

A Random Attention Model^{*}

Matias D. Cattaneo[†] Xinwei Ma[‡] Yusufcan Masatlioglu[§] Elchin Suleymanov[¶]

April 5, 2019

Abstract

We introduce a Random Attention Model (RAM) allowing for a large class of stochastic consideration maps in the context of an otherwise canonical limited attention model for decision theory. The model relies on a new restriction on the unobserved, possibly stochastic consideration map, termed *Monotonic Attention*, which is intuitive and nests many recent contributions in the literature on limited attention. We develop revealed preference theory within RAM and obtain precise testable implications for observable choice probabilities. Using these results, we show that a set (possibly a singleton) of strict preference orderings compatible with RAM is identifiable from the decision maker’s choice probabilities, and establish a representation of this identified set of unobserved preferences as a collection of inequality constraints on her choice probabilities. Given this nonparametric identification result, we develop uniformly valid inference methods for the (partially) identifiable preferences. Furthermore, to illustrate the applicability of our results and their concrete empirical content in specific settings, we also develop revealed preference theory and accompanying econometric methods under additional nonparametric assumptions on the consideration map at binary choice problems. Finally, we showcase the performance of our proposed methods using simulations, and provide general-purpose software implementation of our estimation and inference results in the R software package `ramchoice`. Our proposed econometric methods are computationally very fast to implement.

Keywords: revealed preference, limited attention models, random utility models, nonparametric identification, partial identification.

^{*}We thank Ivan Canay, Ignacio Esponda, Rosa Matzkin, Francesca Molinari, Jose Luis Montiel-Olea, Pietro Ortoleva, Kota Saito, Joerg Stoye, and Rocio Titiunik for very helpful comments and suggestions that improved this paper. We also thank the Editor, Emir Kamenica, an anonymous Associate Editor, and three reviewers for their constructive criticism of our paper, which led to substantial improvements. Financial support from the National Science Foundation through grant SES-1628883 is gratefully acknowledged.

[†]Department of Operations Research and Financial Engineering, Princeton University.

[‡]Department of Economics, University of California, San Diego.

[§]Department of Economics, University of Maryland.

[¶]Department of Economics, University of Michigan.

1 Introduction

Revealed preference theory is not only a cornerstone of modern economics, but is also the source of important theoretical, methodological and policy implications for many social and behavioral sciences. This theory aims to identify the preferences of a decision maker (e.g., an individual or a firm) from her observed choices (e.g., buying a house or hiring a worker). In its classical formulation, revealed preference theory assumes that the decision maker selects the best available option after full consideration of all possible alternatives presented to her. This assumption leads to specific testable implications based on observed choice probabilities but, unfortunately, empirical testing of classical revealed preference theory shows that it is not always compatible with observed choice behavior (Hauser and Wernerfelt, 1990; Goeree, 2008; van Nierop et al., 2010; Honka et al., 2017). For example, Reutskaja, Nagel, Camerer, and Rangel (2011) provides interesting experimental evidence against the full attention assumption using eye tracking and choice data.

Motivated by these findings, and the fact that certain theoretically important and empirically relevant choice patterns can not be explained using classical revealed preference theory based on full attention, scholars have proposed other economic models of choice behavior. An alternative is the limited attention model (Masatlioglu, Nakajima, and Ozbay, 2012; Lleras, Masatlioglu, Nakajima, and Ozbay, 2017; Dean, Kibris, and Masatlioglu, 2017), where decision makers are assumed to select the best available option from a subset of all possible alternatives, known as the consideration set. This framework takes the formation of the consideration set, sometimes called attention rule or consideration map, as unobservable and hence as an intrinsic feature of the decision maker. Nonetheless, it is possible to develop a fruitful theory of revealed preference within this framework, employing only mild and intuitive nonparametric restrictions on how the decision maker decides to focus attention on specific subsets of all possible alternatives presented to her.

Until very recently, limited attention models have been deterministic, a feature that diminished their empirical applicability: testable implications via revealed preference have relied on the assumption that the decision maker pays attention to the same subset of options every time she is confronted with the same set of available alternatives. This requires that, for example, an online shopper uses always the same keyword and the same search engine (e.g. Google) on the same platform (e.g. tablet) to look for a product. This is obviously restrictive, and can lead to predic-

tions that are inconsistent with observed choice behavior. Aware of this fact, a few scholars have improved deterministic limited attention models by allowing for stochastic attention (Manzini and Mariotti, 2014; Aguiar, 2015; Brady and Rehbeck, 2016; Horan, 2018), which permits the decision maker to pay attention to different subsets with some non-zero probability given the same set of alternatives to choose from. All available results in this literature proceed by first parameterizing the consideration set formation (i.e., committing to a particular parametric attention rule), and then studying the revealed preference implications of these parametric models.

In contrast to earlier approaches, we introduce a Random Attention Model (RAM) where we abstain from any specific parametric (stochastic) consideration set formation, and instead consider a large class of nonparametric random attention rules. Our model imposes one intuitive condition, termed *Monotonic Attention*, which is satisfied by many stochastic consideration set formations. Given that consideration sets are unobservable, this feature is crucial for applicability of our revealed preference results, as our findings and empirical implications are valid under many different, particular attention rules that could be operating in the background. In other words, our revealed preference results are derived from nonparametric restrictions on the consideration set formation and hence are more robust to misspecification biases.

Another important benefit of imposing a weak nonparametric condition on the consideration set formation is that it accommodates seemingly anomalous but frequently observed behaviors that cannot be captured by the standard random utility models (RUM). Rieskamp, Busemeyer, and Mellers (2006) and Busemeyer and Rieskamp (2014) review the empirical evidence on stochastic choice data accumulated over the last 50 years from economics, marketing, and psychology. They find strong evidence against RUM. The main finding in this literature is the violation of the principle of regularity, which states that the probability of choosing an alternative never increase when more products become available.¹ Our explanations for these choice patterns depend solely on random attention, hence seemingly irrational behavior can be rationalized without introducing random preferences.

RAM is best suited for eliciting information about the preference ordering of a single decision-

¹The regularity violation has been illustrated not only with human beings but also with a wide range of animals, including honeybees, gray jays, and hummingbirds (Hurly and Oseen, 1999; Shafir, Waite, and Smith, 2002; Bateson, Healy, and Hurly, 2003).

making unit when her choices are observed repeatedly.² For example, scanner data keeps track of the same single consumer’s purchases across repeated visits, where the grocery store adjusts product varieties and arrangements regularly. Another example is web advertising on digital platforms, such as search engines or shopping sites, where not only abundant records from each individual decision maker is available, but also is common to see manipulations/experiments altering the options offered to them. A third example is given in [Kawaguchi, Uetake, and Watanabe \(2016\)](#), where large data on each consumer’s choices from vending machines (with varying product availability) is analyzed. In addition, our model can be used empirically with aggregate data on a group of distinct decision makers, provided each of them may differ on what they pay attention to but all share the same preference (e.g., searching for the cheapest option).

Our key identifying assumption, *Monotonic Attention*, restricts the possibly stochastic attention formation process in a very intuitive way: each consideration set competes for the decision maker’s attention, and hence the probability of paying attention to a particular subset is assumed not to decrease when the total number of possible consideration sets decreases. We show that this single nonparametric assumption is general enough to nest most (if not all) previously proposed deterministic and random limited attention models. Furthermore, under our proposed monotonic attention assumption, we are able to develop a theory of revealed preference, obtain specific testable implications, and (partially) identify the underlying preferences of the decision maker by investigating her observed choice probabilities. Our revealed preference results are applicable to a wide range of consideration set formations, including the parametric ones currently available in the literature which, as we show, satisfy *Monotonic Attention*.

Based on these theoretical findings, we also develop econometric results for identification, estimation, and inference of the decision maker’s preferences, as well as specification testing of RAM. To be specific, we show that RAM implies that the set of partially identified preference orderings containing the decision maker’s true preferences is equivalent to a set of inequality restrictions on the choice probabilities (one for each preference ordering in the identified set). This result allows us to employ the identifiable/estimable choice probabilities to (i) develop a model specification test

²The finding that individual choices frequently exhibit randomness was first reported in [Tversky \(1969\)](#) and has now been illustrated by [Agranov and Ortoleva \(2017\)](#) and numerous other studies. Similar to our work, [Manzini and Mariotti \(2014\)](#), [Fudenberg, Iijima, and Strzalecki \(2015\)](#), and [Brady and Rehbeck \(2016\)](#) among others have developed models which allow the analyst to reveal information about the agent’s fixed preferences from her observed random choices.

(i.e., test whether there exists a non-empty set of preference orderings compatible with RAM), (ii) conduct hypothesis testing on specific preference orderings (i.e., test whether the inequality constraints on the choice probabilities are satisfied), and (iii) develop confidence sets containing the decision maker’s true preferences with pre-specified coverage (i.e., via test inversion). Our econometric methods rely on ideas and results from the literature on partially identified models and moment inequality testing: see [Canay and Shaikh \(2017\)](#), [Ho and Rosen \(2017\)](#) and [Molinari \(2019\)](#) for recent reviews and further references.

RAM is fully nonparametric and agnostic because it relies on the *Monotonic Attention* assumption only. As a consequence, it may lead to relatively weak testable implications in some applications, that is, “little” revelation or a “large” identified set of preferences. However, RAM also provides the basis for incorporating additional (parametric and) nonparametric restrictions that can substantially improve identification of preferences. In this paper, we illustrate how RAM can be combined with additional, mild nonparametric restrictions to tighten identification in non-trivial ways: In Section 5.1, we incorporate an additional restriction on consideration maps for binary sets, and show that this alone leads to important revelation improvements within RAM. We also illustrate this result numerically in our simulation study.

Finally, we implement our estimation and inference methods in the general-purpose software package `ramchoice` for R—see <https://cran.r-project.org/package=ramchoice> for details. Our novel identification results allow us to develop inference methods that avoid optimization over a possibly high-dimensional parameter space, leading to methods that are very fast and easy to implement when applied to realistic empirical problems. See the Supplemental Appendix for numerical evidence.

Our work contributes to both economic theory and econometrics. We discuss in detail the connections and distinctions between this paper and the economic theory literature in Section SA-1 of the Supplemental Appendix, after we introduce our proposed RAM, describe several examples covered by our model, and develop revealed preference theory. In particular, we show how RAM nests and/or connects to the recent work by [Manzini and Mariotti \(2014\)](#), [Brady and Rehbeck \(2016\)](#), [Gul, Natenzon, and Pesendorfer \(2014\)](#), [Echenique, Saito, and Tserenjigmid \(2018\)](#), [Echenique and Saito \(2019\)](#), [Fudenberg, Iijima, and Strzalecki \(2015\)](#), and [Aguiar, Boccardi, and Dean \(2016\)](#), among others.

This paper is also related to a rich econometric literature on nonparametric identification, estimation and inference both in the specific context of Random Utility Models (RUMs), and more generally. See [Matzkin \(2013\)](#) for a review and further references on nonparametric identification, [Hausman and Newey \(2017\)](#) for a recent review and further references on nonparametric welfare analysis, and [Blundell, Kristensen, and Matzkin \(2014\)](#), [Kawaguchi \(2017\)](#), [Kitamura and Stoye \(2018\)](#), and [Deb, Kitamura, Quah, and Stoye \(2018\)](#) for a sample of recent contributions and further references. As mentioned above, a key feature of RAM is that our proposed Monotonic Attention condition on consideration formation nests previous models as special cases, and also covers many new models of choice behavior. In particular, RAM is more general than RUM, which is important because numerous studies in psychology, finance and marketing have shown that decision makers exhibit limited attention when making choices: they only compare (and choose from) a subset of all available options. Whenever decision makers do not pay full attention to all options, implications from revealed preference theory under RUM no longer hold in general, implying that empirical testing of substantive hypotheses as well as policy recommendations based on RUM will be invalid. On the other hand, our results may remain valid because RAM is a strict, non-trivial generalization of RUM.

Finally, in contemporaneous work, a few scholars have also developed identification and inference results under (random) limited attention, trying to connect behavioral theory and econometric methods, as we do in this paper. Three recent examples of this new research area include [Abaluck and Adams \(2017\)](#), [Dardanoni, Manzini, Mariotti, and Tyson \(2018\)](#), and [Barseghyan, Coughlin, Molinari, and Teitelbaum \(2018\)](#). These papers are complementary to ours insofar different assumptions on the random attention rule and preference(s) are imposed, which lead to different levels of (partial) identification of preference(s) and (random) attention map(s).

The rest of the paper proceeds as follows. Section 2 presents the basic setup, where our key monotonicity assumption on the decision maker’s stochastic consideration map is presented in Section 2.1. Section 3 discusses in detail our random attention model, including the main revealed preference results. Section 4 presents our main econometrics methods, including nonparametric (partial) identification, estimation, and inference results. In Section 5.1, we consider additional restrictions on the consideration map for binary choice problems, which can help improve our identification and inference results considerably. We also consider random attention filters in Section

5.2, which are one of the motivating examples of monotonic attention rules. In this case, however, there is no additional identification. Section 6 summarizes the findings from a simulation study. Finally, Section 7 concludes with a discussion of directions for future research. A companion on-line Supplemental Appendix includes more examples, extensions and other methodological results, omitted proofs, and additional simulation evidence.

2 Setup

We designate a finite set X to act as the universal set of all mutually exclusive alternatives. This set is thus viewed as the grand alternative space, and is kept fixed throughout. A typical element of X is denoted by a and its cardinality is $|X| = K$. We let \mathcal{X} denote the set of all non-empty subsets of X . Each member of \mathcal{X} defines a choice problem.

Definition 1 (Choice Rule). A choice rule is a map $\pi : X \times \mathcal{X} \rightarrow [0, 1]$ 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 $\sum_{a \in S} \pi(a|S) = 1$.

Thus, $\pi(a|S)$ represents the probability that the decision maker chooses alternative a from the choice problem S . Our formulation allows both stochastic and deterministic choice rules. If $\pi(a|S)$ is either 0 or 1, then choices are deterministic. For simplicity in the exposition, we assume that all choice problems are potentially observable throughout the main paper, but this assumption is relaxed in Section SA-3 of the Supplemental Appendix to account for cases where only data on a subcollection of choice problems is available.

The key ingredient in our model is probabilistic consideration sets. Given a choice problem S , each non-empty subset of S could be a consideration set with certain probability. We impose that each frequency is between 0 and 1 and that the total frequency adds up to 1. Formally,

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

Thus, $\mu(T|S)$ represents the probability of paying attention to the consideration set $T \subset S$ when the choice problem is S . This formulation captures both deterministic and stochastic attention rules. For example, $\mu(S|S) = 1$ represents an agent with full attention. Given our approach, we can always extract the probability of paying attention to a specific alternative: For a given $a \in S$,

$\sum_{a \in T \subset S} \mu(T|S)$ is the probability of paying attention to a in the choice problem S . The probabilities on consideration sets allow us derive the attention probabilities on alternatives uniquely.

2.1 Monotonic Attention

We consider a choice model where a decision maker picks the maximal alternative with respect to her preference among the alternatives she pays attention to. Our ultimate goal is to elicit her preferences from observed choice behavior without requiring any information on consideration sets. Of course, this is impossible without any restrictions on her (possibly random) attention rule. For example, a decision maker’s choice can always be rationalized by assuming she only pays attention to singleton sets. Because the consumer never considers two alternatives together, one cannot infer her preferences at all.

We propose a property (i.e., an identifying restriction) on how stochastic consideration sets change as choice problems change, as opposed to explicitly modeling how the choice problem determines the consideration set. We argue below that this nonparametric property is indeed satisfied by many problems of interest and mimics heuristics people use in real life (see examples below and in Section SA-2 of the Supplemental Appendix). This approach makes it possible to apply our method to elicit preference without relying on a particular formation mechanism of consideration sets.

Assumption 1. (Monotonic Attention) For any $a \in S - T$, $\mu(T|S) \leq \mu(T|S - a)$.

Monotonic μ captures the idea that each consideration set competes for consumers’ attention: The probability of a particular consideration set does not shrink when the number of possible consideration sets decreases. Removing an alternative that does not belong to the consideration set T results in less competition for T , hence the probability of T being the consideration set in the new choice problem is weakly higher. Our assumption is similar to the regularity condition proposed by [Suppes and Luce \(1965\)](#). The key difference is that their regularity condition is defined on choice probabilities, while our assumption is defined on attention probabilities.

To demonstrate the richness of the framework and motivate the analysis to follow, we discuss six leading examples of families of monotonic attention rules, that is, attention rules satisfying Assumption 1. We offer several more examples in Section SA-2 of the Supplemental Appendix. The first example is deterministic (i.e., $\mu(T|S)$ is either 0 or 1), but the others are all stochastic.

1. (ATTENTION FILTER; Masatlioglu, Nakajima, and Ozbay, 2012) A large class of deterministic attention rules, leading to consideration sets that do not change if an item not attracting attention is made unavailable (Attention Filter), was introduced by Masatlioglu et al. (2012). A classical example in this class is when a decision maker considers all the items appearing in the first page of search results and overlooks the rest. Formally, let $\Gamma(S)$ be the deterministic consideration set when the choice problem is S , and hence $\Gamma(S) \subset S$. Then, Γ is an Attention Filter if when $a \notin \Gamma(S)$, then $\Gamma(S - a) = \Gamma(S)$. In our framework, this class corresponds to the case $\mu(T|S) = 1$ if $T = \Gamma(S)$, and 0 otherwise.

2. (RANDOM ATTENTION FILTERS) Consider a decision maker whose attention is deterministic but utilizes different deterministic attention filters on different occasions. For example, it is well-known that search behavior on distinct platforms (mobile, tablet, and desktop) is drastically different (e.g., the same search engine produces different first page lists depending on the platform, or different platforms utilize different search algorithms). In such cases, while the consideration set comes from a (deterministic) attention filter for each platform, the resulting consideration set is random. Formally, if a decision maker utilizes each attention filter Γ_j with probability ψ_j , then the attention rule can be written as

$$\mu(T|S) = \sum_j \mathbb{1}(\Gamma_j(S) = T) \cdot \psi_j.$$

We will pay special attention to this class of attention rules in Section 5.2.

3. (INDEPENDENT CONSIDERATION; Manzini and Mariotti, 2014) Consider a decision maker who pays attention to each alternative a with a fixed probability $\gamma(a) \in (0, 1)$. γ represents the degree of brand awareness for a product, or the willingness of an agent to seriously evaluate a political candidate. The frequency of each set being the consideration set can be expressed as follows: for all $T \subset S$,

$$\mu(T|S) = \frac{1}{\beta_S} \prod_{a \in T} \gamma(a) \prod_{a \in S-T} (1 - \gamma(a)),$$

where $\beta_S = 1 - \prod_{a \in S} (1 - \gamma(a))$, which represents the probability that the decision maker

considers no alternative in S , is used to adjust each probability so that they sum up to 1.

4. (LOGIT ATTENTION; [Brady and Rehbeck, 2016](#)) Consider a decision maker who assigns a positive weight for each non-empty subset of X . Psychologically w_T is a strength associated with the subset T . The probability of considering T in S can be written as follows:

$$\mu(T|S) = \frac{w_T}{\sum_{T' \subset S} w_{T'}}.$$

Even though there is no structure on weights in the general version of this model, there are two interesting special cases where weights solely depend on the size of the set. These are $w_T = |T|$ and $w_T = \frac{1}{|T|}$, which are conceptually different. In the latter, the decision maker tends to have smaller consideration sets, while larger consideration sets are more likely in the former.

5. (DOGIT ATTENTION) This example is a generalization of Logit Attention, and is based on the idea of the Dogit model ([Gaundry and Dagenais, 1979](#)). A decision maker is captive to a particular consideration set with certain probability, to the extent that she pays attention to that consideration set regardless of the weights of other possible consideration sets. Formally, let

$$\mu(T|S) = \frac{1}{1 + \sum_{T' \subset S} \theta_{T'}} \frac{w_T}{\sum_{T' \subset S} w_{T'}} + \frac{\theta_T}{1 + \sum_{T' \subset S} \theta_{T'}},$$

where $\theta_T \geq 0$ represents the degree of captivity (impulsivity) of T . The “captivity parameter” reflects the attachment of a decision maker to a certain consideration set. Since w_T are non-negative, the second term, which is independent of w_T , is the smallest lower bound for $\mu(T|S)$. The larger θ_T , the more likely the decision maker is to be captive to T and pay attention to it. When $\theta_T = 0$ for all T , this model becomes Logit Attention. This formulation is able to distinguish between impulsive and deliberate attention behavior.

6. (ELIMINATION BY ASPECTS) Consider a decision maker who intentionally or unintentionally focuses on a certain “aspect” of alternatives, and then refuses or ignores those alternatives that do not possess that aspect. This model is similar in spirit to [Tversky \(1972\)](#). Let $\{j, k, \ell, \dots\}$ be the set of aspects. Let ω_j represent the probability that aspect j “draws

attention to itself.” It reflects the salience and/or importance of aspect j . All alternatives without that aspect fail to receive attention. Let B_j be the set of alternatives that possess aspect j . We assume that each alternative must belong to at least one B_j with $\omega_j > 0$. If aspect j is the salient aspect, the consideration set is $B_j \cap S$ when S is the set of feasible alternatives. The total probability of T being the consideration set is the sum of ω_j such that $T = B_j \cap S$. When there is no alternative in S possessing the salient aspect, a new aspect will be drawn. Formally, the probability of T being the consideration set under S is given by

$$\mu(T|S) = \sum_{B_j \cap S = T} \frac{\omega_j}{\sum_{B_k \cap S \neq \emptyset} \omega_k}.$$

These six examples give a sample of different limited attention models of interest in economics, psychology, marketing, and many other disciplines. While these examples are quite distinct from each other, all of them are monotonic attention rules.³ As a consequence, our revealed preference characterization will be applicable to a wide range of choice rules without committing to a particular attention mechanism, which is not observable in practice and hence untestable. Furthermore, as illustrated by the examples above (and those in Section SA-2 of the Supplemental Appendix), our upcoming characterization, identification, estimation and inference results nest important previous contributions in the literature.

3 A Random Attention Model (RAM)

We are ready to introduce our random attention model based on Assumption 1. We assume the decision maker has a strict preference ordering \succ on X . To be precise, we assume the preference ordering is an *asymmetric*, *transitive* and *complete* binary relation. A binary relation \succ on a set X is (i) asymmetric, if for all $x, y \in X$, $x \succ y$ implies $y \not\succeq x$; (ii) transitive, if for all $x, y, z \in X$, $x \succ y$ and $y \succ z$ imply $x \succ z$; and (iii) complete, if for all $x \neq y \in X$, either $x \succ y$ or $y \succ x$ is true. Consequently, the decision maker always picks the maximal alternative with respect to her preference among the alternatives she pays attention to. Formally,

³To provide an example where Assumption 1 might be violated, consider a generalization of Independent Consideration of [Manzini and Mariotti \(2014\)](#). In this generalization, the degree of brand awareness for a product is not only a function of the product but also a function of the context, that is, $\gamma_S(a)$. Then the frequency of each set being the consideration set is calculated as in Independent Consideration rule.

Definition 3 (Random Attention Representation). A choice rule π has a *random attention representation* if there exists a preference ordering \succ over X and a monotonic attention rule μ (Assumption 1) such that

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

for all $a \in S$ and $S \in \mathcal{X}$. In this case, we say π is represented by (\succ, μ) . We may also say that \succ represents π , which means that there exists some monotonic attention rule μ such that (\succ, μ) represents π . We also say π is a *Random Attention Model (RAM)*.

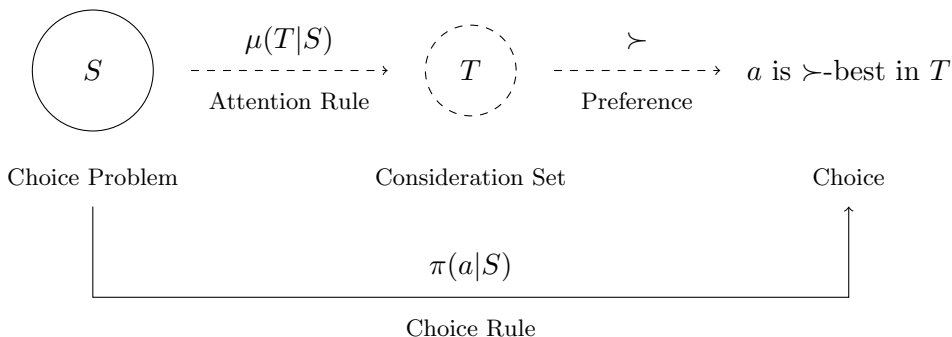


Figure 1. Illustration of a RAM. *Observable:* choice problem and choice (solid line). *Unobservable:* attention rule, consideration set and preference (dashed line).

While our framework is designed to model stochastic choices, our model captures deterministic choices as well. In classical choice theory, a decision maker chooses the best alternative according to her preferences with probability 1, hence choice is deterministic. In our framework, this case is captured by a monotone attention rule with $\mu(S|S) = 1$. Figure 1 gives a graphical representation of RAM.

In all the examples we provided above, the attention rule is independent of the underlying preference. However, RAM is flexible enough to allow for such dependence. For example, consider a decision maker who pays attention to the top K alternatives according to a ranking, which is correlated with her own preferences. For example, Google's personalized first page search results are highly correlated individual preferences. This example illustrates that different individuals might have different consideration sets even though the formation of consideration sets are common among these individuals.

We now derive some implications for our random attention model. They can be used to test

the model in terms of observed choice rules/probabilities.⁴ In the literature, there is a principle called *regularity* (see [Suppes and Luce, 1965](#)), according to which adding a new alternative should only decrease the probability of choosing one of the existing alternatives. However, empirical findings suggest otherwise. [Rieskamp, Busemeyer, and Mellers \(2006\)](#) provide a detailed review of empirical evidence on violations of regularity and alternative theories explaining these violations. Importantly, our model allows regularity violations.

The next example illustrates that adding an alternative to the feasible set can increase the likelihood that an existing alternative is selected. This cannot be accommodated in the Luce model, nor in any Random Utility Model (RUM). In RAM, the addition of an alternative changes the choice set and therefore the decision maker’s attention, which could increase the probability of an existing alternative being chosen.

Example 1 (Regularity Violation). Let $X = \{a, b, c\}$ and consider two nested choice problems $\{a, b, c\}$ and $\{a, b\}$. Imagine a decision maker with $a \succ b \succ c$ and the following monotonic attention rule μ . Each row corresponds to a different choice problem and columns represent possible consideration sets.

$\mu(T S)$	$T = \{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$	$\{a\}$	$\{b\}$	$\{c\}$
$S = \{a, b, c\}$	2/3	0	0	1/6	0	0	1/6
$\{a, b\}$		1/2			0	1/2	
$\{a, c\}$			1/2		0		1/2
$\{b, c\}$				1/2		0	1/2

Then $\pi(a|\{a, b, c\}) = 2/3 > 1/2 = \pi(a|\{a, b\}) = \pi(a|\{a, c\})$. ┘

This example shows that RAM can explain choice patterns that cannot be explained by the classical RUM. Given that the model allows regularity violations, one might think that the model has very limited empirical implications, i.e. that it is too general to have empirical content. However, it is easy to find a choice rule π that lies outside RAM with only three alternatives. We would like to provide an example where our model makes very sharp predictions.

⁴In this section, we treat the choice rule as known/observed, in order to facilitate the discussion of preference elicitation. In practice, the researcher may only observe a set of choice problems and choices thereof. We discuss the econometric implementation in Section 4. We note that even if the choice rule is not directly observed, it is identified (i.e., consistently estimable) from the choice data.

Example 2 (RAM Violation). The following choice rule π is not compatible with our random attention model as long as the decision maker chooses each alternative with positive probability from the set $\{a, b, c\}$, i.e., $\lambda_a \lambda_b \lambda_c > 0$. Each column corresponds to a different choice problem.

$\pi(\cdot S)$	$S = \{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$
a	λ_a	1	0	
b	λ_b	0		1
c	λ_c		1	0

We now illustrate that π is not a RAM. Since the choice behavior is symmetric among all binary choices, without loss of generality, assume $a \succ b \succ c$. Given that c is the worst alternative, $\{c\}$ is the only consideration set in which c can be chosen. Hence the decision maker must consider the consideration set $\{c\}$ with probability λ_c (i.e., $\mu(\{c\}|\{a, b, c\}) = \lambda_c$). Assumption 1 implies that $\mu(\{c\}|\{b, c\})$ must be greater than $\lambda_c > 0$. This yields a contradiction since $\pi(c|\{b, c\}) = 0$. In sum, given the above binary choices, our model predicts that the decision maker must choose at least one alternative with 0 probability, which is a stark prediction in probabilistic choice. \perp

One might wonder that the model makes strong assumptions due to the cyclical binary choices. that is, i.e., $\pi(a|\{a, b\}) = \pi(b|\{b, c\}) = \pi(c|\{a, c\}) = 1$. We can generate a similar prediction where the individual is perfectly rational in the binary choices, i.e., $\pi(a|\{a, b\}) = \pi(a|\{a, c\}) = \pi(b|\{b, c\}) = 1$. In this case, our model predicts that the individual cannot chose both b and c with strictly positive probability when the choice problem is $\{a, b, c\}$. Therefore, we get similar predictions. Given that RAM has non-trivial empirical content, it is natural to investigate to what extent Assumption 1 can be used to elicit (unobserved) strict preference orderings given (observed) choices of decision makers.

3.1 Revealed Preference

In general, a choice rule can have multiple RAM representations with different preference orderings and different attention rules. When multiple representations are possible, we say that a is revealed to be preferred to b if and only if a is preferred to b in all possible RAM representations. This is a very conservative approach as it makes sure we never make false claims about the preference of

the decision maker. The same approach is used in [Masatlioglu, Nakajima, and Ozbay \(2012\)](#).

Definition 4 (Revealed Preference). Let $\{(\succ_j, \mu_j)\}_{j=1, \dots, J}$ be all random attention representations of π . We say that a is *revealed to be preferred* to b if $a \succ_j b$ for all j .

We now show how revealed preference theory can still be developed successfully in our RAM framework. If all representations share the same preferences \succ (or there is a unique representation), then the revealed preference will be equal to \succ . In general, if one wants to know whether a is revealed to be preferred to b , it would appear necessary to identify all possible (\succ_j, μ_j) representations. However, this is not practical, especially when there are many alternatives. Instead, we shall now provide a handy method to obtain the revealed preference completely.

Our theoretical strategy nicely parallels that of [Masatlioglu, Nakajima, and Ozbay \(2012\)](#) (MNO) in their study of a deterministic model of inattention. MNO identifies a as revealed preferred to b whenever a is chosen in the presence of b , and removing b causes a choice reversal. This particular observation, in conjunction with the structure of attention filters, ensures that the decision maker considers b while choosing a . Here, we show that a as revealed preferred to b if removing b causes a regularity violation, that is, $\pi(a|S) > \pi(a|S - b)$. To see this, assume (\succ, μ) represents π and $\pi(a|S) > \pi(a|S - b)$. By definition π , we have

$$\begin{aligned}
\pi(a|S) &= \sum_{\substack{T \subset S, \\ a \text{ is } \succ\text{-best in } T}} \mu(T|S) \\
&= \sum_{\substack{b \in T \subset S, \\ a \text{ is } \succ\text{-best in } T}} \mu(T|S) + \sum_{\substack{b \notin T \subset S, \\ a \text{ is } \succ\text{-best in } T}} \mu(T|S) \\
&\leq \sum_{\substack{b \in T \subset S, \\ a \text{ is } \succ\text{-best in } T}} \mu(T|S) + \sum_{\substack{T \subset S - b, \\ a \text{ is } \succ\text{-best in } T}} \mu(T|S - b) \quad (\text{by monotonicity}) \\
&= \sum_{\substack{b \in T \subset S, \\ a \text{ is } \succ\text{-best in } T}} \mu(T|S) + \pi(a|S - b)
\end{aligned}$$

Hence, we have the following inequality:

$$\pi(a|S) - \pi(a|S - b) \leq \sum_{\substack{b \in T \subset S, \\ a \text{ is } \succ\text{-best in } T}} \mu(T|S)$$

Since $\pi(a|S) - \pi(a|S - b) > 0$, there must exist at least one T such that i) $b \in T$, ii) a is \succ -best in T ,

and iii) $\mu(T|S) \neq 0$. Therefore, there exists at least one occasion that the decision maker pays attention to b while choosing a (Revealed Preference). The next lemma summarizes this interesting relationship between regularity violations and revealed preferences. It simply illustrates that the existence of a regularity violation informs us about the underlying preference.

Lemma 1. Let π be a RAM. If $\pi(a|S) > \pi(a|S - b)$, then a is revealed to be preferred to b .

Lemma 1 allows us to define the following binary relation. For any distinct a and b , define:

$$aPb, \text{ if there exists } S \in \mathcal{X} \text{ including } a \text{ and } b \text{ such that } \pi(a|S) > \pi(a|S - b).$$

By Lemma 1, if aPb then a is revealed to be preferred to b . In other words, this condition is sufficient to reveal preference. In addition, since the underlying preference is transitive, we also conclude that she prefers a to c if aPb and bPc for some b , even when aPc is not directly revealed from her choices. Therefore, the transitive closure of P , denoted by P_R , must also be part of her revealed preference. One may wonder whether some revealed preference is overlooked by P_R . The following theorem, which is our first main result, shows that P_R includes all preference information given the observed choice probabilities, under only Assumption 1.

Theorem 1 (Revealed Preference). Let π be a RAM. Then a is revealed to be preferred to b if and only if aP_Rb .

Proof. The “if” part of the Theorem follows from Lemma 1. To prove the “only if” part, we show that given any preference \succ that includes P_R , there exists a monotonic attention rule μ such that (\succ, μ) represents π . The details of the construction can be found in the proof of Theorem 2. ■

We now illustrate that Theorem 1 could be very useful to understand the attraction effect phenomena. The attraction effect introduced by [Huber, Payne, and Puto \(1982\)](#) was the first evidence against the regularity condition. It refers to an inferior product’s ability to increase the attractiveness of another alternative when this inferior product is added to a choice set. In a typical attraction effect experiment, we observe $\pi(a|\{a, b, c\}) > \pi(a|\{a, b\})$. Let assume we have no information about the alternatives other than the frequency of choices. Then, by simply using observed choice, Theorem 1 informs us that the third product c is indeed an inferior alternative compared

to a ($a \succ c$). This is exactly how these alternatives are chosen in these experiments. While alternatives a and b are not comparable, c is dominated by a , but is not comparable to b . Theorem 1 informs us about the nature of products by just observing choice frequencies.

Our revealed preference result includes the one given in Masatlioglu, Nakajima, and Ozbay (2012) for non-random attention filters. In their model, a is revealed to be preferred to b if there is a choice problem such that a is chosen and b is available, but it is no longer chosen when b is removed from the choice problem. This means we have $1 = \pi(a|S) > \pi(a|S - b) = 0$. Given Theorem 1, this reveals that a is better than b . On the other hand, generalizing to non-deterministic attention rules allows for a broader class of empirical and theoretical settings to be analyzed, hence our revealed preference result (Theorem 1) is strictly richer than those obtained in previous work. For example, in a deterministic world with three alternatives, there is no data revealing the entire preference. On the other hand, we illustrate that it is possible to reveal the entire preference in RAM with only three alternatives. This discussion makes clear the connection between deterministic and probabilistic choice in terms of revealed preference.

Example 3 (Full Revelation). Consider the following stochastic choice with three alternatives:

$\pi(\cdot S)$	$S = \{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$
a	λ	$1 - \lambda_b$	λ_a	
b	$1 - \lambda$	λ_b		$1 - \lambda_c$
c	0		$1 - \lambda_a$	λ_c

If $1 - \lambda_b > \lambda > \lambda_a, \lambda_c$, then we can verify that π has a random attention representation (see Theorem 2). Now we show that in all possible representations of π , $a \succ b \succ c$ must hold. By Lemma 1, $\pi(a|\{a, b, c\}) > \pi(a|\{a, c\})$ implies that a is revealed to be preferred to b . Similarly, $\pi(b|\{a, b, c\}) > \pi(b|\{a, b\})$ implies b is revealed to be preferred to c . Hence preference is uniquely identified. ┘

3.2 A Characterization

Theorem 1 characterizes the revealed preference for our model. However, it is not applicable unless the observed choice behavior has a random attention representation, which motivates the following

question: how can we test whether a choice rule is consistent with RAM? It turns out that RAM can be simply characterized by only one behavioral postulate of choice.

Our characterization is based on an idea similar to [Houthakker \(1950\)](#). Choices reveal information about preferences. If these revelations are consistent in the sense that there is no cyclical preference revelations, the choice behavior has a RAM representation.

Theorem 2 (Characterization). A choice rule π has a random attention representation if and only if P has no cycle.

The idea of the proof is as follows. One direction of the statement follows directly from Lemma 1. For the other direction, we need to construct a preference and a monotonic attention rule representing the choice rule. Given that P has no cycle, there exists a preference relation \succ including P_R . Indeed, we illustrate that any such completion of P_R represents π by an appropriately chosen μ . The construction of μ depends on a particular completion of P_R . We then illustrate that the constructed μ satisfies Assumption 1. At the last step, we show that (\succ, μ) represents π .

Recall that Example 2 is outside of our model. Theorem 2 implies that P_R must have a cycle. Indeed, we have aPb due to the regularity violation $\pi(a|\{a, b, c\}) = \lambda_a > 0 = \pi(a|\{a, c\})$. Similarly, we have bPc by $\pi(b|\{a, b, c\}) = \lambda_b > 0 = \pi(b|\{a, b\})$ and cPa by $(\pi(c|\{a, b, c\}) = \lambda_c > 0 = \pi(c|\{b, c\}))$. Since P_R has a cycle, Example 2 must be outside of our model. Therefore, Theorem 2 provides a very simple test for RAM.

3.3 Further Comments on Revealed Preference

Theorem 1 establishes the empirical content of revealed preferences under monotonic attention only. Our resulting revealed preference results could be incomplete: it may only provide coarse welfare judgments in some cases. At one extreme, there is no preference revelation when there is no regularity violation. This is because the decision maker's behavior can be attributed fully to her preference or to her inattention (i.e., never considering anything other than her actual choice). This highlights the fact that our revealed preference definition is conservative, which guarantees no false claims in terms of revealed preference especially when there are alternative explanations for the same choice behavior.

Nevertheless, a policymaker may want to make a welfare judgment even when our revealed

preference is incomplete. To do so, one can impose additional properties that the attention rule must satisfy, without committing to a particular attention rule (i.e., without imposing parametric assumptions on the attention rule). These additional structure will improve the revealed preference and hence the predictive power of the model. More generally, we discuss several options to further the empirical content of our proposed RAM: (i) imposing additional structures on attention rule; (ii) changing the definition of revealed preference; (iii) committing to a particular attention rule; and/or (iv) looking for additional data other than choice data. In Section 5.1 we illustrate (i) by introducing a measure of full attention at binary choice problems, and discuss in detail how such nonparametric restriction considerably improves the empirical content of our RAM. In the remaining of this section, we briefly discuss the other approaches for completeness.

Changing the Definition of Revealed Preference. While the policymaker is aware of the limited attention issue, she might be willing to attribute choice behavior to preference as much as possible. In other words, the policymaker can choose a pair (\succ, μ) of preference and a monotonic attention rule explaining the data, where μ induces the highest possible consideration.⁵ For example, consider a decision maker satisfying Luce’s IIA axiom, that is $\frac{\pi(a|S)}{\pi(b|S)} = \frac{\pi(a|T)}{\pi(b|T)}$ for all S and T . It is known that this property implies that there is a numerical scale $w(\cdot)$ such that for every $a, b \in S$, the probability of choosing a from S can be expressed as $\frac{w(a)}{\sum_{b \in S} w(b)}$. Psychologically, $w(a)$ represents a preference strength associated with the alternative a . Now consider a RAM representation of the same choice behavior. One can show that the preference in the pair (\succ, μ) providing the highest possible consideration is aligned with the numerical scale w . That is, $w(a) > w(b)$ if and only if $a \succ b$.

Committing a Particular Attention Rule. Another way is committing to a particular attention rule. For example, if the attention rule belongs to the class of independent consideration attention rules, then unique identification of preferences is possible (Horan, 2018). Similarly, under the logit attention rule, the revealed preference is unique up to the two least preferred alternatives (Brady and Rehbeck, 2016). However, in both cases, we commit to a particular attention rule. We must believe that this particular attention rule is the one that the individual utilizes. If it is not the case, the proposed welfare judgment will be misleading. This discussion implies that one needs

⁵Here, one can consider $\frac{\sum_S \sum_T \mu(T|S) f(|T|)}{\sum_S f(|S|)}$, where f is an increasing function, as a measure of the total consideration. The policymaker can choose a representation (\succ, μ) to maximize this measure.

to know *how* attention works in order to reveal preferences uniquely.

Additional Data. In some settings, due to the technological advancement (such as eye-tracking, functional magnetic resonance imaging, and the tracking system in Internet commerce), partial information about consideration sets are available. If the policymaker believes that these sources are trustworthy, he can utilize them to obtain additional information about preferences in addition to what Theorem 1 provides.

4 Econometric Methods

Theorem 1 shows that if the choice probability π is a RAM then preference revelation is possible. Theorem 2 gives a falsification result, based on which a specification test can be designed. The challenge for econometric implementation, however, is that our main assumption, monotonic attention, is imposed on the attention rule, which is not identified from a typical choice data, and has a much higher dimension than the identified (consistently estimable) choice rule. To circumvent this difficulty, we rely on Corollary A.1, which states that if π has a random attention representation (\succ, μ) , then there exists a unique triangular monotonic attention rule⁶ μ' such that (\succ, μ') is also a representation of π . This latter result turns out to be useful for our proposed identification, estimation and inference methods, as it allows us to construct, for each given preference ordering, a mapping from the identified choice rule to a triangular attention rule, for which we can test whether Assumption 1 holds. This test turns out to be a test on moment inequalities.

As we described in the Introduction, RAM is best suited for eliciting information about the preference ordering of a single decision-making unit when her choices are observed repeatedly. Obvious examples include scanner data from grocery stores or web advertising on digital platforms, where multiple choices for each individual are tracked and recorded, and in each instance potentially different options are offered to each of them (e.g., grocery store adjusts product varieties and arrangements regularly, or digital platforms manipulate/experiment the options offered). The econometric methods described in this section are applicable to settings where repeated choices of a single decision maker are observed, provided there is variation in the options offered, in which case our methods allow to (partially) identify, estimate, and test for her preference ordering. In

⁶See Definition A.1 in the Appendix.

addition, when data on choices of multiple decision-making units are considered, our methods can be justified by assuming that all decision makers have the same preference ordering, possibly after conditioning on observed covariates. In this latter case, our methods do allow for unobserved heterogeneous attention rules among the decision makers.

4.1 Nonparametric Identification

We first define the set of partially identified preferences, which mirrors Definition 3, with the only difference that now we fix the choice rule to be that identified (consistently estimable) from data. More precisely, let π be the underlying choice rule/data generating process, then a preference \succ is compatible with π , denoted by $\succ \in \Theta_\pi$, if there exists some monotonic attention rule μ such that (π, \succ, μ) is a RAM.

When π is known, it is possible to employ Theorem 1 directly to construct Θ_π . For example, consider the specific preference ordering $a \succ b$, which can be checked by the following procedure. First, check whether $\pi(b|S) \leq \pi(b|S - a)$ is violated for some S . If so, then we know the preference ordering is not compatible with RAM and hence does not belong to Θ_π (Lemma 1). On the other hand, if the preference ordering is not rejected in the first step, we need to check along “longer chains” (Theorem 1). That is, whether $\pi(b|S) \leq \pi(b|S - c)$ and $\pi(c|T) \leq \pi(c|T - a)$ are simultaneously violated for some S, T and c . If so, the preference ordering is rejected (i.e., incompatible with RAM), while if not then a chain of length three needs to be considered. This process goes on for longer chains until either at some step we are able to reject the preference ordering, or all possibilities are exhausted. In practice, additional comparisons are needed since it is rarely the case that only a specific pair of alternatives is of interest. This algorithm, albeit feasible, can be hard to implement in practice, even when the choice probabilities are known. The fact that π has to be estimated makes the problem even more complicated, since it becomes a sequential multiple hypothesis testing problem.

Another possibility is to employ the J-test approach, which stems from the idea that, given the choice rule, compatibility of a preference is equivalent to the existence of an attention rule satisfying monotonicity. To implement the J-test, one fixes the choice rule (identified/estimated from the data) and the preference ordering (the null hypothesis to be tested), and search the space of all monotonic attention rules and check if Definition 3 applies. The J-test procedure can be quite

computationally demanding, due to the fact that the space of attention rules has high dimension. We discuss the J-test approach in more details in a later section and how it is related to our proposed procedure.

One of the main purposes of this section is to provide an equivalent form of identification, which (i) is simple to implement, and (ii) remains statistically valid even when applied using estimated choice rules. For ease of exposition, we rewrite the choice rule π as a long vector $\boldsymbol{\pi}$, whose elements are simply the probability of each alternative $a \in X$ being chosen from a choice problem $S \in \mathcal{X}$. The specific ordering of elements in $\boldsymbol{\pi}$ does not matter, provided that rows and columns of other matrices (for example, the variance-covariance matrix, the constraint matrix, etc., to be introduced below) have been ordered accordingly. See also Example 4 for a concrete illustration.

Theorem 3 (Nonparametric Identification). Given any preference \succ , there exists a unique matrix \mathbf{R}_\succ such that $\succ \in \Theta_\pi$ if and only if $\mathbf{R}_\succ \boldsymbol{\pi} \leq \mathbf{0}$.

Proof. Recall that (π, \succ) has a RAM representation if and only if there exists a monotonic attention rule $\tilde{\mu}$ such that π is induced by $\tilde{\mu}$ and \succ . In addition, Corollary A.1 in the Appendix states that it is without loss of generality to focus on triangular attention rules (Definition A.1 in the Appendix), which implies $\succ \in \Theta_\pi$ if and only if there exists a monotonic triangular attention rule μ which induces π . The constraint matrix \mathbf{R}_\succ is constructed to take the product form $\mathbf{R}\mathbf{C}_\succ$, where the first matrix, \mathbf{R} , consists of constraints on the attention rules, and the second matrix, \mathbf{C}_\succ , maps the choice rule back to a triangular attention rule.

First consider \mathbf{R} . The only restrictions imposed on attention rules are from the monotonicity assumption (Assumption 1). Again, we represent a generic attention rule μ as a long vector $\boldsymbol{\mu}$. Then \mathbf{R} is easily constructed as a collection of inequalities taking the form $\mu(T|S) - \mu(T|S-a) \leq 0$, for all $a \in S - T$. We note \mathbf{R} does not depend on any preference.

Next, we consider \mathbf{C}_\succ . Using Definition A.1 in the Appendix, and given the choice rule π and some preference \succ , the only possible triangular attention rule that can be constructed is

$$\mu(T|S) = \sum_{k: a_{k,\succ} \in S} \mathbf{1}(T = S \cap L_{k,\succ}) \cdot \pi(a_{k,\succ}|S),$$

where $\{L_{k,\succ} : 1 \leq k \leq K\}$ are the lower contour sets corresponding to the preference ordering \succ .

The above defined the mapping \mathbf{C}_\succ , which represents the triangular attention rule as a linear combination of the choice probabilities. This mapping depends on the preference/hypothesis because the triangular attention rule depends on the preference/hypothesis.

Along the construction, both \mathbf{R} and \mathbf{C}_\succ are unique, hence showing \mathbf{R}_\succ is uniquely determined by the preference \succ . ■

This theorem states that in order to decide whether a preference \succ is compatible with the (identifiable) choice rule π , it suffices to check a collection of inequality constraints. In particular, it is no longer necessary to consider the sequential and multiple testing problems mentioned earlier, or numerically searching in the high dimensional space of attention rules. Moreover, as we discuss below, given the large econometric literature on moment inequality testing, many techniques can be adapted when Theorem 3 is applied to estimated choice rules.

Let \mathbf{R} and \mathbf{C}_\succ be as in the proof of Theorem 3. We combine these matrices and illustrate the form of the final constraint matrix \mathbf{R}_\succ . Fix some preference \succ and let μ be the triangular attention rule shown in the previous proof. Monotonicity (i.e., \mathbf{R}) requires that $\mu(T|S) - \mu(T|S - a_k) \leq 0$; this is trivially satisfied if T is not a lower contour set in S since $\mu(T|S) = 0$. Now assume that $T = S \cap L_{\ell, \succ} \neq \emptyset$ is a lower contour set in S (without loss of generality let $a_{\ell, \succ} \in S$); then it will also be a lower contour set in $S - a_k$. Moreover, $a_k \in S - T$ implies that $a_k \succ a_{\ell, \succ}$, and hence the monotonicity assumption translates into $\pi(a_{\ell, \succ}|S) - \pi(a_{\ell, \succ}|S - a_k) \leq 0$, for all $a_k \succ a_{\ell, \succ}$ in S . We have the following algorithm constructing \mathbf{R}_\succ *directly*, albeit the idea does come from the two auxiliary matrices \mathbf{R} and \mathbf{C}_\succ .

Algorithm 1 Construction of \mathbf{R}_\succ .

Require: Set a preference \succ .

$\mathbf{R}_\succ \leftarrow$ empty matrix

for S in \mathcal{X} **do**

for a in S **do**

for $b \prec a$ in S **do**

$\mathbf{R}_\succ \leftarrow$ add row corresponding to $\pi(b|S) - \pi(b|S - a) \leq 0$.

end for

end for

end for

The only input needed in the previous algorithm is the preference \succ , which we are interested in testing against. Each row of \mathbf{R}_\succ consists of one “+1”, one “-1”, and 0 otherwise. The constraint

matrix \mathbf{R}_{\succ} is nonrandom and does not depend on the estimated choice probabilities, but rather determined by the collection of (fixed, known to the researcher) restrictions on the estimable choice probabilities. Also note that as long as the matrix \mathbf{R}_{\succ} has been constructed for one preference \succ , constraint matrices for other preference orderings can be obtained by column permutations of \mathbf{R}_{\succ} . This is particular useful and saves computation if there are multiple hypotheses to be tested, as the above algorithm only needs to be implemented once. Next we compute the number of constraints (i.e. rows) in \mathbf{R}_{\succ} for the complete data case (i.e., when all choice problems are observed):

$$\#\text{row}(\mathbf{R}_{\succ}) = \sum_{S \in \mathcal{X}} \sum_{a, b \in S} \mathbb{1}(b \prec a) = \sum_{S \in \mathcal{X}, |S| \geq 2} \binom{|S|}{2} = \sum_{k=2}^K \binom{K}{k} \binom{k}{2},$$

where $K = |X|$ is the number of alternatives in the grand set X . Not surprisingly, the number of constraints increases very fast with the size of the grand set.

Finally, we illustrate in the next example that, in simple examples, the constraint matrix \mathbf{R}_{\succ} can be constructed intuitively.

Example 4 (\mathbf{R}_{\succ} with Three Alternatives). Assume there are three alternatives, a , b and c in X , then the choice rule is represented by a vector in \mathbb{R}^9 :

$$\boldsymbol{\pi} = \left[\underbrace{\pi(\cdot|\{a, b, c\})}_{\in \mathbb{R}^3}, \underbrace{\pi(\cdot|\{a, b\})}_{\in \mathbb{R}^2}, \underbrace{\pi(\cdot|\{a, c\})}_{\in \mathbb{R}^2}, \underbrace{\pi(\cdot|\{b, c\})}_{\in \mathbb{R}^2} \right]',$$

where trivial cases such as $\pi(a|\{b, c\}) = 0$ and $\pi(b|\{b\}) = 1$ are ignored. Now consider the preference $b \succ a \succ c$. From Lemma 1, we can reject $b \succ a$ if $\pi(a|\{a, b, c\}) > \pi(a|\{a, c\})$. Therefore, we need the reverse inequality in $\mathbf{R}_{b \succ a \succ c}$, given by a row:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \end{bmatrix}.$$

Similarly, we will be able to reject $a \succ c$ if $\pi(c|\{a, b, c\}) > \pi(c|\{b, c\})$, which implies the following row in the matrix $\mathbf{R}_{b \succ a \succ c}$:

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

The row corresponding to $b \succ c$ is

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix}.$$

Therefore, for this simple problem with three alternatives, we have the following constraint matrix:

$$\mathbf{R}_{b \succ a \succ c} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix}.$$

Note that for problems with more than three alternatives, the above reasoning does not work if implemented naïvely. Consider the case $X = \{a, b, c, d\}$. Then $b \succ a$ can be rejected by $\pi(a|\{a, b, c, d\}) > \pi(a|\{a, c, d\})$, $\pi(a|\{a, b, d\}) > \pi(a|\{a, d\})$ or $\pi(a|\{a, b, c\}) > \pi(a|\{a, c\})$, which correspond to three rows in the constraint matrix. Another complication is that a relation such as $\pi(c|\{a, b, c, d\}) > \pi(c|\{c, d\})$ indicates c is preferred to either a or b , hence rejecting the preference ordering $b \succ a \succ c \succ d$. Later we will provide an algorithm that constructs \mathbf{R}_\succ systematically.

Again we emphasize that, to construct \mathbf{R}_\succ , one does not need to know the numerical value of the choice rule π . The matrix \mathbf{R}_\succ contains restrictions jointly imposed by the monotonicity assumption and the preference \succ that is to be tested. ┘

4.2 Hypothesis Testing

Given the identification result in Theorem 3, we can replace the unobserved but identifiable choice rule with its estimate to conduct estimation and inference of the (partially identifiable) preferences. We can also conduct specification testing by evaluating whether the identified set Θ_π is empty. To proceed, we assume the following data structure.

Assumption 2 (DGP). The data is a random sample of choice problems Y_i and corresponding choices y_i , $\{(y_i, Y_i) : y_i \in Y_i, 1 \leq i \leq N\}$, generated by the underlying choice rule $\mathbb{P}[y_i = a|Y_i = S] = \pi(a|S)$, with $\mathbb{P}[Y_i = S] \geq \underline{p} > 0$ for all $S \in \mathcal{X}$.

We only assume the data is generated from some choice rule π . We allow for the possibility that it is not RAM, since our identification result permits falsifying the RAM representation: π has

a RAM representation if and only if Θ_π is not empty according to Theorem 3. In addition, we only assume that the choice problem Y_i and the corresponding selection $y_i \in Y_i$ are observed for each unit, while the underlying (possibly random) consideration set for the decision maker remains unobserved (i.e., the set T in Definition 2 and Figure 1). For simplicity, we discuss the case of “complete data” where choices for all alternatives are observed, but in Section SA.3 and SA.4 of the Supplemental Appendix we extend our work to the case of incomplete data.

The estimated choice rule is denoted by $\hat{\pi}$,

$$\hat{\pi}(a|S) = \frac{\sum_{1 \leq i \leq N} \mathbf{1}(y_i = a, Y_i = S)}{\sum_{1 \leq i \leq N} \mathbf{1}(Y_i = S)}, \quad a \in S, \quad S \in \mathcal{X}.$$

For convenience, we represent $\hat{\pi}(\cdot|S)$ by the vector $\hat{\pi}_S$, and its population counterpart by π_S . The choice rules are stacked into a long vector, denoted by $\hat{\pi}$, with the population counterpart π .

We consider Studentized test statistics, and hence we introduce some additional notation. Let $\sigma_{\pi, \succ}$ be the standard deviation of $\mathbf{R}_\succ \hat{\pi}$, and $\hat{\sigma}_\succ$ be its plug-in estimate. That is,

$$\sigma_{\pi, \succ} = \sqrt{\text{diag}(\mathbf{R}_\succ \boldsymbol{\Omega}_\pi \mathbf{R}_\succ')} \quad \text{and} \quad \hat{\sigma}_\succ = \sqrt{\text{diag}(\mathbf{R}_\succ \hat{\boldsymbol{\Omega}} \mathbf{R}_\succ')},$$

where $\text{diag}(\cdot)$ denotes the operator that extracts the diagonal elements of a square matrix or constructs a diagonal matrix when applied to a vector. Here $\boldsymbol{\Omega}_\pi$ is block diagonal, with blocks given by $\frac{1}{\mathbb{P}[Y_i=S]} \boldsymbol{\Omega}_{\pi, S}$, and $\boldsymbol{\Omega}_{\pi, S} = \text{diag}(\pi_S) - \pi_S \pi_S'$. The estimator $\hat{\boldsymbol{\Omega}}$ is simply constructed by plugging in the estimated choice rule.

Consider the null hypothesis $H_0 : \succ \in \Theta_\pi$. This null hypothesis is useful if the researcher believes a certain preference represents the underlying data generating process. It also serves as the basis for constructing confidence sets or for ranking preferences according to their (im)plausibility in repeated sampling (for example, via employing associated p-values). Given a specific preference, the test statistic is constructed as the maximum of the Studentized, restricted sample choice probabilities:

$$\mathcal{T}(\succ) = \sqrt{N} \cdot \max \left\{ (\mathbf{R}_\succ \hat{\pi}) \oslash \hat{\sigma}_\succ, 0 \right\},$$

where \oslash denotes elementwise division (i.e, Hadamard division) for conformable matrices. The test statistic is the largest element of the vector $\sqrt{N}(\mathbf{R}_\succ \hat{\pi}) \oslash \hat{\sigma}_\succ$ if it is positive, or zero otherwise.

The reasoning behind such construction is straightforward: If the preference is compatible with the underlying choice rule, then in the population we have $\mathbf{R}_{\succ} \boldsymbol{\pi} \leq \mathbf{0}$, meaning that the test statistic, $\mathcal{T}(\succ)$, should not be “too large.”

Other test statistics have been proposed for testing moment inequalities, and usually the specific choice depends on the context. When many moment inequalities can be potentially violated simultaneously, it is usually preferred to use a statistic based on truncated Euclidean norm. In our problem, however, we expect only a few moment inequalities to be violated, and therefore we prefer to employ $\mathcal{T}(\succ)$. Having said this, the large sample approximation results given in Theorem 4 can be adapted to handle other test statistics commonly encountered in the literature on moment inequalities.

The null hypothesis is rejected whenever the test statistic is “too large,” or more precisely, when it exceeds a critical value, which is chosen to guarantee uniform size control in large samples. We describe how this critical value leading to uniformly valid testing procedures is constructed based on simulating from multivariate normal distributions. Our construction employs the Generalized Moment Selection (GMS) approach of [Andrews and Soares \(2010\)](#); see also [Canay \(2010\)](#) and [Bugni \(2016\)](#) for closely related methods. The literature on moment inequalities testing includes several alternative approaches, some of which we discuss briefly in Appendix SA.4.

To illustrate the intuition behind the construction, first rewrite the test statistic $\mathcal{T}(\succ)$ as the following:

$$\mathcal{T}(\succ) = \max \left\{ (\mathbf{R}_{\succ} \sqrt{N}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) + \sqrt{N} \mathbf{R}_{\succ} \boldsymbol{\pi}) \odot \hat{\boldsymbol{\sigma}}_{\succ}, 0 \right\}.$$

By the central limit theorem, the first component $\sqrt{N}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})$ is approximately distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\pi}})$. The second component, $\mathbf{R}_{\succ} \boldsymbol{\pi}$, although unknown, is bounded above by zero under the null hypothesis. Motivated by these observations, we approximate the distribution of $\mathcal{T}(\succ)$ by simulation as follows:

$$\mathcal{T}^*(\succ) = \sqrt{N} \cdot \max \left\{ (\mathbf{R}_{\succ} \mathbf{z}^*) \odot \hat{\boldsymbol{\sigma}}_{\succ} + \psi_N(\mathbf{R}_{\succ} \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\sigma}}_{\succ}), 0 \right\}.$$

Here \mathbf{z}^* is a random vector simulated from the distribution $\mathcal{N}(\mathbf{0}, \hat{\boldsymbol{\Omega}}/N)$, and $\sqrt{N} \psi_N(\mathbf{R}_{\succ} \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\sigma}}_{\succ})$ is used to replace the unknown moment conditions $(\sqrt{N} \mathbf{R}_{\succ} \boldsymbol{\pi}) \odot \hat{\boldsymbol{\sigma}}_{\succ}$. Several choices of ψ_N have

been proposed. One extreme choice is $\psi_N(\cdot) = 0$, so that the upper bound $\mathbf{0}$ is used to replace the unknown $\mathbf{R}_{\succ}\boldsymbol{\pi}$. Such a choice also delivers uniformly valid inference in large samples, and is usually referred to as “critical value based on the least favorable model.” However, for practical purposes it is better to be less conservative. In our implementation we employ

$$\psi_N(\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\sigma}}_{\succ}) = \frac{1}{\kappa_N} \left(\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}} \odot \hat{\boldsymbol{\sigma}}_{\succ} \right)_-,$$

where $(\mathbf{a})_- = \mathbf{a} \odot \mathbf{1}(\mathbf{a} \leq 0)$, with \odot denoting the Hadamard product, the indicator function $\mathbf{1}(\cdot)$ operating element-wise on the vector \mathbf{a} , and κ_N diverging slowly. That is, the function $\psi_N(\cdot)$ retains the non-positive elements of $(\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}} \odot \hat{\boldsymbol{\sigma}}_{\succ})/\kappa_N$, since under the null hypothesis all moment conditions are nonpositive. We use $\kappa_N = \sqrt{\ln N}$, which turns out to work well in the simulations described in Section 6. For other choices of $\psi_N(\cdot)$, see [Andrews and Soares \(2010\)](#).

In practice, M simulations are conducted to obtain the simulated statistics $\{\mathcal{T}_m^*(\succ) : 1 \leq m \leq M\}$. Then, given some $\alpha \in (0, 1)$, the critical value is constructed as

$$c_\alpha(\succ) = \inf \left\{ t : \frac{1}{M} \sum_{m=1}^M \mathbf{1}(\mathcal{T}_m^*(\succ) \leq t) \geq 1 - \alpha \right\},$$

and the null hypothesis $H_0 : \succ \in \Theta_\pi$ is rejected if and only if $\mathcal{T}(\succ) > c_\alpha(\succ)$. Alternatively, one can compute the p-value as

$$\text{pVal}(\succ) = \frac{1}{M} \sum_{m=1}^M \mathbf{1}(\mathcal{T}_m^*(\succ) > \mathcal{T}(\succ)).$$

To justify the proposed critical values, it is important to address uniformity issues because in finite samples the moment inequalities could be close to binding. A testing procedure is (asymptotically) uniform among a class of data generating processes, if the asymptotic size does not exceed the nominal level across this class. The following theorem shows that conducting inference using the critical values above is uniformly valid.

Theorem 4 (Uniformly Valid Testing). Assume Assumption 2 holds. Let Π be a class of choice rules, and \succ a preference, such that: (i) for each $\pi \in \Pi$, $\succ \in \Theta_\pi$; and (ii) $\inf_{\pi \in \Pi} \min(\boldsymbol{\sigma}_{\pi, \succ}) > 0$.

Then,

$$\limsup_{N \rightarrow \infty} \sup_{\pi \in \Pi} \mathbb{P}[\mathcal{T}(\succ) > c_\alpha(\succ)] \leq \alpha.$$

The proof is postponed to Appendix A.2. The only requirement is that each moment condition is nondegenerate so that the normalized statistics are well-defined in large samples, but no restrictions on correlations among moment conditions are imposed.

4.3 Extensions and Discussion

We discuss some extensions based on Theorem 4, including how to construct uniformly valid confidence sets via test inversion and how to conduct uniformly valid specification testing, both based on testing individual preferences. We also discuss the connection between these methods and alternative methods based on J-tests (i.e., optimization over the high-dimensional space of attention rules).

Confidence Set

Given the uniformly valid hypothesis testing procedure already developed in Theorem 4, we can obtain a uniformly valid confidence set for the (partially) identified preferences by test inversion:

$$\mathcal{C}(\alpha) = \left\{ \succ : \mathcal{T}(\succ) \leq c_\alpha(\succ) \right\}.$$

The resulting confidence set $\mathcal{C}(\alpha)$ exhibits an asymptotic uniform coverage rate of at least $1 - \alpha$:

$$\liminf_{N \rightarrow \infty} \inf_{\pi \in \Pi} \min_{\succ \in \Theta_\pi} \mathbb{P}[\succ \in \mathcal{C}(\alpha)] \geq 1 - \alpha.$$

This inference method offers a uniformly valid confidence set for each member of the partially identified set with prespecified coverage probability, which is a popular approach in the partial identification literature (Imbens and Manski, 2004).

Testing Model Compatibility: $H_0 : \mathcal{P} \cap \Theta_\pi \neq \emptyset$

Given a collection of preferences, an empirically relevant question is whether *any* of them is compatible with the data generating process—a basic model specification question. That is, the question

is whether the null hypothesis $H_0 : \mathcal{P} \cap \Theta_\pi \neq \emptyset$ should be rejected. If the null hypothesis is rejected, then certain features shared by the collection of preferences is incompatible with the underlying decision theory (up to Type I error). See [Bugni, Canay, and Shi \(2015\)](#), [Kaido, Molinari, and Stoye \(2019\)](#) and references therein for further discussion of this idea and related methods.

For a concrete example, consider the question that whether $a \succ b$ is compatible with the data generating process. As long as there are more than 2 alternatives in the grand set, a question like this can be accommodated by setting $\mathcal{P} = \{\succ : a \succ b\}$. Rejection of this null hypothesis provides evidence in favor of b being preferred to a , (up to Type I error). Of course with more preferences included in the collection, it becomes more difficult to reject the null hypothesis.

The test is based on whether the confidence set intersects with \mathcal{P} :

$$H_0 \text{ is rejected if and only if } \mathcal{C}(\alpha) \cap \mathcal{P} = \emptyset.$$

We note that, since $\mathcal{C}(\alpha)$ covers elements in the identified set asymptotically and uniformly with probability $1 - \alpha$, the above testing procedure will have uniform size control. Indeed, if $\mathcal{P} \cap \Theta_\pi \neq \emptyset$, there exists some $\succ \in \mathcal{P} \cap \Theta_\pi$, which will be included in $\mathcal{C}(\alpha)$ with at least $1 - \alpha$ probability asymptotically.

One important application of this idea is to set \mathcal{P} as the collection of all possible preferences, which leads to a specification testing. Then, the null hypothesis becomes $H_0 : \Theta_\pi \neq \emptyset$, and is rejected based on the following rule:

$$H_0 \text{ is rejected if and only if } \mathcal{C}(\alpha) = \emptyset.$$

Rejection in this case implies that at least one of the underlying assumptions is violated, and the data generating process cannot be represented by a RAM (up to Type I error).

Testing Against a Collection of Preferences: $H_0 : \mathcal{P} \subset \Theta_\pi$

We consider this testing problem because it is a natural generalization, and because it can be easily accommodated by our methodological results. Despite being more general, this approach may have limited use in practice: even if the null hypothesis is rejected, it is unclear how such information

should be incorporated into data analysis and decision theoretic modeling. Further, it may suffer from low power in finite samples because of testing against many preferences, especially when only one or two preferences in the collection are incompatible with the underlying choice rule.

Let \mathcal{P} be a collection of preferences; then the test statistic is constructed as

$$\mathcal{T}(\mathcal{P}_V) = \max_{\succ \in \mathcal{P}} \mathcal{T}(\succ) = \max_{\succ \in \mathcal{P}} \left[\sqrt{N} \cdot \max \left\{ (\mathbf{R}_\succ \hat{\boldsymbol{\pi}}) \odot \hat{\boldsymbol{\sigma}}_\succ, 0 \right\} \right].$$

We use \mathcal{P}_V to emphasize that the null hypothesis is “all preferences in \mathcal{P} are compatible with the underlying choice rule.” The critical value is simply obtained as the $1 - \alpha$ quantile of the corresponding simulated statistic:

$$c_\alpha(\mathcal{P}_V) = \inf \left\{ t : \frac{1}{M} \sum_{m=1}^M \mathbf{1} \left(\max_{\succ \in \mathcal{P}} \mathcal{T}_m^*(\succ) \leq t \right) \geq 1 - \alpha \right\}.$$

The validity of the above critical value follows from Theorem 4.

5 Applications

Our identification and inference results so far are obtained using RAM only, that is, all empirical content of our revealed preference theory comes from the weak nonparametric Assumption 1. As mentioned before, our model provides a minimum benchmark for preference revelation, which sometimes may not deliver enough empirical content. However, it is easy to incorporate additional (nonparametric) assumptions in specific settings. In this section, we first illustrate one such possibility, where additional restrictions on the attentional rule are imposed for binary choice problems. This will improve our identification and inference results considerably. We then also consider random attention filters, which are one of the motivating examples of monotonic attention rules, and show that in this case there is no identification improvement relative to the baseline RAM.

5.1 Attentive at Binaries

To motivate our approach, a policy maker may want to conclude that a is revealed to be preferred to b if the decision maker chooses a over b “frequently enough” in binary choice problems. “Frequently Enough” is measured by a constant $\phi \geq 1/2$. For example, when $\phi = 2/3$, it means that choosing

a twice more often than choosing b implies a is better than b . ϕ represents how cautious the policy maker is. Denote by

$$a \mathbf{P}^\phi b \quad \text{if and only if} \quad \pi(a|\{a, b\}) > \phi.$$

To justify \mathbf{P}^ϕ as preference revelation, the policy maker inherently assumes that the decision maker pays attention to the entire set “frequently enough.” This is captured by the following assumption on the attention rule.

Assumption 3 (ϕ -Attentive at Binaries). For all $a, b \in X$,

$$\mu(\{a, b\}|\{a, b\}) \geq \frac{1-\phi}{\phi} \max \{ \mu(\{a\}|\{a, b\}), \mu(\{b\}|\{a, b\}) \}.$$

The quantity $\frac{1-\phi}{\phi}$ is a measure of full attention at binaries. When $\frac{1-\phi}{\phi} = 0$ (or $\phi = 1$), there is no constraint on $\mu(\{a, b\}|\{a, b\})$. In this case, it is possible that the decision maker only considers singleton consideration sets. When $\frac{1-\phi}{\phi}$ gets larger, the probability of being fully attentive is strictly positive. We now illustrate that, under Assumption 3, if $\pi(a|\{a, b\}) > \phi$ then a is revealed to be preferred to b . Let (\succ, μ) be a RAM representation of π where μ satisfies Assumption 3. First, assumption 3 necessitates that $\mu(\{a\}|\{a, b\})$ cannot be higher than ϕ . (To see this, assume $\mu(\{a\}|\{a, b\}) > \phi$. By Assumption 3, we must have $\mu(\{a, b\}|\{a, b\}) > 1 - \phi$, which is a contradiction.) Then, $\pi(a|\{a, b\}) > \phi$ indicates that a is chosen over b whenever the decision maker pays attention to $\{a, b\}$ (revealed preference). Therefore, $a \succ b$.

Example 5 (Preference Revelation Without Regularity Violation). To illustrate the extra identification power of Assumption 3, consider the following stochastic choice with three alternatives and take $\phi = 1/2$.

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

Note that π satisfies the regularity condition, meaning that there is no preference revelation if only monotonicity (Assumption 1) is imposed on the attention rule. That is, $\mathbf{P} = \emptyset$. On the other hand, by utilizing Assumption 3, we can infer the preference completely. Since $\pi(a|\{a, b\}) > 1/2$

and $\pi(b|\{b, c\}) > 1/2$, we must have $aP^\phi b$ and $bP^\phi c$. Notice that $\pi(a|\{a, c\}) = 1/2$, hence we cannot directly deduce $aP^\phi c$. Since the underlying preference is transitive, we can conclude that the decision maker prefers a to c as $aP^\phi b$ and $bP^\phi c$, even when $aP^\phi c$ is not directly revealed from her choices. Therefore, the transitive closure of P^ϕ , denoted by $P^\phi_{\mathbf{R}}$, must also be part of the revealed preference. In this example, note that the same conclusion can be drawn as long as the policy maker assumes $\phi < 2/3$. \lrcorner

To accommodate the revealed preference defined in the original model (i.e., to combine Assumption 1 and 3), we now define the following binary relation:

$$a(P^\phi \cup P)b \quad \text{if and only if} \\ \text{either (i) for some } S \in \mathcal{S}, \pi(a|S) > \pi(a|S - b), \text{ or (ii) } \pi(a|\{a, b\}) > \phi.$$

$P^\phi \cup P$ includes our original binary relation P , defined under the monotonic attention restriction (Assumption 1), as well as P^ϕ , characterized by the new attentive at binary assumption. Therefore, we can infer more.

The next theorem shows that acyclicity of $P^\phi \cup P$, or its transitive closure $(P^\phi \cup P)_{\mathbf{R}}$, provides a simple characterization of the model we consider in this subsection.

Theorem 5 (Characterization). A choice rule π has a random attention representation (\succ, μ) where μ satisfies Assumption 1 and 3 if and only if $P^\phi \cup P$ has no cycle.

For $\phi < 1$, the model characterized by Theorem 5 has a higher predictive power (i.e., empirical content) compared to the model characterized by Theorem 2. Hence the model will fail to retain some of its explanatory power. For example, Example 3 with $\lambda_a, \lambda_b, \lambda_c < 1 - \phi$ is outside of the model given here.

Theorem 5 illustrates that our framework is flexible enough to reveal preferences almost uniquely while it allows regularity violations.

Remark 1 (Acyclic Stochastic Transitivity). We would like to highlight a close connection between acyclicity of $P^\phi \cup P$ and the acyclic stochastic transitivity (AST) introduced by Fishburn (1973).

The model characterized by Theorem 5 satisfies a weaker version of AST:

$$\pi(a_1|\{a_1, a_2\}) > \phi, \dots, \pi(a_{k-1}|\{a_{k-1}, a_k\}) > \phi \text{ imply } \pi(a_1|\{a_1, a_k\}) \leq \phi.$$

We call this condition ϕ -acyclic stochastic transitivity (ϕ -AST). Note that $\frac{1}{2}$ -AST is equivalent to AST. If we only consider binary choice probabilities, acyclicity of $\mathbb{P}^\phi \cup \mathbb{P}$ becomes equivalent to ϕ -AST. Otherwise, our condition is stronger than ϕ -AST. \perp

Now we discuss the econometric implementation. Recall from Section 4 that, to test if a specific preference ordering is compatible with the observed (identifiable) choice rule and the monotonicity assumption, we first construct a triangular attention rule following Algorithm 1 and then test whether the triangular attention rule satisfies Assumption 1. This is formally justified in the proof of Theorem 3. (See also Corollary A.1 in the Appendix.)

This line of reasoning can be naturally extended to accommodate Assumption 3 in our econometric implementation. Again, the researcher constructs a triangular attention rule based on a specific preference ordering and the identifiable choice rule. She then tests whether the triangular attention rule satisfies Assumption 1 and 3. This is formally justified in the proof of Theorem 5. For testing, only minor changes have to be made when constructing the matrix \mathbf{R}_{\succ} . The precise construction is given in Algorithm 2.

Algorithm 2 Construction of \mathbf{R}_{\succ} .

Require: Set a preference \succ .

$\mathbf{R}_{\succ} \leftarrow$ empty matrix

for S in \mathcal{X} **do**

for a in S **do**

for $b \prec a$ in S **do**

$\mathbf{R}_{\succ} \leftarrow$ add row corresponding to $\pi(b|S) - \pi(b|S - a) \leq 0$.

end for

end for

if $S = \{a, b\}$ is binary and $b \prec a$ **then**

$\mathbf{R}_{\succ} \leftarrow$ add row corresponding to $\frac{1-\phi}{\phi}\pi(b|S) - \pi(a|S) \leq 0$

end if

end for

We can revisit Example 4 to illustrate what additional (identifying) restrictions are imposed by Assumption 3.

Example 6 (Example 4, Continued). Recall that there are three alternatives, a , b and c in X , and the choice rule is represented by a vector in \mathbb{R}^9 :

$$\boldsymbol{\pi} = \left[\underbrace{\pi(\cdot|\{a, b, c\})}_{\in \mathbb{R}^3}, \underbrace{\pi(\cdot|\{a, b\})}_{\in \mathbb{R}^2}, \underbrace{\pi(\cdot|\{a, c\})}_{\in \mathbb{R}^2}, \underbrace{\pi(\cdot|\{b, c\})}_{\in \mathbb{R}^2} \right]'$$

For the preference $b \succ a \succ c$, the constraint matrix $\mathbf{R}_{b \succ a \succ c}$ takes the following form if only Assumption 1 is imposed:

$$\text{Assumption 1 only: } \mathbf{R}_{b \succ a \succ c} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix}.$$

With our new restriction on the attention rule for binary choice problems, $\mathbf{R}_{b \succ a \succ c}$ is further augmented:

$$\text{Assumption 1 and 3: } \mathbf{R}_{b \succ a \succ c} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ \hline 0 & 0 & 0 & \frac{1-\phi}{\phi} & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & \frac{1-\phi}{\phi} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & \frac{1-\phi}{\phi} \end{bmatrix}.$$

┘

Assumption 3 improves considerably the empirical content of our benchmark RAM (Assumption 1). However, this assumption is just one of many possible assumptions that could be used in addition to our general RAM. The main takeaway is that our proposed RAM offers a baseline for specific, empirically relevant models of choice under random limited attention. In Section 6 we compare using simulations the empirical content of our benchmark RAM, which employs only Assumption 1, and the model that incorporates Assumption 3 as well.

5.2 Random Attention Filter

We focus on random attention filters, which are one of the motivating examples of monotonic attention rules. Recall from Section 2.1 that an attention filter is a deterministic attention rule that satisfies Assumption 1, and a random attention filter is a convex combination of attention filters, and hence a random attention filter will also satisfy Assumption 1. For example, the same individual might be utilizing different platforms during her Internet search. Each platform yields a different attention filter, and the usage frequency of each platform is equal to the weight of that attention filter. Random attention filters also give a different interpretation of our model.

The set of all random attention filters is a strict subset of monotonic attention rules. This is not surprising given that the class of monotonic attention rules is very large. What is (arguably) surprising is the following fact that we are able to show: if (π, \succ, μ) is a RAM with μ being a monotonic attention rule, there exists a random attention filter μ' such that (π, \succ, μ') is still a RAM (see Remark 2 and Figure 2). Before presenting this result, however, we observe that μ and μ' need not be the same, which means that there are monotonic attention rules that cannot be written as a linear combination of attention filters.

Example 7. Let $X = \{a_1, a_2, a_3, a_4\}$. Consider a monotonic attention rule μ such that (i) $\mu(T|S)$ is either 0 or 0.5, (ii) $\mu(T|S) = 0$ if $|T| > 1$, and (iii) if $\mu(\{a_j\}|S) = 0$ and $k < j$ then $\mu(\{a_k\}|S) = 0$. Then we must have $\mu(\{a_3\}|\{a_1, a_2, a_3, a_4\}) = \mu(\{a_4\}|\{a_1, a_2, a_3, a_4\}) = 0.5$. We now show that μ is not a random attention filter.

Suppose μ can be written as a linear combination of attention filters. Then $\mu(\{a_3\}|\{a_1, a_2, a_3, a_4\}) = \mu(\{a_4\}|\{a_1, a_2, a_3, a_4\}) = 0.5$ implies that only attention filters for which $\Gamma(\{a_1, a_2, a_3, a_4\}) = \{a_3\}$ or $\Gamma(\{a_1, a_2, a_3, a_4\}) = \{a_4\}$ must be assigned positive probability. On the other hand, $\mu(\{a_2\}|\{a_1, a_2, a_3\}) = 0.5 = \mu(\{a_2\}|\{a_1, a_2, a_4\}) = 0.5$ imply that for all Γ which are assigned positive probability $\Gamma(\{a_1, a_2, a_3\}) = \{a_2\}$ whenever $\Gamma(\{a_1, a_2, a_3, a_4\}) = \{a_4\}$ and $\Gamma(\{a_1, a_2, a_4\}) = \{a_2\}$ whenever $\Gamma(\{a_1, a_2, a_3, a_4\}) = \{a_3\}$. To see this, notice that the attention filter property implies $\Gamma(\{a_1, a_2, a_3\}) = \{a_3\}$ for all Γ with $\Gamma(\{a_1, a_2, a_3, a_4\}) = \{a_3\}$ and $\Gamma(\{a_1, a_2, a_4\}) = \{a_4\}$ for all Γ with $\Gamma(\{a_1, a_2, a_3, a_4\}) = \{a_4\}$. But then it must be the case that $\Gamma(\{a_1, a_2\}) = \{a_2\}$ for all Γ which are assigned positive probability, or that $\mu(\{a_2\}|\{a_1, a_2\}) = 1$, a contradiction. \lrcorner

We now show that if we restrict our attention to a certain type of monotonic attention rules, then

we can show that within that class every attention rule is a random attention filter (i.e., convex combination of deterministic attention filters). Let $\mathcal{MT}(\succ)$ denote the set of all attention rules that are both monotonic (Assumption 1) and triangular with respect to \succ (Definition A.1 in the Appendix), and let $\mathcal{AF}(\succ)$ denote all attention filters that are triangular with respect to \succ . We are now ready to state the main result of this section.

Theorem 6 (Random Attention Filter). For any $\mu \in \mathcal{MT}(\succ)$, there exists a probability law ψ on $\mathcal{AF}(\succ)$ such that for any $S \in \mathcal{X}$ and $T \subset S$

$$\mu(T|S) = \sum_{\Gamma \in \mathcal{AF}(\succ)} \mathbb{1}(\Gamma(S) = T) \cdot \psi(\Gamma).$$

Remark 2 (Triangular Random Attention Filter Representation). Combining this theorem and Corollary A.1 in the Appendix, we easily reach the following conclusion: If π has a random attention representation (\succ, μ) , then there exists a triangular random attention filter μ' such that (\succ, μ') also represents π . See Figure 2 for an illustration. \perp

The proof of Theorem 6 is long and hence left to Appendix, but here we provide a sketch of it. First, $\mathcal{MT}(\succ)$ is a compact and convex set, and thus the above theorem can alternatively be stated as follows: The set of extreme points of $\mathcal{MT}(\succ)$ is $\mathcal{AF}(\succ)$. (An attention rule $\mu \in \mathcal{MT}(\succ)$ is an extreme point of $\mathcal{MT}(\succ)$ if it cannot be written as a nondegenerate convex combination of any $\mu', \mu'' \in \mathcal{MT}(\succ)$.) Then, Minkowski's Theorem guarantees that every element of $\mathcal{MT}(\succ)$ lies in the convex hull of $\mathcal{AF}(\succ)$.

Obviously, every element of $\mathcal{AF}(\succ)$ is an extreme point of $\mathcal{MT}(\succ)$. We then show that non-deterministic triangular attention rules cannot be extreme points, i.e. given any $\mu \in \mathcal{MT}(\succ) - \mathcal{AF}(\succ)$ we can construct $\mu', \mu'' \in \mathcal{MT}(\succ)$ such that $\mu = \frac{1}{2}\mu' + \frac{1}{2}\mu''$. The key step is to show that both μ' and μ'' that we construct are monotonic. After this step, we have shown that no $\mu \in \mathcal{MT}(\succ) - \mathcal{AF}(\succ)$ can be an extreme point, thus concluding the proof.

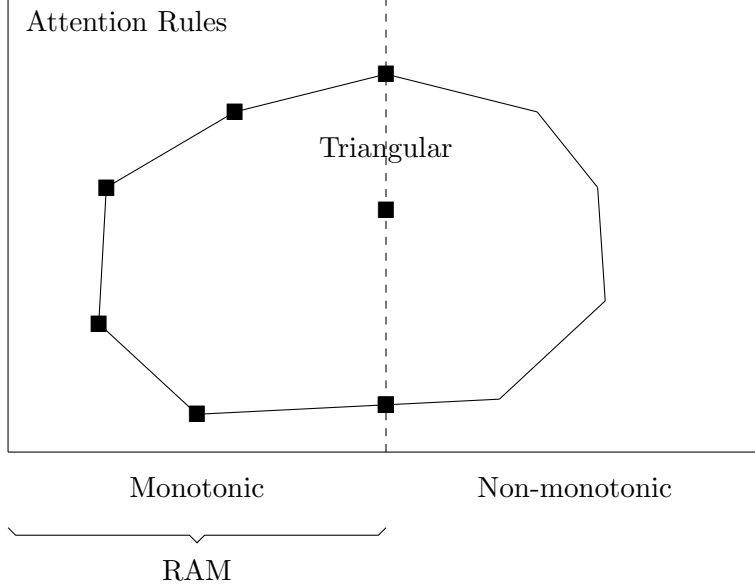


Figure 2. Illustration of attention rules. The *baseline RAM* (Assumption 1) is denoted by the left rectangular region. Corollary A.1 in the Appendix shows that RAMs can also be represented by *triangular monotonic attention rules* (the polygon to the left of the dashed line). Theorem 6 further illustrates that the set of triangular monotonic attention rules coincides with the convex hull of deterministic attention filters (solid black squares).

6 Simulation Evidence

This section gives a summary of a simulation study conducted to assess the finite sample properties of our proposed econometric methods. We consider a class of logit attention rules indexed by ς :

$$\mu_{\varsigma}(T|S) = \frac{w_{T,\varsigma}}{\sum_{T' \subset S} w_{T',\varsigma}}, \quad w_{T,\varsigma} = |T|^{\varsigma},$$

where $|T|$ is the cardinality of T . Thus the decision maker pays more attention to larger sets if $\varsigma > 0$, and pays more attention to smaller sets if $\varsigma < 0$. When ς is very small (negative and large in absolute magnitude), the decision maker almost always pays attention to singleton sets, hence nothing will be learned about the underlying preference from the choice data.

Other details on the data generating process used in the simulation study are as follows. First, the grand set X consists of five alternatives, a_1, a_2, a_3, a_4 and a_5 . Without loss of generality, assume the underlying preference is $a_1 \succ a_2 \succ a_3 \succ a_4 \succ a_5$. Second, the data consists of choice problems of size two, three, four and five. That is, there are in total 26 choice problems. Third, given a specific realization of Y_i , a consideration set is generated from the logit attention model

with $\varsigma = 2$, after which the choice y_i is determined by the aforementioned preference.⁷ Finally, the observed data is a random sample $\{(y_i, Y_i) : 1 \leq i \leq N\}$, where the effective sample size can be 50, 100, 200, 300 and 400.⁸

For inference, we employ the procedure introduced in Section 4 and test whether a specific preference ordering is compatible with the basic RAM (Assumption 1). We also incorporate the attentive at binaries assumption introduced in Section 5.1. Recall from Assumption 3 that $(1-\phi)/\phi$ is a measure of full attention at binaries, and specifying a larger value (i.e., a smaller value of ϕ) implies that the researcher is more willing to draw information from binary comparisons. Note that with $\phi = 1$, imposing Assumption 3 does not bring any additional identification power. Before proceeding, we list five hypotheses (preference orderings), and whether they are compatible with our RAM and specific values of ϕ .

	ϕ										
	1	.95	.90	.85	.80	.75	.70	.65	.60	.55	.50
$H_{0,1} : a_1 \succ a_2 \succ a_3 \succ a_4 \succ a_5$	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
$H_{0,2} : a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_1$	✓	✓	✓	✓	×	×	×	×	×	×	×
$H_{0,3} : a_3 \succ a_4 \succ a_5 \succ a_2 \succ a_1$	×	×	×	×	×	×	×	×	×	×	×
$H_{0,4} : a_4 \succ a_5 \succ a_3 \succ a_2 \succ a_1$	×	×	×	×	×	×	×	×	×	×	×
$H_{0,5} : a_5 \succ a_4 \succ a_3 \succ a_2 \succ a_1$	×	×	×	×	×	×	×	×	×	×	×

As can be seen, $H_{0,1}$ always belongs to the identified set of preferences, as it is the preference ordering used in the underlying data generating process. $H_{0,2}$, however, may or may not belong to the identified set depending on the value of ϕ : with ϕ close to 0.5, the researcher is confident enough using information from binary comparisons, and she will be able to reject this hypothesis; for ϕ close to 1, Assumption 3 no longer brings too much additional identification power beyond the monotonic attention assumption, and monotonic attention along is not strong enough to reject this hypothesis. Indeed, with $\phi = 1$ (i.e., Assumption 1 alone), the set of identified preference is $\{\succ : a_2 \succ a_3 \succ a_4 \succ a_5\}$, which contains $H_{0,2}$. The other three hypotheses, $H_{0,3}$, $H_{0,4}$ and $H_{0,5}$, do not belong to the identified set even with $\phi = 1$.

⁷In the Supplemental Appendix, we also report simulation evidence for $\varsigma \in \{0, 1\}$.

⁸Effective sample size refers to the number of observations for each choice problem. Because there are 26 choice problems, the overall sample size is $N \in \{1300, 2600, 5200, 7800, 10400\}$.

Overall, our simulation has 5 (different N) \times 5 (different preference orderings) \times 11 (different ϕ) = 275 designs. For each design, $5,000$ simulation repetitions are used, and the five null hypotheses are tested using our proposed method at the 5% nominal level. Simulation results are summarized in Figure 3.

We first focus on $H_{0,1}$ (panel a). As this preference ordering is compatible with our RAM, one should expect the rejection probability to be less than the nominal level. Indeed, the rejection probability is far below 0.05 : this illustrates a generic feature of any (reasonable) procedure for testing moment inequalities – to maintain uniform asymptotic size control, empirical rejection probability is below the nominal level when the inequalities are far from binding. Next consider $H_{0,2}$ (panel b). For ϕ larger than 0.85 , the rejection probability is below the nominal size, which is compatible with our theory, because this preference belongs to the identified set when only Assumption 1 is imposed. With smaller ϕ , the researcher relies more heavily on information from binary comparisons/choice problems, and she is able to reject this hypothesis much more frequently. This demonstrates how additional restrictions on the attention rule can be easily accommodated by our basic RAM, which, in turn, can bring additional identification power. The other three hypotheses (panel c–e) are not compatible with our RAM, and we do see that the rejection probability is much higher than the nominal size even for $\phi = 1$, showing that even our basic RAM has non-trivial empirical content in this case.

In the online Supplemental Appendix, we also report simulation evidence for $\varsigma \in \{0, 1\}$.

7 Conclusion

We introduce a limited attention model allowing for a general class of monotonic stochastic consideration rules, which we call a Random Attention Model (RAM). We show that this model nests several important recent contributions in both economic theory and econometrics, in addition to other classical results from decision theory. Using our RAM framework, we obtain a testable theory of revealed preferences and develop partial identification results for the decision maker’s unobserved strict preference ordering. Our results include a precise constructive characterization of the identified set for preferences, as well as uniformly valid inference methods based on this characterization. We illustrate good finite sample performance of our methods in a simulation experiment. Finally,

we provide the general-purpose R software package `ramchoice`, which allows other researchers to easily employ our econometric methods in empirical applications.

We regard these results as a first step in a research program, since many open questions and extensions within our proposed RAM framework are worth investigating. For example, from an economic theory perspective, two natural lines of inquiry are (i) to develop a theory of stochastic choice revealing random attention maps from observed choice behavior, and (ii) to expand the domain of the choice problem to study the effect of framing (e.g., presentation or advertising) on random attention maps. Similarly, from an econometric perspective, two other natural extensions of our work include (i) allowing for observed covariates to enter RAM and hence the use of other (conditional) choice probability estimates, and (ii) learning about unobserved heterogeneity in order to develop counterfactual policy analysis (Theorem 6 gives a first result along this line). Research along these lines is underway.

Appendix: Omitted Proofs

This appendix collects proofs that are omitted from the main text to improve the exposition. We make one additional definition.

Definition A.1 (Lower Contour Set; Triangular Attention Rule). Given a preference ordering \succ of the alternatives in X : $a_{1,\succ} \succ a_{2,\succ} \succ \cdots \succ a_{K,\succ}$, a *lower contour set* is defined as $L_{k,\succ} = \{a_{j,\succ} : j \geq k\} = \{a \in X : a \preceq a_{k,\succ}\}$.

A *triangular attention rule* is an attention rule which puts weights only on lower contour sets. That is, $\mu(T|S) > 0$ implies $T = L_{k,\succ} \cap S$ for some k such that $a_{k,\succ} \in S$.

A.1 Proof of Theorem 2

Suppose π has a random attention representation (\succ, μ) . Then Lemma 1 implies that \succ must include P so P must be acyclic.

For the other direction, suppose that P has no cycle. Pick any preference \succ that includes P_R and enumerate all alternatives with respect to \succ : $a_{1,\succ} \succ a_{2,\succ} \succ \cdots \succ a_{K,\succ}$. Let $\{L_{k,\succ} : 1 \leq k \leq K\}$ be the corresponding lower contour sets (Definition A.1). Then we specify $\tilde{\mu}$ as

$$\tilde{\mu}(T|S) = \begin{cases} \pi(a_{k,\succ}|S) & \text{if } a_{k,\succ} \in S \text{ and } T = L_{k,\succ} \cap S \\ 0 & \text{otherwise} \end{cases}.$$

It is trivial to verify that $(\succ, \tilde{\mu})$ represents π , since $(\succ, \tilde{\mu})$ induces the following choice rule:

$$\begin{aligned}
\sum_{T \subset S} \mathbf{1}[a \text{ is } \succ\text{-best in } T] \tilde{\mu}(T|S) &= \sum_{a_{k,\succ} \in S} \mathbf{1}[a \text{ is } \succ\text{-best in } L_{k,\succ} \cap S] \tilde{\mu}(L_{k,\succ} \cap S|S) \\
&= \sum_{a_{k,\succ} \in S} \mathbf{1}[a \text{ is } \succ\text{-best in } L_{k,\succ} \cap S] \pi(a_{k,\succ}|S) \\
&= \sum_{a_{k,\succ} \in S} \mathbf{1}[a = a_{k,\succ}] \pi(a_{k,\succ}|S) \\
&= \pi(a|S),
\end{aligned}$$

which is the same as π . For the first equality, we use the definition that a triangular attention rule only puts weights on lower contour sets; for the second equality, we apply the definition/construction of $\tilde{\mu}$; the third equality follows from the definition of lower contour sets.

Now we verify that $\tilde{\mu}$ satisfies Assumption 1. Assume this is not the case, then it means there exists some S , $a_{k,\succ}, a_{\ell,\succ} \in S$, such that (i) $L_{k,\succ} \cap S = L_{k,\succ} \cap (S - a_{\ell,\succ})$, and (ii) $\tilde{\mu}(L_{k,\succ} \cap S|S) > \tilde{\mu}(L_{k,\succ} \cap (S - a_{\ell,\succ})|S - a_{\ell,\succ})$. By the definition of lower contour sets, (i) implies $a_{\ell,\succ} \succ a_{k,\succ}$. Then (ii) implies

$$\tilde{\mu}(L_{k,\succ} \cap S|S) = \pi(a_{k,\succ}|S) > \tilde{\mu}(L_{k,\succ} \cap (S - a_{\ell,\succ})|S - a_{\ell,\succ}) = \pi(a_{k,\succ}|S - a_{\ell,\succ}).$$

The above, however, implies that $a_{k,\succ} \mathbf{P} a_{\ell,\succ}$, which contradicts the implication of (i) that $a_{\ell,\succ} \succ a_{k,\succ}$. This closes the proof.

Remark A.1. The previous proof has a nice implication that, a choice rule can be represented by a monotonic attention rule if and only if it can also be represented by a monotonic triangular attention rule. Formally, if π has a random attention representation, (\succ, μ) , then $(\succ, \tilde{\mu})$ also represents π where $\tilde{\mu}$ is monotonic and triangular with respect to \succ . Hence, we can focus on monotonic triangular attention rules without loss of generality. This is formally summarized in Corollary A.1. \square

Corollary A.1 (Monotonic Triangular Attention Rule Representation). Assume (\succ, μ) is a representation of π with μ satisfying Assumption 1. Then it is possible to construct a triangular attention rule $\tilde{\mu}$ that also satisfies Assumption 1, and that $(\succ, \tilde{\mu})$ is a representation of π .

A.2 Proof of Theorem 4

See Section SA.4.1 of the Supplemental Appendix.

A.3 Proof of Theorem 5

The “only if” part is trivial and is omitted. We illustrate the “if” part. Assume that $\mathbf{P}^\phi \cup \mathbf{P}$ has no cycle (or equivalently, its transitive closure $(\mathbf{P}^\phi \cup \mathbf{P})_{\mathbf{R}}$ has no cycle), then there exists some preference ordering that embeds $\mathbf{P}^\phi \cup \mathbf{P}$. Fix one such preference \succ . With the same argument used in the proof of Theorem 2, we can construct a triangular attention rule $\mu(T|S)$ and show that it satisfies Assumption 1.

We then show that $\mu(T|S)$ satisfies Assumption 3. Take binary $S = \{a, b\}$ and assume without loss of generality that $a \succ b$. Then $\mu(\{a, b\}|\{a, b\}) = \pi(a|S)$ and $\mu(\{b\}|\{a, b\}) = \pi(b|S)$. Violation of Assumption 3 implies $\pi(a|\{a, b\}) < \frac{1-\phi}{\phi} \pi(b|\{a, b\})$, and equivalently, $\pi(b|\{a, b\}) > \phi$. This means that $b \mathbf{P}^\phi a$, which violates our definition of \succ .

A.4 Proof of Theorem 6

We show that the set of extreme points of $\mathcal{MT}(\succ)$ is $\mathcal{AF}(\succ)$. Clearly, any $\Gamma \in \mathcal{AF}(\succ)$ is an extreme point. Pick a non-deterministic attention rule $\mu \in \mathcal{MT}(\succ)$. We show that μ cannot be an extreme point. Let $\mathcal{X}_\mu \subset \mathcal{X}$ stand for all sets for which $\mu(T|S) = 1$ for no $T \subset S$. We start by choosing $\varepsilon > 0$ small enough so that none of the non-binding constraints are affected whenever ε is added to or subtracted from $\mu(T|S)$ for all $T \subset S$ and $S \in \mathcal{X}$. Let $k_\mu = \min_{S \in \mathcal{X}_\mu} |S|$. Since μ is not deterministic, such k_μ exists.

We begin with the following simple observation that given S with $|S| = k_\mu$ we can have at most two subsets of S with $\mu(T|S) \in (0, 1)$. Moreover, it must be the case that $\mu(S|S) \in (0, 1)$.

Lemma A.1. Let S with $|S| = k_\mu$ be given. Then there exist at most two $T \subset S$ such that $\mu(T|S) \in (0, 1)$. Furthermore, $\mu(S|S) \in (0, 1)$.

Proof. Suppose there exist three such subsets: T_1 , T_2 , and T_3 . Since μ is triangular the subsets which are considered with positive probability can be ordered by set inclusion. Hence, we can without loss of generality assume $T_1 \subset T_2 \subset T_3$. But then since μ is monotonic and $T_1 \subset T_2 \subset S$ it must be that $\mu(T_1|T_2) \in (0, 1)$ and $\mu(T_2|T_2) \in (0, 1)$. This contradicts the definition of k_μ . Hence there can be at most two subsets T_1 and T_2 with positive probability. The same contradiction appears as long as $T_2 \subsetneq S$. Hence, $T_2 = S$. \blacksquare

Now for all sets $S \in \mathcal{X}_\mu$ with $|S| = k_\mu$, we define μ' and μ'' as follows:

$$\mu'(T|S) = \mu(T|S) + \varepsilon,$$

$$\mu'(S|S) = \mu(S|S) - \varepsilon,$$

and

$$\mu''(T|S) = \mu(T|S) - \varepsilon,$$

$$\mu''(S|S) = \mu(S|S) + \varepsilon$$

where $T \subsetneq S$ with $\mu(T|S) \in (0, 1)$.

Suppose we have defined μ' and μ'' for all sets with $|S| \leq l$ and let S with $|S| = l + 1$ be given. If there exist no $T \subset S$ and $S_T \subset S$ such that $\mu'(T|S_T) \neq \mu''(T|S_T)$ and $\mu(T|S) = \mu(T|S_T)$, then we set $\mu(T|S) = \mu'(T|S) = \mu''(T|S)$ for all $T \subset S$. Otherwise, pick the smallest T for which such S_T exists. If $\mu'(T|S_T) > \mu''(T|S_T)$, then let $\mu'(T|S) = \mu(T|S) + \varepsilon$ and $\mu''(T|S) = \mu(T|S) - \varepsilon$ and if $\mu'(T|S_T) < \mu''(T|S_T)$, then let $\mu'(T|S) = \mu(T|S) - \varepsilon$ and $\mu''(T|S) = \mu(T|S) + \varepsilon$. If T is the only set for which such S_T exists, then let T' be the largest set for which $\mu(T'|S) \in (0, 1)$. Otherwise T' denotes the other set for which $S_{T'}$ satisfying the description exists. If $\mu'(T|S_T) > \mu''(T|S_T)$, then let $\mu'(T'|S) = \mu(T'|S) - \varepsilon$ and $\mu''(T'|S) = \mu(T'|S) + \varepsilon$ and if $\mu'(T|S_T) < \mu''(T|S_T)$, then let $\mu'(T'|S) = \mu(T'|S) + \varepsilon$ and $\mu''(T'|S) = \mu(T'|S) - \varepsilon$. For all other subsets μ , μ' , and μ'' agree. We proceed iteratively.

Lemma A.2. Suppose there exist $T \subset S$ and $S_T \subset S$ such that $\mu'(T|S_T) \neq \mu''(T|S_T)$ and $\mu(T|S) = \mu(T|S_T)$. Then either T is the smallest set in S satisfying the description or we can set $S_T = T$.

Proof. The claim follows from Lemma A.1 when $|S| = k_\mu + 1$. Suppose the claim holds whenever $|S| \leq l$. We show that the claim holds when $|S| = l + 1$. Let $T \subset S$ and $S_T \subset S$ satisfy the description and suppose T is not the smallest set in S satisfying the description. Since $\mu'(T|S_T) \neq \mu''(T|S_T)$, by construction, either T is the largest set satisfying $\mu(T|S_T) \in (0, 1)$ or there exists

$S_{S_T} \subset S_T$ such that $\mu'(T|S_{S_T}) \neq \mu''(T|S_{S_T})$ and $\mu(T|S_T) = \mu(T|S_{S_T})$. If the first case is true, then since μ is monotonic, it must be the case that $\mu(T'|T) = \mu(T'|S_T)$ for all $T' \subset T$, and hence we are done. In the second case, the claim follows from induction. ■

Lemma A.3. For any S , there exist either zero or two subsets satisfying $\mu'(T|S) \neq \mu''(T|S)$. Moreover if there are two sets satisfying the description, then $\mu'(T_1|S) > \mu''(T_1|S)$ if and only if $\mu'(T_2|S) < \mu''(T_2|S)$.

Proof. The claim is trivial when $|S| = k_\mu$. Suppose the claim is true for all S with $|S| \leq l$ and let S with $|S| = l + 1$ be given. If there is no T which satisfies the description in the construction, then no subset will be affected. Suppose there exists only one such T . We show that there exists $T' \supset T$ such that $\mu(T'|S) \in (0, 1)$. To see this notice that by monotonicity property $\mu(T''|S) \leq \mu(T''|S_T)$ for all $T'' \subset T$. Since by induction there are two subsets of S_T for which $\mu'(T|S_T) \neq \mu''(T|S_T)$ either $\mu(T''|S) < \mu(T''|S_T)$ for some $T'' \subset T$ or there exists $T''' \supset T$ such that $\mu(T'''|S_T) \in (0, 1)$. In both cases, $\sum_{T'' \subset T} \mu(T''|S) < 1$ follows. Hence, there is $T' \supset T$ such that $\mu(T'|S) \in (0, 1)$. The construction then guarantees that $\mu'(T'|S) \neq \mu''(T'|S)$ for some $T' \supset T$. Now suppose there are three subsets, T_1, T_2 , and T_3 , satisfying the description. Since μ is triangular, we can without loss of generality assume that $T_1 \subset T_2 \subset T_3$. By the previous lemma, we can without loss of generality assume $S_{T_2} = T_2$ and $S_{T_3} = T_3$. But then since μ is monotonic, 3 subsets of S_{T_3} must satisfy the description, a contradiction to induction hypothesis.

To prove the second part of the claim, notice that the claim follows from construction if $|S| = k_\mu$. Suppose the claim holds whenever $|S| \leq l$ and let $|S| = l + 1$ be given. If $T_2 = S$, then the claim follows from construction. If $T_2 \subsetneq S$, then the claim follows from induction and construction by considering the set T_2 . ■

It is clear that $\mu = \frac{1}{2}\mu' + \frac{1}{2}\mu''$. The previous lemmas also show that both μ' and μ'' are monotonic. Hence, no $\mu \in \mathcal{MT}(\succ) - \mathcal{AF}(\succ)$ can be an extreme point. This concludes the proof of Theorem 6.

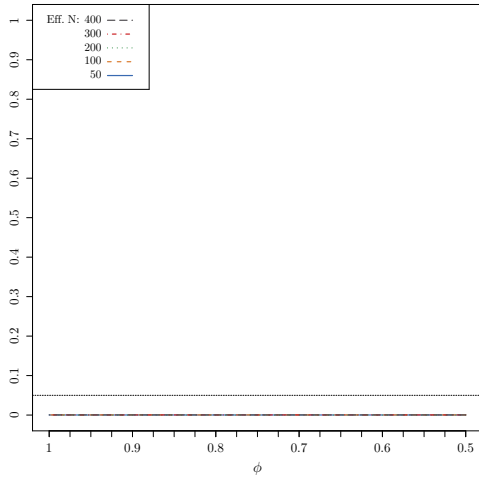
References

- ABALUCK, J. AND A. ADAMS (2017): “What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses,” *NBER Working Paper No. 23566*.
- AGRANOV, M. AND P. ORTOLEVA (2017): “Stochastic Choice and Preferences for Randomization,” *Journal of Political Economy*, 125, 40–68.
- AGUIAR, V. H. (2015): “Stochastic Choice and Attention Capacities: Inferring Preferences from Psychological Biases,” *SSRN Working Paper 2607602*.
- AGUIAR, V. H., M. J. BOCCARDI, AND M. DEAN (2016): “Satisficing and Stochastic Choice,” *Journal of Economic Theory*, 166, 445–482.
- ANDREWS, D. W. AND G. SOARES (2010): “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica*, 78, 119–157.
- BARSEGHYAN, L., M. COUGHLIN, F. MOLINARI, AND J. C. TEITELBAUM (2018): “Heterogeneous Consideration Sets and Preferences,” work in progress, Cornell University.
- BATESON, M., S. D. HEALY, AND T. A. HURLY (2003): “Context-Dependent Foraging Decisions in Rufous Hummingbirds,” *Proceedings of the Royal Society of London B: Biological Sciences*, 270, 1271–1276.

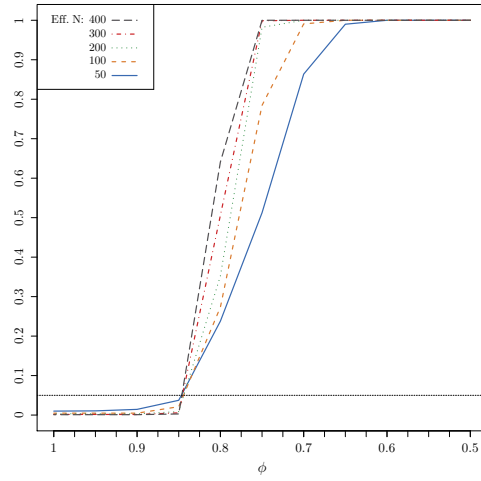
- BLUNDELL, R., D. KRISTENSEN, AND R. MATZKIN (2014): “Bounding Quantile Demand Functions Using Revealed Preference Inequalities,” *Journal of Econometrics*, 179, 112–127.
- BRADY, R. L. AND J. REHBECK (2016): “Menu-Dependent Stochastic Feasibility,” *Econometrica*, 84, 1203–1223.
- BUGNI, F. A. (2016): “Comparison of Inferential Methods in Partially Identified Models in terms of Error in Coverage Probability,” *Econometric Theory*, 32, 187–242.
- BUGNI, F. A., I. A. CANAY, AND X. SHI (2015): “Specification Tests for Partially Identified Models Defined by Moment Inequalities,” *Journal of Econometrics*, 185, 259–282.
- BUSEMEYER, J. R. AND J. RIESKAMP (2014): “Psychological Research and Theories on Preferential Choice,” in *Handbook of Choice Modeling*, ed. by S. Hess and A. Daly, Edward Elgar Publishing, 49–72.
- CANAY, I. A. (2010): “EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity,” *Journal of Econometrics*, 156, 408–425.
- CANAY, I. A. AND A. M. SHAIKH (2017): “Practical and Theoretical Advances for Inference in Partially Identified Models,” in *Advances in Economics and Econometrics: Volume 2: Eleventh World Congress*, ed. by B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge University Press, 271–306.
- DARDANONI, V., P. MANZINI, M. MARIOTTI, AND C. J. TYSON (2018): “Inferring Cognitive Heterogeneity from Aggregate Choices,” *Working Paper Series 1018, Department of Economics, University of Sussex*.
- DEAN, M., Ö. KIBRIS, AND Y. MASATLIOGLU (2017): “Limited Attention and Status Quo Bias,” *Journal of Economic Theory*, 169, 93–127.
- DEB, R., Y. KITAMURA, J. K. QUAH, AND J. STOYE (2018): “Revealed Price Preference: Theory and Stochastic Testing,” *Cowles Foundation Discussion Paper No. 2087*.
- ECHENIQUE, F. AND K. SAITO (2019): “General Luce Model,” *Economic Theory, forthcoming*.
- ECHENIQUE, F., K. SAITO, AND G. TSERENJIGMID (2018): “The Perception-Adjusted Luce Model,” *Mathematical Social Sciences*, 93, 67–76.
- FISHBURN, P. C. (1973): “Binary Choice Probabilities: On the Varieties of Stochastic Transitivity,” *Journal of Mathematical Psychology*, 10, 327–352.
- FUDENBERG, D., R. IJIMA, AND T. STRZALECKI (2015): “Stochastic Choice and Revealed Perturbed Utility,” *Econometrica*, 83, 2371–2409.
- GAUNDRY, M. J. AND M. G. DAGENAIS (1979): “The Dogit Model,” *Transportation Research Part B: Methodological*, 13, 105–111.
- GOEREE, M. S. (2008): “Limited Information and Advertising in the US Personal Computer Industry,” *Econometrica*, 76, 1017–1074.
- GUL, F., P. NATENZON, AND W. PESENDORFER (2014): “Random Choice as Behavioral Optimization,” *Econometrica*, 82, 1873–1912.

- HAUSER, J. R. AND B. WERNERFELT (1990): “An Evaluation Cost Model of Consideration Sets,” *Journal of Consumer Research*, 16, 393–408.
- HAUSMAN, J. A. AND W. K. NEWEY (2017): “Nonparametric Welfare Analysis,” *Annual Review of Economics*, 9, 521–46.
- HO, K. AND A. M. ROSEN (2017): “Partial Identification in Applied Research: Benefits and Challenges,” in *Advances in Economics and Econometrics: Volume 2: Eleventh World Congress*, ed. by B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge University Press, 307–359.
- HONKA, E., A. HORTAÇSU, AND M. A. VITORINO (2017): “Advertising, Consumer Awareness, and Choice: Evidence from the US Banking Industry,” *The RAND Journal of Economics*, 48, 611–646.
- HORAN, S. (2018): “Random Consideration and Choice: A Case Study of “Default” Options,” *Working Paper, Université de Montréal and CIREQ*.
- HOUTHAKKER, H. S. (1950): “Revealed Preference and the Utility Function,” *Economica*, 17, 159–174.
- HUBER, J., J. W. PAYNE, AND C. PUTO (1982): “Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis,” *Journal of Consumer Research*, 9, 90–98.
- HURLY, T. A. AND M. D. OSEEN (1999): “Context-Dependent, Risk-Sensitive Foraging Preferences in Wild Rufous Hummingbirds,” *Animal Behaviour*, 58, 59–66.
- IMBENS, G. W. AND C. F. MANSKI (2004): “Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 72, 1845–1857.
- KAIDO, H., F. MOLINARI, AND J. STOYE (2019): “Confidence Intervals for Projections of Partially Identified Parameters,” *Econometrica*, , *forthcoming*.
- KAWAGUCHI, K. (2017): “Testing Rationality without Restricting Heterogeneity,” *Journal of Econometrics*, 197, 153–171.
- KAWAGUCHI, K., K. UETAKE, AND Y. WATANABE (2016): “Identifying Consumer Attention: A Product-Availability Approach,” *SSRN Working Paper 2529294*.
- KITAMURA, Y. AND J. STOYE (2018): “Nonparametric Analysis of Random Utility Models,” *Econometrica*, 86, 1883–1909.
- LLERAS, J. S., Y. MASATLIOGLU, D. NAKAJIMA, AND E. Y. OZBAY (2017): “When More Is Less: Limited Consideration,” *Journal of Economic Theory*, 170, 70–85.
- MANZINI, P. AND M. MARIOTTI (2014): “Stochastic Choice and Consideration Sets,” *Econometrica*, 82, 1153–1176.
- MASATLIOGLU, Y., D. NAKAJIMA, AND E. Y. OZBAY (2012): “Revealed Attention,” *American Economic Review*, 102, 2183–2205.
- MATZKIN, R. L. (2013): “Nonparametric Identification in Structural Economic Models,” *Annual Review of Economics*, 5, 457–486.

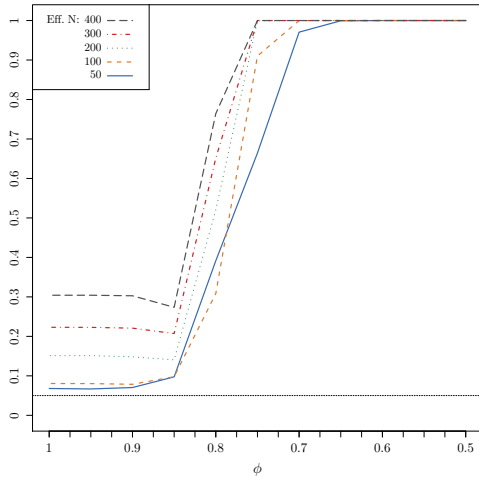
- MOLINARI, F. (2019): “Econometrics with Partial Identification,” in Working Paper.
- REUTSKAJA, E., R. NAGEL, C. F. CAMERER, AND A. RANGEL (2011): “Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study,” *American Economic Review*, 101, 900–926.
- RIESKAMP, J., J. R. BUSEMEYER, AND B. A. MELLERS (2006): “Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice,” *Journal of Economic Literature*, 44, 631–661.
- SHAFIR, S., T. A. WAITE, AND B. H. SMITH (2002): “Context-Dependent Violations of Rational Choice in Honeybees (*Apis Mellifera*) and Gray Jays (*Perisoreus Vanadensis*),” *Behavioral Ecology and Sociobiology*, 51, 180–187.
- SUPPES, P. AND R. D. LUCE (1965): “Preference, Utility, and Subjective Probability,” *Handbook of Mathematical Psychology*, 3, 249–410.
- TVERSKY, A. (1969): “Intransitivity of Preferences.” *Psychological Review*, 76, 31–48.
- (1972): “Elimination by Aspects: A Theory of Choice.” *Psychological Review*, 79, 281.
- VAN NIEROP, E., B. BRONNENBERG, R. PAAP, M. WEDEL, AND P. H. FRANSES (2010): “Retrieving Unobserved Consideration Sets from Household Panel Data,” *Journal of Marketing Research*, 47, 63–74.



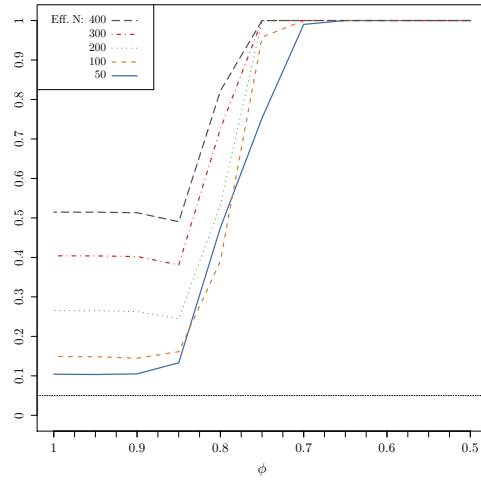
(a) $H_{0,1} : a_1 \succ a_2 \succ a_3 \succ a_4 \succ a_5$



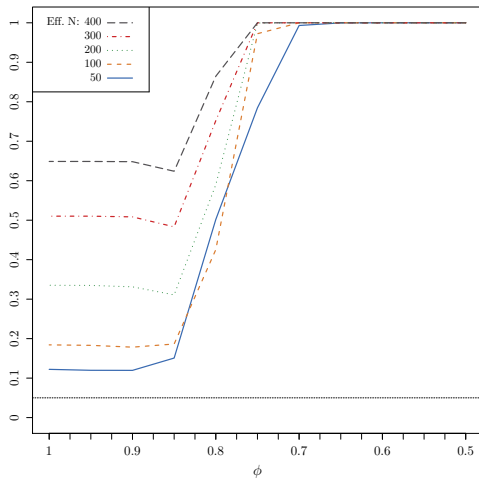
(b) $H_{0,2} : a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_1$



(c) $H_{0,3} : a_3 \succ a_4 \succ a_5 \succ a_2 \succ a_1$



(d) $H_{0,4} : a_4 \succ a_5 \succ a_3 \succ a_2 \succ a_1$



(e) $H_{0,5} : a_5 \succ a_4 \succ a_3 \succ a_2 \succ a_1$

Shown in the figure are empirical rejection probabilities testing the five null hypothesis through 2,000 simulations, with nominal size 0.05. Logit attention rule with $\zeta = 2$ is used, as described in the text. For each simulation repetition, five effective sample sizes are considered 50, 100, 200, 300 and 400.

Figure 3. Empirical Rejection Probabilities