisfying Assumption 1, and also discusses our main results in light of the related literature. For example, in Manzini and Mariotti (2014) the attention frequency is held fixed. Another example is the random competition filter, which is a major special case of our model.

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

Definition 3 (Attention Overload Representation)**.** A choice rule π has an attention overload representation if there exists a preference ordering $\succ$ over X and an attention rule μ 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 $S \in \mathcal{X}$. In this case, we say π is represented by $(\succ, \mu)$. We also say π is an *Attention Overload Model* (AOM).

It should be clear that the model primitives are the attention rule μ and the preference ordering $\succ$. In this paper, however, we only assume that the choice rule π is estimable (i.e., point identifiable in econometrics language). In other words, we do not require any additional information beyond standard choice data. We will show in the next subsection that it is possible to partially identify the underlying preference ordering by exploiting the attention overload Assumption 1. In Section 2.2, we present novel partial identification results on the attention frequency.

2.1 Revealed Preference

Given choice data that satisfies the AOM, it possible to (partially) identify the decision maker's underlying preference ordering. We exploit the fact that the attention frequency ϕ satisfies attention overload, meaning that each alternative gets more attention when the choice problem is smaller. It is then natural to *expect* that each alternative would be more likely to be picked when there are fewer alternatives available. However, if we observe the opposite (i.e. an alternative has a lower probability of being selected in a smaller choice problem), we could then deduce that there must be something better than the original choice in the smaller choice problem. To formalize this identification idea, we first define revealed preference in the model.

Definition 4 (Revealed Preference)**.** Let $\{(\succ_\ell, \mu_\ell)\}_{\ell \in \mathscr{L}}$ be all attention overload representations of π. We say that b is revealed to be preferred to a if $b \succ_\ell a$ for all $\ell \in \mathscr{L}$.

This revealed preference definition checks all possible representations for the choice rule π and concludes that b is revealed to be preferred to a only if all possible representations agree on this binary comparison. This conservative specification is employed in most (if not all) limited consideration behavioral models; see Section 5 for references and further discussion. According to this definition, if one wants to know whether a is revealed to be preferred to b, one must uncover all possible $\{(\succ_\ell, \mu_\ell)\}$ representations. To see whether a particular preference, $\succ$, can represent the data, one needs to construct an attention rule, μ satisfying AOM and $(\succ, \mu)$ represents the data. When there are many alternatives, the set of all representations would be quite large and constructing an attention rule satisfying attention overload could be a tedious task. Hence, this method is not practical to identify the preference. Instead, we shall now provide a practical method to uncover the revealed preference completely.

Our first observation on the identification of preference is that *regularity violations* at binary choice problems reveal the decision maker's preference. More specifically, we illustrate that if $a, b \in S$ and $\pi(a|S) > \pi(a|\{a,b\})$, then b must be preferred to a. To reach such a conclusion, assume the contrary: there exists $(\succ, \mu)$ representing π such that $a \succ b$ and μ satisfies attention overload. First note that attention is a perquisite for choice: the decision maker must first pay attention to an alternative in order to be able to choose it. Hence, 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 with our attention overload assumption, thereby providing preference identification power. The next lemma formally states this observation.

Lemma 1 (Regularity Violation at Binary Comparisons)**.** Let π be an AOM, and $a, b \in S$. If $\pi(a|S) > \pi(a|\{a,b\})$, then b is revealed to be preferred to a.

Note that Lemma 1 provides a guideline to reveal preference without knowledge about each particular representation. Hence, it provides a handy method. Now, the key question is whether we can generalize the implication of Lemma 1 for an arbitrary set $T \subseteq S$ instead of only for binary sets. The example below demonstrates why the answer is not straightforward.

Example 1. Consider the following choice data.

$\pi(\cdot|S)$	a	b	c	d
$\{a,b,c,d\}$	0	0.2	0.3	0.5
$\{b,c,d\}$		0.25	0	0.75

Notice that $\pi(c|\{a,b,c,d\}) > \pi(c|\{b,c,d\})$. By attention overload, we have $\phi(c|\{b,c,d\}) \geq \phi(c|\{a,b,c,d\})$. Next, we further split the attention frequency according to whether c is actually chosen:

$$\Rightarrow \quad \pi(c|\{b,c,d\}) + \sum_{\substack{T:\ c\in T\subseteq\{b,c,d\} \\ c \text{ is not } \succ\text{-best in } T}} \mu(T|\{b,c,d\}) \;\geq\; \pi(c|\{a,b,c,d\}) + \sum_{\substack{T:\ c\in T\subseteq\{a,b,c,d\} \\ c \text{ is not } \succ\text{-best in } T}} \mu(T|\{a,b,c,d\})$$

$$\Rightarrow \qquad \sum_{\substack{T:\ c\in T\subseteq\{b,c,d\}\\ c \text{ is not } \succ\text{-best in } T}} \mu(T|\{b,c,d\}) \ \geq \pi(c|\{a,b,c,d\}) - \pi(c|\{b,c,d\}) = 0.3 > 0.$$

This implies that there exists something better than c in the set $\{b, c, d\}$. However, we could not identify whether this alternative is b or d (might be both). ⌟

As the above example demonstrates, identification from regularity violation may not be informative when there are more than two alternatives in the smaller set. Notice that from Lemma 1 we are able to claim revealed preference between two alternatives. However, when the smaller set contains more than two alternatives, we only know there is some alternative better than c in the smaller set. We put this observation in the following Lemma.

Lemma 2 (Regularity Violation at Bigger Choice Problems)**.** Let π 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 Lemma 1 and 2 are based on regularity violations. We now demonstrate that regularity violation alone does not exhaust all the identification power of our attention overload assumption. Recall that in the above example, Lemma 2 indicates that either b or d (or both) should be preferred to c. Hence, there can be four possible preferences compatible with the choice data: $b \succ_1 c \succ_1 d$, $d \succ_2 c \succ_2 b$, $b \succ_3 d \succ_3 c$, and $d \succ_4 b \succ_4 c$. Notice that the first one is the only ordering which ranks c above d. We now show that this ordering is not compatible with the data, hence, d must be *revealed preferred* to c. This observation illustrates that the additional revelation is overlooked by regularity violation and the argument in Example 1. To see this, we suppose the opposite, that is, there exists a AOM which represents the data with $c \succ d$. It gives $b \succ c \succ d$. Then, from the last line in Example 1, the left-hand side reduces to the frequency of paying attention to a set containing both b and c, and this frequency is bounded from above by $\pi(b|\{b, c, d\}) = 0.25$. That is,

$$0.25 \geq \mu(\{b, c\}|\{b, c, d\}) \geq \pi(c|\{a, b, c, d\}) - \pi(c|\{b, c, d\}) = 0.3.$$

which leads to a contradiction. Hence, every representation must agree that $d \succ c$.

To capture this hidden revelation and provide a sharper identification result, we introduce a new condition on a pair of choice data and preference. With a slight abuse of notation, we define

$$\pi(U_{\succ}(a)|S) = \sum_{b\in S:\ b\succ a} \pi(b|S),$$

which is the probability of choosing an alternative from the strict upper contour set of a in S, where $U_{\succ}(a)$ denotes this strict upper contour set. Similarly, let $\pi(U_{\succeq}(a)|S) = \pi(U_{\succ}(a)|S) + \pi(a|S)$, and $U_{\succeq}(a)$ is the weak upper contour set of a. Intuitively, the following definition says that the weak upper contour set of a in the smaller choice problem must *compensate* the (minimal) attention a receives in the bigger choice problem.

Definition 5 (Attention Compensation). $(\pi, \succ)$ satisfies Attention Compensation (AC) if for all $a \in T \subseteq S$, $\pi(U_{\succeq}(a)|T) \geq \pi(a|S)$.

Notice that if π satisfies regularity, it also satisfies AC for any preference ordering. On the other hand, the existence of regularity violation restricts the set of possible preferences that satisfy AC with π. In the choice data provided in Example 1, among the four possible rankings on $\{b, c, d\}$, $\succ_1$ is the only one violating AC with the choice data. To see this,

$$\pi(U_{\succeq_1}(c)|\{b, c, d\}) = \pi(\{b, c\}|\{b, c, d\}) = 0.25 < 0.30 = \pi(c|\{a, b, c, d\})$$

Hence, AC correctly identifies the set of orderings compatible with the data in this example, namely $\succ_2, \succ_3$ and $\succ_4$. Note that all these orderings rank d above c, which is the *revealed preference*. This suggests that AC could be a handy method to identify the underlying preference. We first define a binary relation based on AC property:

$$b\mathrm{P}^{\mathtt{AC}}a \quad \text{if } b \succ a \text{ for all preference orderings such that } (\pi, \succ) \text{ satisfies AC,} \tag{1}$$

which is in spirit similar to Definition 4. However, $\mathrm{P}^{\mathtt{AC}}$ does not suffer from the critique for Definition 4, which requires to construct all AOM representations. Given preferences, finding the corresponding attention rule satisfying attention overload could be a daunting task. On the other hand, checking whether $(\pi, \succ)$ satisfy AC is straightforward. This is because AC bypasses the requirement that one must find the corresponding attention rule. Indeed, we utilize this fact in Section 3.1 to develop econometric methods. The next theorem states a surprising result that the revealed preference given by Definition 4 is equivalent to $\mathrm{P}^{\mathtt{AC}}$.

Theorem 1 (Revealed Preference). Let π be an AOM. Then, b is revealed preferred to a if and only if $b\mathrm{P}^{\mathtt{AC}}a$.

Theorem 1 can capture all binary revelations from the data. As Lemma 2 states, we might infer more even when $\mathrm{P}^{\mathtt{AC}}$ is empty. For example, consider the modified version of Example 1.

$\pi(\cdot\|S)$	a	b	c	d
$\{a, b, c, d\}$	0	0.2	0.3	0.5
$\{b, c, d\}$		0.5	0	0.5

While $\mathrm{P}^{\mathtt{AC}}$ is empty, Lemma 2 still informs us that the underlying preference cannot be $c \succ b \succ d$ or $c \succ d \succ b$, because AC does not hold for these preferences. Hence, AC could be a useful tool to identify all compatible preference orderings. Indeed, we will utilize AC when we provide a characterization for AOM.

2.2 Revealed Attention

The attention overload model is built upon a simple nonparametric requirement that each alternative gets weakly less attention in bigger choice problems, which is captured by the monotonicity in attention frequency. 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; policy markers could be interested in assessing whether consumers allocate their attention on better products. However, since we do not put any restriction on the formation of attention rule, the attention frequency can vary depending on the actual attention rule that the decision maker has. Nonetheless, we show that it is possible to develop upper and lower bounds for the attention frequency, even without a specification of the attention rule.

As a first step towards this goal, consider bounding ϕ from below, which can be achieved as follows. Take any superset $R \supseteq S$, then by our attention overload assumption, one has $\pi(a|R) \leq \phi(a|R) \leq \phi(a|S)$. 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 is not built upon any particular AOM representation, that is, it does not require knowledge on the underlying attention rule.

It is also possible to derive an upper bound, although in this case the bound will depend on a preference. Consider a preference ordering $\succ$ and an attention rule satisfying attention overload, so that π is an AOM with $(\succ, \mu)$. Then, for any subset $T \subseteq S$, $\phi(a|S) \leq \phi(a|T) \leq \pi(U_{\succeq}(a)|T)$, which implies that $\phi(a|S) \leq \min_{T \subseteq S} \pi(U_{\succeq}(a)|T)$. We summarize these observations into the following Theorem.

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

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

We now consider three extreme cases of this theorem. If both the lower bound and the upper bound is 1, we say a attracts full attention at S (Revealed Full Attention). If each bound is 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 is 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. Theorem 2 is a novel contribution in the competing attention framework.

Since the stochastic data is richer, Theorem 2 covers another interesting case, which we call partial identification. That is, the upper bound is strictly below one and/or the lower bound is strictly above zero (Partial Revealed Attention). In some cases, the attention frequency will be uniquely identified for certain alternative a and set S. For example, for data $\pi(b|\{b,c,d\}) = \pi(c|\{b,c,d\}) = 0.3$ and $\pi(c|\{a,b,c,d\}) = 0.6$ with candidate preference $a \succ b \succ c \succ d$ (possibly coming from the revealed preference analysis on the data), we immediately conclude that c's attention frequency is 0.6 in the set $\{b,c,d\}$. It is because the upper bound for $\phi(c|\{b,c,d\})$ is $\pi(U_{\succeq}(c)|\{b,c,d\}) = \phi(\{b,c\}|\{b,c,d\}) = 0.6$ and the lower bound is $\pi(c|\{a,b,c,d\}) = 0.6$.

We must highlight the fact that these bounds are tight. In other words, for any number, say c, in this interval, there exists an attention rule, μ, satisfying attention overload such that the attention frequency of a at S is c.

Theorem 2 is useful in real world applications where it informs a firm (a government) how much attention each product (policy) receives among other options. While the lower bounds 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, μ, 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, it turns out that the answer is affirmative. We state this result in the following theorem.

Theorem 3 (Pessimistic Evaluation for Attention)**.** Let π be an AOM and $(\succ, \mu)$ represent π. Then there exists a pessimistic attention rule μ^* such that $(\succ, \mu^*)$ is also an AOM representation of π. 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 well.

2.3 Characterization

We have shown how to achieve revealed preference and attention for our model, and to do this, we assumed that the data is consistent with the AOM. In other words, our revelation methods are only applicable when the observed choice behavior has an AOM representation. It turns out that AC also plays a critical role to test our model: by taking a stochastic choice and a preference ordering as inputs, AC determines whether the data is compatible with the AOM. For this reason, we now state AC as a behavioral postulate.

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

Our axiom applies to the stochastic choice data, π, an observable object, and is stated in terms of preferences, $\succ$, an unobservable primitive of the model. Given $\succ$, it is routine to check whether $(\pi, \succ)$ satisfies AC. This axiom is closely related to, but different from, the classical regularity condition. This axiom trivially implies the regularity condition for the best alternative, a^*, 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 will be partially relaxed. At the other extreme, AC does not restrict the choice probabilities for the worst alternative, a_*, since $U_{\succeq}(a_*) = X$, and hence for all T,

$$\pi(U_{\succeq}(a_*)|T) = 1 \geq \pi(a_*|S),$$

meaning that AC holds trivially.

We now establish the necessity of AC for AOM. Assume that $(\succ, \mu)$ represents π and $T \subset S$. Since attention is a requirement for the choice, the 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 $\phi(a|T) \geq \phi(a|S) \geq \pi(a|S)$. The difference $\phi(a|T) - \pi(a|T)$ captures the probability that a receives attention but not chosen in T. As a consequence, in these cases, a better option must be chosen in T, which implies

$$\phi(a|T) - \pi(a|T) \leq \pi(U_{\succ}(a)|T).$$

Combining all these observations, we get

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

Therefore, AC must be necessary for AOM. The next theorem states that it is also sufficient.

Theorem 4 (Characterization)**.** The pair $(\pi, \succ)$ is an AOM if and only if $(\pi, \succ)$ satisfies AC.

An immediate corollary of this result is that π is AOM if and only if there exists $\succ$ such that AC is satisfied. We provide above the proof of the necessity of AC. The sufficiency of AC, which relies on Farkas's Lemma, is given in the appendix. AC informs us whether π has an AOM representation with $\succ$. Of course, it is possible that AC can be violated for $\succ$ but is satisfied for another preference $\succ'$. Hence, AC allows us to identify all possible preference orderings without constructing the underlying attention rule. This theorem also justifies the use of AC for revealed preference purposes when the choice rule is an AOM.

3 Econometric Methods

We obtained several testable implications and related empirically relevant results for the AOM (Theorems 1 and 4), under essentially only one nonparametric identifying restriction (Assumption 1). We also showed how sharp bounds on the attention frequency can be derived from the choice probabilities (Theorem 2). Our next goal is to develop appropriate 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 $i = 1, 2, \cdots, n$. 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.

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

First and foremost, we do not observe the unit's consideration set, and our econometric methods will not rely on any information on consideration set formation beyond attention overload. Second, we accommodate the possibility that only a subset of choices problems (denoted by $\mathcal{S}$ in the assumption) are available, since the analyst may not observe all choice problems in a specific application/sample. In other words, our analyses and econometric methods do not require "complete" data. Lastly, both our revealed preference and revealed attention results rely on having sufficient variation in choice problems. Fortunately, such variation is widely available: for instance, online retailers constantly change their product catalogs, and grocery stores may rearrange shelves and aisles. Even for a single decision-making unit using the same online search engine, she may face a different set of alternatives depending on the platform (e.g., desktop computers vs. smart phones). In addition, advertising sources such as Google or Amazon routinely conduct marketing experiments, which can also lead to variation in the choice problem.

3.1 Revealed Preference

Recall that $\pi(a|S)$ denotes the probability of choosing a in the choice problem S. We first discuss the testing problem for preference elicitation. Given a preference ordering $\succ$ and the (consistently estimable) choice probability π, we introduce the null hypothesis that corresponds to the testable implications of Theorem 4.

To begin with, we express attention compensation (AC in Definition 3) as $D(a|S,T) = \pi(a|S) - \pi(U_{\succeq}(a)|T) \leq 0$. Theorem 4 then implies that a preference ordering is compatible with our AOM *if and only if* the choice probabilities satisfy AC, that is,

$$\mathsf{H}_0 : \max_{\substack{a \in T \subset S \\ T,S \in \mathcal{S}}} D(a|S,T) \leq 0 \qquad \text{vs.} \qquad \mathsf{H}_1 : \max_{\substack{a \in T \subset S \\ T,S \in \mathcal{S}}} D(a|S,T) > 0. \tag{2}$$

The max operation in (2) enumerates all pairs of nested choice problems $T \subset S$ in $\mathcal{S}$, and all alternatives $a \in T$. Rejecting the null hypothesis in (2) is equivalent to rejecting the AOM for a preference ordering. Of course, if one is able to reject the null hypothesis for all possible preference orderings, then it means the choice probabilities are not compatible with our AOM, which can be taken as a model specification test.

One distinguishing feature of our estimation and inference procedures is that we allow the complexity of the problem to be "large," that is, we develop valid distributional approximations and related econometric methods where the complexity can be exponentially large relative to the sample size. In particular, we can consider a large-sample framework where both the number of alternatives in the grand set and the number of choice problems (i.e., $|X|$ and $|\mathcal{S}|$) can grow with the sample size. To illustrate why it is crucial to allow for this "increasing dimension" in our analysis, suppose that there are ten alternatives in the grand set, and that we can observe choice problems of sizes 7 and 8 in the sample (so that there are in total $\binom{10}{8}+\binom{10}{7}=165$ choice problems). Then, the null hypothesis in (2) will involve comparing $\binom{10}{8}\binom{8}{7}6 = 2,160$ pairs of choice probabilities, which is not "small" in a sample of, say, $5,000$ observed choices. Large-sample analysis assuming fixed dimensions may lead to overly optimistic conclusions, and the resulting inference procedure may perform poorly in finite samples. By explicitly taking into account the complexity of the inference problem, our large-sample analysis will shed light on how the error in the distributional approximation relates to not only the sample size but also other features of the data generating process (see Theorem 5 below).

Constructing a test statistic for the competing hypotheses in (2) is straightforward, as we can replace the unknown choice probabilities by some estimates constructed from the data. We estimate $\pi(a|S)$ by

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

where N_S is the effective sample size for the choice problem S. Similarly,

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

leading to $\widehat{D}(a|S,T) = \widehat{\pi}(a|S) - \widehat{\pi}(U_{\succeq}(a)|T)$. To improve power, we consider the Studentized test statistic using the estimated differences $\widehat{D}(a|S,T)$:

$$\mathcal{T}(\succ) = \max\left\{\max_{\substack{a\in T\subset S\\ T,S\in\mathcal{S}}}\frac{\widehat{D}(a|S,T)}{\widehat{\sigma}(a|S,T)}\ ,\ 0\right\},$$
$$\widehat{\sigma}(a|S,T) = \sqrt{\frac{1}{N_S}\widehat{\pi}(a|S)\big(1-\widehat{\pi}(a|S)\big) + \frac{1}{N_T}\widehat{\pi}(U_{\succeq}(a)|T)\big(1-\widehat{\pi}(U_{\succeq}(a)|T)\big)}.$$

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

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 (2021) for many moment inequality testing. (See also Molinari (2020) for an overview and further references.) More specifically, we seek for a critical value, denoted by $\mathrm{cv}(\alpha, \succ)$, such that under the null hypothesis (i.e., when the preference is compatible with our AOM),

$$\mathbb{P}\left[\mathcal{T}(\succ) > \mathrm{cv}(\alpha, \succ)\right] \leq \alpha + \mathfrak{r}_{\succ}, \tag{3}$$

where $\alpha \in (0,1)$ denotes the desired significance level of the test (usually 5% or 10%), and $\mathfrak{r}_{\succ}$ denotes a quantifiable error of approximation (which should vanish as the same size increases, ideally allowing for "large" complexity). The critical value is specific to the preference ordering under consideration by the null hypothesis.

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

$$\frac{\widehat{D}(a|S,T)}{\widehat{\sigma}(a|S,T)} \overset{\mathrm{a}}{\sim} \mathcal{N}\left(\frac{D(a|S,T)}{\sigma(a|S,T)},\ 1\right).$$

Since $D(a|S,T) \leq 0$ under the null hypothesis (i.e., whenever the preference ordering is compatible with our AOM), the above Gaussian distribution will be first-order stochastically dominated by the standard Gaussian 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 $\mathcal{T}(\succ)$ 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 $\{\widehat{\pi}(a|S) : a \in S \in \mathcal{S}\}$ and the effective sample sizes $\{N_S : S \in \mathcal{S}\}$, which we denote by $\widehat{\mathbf{\Omega}}$. The critical value is then the $(1-\alpha)$-quantile of the maximum of a Gaussian vector, and is precisely defined as

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

where the inner max operation computes the maximum over the elements of $\widehat{\mathbf{\Omega}}^{1/2}\mathbf{z}$, and $\mathbf{z}$ denotes a standard Gaussian random vector of dimension equal to the number of comparisons (i.e., inequalities) in (2). The following theorem presents our first econometric result.

Theorem 5 (Preference Elicitation)**.** Assume Assumption 2 holds. Let $\mathfrak{c}_1$ be the number of comparisons (i.e., inequalities) in (2), and

$$\mathfrak{c}_2 = \left(\min_{S \in \mathcal{S}} N_S\right) \cdot \left(\min_{\substack{a \in T \subset S \\ T,S \in \mathcal{S}}} \sigma(a|S,T)\right).$$

Then, under the null hypothesis H_0, (3) holds with

$$\mathfrak{r}_{\succ} = C \cdot \left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{1/4},$$

where C denotes an absolute constant.

This theorem shows that the error in distributional approximation, $\mathfrak{r}_{\succ}$, only depends on the dimension of the problem (i.e., $\mathfrak{c}_1$) logarithmically, and therefore our estimation and inference procedures remain valid even if the test statistic $\mathcal{T}(\succ)$ involves comparing "many" pairs of estimated choice probabilities. By providing a statistical guarantee as in (3), 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.

Theorem 5 also demonstrates that the distributional approximation error is inversely related to $\mathfrak{c}_2$, which can be regarded as the effective sample size. To gain some intuition, consider the special case of balanced sample sizes: If $\min_{S\in\mathcal{S}} N_S \asymp \max_{S\in\mathcal{S}} N_S \asymp \mathfrak{n}$, meaning that the sample size for each choice problem is of the same order, then

$$\mathfrak{c}_2 \propto \sqrt{\mathfrak{n}\left[\pi(a|S)\big(1-\pi(a|S)\big)+\pi(U_{\succeq}(a)|T)\big(1-\pi(U_{\succeq}(a)|T)\big)\right]}.$$

This quantity involves the choice probabilities $\pi(a|S)$ and $\pi(U_{\succeq}(a)|T)$. These two choice probabilities arise because they are related to the variance of the events $\mathbb{1}(y_i = a, Y_i = S)$ and $\mathbb{1}(y_i \succeq a, Y_i = T)$, which we use to form an estimate of the inequality constraint $D(a|S,T)$. Naturally, $\mathfrak{c}_2$ needs to be large (i.e., diverge as the sample size increases) in order for the distributional approximation error in (3) to become sufficient accurate (i.e., $\mathfrak{r}_{\succ} \to 0$). Such a condition is usually required for a valid normal approximation to an average of Bernoulli trials. Having said this, our result allows one of the two choice probabilities to be close to 1 or 0, and hence our distributional approximation result can accommodate situations where certain alternatives in a choice problem are always or never chosen.

One of our main econometric contributions is a careful study of the properties of the estimated correlation matrix, $\widehat{\boldsymbol{\Omega}}$, where we provide an explicit bound on the supremum of the entry-wise estimation error $\|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}\|_\infty$. This is further combined with the results in Chernozhukov, Chetverikov, Kato, and Koike (2021) to establish a normal approximation for the centered and normalized inequality constraints, $(\hat{D}(a|S,T)-D(a|S,T))/\hat{\sigma}(a|S,T)$. Similar normal approximations are also employed in Section 3.2 for revealed attention econometric analysis. Chernozhukov, Chetverikov, and Kato (2019) also study the problem of testing moment inequalities but their results do not seem to apply directly to our setting, as the moment inequalities we consider involve probabilities estimated from different choice problems, and hence they cannot be expressed as averages of i.i.d. observations. 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 (2). 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. As a consequence, relative to prior work in the literature, our proposed new econometric results formalized in Theorem 5 offer major improvements for preference elicitation, both theoretically and methodologically.

Before closing this subsection, we note that critical values used in Theorem 5 can sometimes be conservative, as this choice corresponds to the least favorable scenario where all of the inequalities are binding. Various approaches have been proposed in the literature on moment inequality testing to address this concern, which is related to power. In this paper we also study a two-step method based on pre-selecting the inequality constraints, originally proposed by Chernozhukov, Chetverikov, and Kato (2019). In the first step, we obtain the least favorable critical value $\mathrm{cv}(\mathfrak{c}_3, \succ)$ in Theorem 5 by choosing an $\mathfrak{c}_3$ usually much smaller than the final desired significance level α. The preliminary significance level $\mathfrak{c}_3$ will reflect the uncertainty in the (moment inequality binding) preliminary selection step. In the second and final step, an adjusted critical value is constructed as follows:

$$\mathrm{cv}_{\mathtt{adj}}(\alpha, \succ) = \inf\Big\{t : \mathbb{P}\big[\mathcal{T}^{\mathtt{G}}(\succ) \le t|\mathrm{Data}\big] \ge 1-\alpha\Big\}, \quad \mathcal{T}^{\mathtt{G}}(\succ) = \max\Big\{\max\big(\widehat{\boldsymbol{\Omega}}^{1/2}\mathbf{z} + \widehat{\boldsymbol{\psi}}\big)\ ,\ 0\Big\},$$

where the data-dependent moment selection adjustment, $\widehat{\boldsymbol{\psi}}$, has typical element

$$\hat{\psi}(a|S,T) = \begin{cases} 0 & \text{if } \hat{D}(a|S,T)/\hat{\sigma}(a|S,T) \ge -2\mathrm{cv}(\mathfrak{c}_3, \succ) \\ -\infty & \text{otherwise} \end{cases}.$$

In other words, we drop a specific inequality constraint if the ratio $\hat{D}(a|S,T)/\hat{\sigma}(a|S,T)$ is negative and large in magnitude. We show in the online supplemental appendix that, with the two-step moment selection method, (3) holds with

$$\mathfrak{r}_{\succ} = 2\mathfrak{c}_3 + C \cdot \left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{1/4}.$$

Therefore, for proper size control, we can either use some sequence $\mathfrak{c}_3 \to 0$, or we can compensate by employing the critical value $\mathrm{cv}_{\mathtt{adj}}(\alpha - 2\mathfrak{c}_3, \succ)$ for the two-step moment selection hypothesis test.

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

$$\mathcal{C}(1-\alpha) = \Big\{\succ :\ \mathcal{T}(\succ) \le \mathrm{cv}(\alpha, \succ)\Big\},$$

and therefore, for any preference $\succ$ that is compatible with our AOM, we have:

$$\mathbb{P}\left[\succ \in \mathcal{C}(1-\alpha)\right] \ge 1 - \alpha - \mathfrak{r}_{\succ}.$$

We illustrate these methods with simulated data in Section 6.

3.2 Revealed Attention

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. In fact, even if only one superset of S is observed in the data—which makes the maximum operation unnecessary—the resulting plug-in lower bound will not deliver a satisfactory empirical coverage rate.

Our aim is to provide a construction of the lower bound, denoted as $\hat{\phi}_{\mathtt{L}}(a|S)$, where the subscript $\mathtt{L}$ stands for "lower bound", such that

$$\mathbb{P}\left[\phi(a|S) \geq \hat{\phi}_{\mathtt{L}}(a|S)\right] \geq 1 - \alpha + \mathfrak{r}_{\phi_{\mathtt{L}}(a|S)} \tag{4}$$

with $\alpha \in (0, 1)$ denoting the desired significance level, and $\mathfrak{r}_{\phi_{\mathtt{L}}(a|S)}$ 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

$$\hat{\phi}_{\mathtt{L}}(a|S) = \max_{R\supseteq S, R\in\mathcal{S}} \left\{\hat{\pi}(a|R) - \mathrm{cv}(\alpha, \phi_{\mathtt{L}}(a|S)) \cdot \hat{\sigma}(a|R)\right\}.$$

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

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

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

To gain some intuition, we begin with the normal approximation: $\hat{\pi}(a|R) \overset{\mathrm{a}}{\sim} \mathcal{N}\left(\pi(a|R),\ \sigma(a|R)\right)$. In addition, the estimated choice probabilities are mutually independent since they are constructed from different subsamples. Then,

$$\begin{aligned}
&\mathbb{P}\Big[\hat{\pi}(a|R) - \mathrm{cv}(\alpha, \phi_{\mathtt{L}}(a|S)) \cdot \hat{\sigma}(a|R) \leq \pi(a|R),\ \forall R \supseteq S,\ R \in \mathcal{S}\Big] \\
&= \mathbb{P}\Big[\max_{R\supseteq S, R\in\mathcal{S}} \frac{\hat{\pi}(a|R) - \pi(a|R)}{\hat{\sigma}(a|R)} \leq \mathrm{cv}(\alpha, \phi_{\mathtt{L}}(a|S))\Big] \\
&\approx \mathbb{P}\Big[\max(\mathbf{z}) \leq \mathrm{cv}(\alpha, \phi_{\mathtt{L}}(a|S))\Big] = 1 - \alpha.
\end{aligned}$$

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 $\hat{\pi}(a|R) - \mathrm{cv}(\alpha, \phi_{\mathtt{L}}(a|S)) \cdot \hat{\sigma}(a|R)$. 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 6 (Attention Frequency Elicitation)**.** Assume Assumptions 1 and 2 hold. Let $\mathfrak{c}_1 = |\{R \in \mathcal{S} : R \supseteq S\}|$ be the number of supersets of S, and

$$\mathfrak{c}_2 = \left(\min_{R\supseteq S, R\in\mathcal{S}} N_R\right) \cdot \left(\min_{R\supseteq S, R\in\mathcal{S}} \sigma(a|R)\right).$$

Then (4) holds with

$$\mathfrak{r}_{\phi_{\mathtt{L}}(a|S)} = C \cdot \left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{1/4},$$

where C denotes an absolute constant.

In applications, the researcher might be interested in bounding attention frequencies across multiple choice problems. For example, providing lowers bounds for an alternative at different choice problems can help reveal the effectiveness of an advertisement strategy. We illustrate how the proposed econometric techniques can be easily adapted, focusing only on the lower bound for brevity. Assume the goal is to construct a pair, $\hat{\phi}_{\mathtt{L}}(a|S)$ and $\hat{\phi}_{\mathtt{L}}(a'|S')$, such that the joint coverage is well-controlled asymptotically: $\mathbb{P}[\phi(a|S) \geq \hat{\phi}_{\mathtt{L}}(a|S),\ \phi(a'|S') \geq \hat{\phi}_{\mathtt{L}}(a'|S')] \geq 1-\alpha$ (up to an error term). Given our earlier discussions, the two lower bounds are estimated by the maxima of two adjusted processes:

$$\begin{aligned}
\hat{\phi}_{\mathtt{L}}(a|S) &= \max_{R\supseteq S, R\in\mathcal{S}} \left\{\hat{\pi}(a|R) \quad - \mathrm{cv}(\alpha, \phi_{\mathtt{L}}(a|S; a'|S')) \cdot \hat{\sigma}(a|R)\right\} \\
\hat{\phi}_{\mathtt{L}}(a'|S') &= \max_{R\supseteq S', R\in\mathcal{S}} \left\{\hat{\pi}(a'|R) - \mathrm{cv}(\alpha, \phi_{\mathtt{L}}(a|S; a'|S')) \cdot \hat{\sigma}(a'|R)\right\}.
\end{aligned}$$

As it should be clear from the notation, the only change happens in the critical value: while we employ the maximum of a standard normal random vector before, this is no longer appropriate because overlapping choice problems may induce correlation in the estimated choice probabilities.

To construct valid critical values, let $\hat{\mathbf{\Omega}}$ be the estimated correlation matrix of $\{\hat{\pi}(a|R), \hat{\pi}(a'|R') : R \supseteq S, R' \supseteq S'\}$, then the critical value can be chosen as the $1-\alpha$ quantile:

$$\mathrm{cv}(\alpha, \phi_{\mathtt{L}}(a|S; a'|S')) = \inf\left\{t:\ \mathbb{P}\left[\max(\hat{\mathbf{\Omega}}^{1/2}\mathbf{z}) \leq t|\mathrm{Data}\right] \geq 1-\alpha\right\},$$

where $\mathbf{z}$ is a standard normal random vector of conformable dimension.

Section 6 illustrate how the above methods perform empirically using simulated data.

4 Generalizations and Extensions

Our main characterization results (Theorems 2 and 4) and econometric methods (Theorems 5 and 6) so far build on the AOM only (i.e., Assumption 1). We now discuss generalizations and extensions to address empirical behavioral phenomena. Firstly, we consider an extension of our baseline AOM which can accommodate a default option. We show that our AOM is fully compatible with the observation that the decision maker tend to choose the default option more often when there are more alternatives available in the choice problem. Next, we illustrate how additional restrictions on the attention rule can be incorporated so that the augmented model may deliver more empirical content. In particular, we consider preference and attention revelation by binary comparisons. Another possibility is to combine our AOM with the random attention model proposed by Cattaneo, Ma, Masatlioglu, and Suleymanov (2020), which we discuss in more detail in Section 5.

4.1 Default Option and Choice Overload

The literature relies on the default option to identify choice overload. Such data is often assumed to be available (see, for example, Manzini and Mariotti 2014; Brady and Rehbeck 2016; Echenique, Saito, and Tserenjigmid 2018). Therefore, to explain choice overload, we extend the AOM to accommodate a default option. A default option is an alternative, o^*, which always receives attention. Alternatively, it can be interpreted as "choosing nothing." Let $X^* = X \cup \{o^*\}$ and $S^* = S \cup \{o^*\}$ for all $S \in \mathcal{X}$. We require that the choice rule satisfies $\sum_{a \in S^*} \pi(a|S^*) = 1$ and $\pi(a|S^*) \geq 0$ for all $a \in S^*$. Following the literature, we assume the default option is inferior to all other alternatives and is chosen only when the DM considers no other alternatives, which implies that $\pi(o^*|S^*) = \mu(\{o^*\}|S^*)$.

The phenomenon "choice overload," termed by Iyengar and Lepper (2000), states that the default option is chosen more often as the choice problem gets bigger. Formally, we say π satisfies choice overload if $\pi(o^*|S^*) \geq \pi(o^*|T^*)$ whenever $T \subset S$. Interestingly, many existing models of random attention rule predict the opposite way (Manzini and Mariotti, 2014; Brady and Rehbeck, 2016; Cattaneo, Ma, Masatlioglu, and Suleymanov, 2020). In stark contrast, our AOM can accommodate choice overload: while the competition between alternatives gets more fierce as the decision problems get bigger, the decision maker tends to not consider any alternative and choose the default option.

Although it is relatively straightforward to see that our earlier characterization result (Theorem 4) continues to hold, econometric implementation may not be straightforward. To be more specific, some datasets may not register the default option, which effectively leads to a missing data problem. For example, an online advertising platform may only record consumers who click one of the links it provides. In this case, what is available (consistently estimable) from the data is the normalized choice rule without any information on the default option:

$$\left\{\pi^*(a|S) = \frac{\pi(a|S^*)}{\sum_{b\in S}\pi(b|S^*)} : \; a \in S, S \in \mathcal{S}\right\}.$$

To understand the relationship between π^* and the (unobservable) choice rule π, we provide the following example.

Example 2. Consider the following choice data and the normalized version.

$\pi(\cdot\|S^*)$	a	b	c	d	o^*
$\{a,b,c,d,o^*\}$	0.1	0.1	0.1	0.1	0.6
$\{b,c,d,o^*\}$		0.2	0.2	0.6	0

$\pi^*(\cdot\|S)$	a	b	c	d
$\{a,b,c,d\}$	0.25	0.25	0.25	0.25
$\{b,c,d\}$		0.2	0.2	0.6

Then it should be clear that the choice rule with the default option, π, satisfies choice overload and is compatible with the preference ordering $a \succ b \succ c \succ d \succ o^*$. However, if one only has access to the normalized choice rule, π^*, she may (falsely) rule out this preference ordering as $\pi^*(b|\{a,b,c,d\}) > \pi^*(b|\{b,c,d\})$, which implies b cannot be the most preferred alternative in the set $\{b,c,d\}$. ┘

The above example demonstrates that if the data on the default option is not available, a naïve implementation of our characterization result (Theorem 4) to the normalized choice rule can lead to false rejection in terms of revealed preference. It is possible such revealed preference inference happens across different choice set such that it leads to false rejection of the model. However, from another perspective, if the data without the default option satisfies AC, then we can safely claim that the full data *will* satisfy AC, under the condition that it satisfies choice overload. The reason is because normalization under the condition that the default option satisfies choice overload makes it more difficult for the data to satisfy AC. We put this observation in the following proposition.

Proposition 1 (Default Option and Normalization)**.** Assume $(\pi^*, \succ)$ satisfies AC. Then, for every choice rule π which satisfies choice overload and its normalization is equal to π^*, $(\pi, \succ)$ also satisfies AC.

4.2 Attentive at Binaries

When an alternative is chosen frequently enough in a binary choice problem, a policymaker may want to conclude that the alternative is better than the other. It is up to the policymaker to decide what frequency is sufficient. We first denote this threshold frequency as η, and will say that $(\pi, \succ)$ satisfies η-constrained revealed preference if $a \succ b$ whenever $\pi(a|\{a,b\}) > \eta$.

As discussed before, by considering $(\pi, \succ)$, we are matching a preference ordering to a choice data. The definition fulfills exactly our needs in revealing the preference if information on binary comparisons is available. The parameter η measures how cautious the policy maker is when making a welfare judgement: if $\eta = 1$ the policymaker would not draw any conclusion from binary comparisons only, while η being too low may lead to a cyclic preference ordering which does not help with policy-making. Hence, the question is how one can safely incorporate the new η-constrained revealed preference into Definition 4. We propose to restrict η as[4]

$$\eta \geq \max \Big\{ \mu(\{a\}|\{a,b\}), \mu(\{b\}|\{a,b\}) \Big\} \quad \text{for all } a, b \in X. \tag{5}$$

The above restriction is simple and intuitive. Consider $\eta = 0.6$. Whenever we observe $\pi(a|\{a,b\}) > 0.6$, it is immediate that $\mu(\{a,b\}|\{a,b\}) > 0$ and that the consumer chooses a over b when she pays attention to both alternatives, because the choice probability of a cannot be entirely attributed to the attention on the singleton set, $\mu(\{a\}|\{a,b\})$.

We can also gain intuition via attention frequency. Consider the same setting above. Observing $\pi(a|\{a,b\}) > 0.6$ implies that $\pi(b|\{a,b\}) < 0.4$. On the other hand, we must have $\phi(b|\{a,b\}) = 1 - \mu(\{a\}|\{a,b\}) \geq 1 - 0.6 = 0.4$. In other words, there is a strictly positive probability that the decision maker pays attention to b but ends up not choosing it, and hence a must be the preferred alternative.

Based on the previous discussion, we immediately have the following result: Assume $(\pi, \succ, \mu)$ is an AOM and that μ also satisfies (5), then AC holds, and $\succ$ satisfies η-constrained revealed preference. The reverse is also true. That is, given a choice data, we know "when" the data can be represented by the AOM with the new assumption of η-attentive at binaries. this is summarized in the following theorem.

Proposition 2 (Characterization with Attentive at Binaries)**.** $(\pi, \mu, \succ)$ is an AOM with μ satisfying (5) if and only if $(\pi, \succ)$ satisfies AC and η-constrained revealed preference.

By providing a joint characterization result, the above theorem also shows that it is very easy to incorporate the new η-attentive at binaries assumption into our econometric implementation. To be precise, the test statistic and the critical value we introduced in Section 3 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 η-constrained revealed preference.

The new restriction in (5) also brings additional empirical content to our revealed attention analysis, as it bounds the attention frequency from below in binary comparisons: $\phi(a|\{a,b\}) \geq 1-\eta$. As a result, the lower bound in Theorem 2 will take the form $\phi(a|\{a,b\}) \geq \max\{\max_{R \supseteq \{a,b\}} \pi(a|R), 1-\eta\}$.

[4]This is strictly weaker than a similar assumption proposed by Cattaneo, Ma, Masatlioglu, and Suleymanov (2020): any attention rule that satisfies their Assumption 3 would automatically satisfy our restriction.

5 Related Literature

The AOM is a crucial missing piece of the puzzle in the limited consideration models literature. We first briefly survey the experimental literature which leads to our key identifying assumption (Assumption 1). Then, we discuss the connections and differences between our proposed AOM and several other models featuring (deterministic or random) limited consideration, both under parametric and nonparametric identifying restrictions.

Experimental Evidence. There are a number of experiments studying the effect on attention when the number of products increases. To begin with, using eye-tracking technique, Visschers, Hess, and Siegrist (2010) found that consumers tend to ignore product information, e.g. nutrition labels, when there are more products available. Moreover, Reutskaja, Nagel, Camerer, and Rangel (2011) showed that subjects considered proportionally less items when there are more items. In a field experiment, Iyengar and Lepper (2000) found that subjects tend to choose not to make any purchase when choices are overloading. The evidence altogether suggests that each alternative should get less attention when there are more of them available, which is precisely encoded in our novel attention overload Assumption 1. See also Section 4.1 for an extension to our framework that accommodates default options.

Although extensive experimental evidence has been documented, existing models in the literature have not been able to capture the idea of attention overload. Some models (e.g. Manzini and Mariotti 2014 and Aguiar 2017) implicitly assume constant attention frequency. While being a knife-edge special case of our AOM, these models would not be able to provide insight on consumer behavior when attention frequency is strictly monotonic. On the other hand, some models (e.g. Brady and Rehbeck 2016 and Cattaneo, Ma, Masatlioglu, and Suleymanov 2020) allow for non-monotonic attention frequency, which also will not help understand the behavioral implication of attention overload. In the following, we discuss some of these related models in more detail.

Random Attention Model. Cattaneo, Ma, Masatlioglu, and Suleymanov (2020) proposed a random attention model (RAM), which generalizes the (deterministic) attention filter in Masatlioglu, Nakajima, and Ozbay (2012). Our proposed AOM can be understood as a generalization of the competition filter in Lleras, Masatlioglu, Nakajima, and Ozbay (2017). Both competition filter and attention filter give revealed preference by considering removing an element from the choice set, but the intuition behind each case is different: when we observe choice reversal of an alternative, the attention filter says that the item is better than another item in the bigger choice set, while the competition filter says that the item is worse than another item in the smaller choice set.

These two ideas naturally extend to probabilistic settings: when one observes a violation of regularity of an alternative, RAM says the alternative is better than another alternative in the bigger choice set, and AOM suggests the alternative is worse than another alternative in the smaller choice set. Therefore AOM and RAM are not nested. In terms of attention rules, we provide two examples with three alternatives in the following to point out their differences: the first example,

$\mu_{\texttt{RAM}}$, satisfies monotonic attention in Cattaneo, Ma, Masatlioglu, and Suleymanov (2020) but not attention overload. The second example, $\mu_{\texttt{AOM}}$, satisfies attention overload but not monotonic attention. This attention rule highlights the idea of "less is more," meaning that the decision maker has trouble dealing with larger choice sets and only considers singleton consideration sets.

$\mu_{\texttt{RAM}}(T\|S)$	$T = \{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$	$\{a\}$	$\{b\}$	$\{c\}$
$S = \{a, b, c\}$	1	0	0	0	0	0	0
$\{a, b\}$		0			1/2	1/2	
$\{a, c\}$			0		1/2		1/2
$\{b, c\}$				0		1/2	1/2
$\mu_{\texttt{AOM}}(T\|S)$	$T = \{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$	$\{a\}$	$\{b\}$	$\{c\}$
$S = \{a, b, c\}$	0	0	0	0	1/3	1/3	1/3
$\{a, b\}$		1			0	0	
$\{a, c\}$			1		0		0
$\{b, c\}$				1		0	0

On the other hand, one might wonder whether these two models are behaviorally different (in terms of the choice rule). Below we illustrates that there exists choice data which has an AOM representation but not a RAM representation, and vise versa. In the first (left) example, the RAM predicts that that $b \succ c$ and $c \succ b$, and hence, it could not be explained by RAM. Our AOM suggests $a \succ c$ and $a \succ b$. Interestingly, there does not exists an example of three alternatives which RAM can explain but AOM cannot. In other words, AOM explains more choice data than RAM in 3-alternative cases. With more than three alternatives, however, it is possible that some choice data is compatible with RAM but not AOM, as the second (right) example illustrates. In the AOM, revealed preference says that $a \succ b$ and $b \succ a$. Hence, it could not be explained by AOM. In the RAM, however, revealed preference says $a \succ d$ and $b \succ c$.

$\pi(\cdot\|S)$	a	b	c
$\{a, b, c\}$	0.4	0.3	0.3
$\{a, b\}$	0.8	0.2	
$\{a, c\}$	0.8		0.2
$\{b, c\}$		0.5	0.5

$\pi(\cdot\|S)$	a	b	c	d
$\{a, b, c, d\}$	1/2	1/2	0	0
$\{a, b, c\}$	0	2/3	1/3	
$\{a, b\}$	1/2	1/2		

More generally, while both our AOM and the RAM provide testable restrictions (inequalities) in terms of the choice rule, they can take very different forms. To compare, first recall from our main characterization result (Theorem 4) that a preference ordering has an AOM representation if and only if the AC condition holds: $\pi(a|S) \leq \pi(U_{\succeq}(a)|T)$ for $T \subset S$. The RAM, on the other hand, restricts the choice rule for lower contour sets. That is, take $a \in T \subset S$ such that the lower contour sets agree: $\{b \preceq a : b \in S\} = \{b \preceq a : b \in T\}$, then RAM requires $\pi(a|S) \leq \pi(a|T)$. In Section 6 we provide an example in which the two models complement each other: our AOM turns out to be very powerful at testing mistakes on the best alternative, while the RAM is best at identifying the less preferred alternatives. As a result, they collectively deliver powerful revealed preference results.

Independent Consideration. We mentioned Manzini and Mariotti (2014) as a special case of our proposed AOM. They assume that each alternative a has a fixed probability, $\gamma(a)$, to be considered. In other words, the attention frequency, $\phi_{\texttt{MM}}(a|S)$, is held fixed across different choice problems in their model, i.e., $\phi_{\texttt{MM}}(a|S) = \gamma(a)$. Since attention overload requires only weak inequality, the model falls into AOM.

Attention by Category. Aguiar (2017) assumes that each category D has a fixed probability $m(D)$. If the category is *available*, the decision maker picks the best alternative out of it. If not, she chooses the default option. Let the set of all categories be $\mathcal{D}$. Then the attention frequency is given by

$$\phi_{\texttt{Aguiar}}(a|S) = \sum_{a \in D \in \mathcal{D}} m(D).$$

It is straightforward to see that this attention rule satisfies attention overload.

Random Competition Filter. The random competition filter (RCF) is a major special case of our model. Let $\Gamma_j(\cdot)$ be deterministic consideration set mappings which satisfy competition filter, and $\sum_{j=1}^{J} \alpha_j = 1$. A random competition filter model is specified by the following attention rule with the respective attention frequency,

$$\mu_{\texttt{RCF}}(T|S) = \sum_{j=1}^{J} \alpha_j \mathbb{1}(\Gamma_j(S) = T), \quad \phi_{\texttt{RCF}}(a|S) = \sum_{j: a \in \Gamma_j(S)} \alpha_j.$$

Random competition filters satisfy attention overload because $a \in \Gamma_j(T)$ for all $T \subseteq S$ if $a \in \Gamma_j(S)$.

The random competition filter model nests two other models, bounded rationalization and imprecise narrowing down. Bounded rationalization is a generalization of Cherepanov, Feddersen, and Sandroni (2013). It states that the decision maker does not always stick to the same set of rationale given the same choice set. Hence, it is as if the decision maker assigns a probability distribution over the power set on the set of rationale. Since Cherepanov, Feddersen, and Sandroni (2013) is a special case of Lleras, Masatlioglu, Nakajima, and Ozbay (2017), it follows that the bounded rationalization model is a special case of random competition filter. Imprecise narrowing down shares a similar idea. Given the same choice set, the decision maker does not necessarily follow the same procedure on setting up criteria. Thus, it is as if the decision maker assigns a probability distribution over the set of all possible procedures. It makes imprecise narrow down again a special case of random competition filter.

Other Stochastic Choice Models. There exist several stochastic choice models that do not involve attention, a number of them respecting regularity. The seminal work of the random utility model (RUM) is of course a prime example. By previous discussion, the condition AC is automatically satisfied when a model satisfies regularity. Hence, a standard random utility model can be represented by an AOM. Note that there are a number of models nested in RUM. For example, Gul, Natenzon, and Pesendorfer (2014) consider an attribute rule in which the decision maker first draws an attribute and then picks an alternative which contains such attribute. They show that every attribute rule is a RUM, and hence every attribute rule can be represented by AOM. Fudenberg, Iijima, and Strzalecki (2015) introduce the additive perturbed utility model where the decision maker intentionally randomizes as deterministic choices can be costly. Since the choices in their model always satisfy regularity, any choice rule in the additive perturbed utility model has an AOM representation.

There are several stochastic choice models which allow for violation of regularity. Intriguingly, we can show that some of them are AOM by directly checking the condition AC. Important examples include Echenique, Saito, and Tserenjigmid (2018) and Echenique and Saito (2019).[5] Lastly, Filiz-Ozbay and Masatlioglu (2020) introduces the Less-is-more Progressive Random Choice model relying on the less-is-more choice function from Lleras, Masatlioglu, Nakajima, and Ozbay (2017). Since each less-is-more choice function can be mapped back to a competition filter, the model is essentially a special case of random competition filter. Therefore, it is included in AOM.

Discrete Choice and Attention. The literature on discrete choice and attention is vast and spans many disciplines. We discuss a few recent contributions to this literature and how they differ from our AOM. Barseghyan, Coughlin, Molinari, and Teitelbaum (2021) also study the partial identification of preference and consideration set formation. Unlike our AOM, which imposes monotonicity of the attention frequency across nested choice problems, they restrict the size of consideration sets (or alternatively, assume that the decision maker cannot pay attention to singleton sets too often). This nonparametric identifying assumption is similar in spirit to our attentive at binary assumption (Section 4.2), and hence can bring additional empirical content when combined with our AOM. Abaluck and Adams (2021) exploit asymmetries in cross-partial derivatives and show that consideration set formation and preference distribution can be separately identified from observed choices when there is rich exogenous variation in observed covariates. Barseghyan, Molinari, and Thirkettle (2021), on the other hand, provide identification results for risk preference when exogenous variation in observed covariates is more restricted. They also demonstrate the tradeoff between the exclusion restrictions and the assumptions on choice set formation. Finally, Dardanoni, Manzini, Mariotti, and Tyson (2020) consider a model where individuals might differ both in terms of their consideration capacities and preferences, and assume that only aggregate choice shares are observed.

[5] Available upon request.

6 Simulation Evidence

This section presents simulation evidence on the finite sample properties of our econometric methods. We first introduce the simulation setting. The grand set contains six alternatives, and without loss of generality, we assume the preference ordering $a_1 \succ a_2 \succ \cdots \succ a_6$. We employ the logit attention rule of Brady and Rehbeck (2016), which takes the form

$$\mu(T|S) = \frac{|T|^{\varsigma}}{\sum_{T' \subseteq S} |T'|^{\varsigma}},$$

where recall that $|T|$ denotes the size (cardinality) of a set. For specificity, we set $\varsigma = 2$ in our simulation study. Explicit calculation shows that this logit attention rule satisfies attention overload (Assumption 1).[6] With the preference and the attention rule introduced above, we are able to find the choice rule according to Definition 1. We then generate the choice data. In our simulation studies, we assume each choice problem has the same effective sample size $N_S = 200$.

We implement our test (Section 3.1) against four hypothesized preferences. The first preference, $a_1 \succ a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6$, is compatible with our AOM, and hence we do not expect rejecting the null hypothesis very often. In other words, the rejection probability in this case corresponds to the size of our test. The other three preferences, however, cannot be represented by the AOM, and therefore the rejection probabilities will shed light on the power of the test.

In each simulation setting, we conduct 2,000 Monte Carlo repetitions and the average rejection probabilities are reported in Table 1 (row "**rej prob**"). The nominal size is set to be 0.05. To employ the two-step moment selection method, we set $\mathfrak{c}_3$ to 0.005. We first focus on the column, "**AOM**," where we only employ the inequality constrains implied by our AOM (i.e., the characterization results in Theorem 4 using AC). There are 664 inequality constraints in total (row "**# restrictions**"). As the first preference ordering satisfies our AOM, none of these constraints will be violated. On the other hand, for the other three preferences, 90, 6 and 23 out of the 664 inequalities are strictly positive (row "**# violations**"). We also show the largest inequality constraint (row "**max inequality**"). A larger number indicates that the preference is further away from the null space, and hence it should be easier for our test to detect. To be more precise, for the specific logit attention rule that we consider, the AOM is best at eliciting the "best" alternative, that is, it is most powerful against mistakes regarding the most preferred alternative, a_1. As we can see from the table (row "**rej prob**"), among the three preferences in the alternative space, the AOM has the highest power for testing the second preference, $a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6 \succ a_1$. This is because by placing a_1 in the last position, lots of the inequalities will be violated according to our AOM. As a comparison, although the third and fourth preferences are also rejected by our AOM, they are relatively close to the null space due to the small fraction of violated inequality constraints, and hence the power of our test is not very high.

[6] With the logit attention rule, the attention frequency only depends on the size of the choice problem but not the alternatives. For $|S| = 2, 3, 4, 5, 6$, the attention frequencies are 0.833, 0.750, 0.700, 0.667 and 0.643, respectively.

Table 1: Empirical Rejection Probabilities.

	AOM	RAM	AOM +RAM	+ Attentive at Binaries ($\eta =$) 0.9	0.8	0.7	0.6
# restrictions	664	416	1080	1095	1095	1095	1095
$a_1 \succ a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6$							
# violations	0	0	0	0	0	0	0
max inequality	−0.024	−0.003	−0.003	−0.003	−0.003	−0.003	−0.003
rej prob	0.001	0.000	0.000	0.000	0.000	0.000	0.000
rej prob (size adj)	0.052	0.048	0.046	0.053	0.046	0.046	0.049
$a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6 \succ a_1$							
# violations	90	0	90	90	95	95	95
max inequality	0.071	−0.003	0.071	0.071	0.071	0.190	0.389
rej prob	0.250	0.000	0.173	0.172	0.216	1.000	1.000
rej prob (size adj)	0.478	0.047	0.472	0.471	0.529	1.000	1.000
$a_1 \succ a_2 \succ a_6 \succ a_5 \succ a_4 \succ a_3$							
# violations	6	107	113	113	119	119	119
max inequality	0.058	0.067	0.067	0.067	0.067	0.190	0.389
rej prob	0.045	0.315	0.210	0.210	0.306	1.000	1.000
rej prob (size adj)	0.350	0.577	0.570	0.569	0.643	1.000	1.000
$a_1 \succ a_6 \succ a_5 \succ a_4 \succ a_3 \succ a_2$							
# violations	23	172	195	195	205	205	205
max inequality	0.067	0.071	0.071	0.071	0.071	0.190	0.389
rej prob	0.130	0.408	0.278	0.278	0.384	1.000	1.000
rej prob (size adj)	0.548	0.600	0.594	0.590	0.701	1.000	1.000

Note. Shown in the table are the empirical rejection probabilities of our test for four preference orderings and different sets of inequality constraints. The results are based on 2,000 Monte Carlo simulations with nominal size 0.05. The effective sample size for each choice problem is 200. Other features of the simulation designs are discussed in the main text.

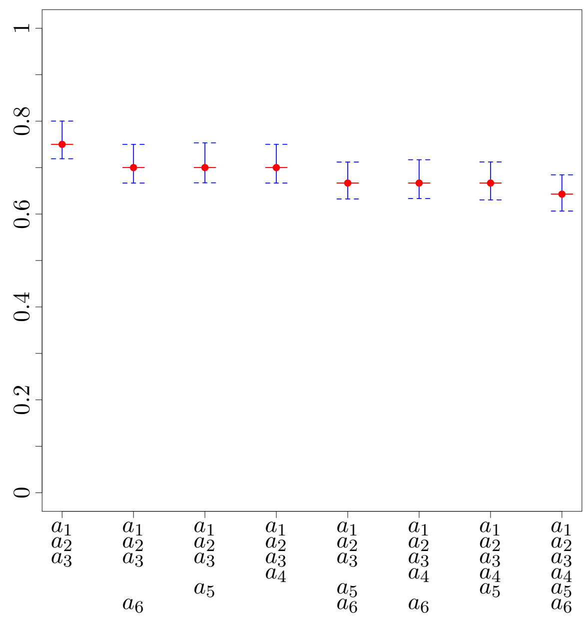
While in this particular setting it is difficult for our AOM to elicit preference on alternatives other than a_1, the random attention model (RAM) of Cattaneo, Ma, Masatlioglu, and Suleymanov (2020) can be powerful at identifying the less preferred alternatives. In other words, the two models, AOM and RAM, can complement each other and, when combined, may provide powerful identification on the decision maker's preference. In column "**RAM**," we list the number of restrictions imposed by the RAM, as well as the number of inequality violations implied for each preference ordering (See Section 5 for more details on the RAM). As before, the first preference is compatible with both our AOM and the RAM, which means it will not be rejected by any of the two models. The second preference is an interesting example which is compatible with the RAM but does not have an AOM representation. Indeed, the RAM alone does not lead to any inequality violation. The third and fourth preferences are instances where the RAM brings more empirical content. For example, the RAM strongly rejects both preferences but our AOM does not.

In the third column (labeled "**AOM+RAM**") we demonstrate the performance of our test when the two models are combined. Consider the second preference as an example. The RAM alone will not be able to reject this preference ordering, and hence the corresponding rejection probabilities are well below the nominal size. With AOM, however, we will be able to rule out this hypothesized preference with nontrivial power. The same holds for the other two preferences: although the AOM will find it challenging to falsify these preferences, incorporating the RAM will significantly improve the power of the test. In the last four columns, we further incorporate constraints implied by the attentive-at-binaries condition in (5). These additional restrictions also help improve the power of the test. See Section 4.2 for details.

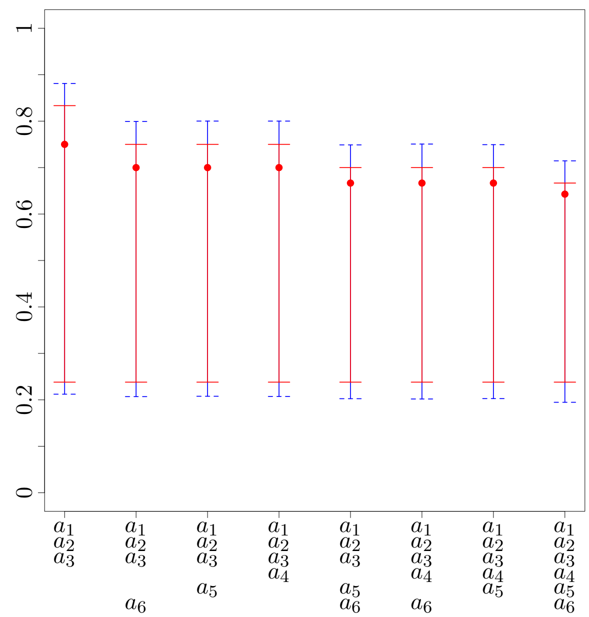
As our testing procedure involves estimating and evaluating a large number of inequality constraints, it can be conservative even with a pre-selection step. For theoretical comparison, we also show the size adjusted rejection probabilities (row "**rej prob (size adj)**"). These are obtained by employing the (infeasible) critical values which are simulated from the correctly centered multivariate normal distribution. That is, we set $\hat{\psi}(a|S,T) = \min\{D(a|S,T), 0\}/\sigma(a|S,T)$ (see Section 3.1). It should not come as a surprise that the empirical rejection probabilities are much closer to the nominal size for the first preference (or for the second preference when only the RAM restrictions are imposed), and that the power of the test becomes much higher.

To showcase our econometric methods for revealed attention, we compute the upper and lower bounds (Section 3.2) for the attention frequencies, $\phi(a_1|\cdot)$, $\phi(a_2|\cdot)$, and $\phi(a_3|\cdot)$ in the same simulation setting. In Figure 1, we first show the true attention frequency (red dots), which can be easily derived from the attention rule we employ. As expected, it decreases as the choice problem gets larger. For comparison, we plot the theoretical upper and lower bounds in Theorem 2 (i.e., the sharp identification region, represented by the red vertical lines). The bounds coincide with the true attention frequencies in panel (a), because a_1 is the most preferred alternative, meaning that it will be chosen whenever it attracts attention. In other words, the attention frequency of a_1 is point identified. For a_2 and a_3, however, the bounds no longer agree with the true attention frequency. For example, in panel (c), we see that the theoretical upper bound of the attention frequency $\phi(a_3|\{a_1, a_2, a_3\})$ is 1, meaning that it is completely uninformative. This is because a_3 is least preferred among the three alternatives, and hence it may not be chosen very often even if this alternative always attracts attention. Once we consider bigger choice problems, say $\phi(a_3|\{a_1, a_2, a_3, a_4\})$, then the upper bound will be strictly smaller than 1, and hence it is now informative. The reason is that the decision maker cannot always pay attention to a_3, otherwise she will never pick an inferior option such as a_4.

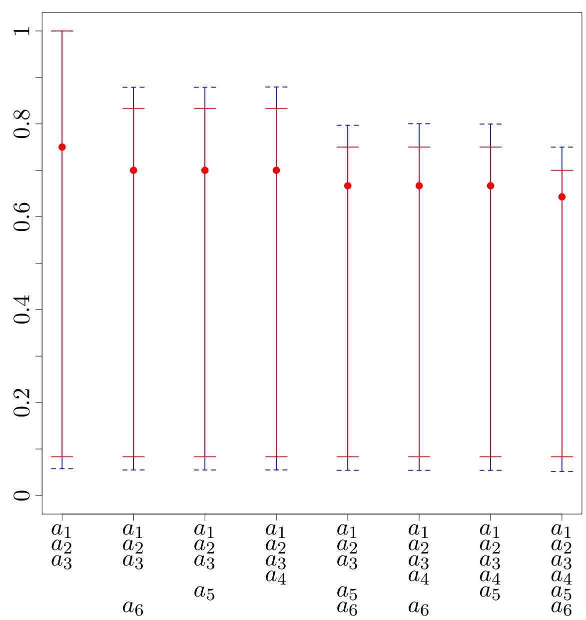
We also conducted 2,000 Monte Carlo simulations to investigate attention frequency elicitation from observed data. In each simulation iteration, we estimate the upper and lower bounds using the methods proposed in Section 3.2. We set $\alpha = 0.05$, which means the estimates should not cross the theoretical bounds more than 5%. In the same figure, we plot the 5th percentile of the



(a) $\phi(a_1|S)$



(b) $\phi(a_2|S)$



(c) $\phi(a_3|S)$

Note. Shown in the figure are theoretical and empirical attention frequencies of a_1, a_2, and a_3 across a range of choice problems (x-axis). The true attention frequencies are depicted by red dots, while the sharp identification regions are represented by the red vertical lines. The blue vertical lines are the 5th percentiles of the estimated upper bounds and the 95th percentiles of the estimated lower bounds, which are calculated from 2,000 Monte Carlo simulations. In each simulation repetition, we set the effective sample size to 200. We also set $\alpha = 0.05$ for simulating the critical values.

Figure 1: Theoretical and empirical attention frequencies.

estimated upper bounds and the 95th percentile of the estimated lower bounds (blue vertical lines). Indeed, they are always above and below the theoretical upper and lower bounds, respectively.

7 Conclusion

We introduced an attention overload model in which alternatives compete for the decision maker's attention. We showed that our model nests several important recent work on limited random attention, and can accommodate empirical behavioral phenomena such as choice overload. Despite being very general, we demonstrated that the nonparametric attention overload assumption still allows the development of a fruitful revealed preference theory, and we obtained testable implications on the choice probabilities which can be applied for preference revelation. Based on these testable implications, we proposed test statistics and critical values. We showed that our econometric methods remain valid even when "many" pairs of estimated choice probabilities are involved. Furthermore, we studied how additional restrictions on the attention rule can be incorporated to strengthen identification. We also considered (point and partial) identification and related econometric methods for the unobserved attention frequency.

Appendix A Proofs

This appendix contains the proofs of our main results. Complementary material, such as additional discussions and proofs of the technical lemmas are available in the online supplemental appendix.

A.1 Proof of Theorems 1, 3 and 4

First and foremost, we note that Theorem 1 is implied by Theorem 4, that is, AC captures all of the empirical content that our AOM delivers for revealed preference. In addition, we have already shown in the main paper the necessity of AC. Therefore, we only need to demonstrate the sufficiency. On the other hand, Theorem 3 will be shown in the following proof since we are exactly proving the existence of a attention rule that uses pessimistic evaluation. For optimistic evaluation, see footnote 7.

The proof is mainly divided into two parts. The first part sets up a system of linear equation which pins down the attention rule that satisfies the desired property. Some algebraic operations are devoted into lining up the system in a way to prepare for the second part. The second part shows how we can utilize the Farkas's Lemma to prove the existence of a solution to the system of equations for any parameter value which satisfies the attention compensation (AC) property.

Assume $(\pi, \succ)$ satisfies property AC. 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., $\phi(x|S) = \max_{R \supseteq S} \pi(x|R)$.[7] 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 $x_1 \succ x_2 \succ \cdots \succ x_n$. Then, we have for $i = 1, \cdots, n$

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

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

Note that for x_1, AC 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 $i = 1, \cdots, n-1$ and $\mathcal{M}_i$ for $i = 2, \cdots, n$. Then, we create $\mathcal{M}'_i \equiv \sum_{j \leq i} \mathcal{P}_j - \mathcal{M}_i$ for every $i = 2, \cdots, n$, that is,

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

The above makes sense because $\sum_{j \leq i} \pi(x_j|S) - \max_{R \supseteq S} \pi(x_i|R) \geq 0$ for $i = 2, \cdots, n$, which is

[7] 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., $\phi_\mu(a|S) = \min_{T \subseteq S} \pi(U_{\succeq}(a)|T)$, and same proof strategy goes through. In fact, for any attention frequency between these bounds, our proof remains valid if we choose ϕ such that it satisfies attention overload.

required by AC. Lastly, we define $\mathcal{P}_1' \equiv \mathcal{P}_1 - \sum_{j>1} \mathcal{M}_j$. Hence, we are left with $\mathcal{P}_1'$, $\mathcal{P}_i$ for $i = 2 \cdots, n-1$, and $\mathcal{M}_i'$ for $i = 2, \cdots, n$. We utilize the 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 only on the case $n \geq 3$.

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, $\mathbf{A} = \left(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_{2n-2}\right)^\top$, and $\mathbf{b} = (b_1, b_2, \cdots, b_{2n-2})^\top$, 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 $j = 2, \cdots, n$; $\mathbf{r}_j$ and b_j correspond to the LHS and RHS of $\mathcal{P}_{-n+1+j}$ respectively for $j = n+1, \cdots, 2n-2$. 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 the collection of $\mathbf{b}$ subject to the condition AC:

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

We show that there does not exist $\mathbf{y} = (y_1, y_2, y_3, \cdots, y_{2n-2}) \in \mathbb{R}^{2n-2}$ 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 $\mathbf{b} \in \mathcal{B}$, $\min_{\mathbf{y} \in \mathcal{Y}(\mathbf{A})} \mathbf{y}\mathbf{b} \geq 0$. Note that except for b_1, all the other b_j are positive for all possible $\pi(\cdot|\cdot)$ as long as the choice rule satisfies AC. Hence, the key insight in the following proof is to show that how we can guarantee $\mathbf{y}\mathbf{b} \geq 0$ despite the possibility of b_1 being negative.

Notice that, 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 us $y_j \geq 0$ for all j. In other words, we know that 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 $P \subseteq \{2, 3, \cdots, n-1\}$, and $j = \bar{P}+1, \bar{P}+2, \cdots, 2n-\bar{P}$, 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. $\mathbf{c}_n = (c_2, c_3, \cdots, c_{2n-2})^\top$

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 $\mathbf{C}_n = \{\mathbf{c}_n \in \mathbb{R}_+^{2n-3} | \sum_{i=2}^{n+1-j} c_i + \sum_{i=n+3-j}^{n} c_i + \sum_{i=n+1}^{n-2+j} c_i \geq 1, j = 2, 3, ...n\}$, $\mathbf{Z}_n = \{\mathbf{z}_n \in \mathbb{R}_+^{2n-3} \Big| \sum_{i \in P} z_i + z_j \geq |P|, \forall$ non-empty $P \subseteq \{2, 3, \cdots, n-1\}, j = \bar{P}+1, \bar{P}+2, \cdots, 2n-\bar{P}\}$ 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 $i = 2, \cdots, 2n-2$. 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 $\mathbf{b} \in \mathbf{B}$, $\min_{\mathbf{y} \in \mathcal{Y}(\mathbf{A})} \mathbf{y}\mathbf{b} \geq 0$ 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 $j \in \{\bar{P}+1, \bar{P}+2, \cdots, \bar{P}+n\}$. 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 $j \in \{n+1, n+2, \cdots, 2n-\bar{P}\}$. 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 $\{\bar{P}+1, \bar{P}+2, \cdots, 2n-\bar{P}\}$, 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 $\mathbf{c}_3 \cdot \mathbf{z}_3 = z_2c_2 + z_3c_3 + z_4c_4 \geq z_2(1) + c_3(1-z_2) + c_4(1-z_2) = z_2 + (c_3+c_4)(1-z_2) \geq 1$. 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 λ_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 λ_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 it own unique subscript. We collect all possible subscript labels into the set Λ_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:

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

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 $\mathbf{C}_k$. Note that if $c_i \neq 0$ for all i, it is straight-forward to see that $\sum_{I\in\Lambda_k} \lambda_I \geq 1$. For example, if $c_{2k-2} \neq 0$ and $c_2 \neq 0$, we can get binding constraint due to complementary slackness such that $z_2 = \sum_{2\in I\in\Lambda_k}$ 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

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

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

$$\vdots$$

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

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

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 $i = 2, 3, \cdots, 2(k-1)-2$. 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 $\mathbf{c}_k = (c_2, c_3, ..., c_{2k-2}) := (0, c'_2, c'_3, ..., c'_{2(k-1)-2}, c_{2k-2})$. 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 $i = 3, \cdots, 2(k-1)-2$. Then, we have $\mathbf{c}_k = (c_2, c_3, ..., c_{2k-2}) := (c'_2, 0, c'_3, ..., c'_{2(k-1)-3}, c'_{2(k-1)-2}, 0)$, 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 $i = 4, \cdots, 2(k-1)-3$, and $c'_i = c_{i+2}$ for $i = 2(k-1)-2$. Then, we have $\mathbf{c}_k = (c_2, c_3, ..., c_{2k-2}) := (c'_2, c'_3, 0, c'_5, ..., c'_{2(k-1)-3}, 0, c'_{2(k-1)-2})$, 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 $c_{k+2} = \cdots = c_{2k-3} = c_{2k-2} = 0$. Write $c'_i = c_i$ for $i = 2, ..k-2$, $c'_i = c_{i+1}$ for $i = k-1, k$, and $c'_i = c_{i+2}$ for $i = k+1, .., 2(k-1)-2$. 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)**: $c_2, c_3, c_{k-1} \neq 0$, $c_k = c_{k+1} = \cdots = c_{2k-2} = 0$. Write $c'_i = c_i$ for $i = 2, ..k-1$ and $c'_i = c_{i+2}$ for $i = k, k+1, .., 2(k-1)-2$. 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 cover 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 5

We denote the choice probabilities by the vector $\boldsymbol{\pi}$, and the constraints implied by AC 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. $\oslash$ 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,

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

The next step is to replace the infeasible standard error.

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

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

As a result,

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

As 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 $\xi_2\to 0$. Then,

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

We now state the properties of the estimated correlation matrix.

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

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

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 $\xi_3 \to 0$. Then,

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

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

$$\begin{aligned}
&\mathcal{T}(\succ) = \max\Big\{(\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}}) \oslash \hat{\boldsymbol{\sigma}}_{\succ},\ 0\Big\},\\
&\mathcal{T}(\succ)^{\circ} = \max\Big\{(\mathbf{R}_{\succ}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})) \oslash \hat{\boldsymbol{\sigma}}_{\succ},\ 0\Big\}, \qquad \tilde{\mathcal{T}}(\succ)^{\circ} = \max\Big\{(\mathbf{R}_{\succ}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})) \oslash \boldsymbol{\sigma}_{\succ},\ 0\Big\},\\
&\mathcal{T}^{\mathtt{G}}(\succ) = \max\Big\{\mathbf{z},\ 0\Big\}, \qquad \check{\mathcal{T}}^{\mathtt{G}}(\succ) = \max\Big\{\check{\mathbf{z}},\ 0\Big\}.
\end{aligned}$$

We will also define the following quantiles/critical values.

$$\begin{aligned}
\text{cv}(\alpha, \succ) &= \inf\left\{t \ge 0:\ \mathbb{P}\left[\mathcal{T}^{\mathtt{G}}(\succ) \le t \middle| \text{Data}\right] \ge 1 - \alpha\right\},\\
\check{\text{cv}}(\alpha, \succ) &= \inf_t\left\{t \ge 0:\ \mathbb{P}\left[\check{\mathcal{T}}^{\mathtt{G}}(\succ) \le t\right] \ge 1 - \alpha\right\},\\
\tilde{\text{cv}}(\alpha, \succ) &= \inf_t\left\{t \ge 0:\ \mathbb{P}\left[\max(\check{\mathbf{z}}) \le t\right] \ge 1 - \alpha\right\}.
\end{aligned}$$

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

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

To wrap up the proof of Theorem 5, we rely on a sequence of error bounds. First notice that, under H_0,

$$\begin{aligned}
\mathbb{P}\left[\mathcal{T}(\succ) > \text{cv}(\alpha, \succ)\right] &\le \mathbb{P}\left[\mathcal{T}(\succ) > \check{\text{cv}}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathfrak{c}_1, \succ\right)\right] + c\exp\left\{-\frac{1}{c}\mathfrak{c}_2^2\xi_3^2 + 2\log \mathfrak{c}_1\right\}\\
&\le \mathbb{P}\left[\mathcal{T}(\succ)^{\circ} > \check{\text{cv}}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathfrak{c}_1, \succ\right)\right] + c\exp\left\{-\frac{1}{c}\mathfrak{c}_2^2\xi_3^2 + 2\log \mathfrak{c}_1\right\}. \qquad (6)
\end{aligned}$$

Next, we apply Lemma A.5 and obtain that

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

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

$$\mathbb{P}\left[\tilde{\mathcal{T}}(\succ)^{\circ} > \check{\text{cv}}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathfrak{c}_1, \succ\right) - \xi_1\xi_2\right] \le \mathbb{P}\left[\check{\mathcal{T}}^{\mathtt{G}}(\succ) > \check{\text{cv}}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathfrak{c}_1, \succ\right) - \xi_1\xi_2\right] + c\left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{\frac{1}{4}}.$$

Collecting all pieces, we have

$$\mathbb{P}\left[\mathcal{T}(\succ) > \mathrm{cv}(\alpha, \succ)\right] \le \mathbb{P}\left[\check{\mathcal{T}}^{\mathtt{G}}(\succ) > \check{\mathrm{cv}}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathfrak{c}_1, \succ\right) - \xi_1\xi_2\right]$$
$$+ c\left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{\frac{1}{4}} + c\xi_1^{-1}\sqrt{\log \mathfrak{c}_1} + c\exp\left\{-\frac{1}{c}\mathfrak{c}_2^2(\xi_2 \wedge \xi_3)^2 + 2\log \mathfrak{c}_1\right\}.$$

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

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

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

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

$$\tilde{\mathrm{cv}}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathfrak{c}_1 + 4\xi_1\xi_2(\sqrt{2\log \mathfrak{c}_1} + 1), \succ\right) \le \tilde{\mathrm{cv}}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathfrak{c}_1, \succ\right) - \xi_1\xi_2,$$

which further implies that

$$\mathbb{P}\left[\check{\mathcal{T}}^{\mathtt{G}}(\succ) > \check{\mathrm{cv}}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathfrak{c}_1, \succ\right) - \xi_1\xi_2\right] \le \mathbb{P}\left[\check{\mathcal{T}}^{\mathtt{G}}(\succ) > \tilde{\mathrm{cv}}\left(\alpha + c\xi_3^{\frac{1}{2}} \log \mathfrak{c}_1 + 4\xi_1\xi_2(\sqrt{2\log \mathfrak{c}_1} + 1), \succ\right)\right]$$
$$\le \alpha + c\xi_3^{\frac{1}{2}} \log \mathfrak{c}_1 + 4\xi_1\xi_2(\sqrt{2\log \mathfrak{c}_1} + 1).$$

Finally we have that

$$\mathbb{P}\left[\mathcal{T}(\succ) > \mathrm{cv}(\alpha, \succ)\right] \le \alpha + c\xi_3^{\frac{1}{2}} \log \mathfrak{c}_1 + 4\xi_1\xi_2(\sqrt{2\log \mathfrak{c}_1} + 1)$$
$$+ c\left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{\frac{1}{4}} + c\xi_1^{-1}\sqrt{\log \mathfrak{c}_1} + c\exp\left\{-\frac{1}{c}\mathfrak{c}_2^2(\xi_2 \wedge \xi_3)^2 + 2\log \mathfrak{c}_1\right\}.$$

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

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

Then the last term in the above becomes

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

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

A.5 Proof of Theorem 6

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

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

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

$$\mathfrak{r}_{\phi_{\mathtt{L}}(a|S)} = \left|\mathbb{P}\Big[\frac{\hat{\pi}(a|R) - \pi(a|R)}{\hat{\sigma}(a|R)} \leq \mathrm{cv}(\alpha, \phi_{\mathtt{L}}(a|S)),\ \forall R \supseteq S,\ R \in \mathcal{S}\Big] - \mathbb{P}\Big[\max(\mathbf{z}) \leq \mathrm{cv}(\alpha, \phi_{\mathtt{L}}(a|S)\Big]\right|,$$

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

Lemma A.11. The following normal approximation holds

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

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 $\xi_2 \to 0$. Then

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

As a result,

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

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

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

$$\leq c\xi_1^{-1}\sqrt{\log \mathfrak{c}_1} + c\left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{\frac{1}{4}} + c\exp\left\{-\frac{1}{c}\mathfrak{c}_2^2\xi_2^2 + \log \mathfrak{c}_1\right\} + c\xi_1\xi_2\sqrt{\log \mathfrak{c}_1},$$

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 Proposition 2

The necessity part is immediate. For the sufficiency, we need to be concerned about constructing the μ. For non-binaries choice set, we follow the technique in the proof for Theorem 4. For binaries choice set, without loss of generality, we first let $x \succ y$ throughout the proof. We assume $\mu(\{x,y\}|\{x,y\}) = \pi(x|\{x,y\})$ and $\mu(\{y\}|\{x,y\}) = \pi(y|\{x,y\})$. To check that it fulfill the assumption on η and μ. Suppose not. i.e. there exists y s.t. $\eta < \mu(y|\{x,y\}$. Hence, by η-constrained revealed preference, we know that $y \succ x$, which violates the fact that $x \succ y$. To check that it fulfills attention overload. Note that $\phi(x|\{x,y\}) = \pi(x|\{x,y\}) \geq \pi(x|S)$ for all S (or there would be a contradiction that $y \succ x$). Hence, $\phi(x|\{x,y\}) \geq \phi(x|S)$ for all S since $\phi(x|S) = \max_{R\supseteq S}\pi(x|R)$. On the other hand, $\phi(y|\{x,y\}) = 1$, which automatically satisfies Assumption 1.

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.

Barseghyan, L., M. Coughlin, F. Molinari, and J. C. Teitelbaum (2021): "Heterogeneous Choice Sets and Preferences," *Econometrica*, 89(5), 2015–2048.

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.

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.

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

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 (2021): "Improved Central Limit Theorem and Bootstrap Approximations in High Dimensions," *Working paper*.

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 Peference with Limited Consideration," *American Economic Journal: Microeconomics*, 10(1), 102–131.

Echenique, F., and K. Saito (2019): "General Luce Model," *Economic Theory*, 68, 811–826.

Echenique, F., K. Saito, and G. Tserenjigmid (2018): "The Perception-Adjusted Luce Model," *Mathematical Social Sciences*, 93, 67–76.

Filiz-Ozbay, E., and Y. Masatlioglu (2020): "Progressive Random Choice," *Working Paper*.

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

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*, forthcoming.

Goeree, M. S. (2008): "Limited Information and Advertising in the US Personal Computer Industry," *Econometrica*, 76(5), 1017–1074.

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

Hauser, J. R., and B. Wernerfelt (1990): "An Evaluation Cost Model of Consideration Sets," *Journal of Consumer Research*, 16(4), 393–408.

Horan, S. (2019): "Random Consideration and Choice: A Case Study of "Default" Options," *Mathematical Social Sciences*, 102, 73–84.

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.

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.

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.

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 (2012): "Categorize Then Choose: Boundedly Rational Choice and Welfare," *Journal of the European Economic Association*, 10(5), 1141–1165.

——— (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82(3), 1153–1176.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012): "Revealed Attention," *American Economic Review*, 102(5), 2183–2205.

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.

Reutskaja, E., and R. M. Hogarth (2009): "Satisfaction in Choice as a Function of the Number of Alternatives: When "Goods Satiate"," *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.

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.

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.