A Random Attention Model

Matias D. Cattaneo

Princeton University

Xinwei Ma

University of California San Diego

Yusufcan Masatlioglu

University of Maryland

Elchin Suleymanov

Purdue University

This paper illustrates how one can deduce preference from observed choices when attention is both limited and random. We introduce a random attention model where we abstain from any particular attention formation and instead consider a large class of nonparametric random attention rules. Our intuitive condition, monotonic attention, captures the idea that each consideration set competes for the decision maker's attention. We then develop a revealed preference theory and obtain testable implications. We propose econometric methods for identification, estimation, and inference for the revealed preferences. Finally, we provide a general-purpose software implementation of our estimation and inference results and simulation evidence.

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, and four 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.

Electronically published June 10, 2020
[*Journal of Political Economy*, 2020, vol. 128, no. 7]

I. Introduction

Revealed preference theory is not only a cornerstone of modern economics but 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 patterns, 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, Hortaçsu, and Vitorino 2017). For example, Reutskaja et al. (2011) provide 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 cannot 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; Dean, Kıbrıs, and Masatlioglu 2017; Lleras et al. 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—also known as attention rule or consideration map—as unobservable and hence 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 always uses 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 predictions 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 2019), which permits the decision maker to pay

attention to different subsets with some nonzero probability given the same set of alternatives to choose from. All available results in this literature proceed by first parameterizing the attention rule (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) attention rule 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 attention rules. 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 and particular attention rules that could be operating in the background. In other words, our revealed preference results are derived from nonparametric restrictions on the attention rule and hence are more robust to misspecification biases.

The RAM is best suited for eliciting information about the preference ordering of a single decision-making unit when her choices are observed repeatedly.[1] For example, scanner data keep 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 are abundant records from each individual decision maker available but it is also common to see manipulations or 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) are analyzed. In addition, our model can be used empirically with aggregate data on a group of distinct decision makers, provided that each of them may differ on what they pay attention to but all share the same preference.

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

[1] The 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 that allow the analyst to reveal information about the agent's preferences from her observed random choices.

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 attention rules, including the parametric ones currently available in the literature, which, as we show, satisfy the monotonic attention assumption.

On the basis of 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 the RAM. We show that the 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 (i.e., test whether there exists a nonempty set of preference orderings compatible with the 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 prespecified 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 (2020) for recent reviews and further references.

The 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, the RAM also provides a basis for incorporating additional (parametric and) nonparametric restrictions that can substantially improve identification power. In this paper, we illustrate how the RAM can be combined with additional, mild nonparametric restrictions to tighten identification in nontrivial ways: in section V.A, we incorporate an additional restriction on attention rule for binary choice problems and show that this alone leads to important revelation improvements within the 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 the possibly high-dimensional space of attention rules, leading to methods that are very fast and easy to implement when applied to realistic empirical problems. See appendix B (available online) for numerical evidence.

Our work contributes to both economic theory and econometrics. We describe several examples covered by our model in section II after we introduce our proposed RAM. We also discuss in detail the connections and distinctions between this paper and the economic theory literature in section SA-1 of appendix B. In particular, we show how the RAM nests and/or connects to the recent work by Gul, Natenzon, and Pesendorfer (2014); Manzini and Mariotti (2014); Fudenberg, Iijima, and Strzalecki (2015); Aguiar, Boccardi, and Dean (2016); Brady and Rehbeck (2016); Echenique, Saito, and Tserenjigmid (2018); and Echenique and Saito (2019), 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 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); Deb et al. (2018); and Kitamura and Stoye (2018) for a sample of recent contributions and further references. As mentioned above, a key feature of the RAM is that our proposed monotonic attention condition on attention rule nests previous models as special cases and also covers many new models of choice behavior. In particular, the RAM can accommodate more choice behaviors or patterns than what can be rationalized by random utility models. This is important because numerous studies in psychology, finance, and marketing have shown that decision makers exhibit limited attention when making choices; they compare (and choose from) only a subset of all available options. Whenever decision makers do not pay full attention to all options, implications from revealed preference theory under random utility models no longer hold in general, implying that empirical testing of substantive hypotheses as well as policy recommendations based on random utility models will be invalid. On the other hand, our results may remain valid.

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), Barseghyan et al. (2018), and Dardanoni et al. (2020). These papers are complementary to ours insofar as different assumptions on the random attention rule and preference(s) are imposed, which leads to different levels of (partial) identification of preference(s) and (random) attention rule(s). For further discussion of the relationship with these papers, see section SA-1 of appendix B.

The rest of the paper proceeds as follows. Section II presents the basic setup, where our key monotonicity assumption on the decision maker's stochastic attention rule is presented later in the section. Section III discusses

our RAM in detail, including the main revealed preference results. Section IV presents our main econometrics methods, including nonparametric (partial) identification, estimation, and inference results. In section V.A, we consider additional restrictions on the attention rule for binary choice problems, which can help improve our identification and inference results considerably. We also consider random attention filters in section V.B, which are one of the motivating examples of monotonic attention rules. In this case, however, there is no additional identification. Section VI summarizes the findings from a simulation study. Finally, section VII concludes with a discussion of directions for future research. A companion appendix B includes more examples, extensions, other methodological results, omitted proofs, and additional simulation evidence.

II. 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 nonempty subsets of X. Each member of $\mathcal{X}$ defines a choice problem.

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

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 zero or one, 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 appendix B to account for cases where only data on a subcollection of choice problems are available.

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

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

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 $a \in S$, $\Sigma_{a \in T \subset S} \mu(T|S)$ denotes the probability of paying attention to

a in the choice problem S. The probabilities on consideration sets allow us to derive the attention probabilities on alternatives uniquely.

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 that she pays attention only 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 that people use in real life (see examples below and in sec. SA-2 of app. B). 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, and 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 appendix B. The first example is deterministic (i.e., $\mu(T|S)$ is either zero or one), but the others are all stochastic.

EXAMPLE 1 (Attention filter). 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, Nakajima, and Ozbay (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 zero otherwise.

Example 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

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

where $\mathbb{I}$ denotes the indicator function. We will pay special attention to this class of attention rules in section V.B.

Example 3 (Independent consideration). This example is based on Manzini and Mariotti (2014). Consider a decision maker who pays attention to each alternative a with a fixed probability $\gamma(a) \in (0, 1)$. The parameter 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 to one.

Example 4 (Logit attention). This example is based on Brady and Rehbeck (2016). Consider a decision maker who assigns a positive weight for each nonempty 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

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

Even though there is no structure on weights in the general version of this model, there are two interesting special cases where weights depend solely on the size of the set. These are $w_T = |T|$ and $w_T = 1/|T|$, 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.

EXAMPLE 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 values are nonnegative, the second term, which is independent of w_T, is the smallest lower bound for $\mu(T|S)$. The larger the θ_T values, 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.

EXAMPLE 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 $\{j, k, \ell, \ldots\}$ represent 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 represent 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 represents 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

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

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.[2] As a consequence, our revealed preference

[2] To 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—i.e., $\gamma_S(a)$. Then, the frequency of each set being the consideration set is calculated as an independent consideration rule. Because of this contextual dependence, further restrictions on $\gamma_S(a)$ and $\gamma_{S-b}(a)$ are needed to ensure assumption 1.

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 sec. SA-2 of app. B), our upcoming characterization, identification, estimation, and inference results nest important previous contributions in the literature.

III. A Random Attention Model

We are ready to introduce our RAM based on assumption 1. We assume that the decision maker has a strict preference ordering $\succ$ on X. To be precise, we assume that 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 that $y \not\succ x$; (ii) transitive, if for all $x, y, z \in X$, $x \succ y$ and $y \succ z$ imply that $x \succ z$; 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:

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

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

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

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

We now derive the implications of our RAM. They can be used to test the model in terms of observed choice rules or probabilities. In this section, we treat the choice rule as known/observed to facilitate the discussion of preference elicitation. In practice, the researcher may observe only a set of choice problems and choices thereof. We discuss econometric implementation in section IV: even if the choice rule is not directly observed, it is identified (consistently estimable) from choice data.

In the literature, there is a principle called "regularity" (see Suppes and Luce 1965), according to which adding a new alternative should only

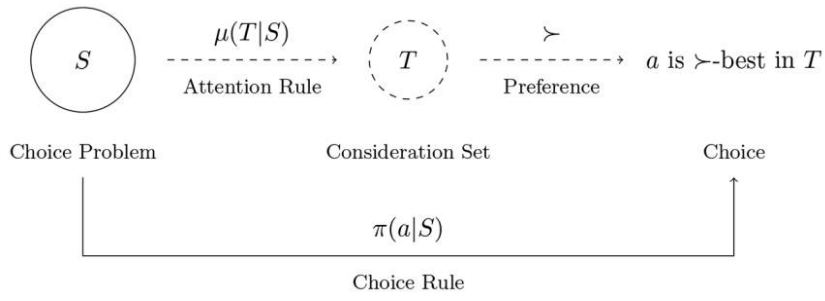

FIG. 1.—Illustration of a RAM. Observable: choice problem and choice (solid line). Unobservable: attention rule, consideration set, and preference (dashed line).

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 (multinomial logit) model or in any random utility model. In the 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 7 (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 the RAM can explain choice patterns that cannot be explained by the classical random utility model. Given that the model allows regularity violations, one might think that the model has very limited empirical implications—that is, it is too general to have empirical content. However, it is easy to find a choice rule π that lies outside the RAM with only three alternatives. Here we provide an example where our model makes very sharp predictions.

EXAMPLE 8 (RAM violation). The following choice rule π is not compatible with our RAM as long as the decision maker chooses each alternative with positive probability from the set $\{a, b, c\}$—that is, $\lambda_a\lambda_b\lambda_c > 0$. Each column corresponds to a different choice problem.

$\pi(\cdot \mid S)$	$S = \{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$
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 that $a \succ b \succ c$. 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 when the choice set is $\{a, b, c\}$ the decision maker must choose at least one alternative with zero probability, which is a stark prediction in probabilistic choice.

One might wonder that the model makes a strong prediction due to the cyclical binary choices—that is, $\pi(a|\{a, b\}) = \pi(b|\{b, c\}) = \pi(c|\{a, c\}) = 1$. We can generate a similar prediction where the individual is perfectly rational in the binary choices—that is, $\pi(a|\{a, b\}) = \pi(a|\{a, c\}) = \pi(b|\{b, c\}) = 1$. 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 obtain similar predictions. Given that the RAM has nontrivial 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.

A. *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 ensures that we never make false claims about the preference of the decision maker.

DEFINITION 4 (Revealed preference). Let $\{(\succ_j, \mu_j)\}_{j=1,\ldots,J}$ represent 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 if 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 $(\succ_j, \mu_j)$ 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 parallels that of Masatlioglu, Nakajima, and Ozbay (2012) in their study of a deterministic model of inattention. Masatlioglu, Nakajima, and Ozbay identify a as revealed to be 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 is revealed to be preferred to b if removing b causes a regularity violation—that is, $\pi(a|S) > \pi(a|S - b)$. To see this, assume that $(\succ, \mu)$ represents π and $\pi(a|S) > \pi(a|S - b)$. By definition, we have

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

where the second term in the third row follows from assumption 1. 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

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

By lemma 1, if $a\mathrm{P}b$, 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 $a\mathrm{P}b$ and $b\mathrm{P}c$ for some b, even when $a\mathrm{P}c$ is not directly revealed from her choices. Therefore, the transitive closure of P, denoted by $\mathrm{P_R}$, must also be part of her revealed preference. One may wonder whether some revealed preference is overlooked by $\mathrm{P_R}$. The following theorem, which is our first main result, shows that $\mathrm{P_R}$ 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 $a\mathrm{P_R}b$.

Proof. The "if" part follows from lemma 1. To prove the "only if" part, we show that given any preference $\succ$ that includes $\mathrm{P_R}$, there exists a monotonic attention rule μ such that $(\succ, \mu)$ represents π. The details of the construction can be found in the proof of theorem 2. QED

Theorem 1 establishes the empirical content of revealed preferences under monotonic attention only. Our resulting revealed preferences could be incomplete: it may provide only 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 fully attributed 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. The following example illustrates that we might make misleading inferences if we wrongly believe that the decision maker uses a particular attention rule.

EXAMPLE 9 (Avoiding misleading inference). We now describe a typical online customer's search behavior. For simplicity, there are three products a, b, and c. She prefers c over a and a over b (not observable). She visits two different search engines: G and Y. Eighty-five percent of her search takes place on engine G across three different platforms: laptop (20%), tablet (50%), and smartphone (15%). Engine G always lists b before a and a before c. Because of screen size, engine G lists up to three, two, and one pieces of product information on laptops, tablets, and smartphones, respectively. The rest of her search is on engine Y (15%), which has a unique platform. In this engine, a is listed first if it is available,

and clicking a's link will provide information about both a and c. If a is not available, b is listed first. In engine Y, she clicks only one link. (When she uses engine Y, her consideration set is $\{a, c\}$ when a and c are both available, $\{a\}$ when a is available but not c, and finally $\{b\}$ when only b and c are available.) On the basis of her underlying preference, the above consideration set formation leads to stochastic choice, the frequencies of which are reported in the following table:

$\pi(\cdot \mid S)$	$S = \{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$
a	.50	.85	.15	
b	.15	.15		.30
c	.35		.85	.70

Assume that we observe the customer's choice data without any knowledge about her underlying search behavior. First, note that the above choice data are consistent with the logit attention model of Brady and Rehbeck (2016).[3] In other words, we can apply their revealed preference result for this choice data. Their model, then, concludes that the unique revealed preference is $a \succ b \succ c$; however, this is not the true one that has generated the data. Therefore, if we make a mistaken assumption that the customer's behavior is in line with the logit model, we will infer that c is the worst alternative when it is the best product for our customer.

Example 9 is an example where a specific consideration set formation model leads to wrong conclusions on the revealed preferences. This example highlights the importance of knowledge about the underlying choice procedure when we conduct welfare analysis. In other words, welfare analysis is more delicate a task than it looks. Notice that in the above example, monotonic attention is satisfied as engines do not change their presentations of first-page results when an alternative outside of the first page becomes unavailable. Hence, theorem 1 is applicable. Since $\pi(a|\{a, b, c\}) > \pi(a|\{a, c\})$, our model correctly identifies her true preference between a and b. However, our model is silent about the relative ranking of c. Therefore, while our revealed preference is conservative, it does not make misleading claims.

We now illustrate that theorem 1 could be very useful for understanding 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 $\pi(a|\{a, b, c\}) > \pi(a|\{a, b\})$. Assume that we have no information about

[3] For example, letting the preference order be $a \succ b \succ c$ and letting weights be given as $w_{\{a\}} = 0$, $w_{\{b\}} = 1/20$, $w_{\{c\}} = 7/20$, $w_{\{a,b\}} = 17/60$, $w_{\{a,c\}} = 21/340$, $w_{\{b,c\}} = 1/10$, $w_{\{a,b,c\}} = 79/510$, it is easy to check that this is a logit attention representation of the choice data given above.

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, alternative c, which is also not comparable to b, is dominated by a. Theorem 1 informs us about the nature of products by observing only choice frequencies.

Our revealed preference result includes the one in Masatlioglu, Nakajima, and Ozbay (2012) for attention filters (i.e., nonrandom monotonic attention rules). 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 that 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 this result to nondeterministic 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 are no data revealing the entire preference. On the other hand, we illustrate that it is possible to reveal the entire preference in the RAM with only three alternatives. This discussion makes clear the connection between deterministic and probabilistic choice in terms of revealed preference.

EXAMPLE 10 (Full revelation). Consider the following stochastic choice with three alternatives:

$\pi(\cdot \mid S)$	$S = \{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$
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 that b is revealed to be preferred to c. Hence, preference is uniquely identified.

Example 10 also illustrates that one can achieve unique identification of preferences by utilizing assumption 1 even when observed choices cannot be explained by well-known models, such as the logit attention model of Brady and Rehbeck (2016) and the independent attention model of Manzini and Mariotti (2014). To see this point, assume that $\max\{1 - \lambda_a, \lambda_c\} > 0$. One can show that neither Brady and Rehbeck (2016) nor Manzini and Mariotti (2014) can explain observed choices in this example. First, notice that since both models satisfy assumption 1 and the preference is uniquely revealed as $a \succ b \succ c$ under assumption 1, if the observed choice data can be explained by either model, then their

revealed preference must also be $a \succ b \succ c$. That is, c must be the worst alternative. On the other hand, c is chosen with zero probability in $\{a, b, c\}$. These models then imply that c must also be chosen with zero probability in $\{a, c\}$ and $\{b, c\}$. This contradicts our assumption that $\max\{1 - \lambda_a, \lambda_c\} > 0$.

B. A Characterization

Theorem 1 characterizes the revealed preference in 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 the RAM? It turns out that the RAM can be simply characterized by only one behavioral postulate of choice: acyclicity. 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 revelation, 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.

Recall that example 8 is outside of our model. Theorem 2 implies that P_R must have a cycle. Indeed, we have $a\mathrm{P}b$ because of the regularity violation $\pi(a|\{a, b, c\}) = \lambda_a > 0 = \pi(a|\{a, c\})$. Similarly, we have $b\mathrm{P}c$ by $\pi(b|\{a, b, c\}) = \lambda_b > 0 = \pi(b|\{a, b\})$ and $c\mathrm{P}a$ by $\pi(c|\{a, b, c\}) = \lambda_c > 0 = \pi(c|\{b, c\})$. Since P has a cycle, example 8 must be outside of our model. Therefore, theorem 2 provides a very simple test of the RAM.

Our characterization result also helps us to understand the relation between our model and random utility models. It is well known in the literature that any choice rule that has a random utility model representation satisfies regularity. On the other hand, for any choice rule that satisfies regularity, P will trivially have no cycle. Hence, any choice rule that has a random utility model representation also has a RAM representation. However, in terms of modeling purposes, the RAM assumes random attention with a deterministic preference, whereas the random utility model assumes random preference and deterministic (full) attention.

Before closing this section, we sketch the proof of theorem 2 and provide a corollary that is used in the next section for developing econometric methods. The "only if" part of theorem 2 follows directly from lemma 1. For the "if" part, 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 and is not unique in general. We then illustrate that the constructed μ satisfies assumption 1. At the last step, we show that $(\succ, \mu)$ represents π. In

corollary 1, we provide one specific construction of the attention rule. We first make a definition.

DEFINITION 5 (Lower contour set; triangular attention rule). Given a preference ordering $\succ$ of the alternatives in X—$a_{1,\succ} \succ a_{2,\succ} \succ \ldots \succ a_{K,\succ}$—a lower contour set is defined as $L_{k,\succ} = \{a_{j,\succ} : j \geq k\} = \{a \in X : a \preccurlyeq a_{k,\succ}\}$. A triangular attention rule is an attention rule that puts weights only on lower contour sets. That is, $\mu(T|S) > 0$ implies that $T = L_{k,\succ} \cap S$ for some k such that $a_{k,\succ} \in S$.

COROLLARY 1 (Monotonic triangular attention rule representation). Assume that $(\succ, \mu)$ is a representation of π, with μ satisfying assumption 1. Then there is a unique triangular attention rule $\tilde{\mu}$ corresponding to $\succ$, which also satisfies assumption 1, such that $(\succ, \tilde{\mu})$ is a representation of π.

IV. 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, which can be used to design a specification test. The challenge for econometric implementation, however, is that our main assumption—monotonic attention—is imposed on the attention rule and that the attention rule is not identified from typical choice data and has a much higher dimension than the identified (consistently estimable) choice rule. To circumvent this difficulty, we rely on corollary 1, which states that if π has a random attention representation $(\succ, \mu)$, then there exists a unique monotonic triangular attention rule $\tilde{\mu}$ such that $(\succ, \tilde{\mu})$ 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.

A. *Nonparametric Identification*

We first define the set of partially identified preferences, which mirrors definition 3, with the only difference being that now we fix the choice rule to be identified/estimated from data. More precisely, let π represent the underlying choice rule/data generating process. Then a preference $\succ$ is compatible with π, denoted by $\succ \in \Theta_\pi$,[4] if there exists some monotonic attention rule μ such that $(\pi, \succ, \mu)$ 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 $a \succ b$,

[4] Θ_π is not the same as P_R (defined in sec. III.A): P_R contains all revealed preferences, while Θ_π is the set of preferences compatible with the choice probability (i.e., all possible

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 that the preference ordering is not compatible with the 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 the 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), searches the space of all monotonic attention rules, and checks whether definition 3 applies. The *J*-test procedure can be quite computationally demanding because the space of attention rules has high dimension. We further discuss the *J*-test approach in section SA.4.3 of appendix B, as well as how it is related to our proposed procedure.

One of the main purposes of this section is to provide an equivalent form of identification that (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 $S \in \mathcal{X}$. For example, one can label the choice problems as S_1, S_2, … and the alternatives as a_1, a_2, …, a_K, and then the vector $\boldsymbol{\pi}$ simply consists of $\pi(a_1|S_1)$, $\pi(a_2|S_1)$, …, $\pi(a_K|S_1)$, $\pi(a_1|S_2)$, $\pi(a_2|S_2)$, and so on. See example 11 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 $(\pi, \succ)$ has a RAM representation if and only if there exists a monotonic and triangular attention rule μ such that π is induced

completions of P_R). For example, when there is no preference revelation, Θ_π contains all preference orderings and P_R will be empty. For the other extreme—that the choice probability is not compatible with our RAM—Θ_π will be empty and P_R will involve cycles.

by μ and $\succ$ (corollary 1). With this fact, we are able to construct the constraint matrix $\mathbf{R}_\succ$ explicitly and write it as a product, $\mathbf{RC}_\succ$. 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 each row of $\mathbf{R}$ will consist of one $+1$, one -1, and zero otherwise. The product $\mathbf{R}\boldsymbol{\mu}$ then corresponds to $\mu(T|S) - \mu(T|S - a)$ for all S, $T \subset S$, and $a \in S - T$. That is, we use $\mathbf{R}\boldsymbol{\mu} \leq 0$ to represent assumption 1. Note that $\mathbf{R}$ does not depend on any preference.

Next consider $\mathbf{C}_\succ$. Given some preference $\succ$ and the choice rule π, the only possible triangular attention rule that can be constructed is

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

(see corollary 1 and the proof of theorem 2 in app. A), where $\{L_{k,\succ} : 1 \leq k \leq K\}$ are the lower contour sets corresponding to the preference ordering $\succ$ (definition 5). The above defines the mapping $\mathbf{C}_\succ$ and 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 that $\mathbf{R}_\succ$ is uniquely determined by the preference $\succ$. QED

This theorem states that 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. An algorithmic construction of the constraint matrix $\mathbf{R}_\succ$ is given in algorithm 1.

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

As can be seen, the only input needed is the preference $\succ$, which we are interested in testing against. Each row of $\mathbf{R}_\succ$ consists of one $+1$, one -1,

and zero otherwise. The constraint matrix $\mathbf{R}_{\succ}$ is nonrandom and does not depend on the estimated choice probabilities but rather is determined by the collection of (fixed, known to the researcher) restrictions on the estimable choice probabilities. 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):

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

where $K = |X|$ denotes 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. However, once the matrix $\mathbf{R}_{\succ}$ has been constructed for one preference $\succ$, the constraint matrices for other preference orderings can be obtained by column permutations of $\mathbf{R}_{\succ}$. This is particularly useful and saves computation if there are multiple hypotheses to be tested, as the above algorithm needs to be implemented only once.

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

EXAMPLE 11 ($\mathbf{R}_{\succ}$ with three alternatives). Assume that there are three alternatives—a, b, and c—in X; then the choice rule is represented by a vector in $\mathbb{R}^9$:

$$\pi = [\underbrace{\pi(\cdot|\{a,b,c\})}_{\in\mathbb{R}^5}, \underbrace{\pi(\cdot|\{a,b\})}_{\in\mathbb{R}^2}, \underbrace{\pi(\cdot|\{a,c\})}_{\in\mathbb{R}^2}, \underbrace{\pi(\cdot|\{b,c\})}_{\in\mathbb{R}^2}]',$$

where, for ease of presentation, trivial cases such as $\pi(a|\{b,c\}) = 0$ and $\pi(b|\{b\}) = 1$ are ignored. Now consider the preference/hypothesis $b \succ a \succ c$. 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:

$$[\,1 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; -1 \;\; 0 \;\; 0 \;\; 0\,].$$

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 $\mathbf{R}_{b\succ a\succ c}$:

$$[\,0 \;\; 0 \;\; 1 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; -1\,].$$

The row corresponding to $b \succ c$ is

$$[\,0 \;\; 0 \;\; 1 \;\; 0 \;\; 0 \;\; 0 \;\; -1 \;\; 0 \;\; 0\,].$$

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. 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.

B. *Hypothesis Testing*

Given the identification result in theorem 3, we can replace the 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 (Data generating process). The data are 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 assume only that the data are generated from some choice rule π. We allow for the possibility that it is not a 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 assume only 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 fig. 1). For simplicity, we discuss the case of "complete data," where all choice problems are potentially observable, but in sections SA.3 and SA.4.4 of appendix B we extend our work to the case of incomplete data.

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

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

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 $\sigma_{\pi,\succ}$ denote the standard deviation of $\mathbf{R}_{\succ}\hat{\pi}$ and $\hat{\sigma}_{\succ}$ denote its plug-in estimate. That is,

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

where $\operatorname{diag}(\cdot)$ denotes the operator that extracts the diagonal elements of a square matrix or constructs a diagonal matrix when applied to a vector. Here Ω_π is block diagonal, with blocks given by $\{1/[\mathbb{P}(Y_i = S)]\}\Omega_{\pi,S}$ and $\Omega_{\pi,S} = \operatorname{diag}(\pi_S) - \pi_S \pi_S'$. The estimator $\hat{\Omega}$ is constructed simply 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 that 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 (e.g., via employing associated p-values). Given a specific preference, the test statistic is constructed as the maximum of the Studentized, restricted sample choice probabilities:

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

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 $\mathbf{R}_\succ \pi \leq \mathbf{0}$, 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 the 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 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 section SA.4.5 of appendix B.

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

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

By the central limit theorem, the first component $\sqrt{N}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})$ is approximately distributed as $\mathcal{N}(\mathbf{0}, \Omega_{\pi})$. 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

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

Here $\mathbf{z}^*$ is a random vector simulated from the distribution $\mathcal{N}(\mathbf{0}, \widehat{\Omega}/N)$, and $\sqrt{N}\psi_N(\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}}, \hat{\sigma}_{\succ})$ is used to replace the unknown moment conditions $(\sqrt{N}\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}}) \oslash \hat{\sigma}_{\succ}$. Several choices of ψ_N have been proposed. One extreme choice is $\psi_N(\cdot) = 0$, so that the upper bound zero 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

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

where $(\mathbf{a})_- = \mathbf{a} \odot \mathbb{I}(\mathbf{a} \le 0)$, with $\odot$ denoting the Hadamard product, the indicator function $\mathbb{I}(\cdot)$ operating elementwise on the vector $\mathbf{a}$, and κ_N diverging slowly. That is, the function $\psi_N(\cdot)$ retains the nonpositive elements of $(\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}} \oslash \hat{\sigma}_{\succ})/\kappa_N$, 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 VI. For other choices of $\psi_N(\cdot)$, see Andrews and Soares (2010).

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

$$c_{\alpha}(\succ) = \inf\left\{t : \frac{1}{M}\sum_{m=1}^{M}\mathbb{I}(\mathcal{T}^*_m(\succ) \le t) \ge 1 - \alpha\right\},$$

and the null hypothesis $\mathrm{H}_0 : \succ \in \Theta_{\pi}$ is rejected if and only if $\mathcal{T}(\succ) > c_{\alpha}(\succ)$. Alternatively, one can compute the p-value as

$$p(\succ) = \frac{1}{M}\sum_{m=1}^{M}\mathbb{I}(\mathcal{T}^*_m(\succ) > \mathcal{T}(\succ)).$$

To justify the proposed critical values, it is important to address uniformity issues. 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. Testing procedures that are valid only

pointwise but not uniformly may yield bad approximations to the finite sample distribution, because in finite samples the moment inequalities could be close to binding. The following theorem shows that conducting inference using the critical values above is uniformly valid.

THEOREM 4 (Uniformly valid testing). Assume that assumption 2 holds. Let Π represent a class of choice rules and $\succ$ a preference, such that (i) for each $\pi \in \Pi$, $\succ \in \Theta_\pi$, and (ii) $\inf_{\pi\in\Pi} \min(\sigma_{\pi,\succ}) > 0$. Then,

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

The proof is given in section B of appendix A. 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.

C. *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.

1. 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:

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

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

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

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).

2. Testing Model Compatibility: $H_0 : \mathcal{P} \cap \Theta_\pi \neq \varnothing$

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

For a concrete example, consider the question of whether $a \succ b$ is compatible with the data generating process. As long as there are more than two 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 $\mathcal{P}$:

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

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 $\mathcal{P} \cap \Theta_\pi \neq \varnothing$, 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 $H_0 : \Theta_\pi \neq \varnothing$ and is rejected on the basis of the following rule:

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

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).

V. Incorporating Additional Restrictions

Our identification and inference results so far are obtained using the 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 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.

A. *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 $\phi \geq 1/2$.[5] For example, when $\phi = 2/3$, it means that choosing a twice as often as choosing b implies that a is better than b. The parameter ϕ represents how cautious the policy maker is. Denote by

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

To justify 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 $a, b \in X$ and $\phi \geq 1/2$,

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

The quantity $(1 - \phi)/\phi$ is a measure of full attention at binaries. When $(1 - \phi)/\phi = 0$ (or $\phi = 1$), there is no constraint on $\mu(\{a, b\}|\{a, b\})$. In this case, it is possible that the decision maker considers only singleton consideration sets. When $(1 - \phi)/\phi$ gets larger (or ϕ gets smaller), the probability of being fully attentive is strictly positive, which creates room for preference revelation. An alternative way to understand assumption 3 is as follows. Take $\phi = \max\{\pi(a|\{a, b\}), \pi(b|\{a, b\})\}$; then $[(1 - \phi)/\phi] \max\{\mu(\{a\}|\{a, b\}), \mu(\{b\}|\{a, b\})\}$ is a strict lower bound on the amount of attention that the decision maker has to pay to both options for revelation to occur.

We now illustrate that, under assumption 3, if $\pi(a|\{a, b\}) > \phi$, then a is revealed to be preferred to b. Let $(\succ, \mu)$ 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 that $\mu(\{a\}|\{a, b\}) > \phi$. By assumption 3, we must have $\mu(\{a, b\}|\{a, b\}) >$

[5] Even when the policy maker is least cautious, we need $\pi(a|\{a, b\}) > \pi(b|\{a, b\})$ to conclude that a is strictly better than b. This implies that $\pi(a|\{a, b\}) > 1/2$. Hence, ϕ must be greater than $1/2$.

$1 - \phi$, 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 12 (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, $\mathrm{P} = \mathrm{P}_R = \varnothing$ (sec. III.A). 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 $a\mathrm{P}^\phi b$ and $b\mathrm{P}^\phi c$. Notice that $\pi(a|\{a, c\}) = 1/2$, and hence we cannot directly deduce $a\mathrm{P}^\phi c$. Since the underlying preference is transitive, we can conclude that the decision maker prefers a to c as $a\mathrm{P}^\phi b$ and $b\mathrm{P}^\phi c$, even when $a\mathrm{P}^\phi c$ is not directly revealed from her choices. Therefore, the transitive closure of P^ϕ, denoted by P^ϕ_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 that $\phi < 2/3$.

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

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

The relation $\mathrm{P}^\phi \cup \mathrm{P}$ 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 $\mathrm{P}^\phi \cup \mathrm{P}$ or its transitive closure $(\mathrm{P}^\phi \cup \mathrm{P})_R$ provides a simple characterization of the model we consider in this subsection.

THEOREM 5 (Characterization). For a given $\phi \geq 1/2$, a choice rule π has a random attention representation $(\succ, \mu)$ where μ satisfies assumptions 1 and 3 if and only if $\mathrm{P}^\phi \cup \mathrm{P}$ has no cycle.

For $\phi < 1$, the model characterized by theorem 5 has a higher predictive power (i.e., empirical content) compared with the model characterized by theorem 2. Hence, the model will fail to retain some of its explanatory

power. For example, example 10 with $\lambda_a, \lambda_b, \lambda_c < 1 - \phi$ is outside of the model given here.

Under the assumption $\phi = 1/2$ and $\pi(a|\{a, b\}) \neq 1/2$ for all a, b, theorem 5 yields that our framework reveals a unique preference while it allows regularity violation.

REMARK 1 (Acyclic stochastic transitivity). We 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:

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

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

Now we discuss the econometric implementation. Recall from section IV that to test whether a specific preference ordering is compatible with the observed (identifiable) choice rule and the monotonicity assumption, we first construct a triangular attention rule and then test whether the triangular attention rule satisfies assumption 1. This is formally justified in the proof of theorem 3.

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 assumptions 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 $[(1 - \phi)/\phi]\pi(b|S) - \pi(a|S) \leq 0$
  end if
end for

We now revisit example 11 to illustrate what additional (identifying) restrictions are imposed by assumption 3.

EXAMPLE 13 (Example 11 continued). Recall that there are three alternatives—a, b, and c—in X, and the choice rule is represented by a vector in $\mathbb{R}^9$. For the preference $b \succ a \succ c$, the matrix $\mathbf{R}_{b\succ a\succ c}$ contains three restrictions if only assumption 1 is imposed. With our new restriction on the attention rule for binary choice problems, $\mathbf{R}_{b\succ a\succ c}$ is further augmented:

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

where the first three rows correspond to restrictions imposed by assumption 1 and the last three rows capture our new assumption 3.

Assumption 3 considerably improves 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 VI, using simulations we compare the empirical content of our benchmark RAM, which employs only assumption 1, and the model that incorporates assumption 3 as well.

B. Random Attention Filter

We now consider random attention filters, which are one of the motivating examples of monotonic attention rules. Recall from section II 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 $(\pi, \succ, \mu)$ is a RAM with μ being a monotonic attention rule, there exists a random attention filter μ' such that $(\pi, \succ, \mu')$ is still a RAM (see remark 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 convex combination of attention filters.

EXAMPLE 14. 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 that 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, $\mu(\{a_2\}|\{a_1, a_2, a_3\}) = 0.5$ and $\mu(\{a_2\}|\{a_1, a_2, a_4\}) = 0.5$ imply that for all Γ that 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 that $\Gamma(\{a_1, a_2, a_3\}) = \{a_3\}$ for all Γ with $\Gamma(\{a_1, a_2, a_3, a_4\}) = \{a_3\}$ and that $\Gamma(\{a_1, a_2, a_4\}) = \{a_4\}$ for all Γ with $\Gamma(\{a_1, a_2, a_3, a_4\}) = \{a_4\}$. However, it must then be the case that $\Gamma(\{a_1, a_2\}) = \{a_2\}$ for all Γ that are assigned positive probability or that $\mu(\{a_2\}|\{a_1, a_2\}) = 1$, which is a contradiction.

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 5), 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 $T \subset S$,

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

REMARK 2 (Triangular random attention filter representation). Combining this theorem and corollary 1 in appendix A, we easily reach the following conclusion: if π has a random attention representation $(\succ, \mu)$,

then there exists a triangular random attention filter μ' such that $(\succ, \mu')$ also represents π.

The proof of theorem 6 is long and hence left to section D of appendix A, 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 $\mu', \mu'' \in \mathcal{MT}(\succ)$.) Minkowski's theorem then 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 nondeterministic triangular attention rules cannot be extreme points; that is, given any $\mu \in \mathcal{MT}(\succ) - \mathcal{AF}(\succ)$, we can construct $\mu', \mu'' \in \mathcal{MT}(\succ)$, such that $\mu = (1/2)\mu' + (1/2)\mu''$. The key step is to show that both the μ' and the 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.

VI. 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:

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

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, and 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 that the underlying preference is $a_1 \succ a_2 \succ a_3 \succ a_4 \succ a_5$. Second, the data consist of choice problems of size two, three, four, and five. That is, there are 26 choice problems in total. 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. We also report simulation evidence for $\varsigma \in \{0, 1\}$ in appendix B. Finally, the observed data are a random sample $\{(y_i, Y_i) : 1 \le i \le N\}$, where the effective sample size can be 50, 100, 200, 300, and 400. (Effective sample

size refers to the number of observations for each choice problem. Because there are 26 choice problems, the overall sample size is $N \in \{1{,}300, 2{,}600, 5{,}200, 7{,}800, 10{,}400\}$.)

For inference, we employ the procedure introduced in section IV 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 V.A. 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 indicate 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. The hypothesis $H_{0,2}$, 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 one, assumption 3 no longer brings too much additional identification power beyond the monotonic attention assumption, and monotonic attention alone is not strong enough to reject this hypothesis. Indeed, with $\phi = 1$ (i.e., assumption 1 alone), the set of identified preferences 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 $\phi = 1$.

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 2.

We first focus on $H_{0,1}$ (fig. 2*A*). As this preference ordering is compatible with our RAM, one should expect the rejection probability to be lower than the nominal level. Indeed, the rejection probability is far below .05: this illustrates a generic feature of any (reasonable) procedure for testing

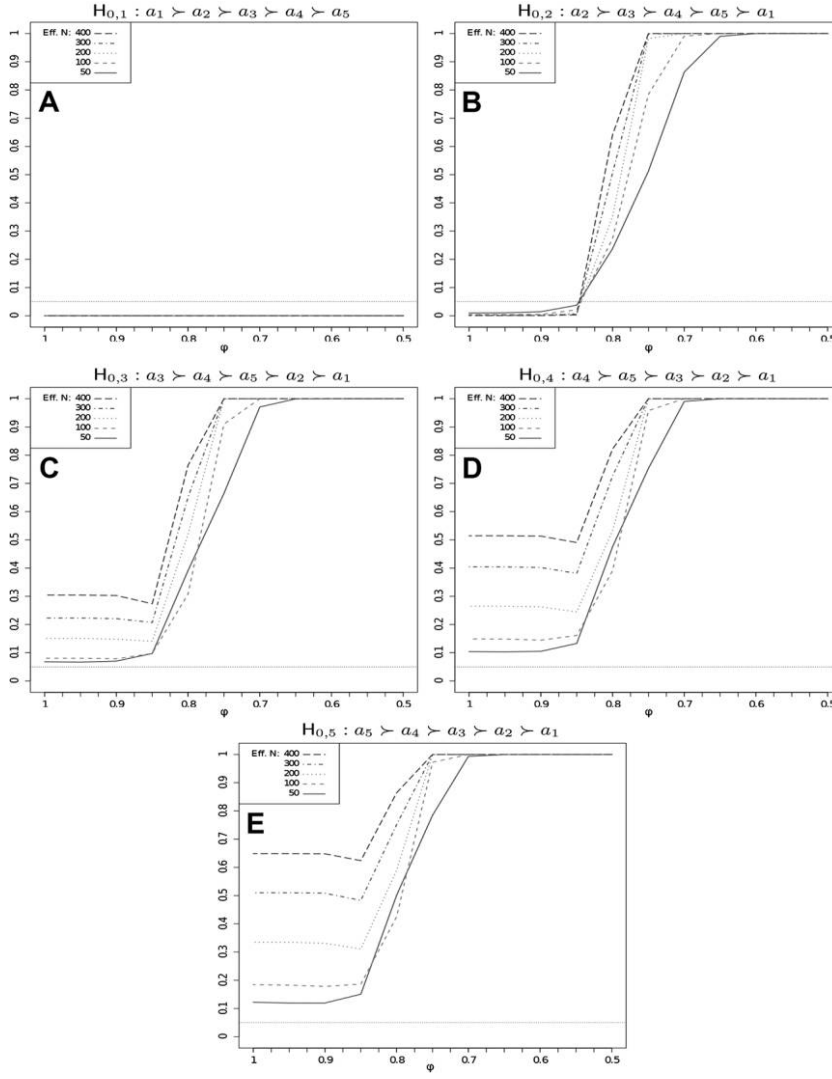

FIG. 2.—Empirical rejection probabilities. Shown in the figure are empirical rejection probabilities testing the five null hypotheses through 5,000 simulations, with nominal size 0.05. Logit attention rule with $\varsigma = 2$ is used, as described in the main text. For each simulation repetition, five effective sample sizes are considered: 50, 100, 200, 300, and 400. A color version of this figure is available online.

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 $H_{0,2}$ (fig. 2*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 (fig. 2*C*–2*E*) are not compatible with our RAM, and we do see that the rejection probability is much larger than the nominal size even for $\phi = 1$, showing that even our basic RAM has nontrivial empirical content in this case.

VII. Conclusion

We introduced a limited attention model allowing for a general class of monotonic (and possibly stochastic) attention rules, which we called a random attention model (RAM). We showed 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, we obtained a testable theory of revealed preferences and developed partial identification results for the decision maker's unobserved strict preference ordering. Our results included a precise constructive characterization of the identified set for preferences, as well as uniformly valid inference methods based on that characterization. Furthermore, we showed how additional nonparametric restriction can be easily incorporated into the RAM to obtain tighter empirical implications and more powerful accompanying econometric procedures. We found good finite sample performance of our econometric methods in a simulation experiment. Last but not least, we provide the general-purpose R software package ramchoice, which allows other researchers to easily employ our econometric methods in empirical applications.

Appendix A

Omitted Proofs

This appendix collects proofs that are omitted from the main text to improve the exposition.

A. *Proof of Theorem 2*

Suppose that π has a random attention representation $(\succ, \mu)$. 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 $\succ$: $a_{1,\succ} \succ a_{2,\succ} \succ \ldots \succ a_{K,\succ}$. Let $\{L_{k,\succ} : 1 \leq k \leq K\}$ be the corresponding lower contour sets (definition 5). Then we specify $\tilde{\mu}$ as

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

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

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

which is the same as π. For the first equality, we use the definition that a triangular attention rule puts weights only 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 that this is not the case; then it means that there exist 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, statement i implies that $a_{\ell,\succ} \succ a_{k,\succ}$. Then statement ii implies that

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

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

Remark A1. 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, $(\succ, \mu)$, 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 1.

B. Proof of Theorem 4

See section SA.4.1 of appendix B.

C. Proof of Theorem 5

The "only if" part is trivial and is omitted. We illustrate the "if" part. Assume that $\mathrm{P}^{\phi} \cup \mathrm{P}$ has no cycle (or equivalently, that its transitive closure $(\mathrm{P}^{\phi} \cup \mathrm{P})_{\mathrm{R}}$ has no cycle); then there exists some preference ordering that embeds $\mathrm{P}^{\phi} \cup \mathrm{P}$. 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 that $\pi(a|\{a, b\}) < [(1 - \phi)/\phi]\pi(b|\{a, b\})$ and equivalently that $\pi(b|\{a, b\}) > \phi$. This means that $b\mathrm{P}^{\phi}a$, which violates our definition of $\succ$.

D. *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 nondeterministic 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 $S \in \mathcal{X}$ for which $\mu(T|S) = 1$ for no $T \subset S$. We start by choosing $\varepsilon > 0$ small enough so that none of the nonbinding 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 D1. 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 that there exist three such subsets: T_1, T_2, and T_3. Since μ is triangular, the subsets that are considered with positive probability can be ordered by set inclusion. Hence, we can assume that $T_1 \subset T_2 \subset T_3$ without loss of generality. 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, $T_2 = S$. QED

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 that 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 D2. Suppose that 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 D1 when $|S| = k_\mu + 1$. Suppose that 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 that 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. QED

LEMMA D3. 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 that 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 that satisfies the description in the construction, then no subset will be affected. Suppose that 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, $\Sigma_{T'' \subset T}\mu(T''|S) < 1$ 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 that there are three subsets, T_1, T_2, and T_3, satisfying the description. Since μ is triangular, we can assume that $T_1 \subset T_2 \subset T_3$ without loss of generality. By the previous lemma, we can assume that $S_{T_2} = T_2$ and $S_{T_3} = T_3$ without loss of generality. But then since μ is monotonic, three subsets of S_{T_3} must satisfy the description, which is 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 that 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. QED

It is clear that $\mu = (1/2)\mu' + (1/2)\mu''$. 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, Jason, and Abi Adams. 2017. "What Do Consumers Consider before They Choose? Identification from Asymmetric Demand Responses." Working Paper no. 23566, NBER, Cambridge, MA.

Agranov, Marina, and Pietro Ortoleva. 2017. "Stochastic Choice and Preferences for Randomization." *J.P.E.* 125 (1): 40–68.

Aguiar, Victor H. 2015. "Stochastic Choice and Attention Capacities: Inferring Preferences from Psychological Biases." SSRN Working Paper no. 2607602, Soc. Sci. Res. Network.

Aguiar, Victor H., María José Boccardi, and Mark Dean. 2016. "Satisficing and Stochastic Choice." *J. Econ. Theory* 166:445–82.

Andrews, Donald W. K., and Gustavo Soares. 2010. "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection." *Econometrica* 78 (1): 119–57.

Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C. Teitelbaum. 2018. "Heterogeneous Choice Sets and Preferences." CEMMAP Working Paper no. CWP37/19, Centre Microdata Methods and Practice, London.

Blundell, Richard, Dennis Kristensen, and Rosa Matzkin. 2014. "Bounding Quantile Demand Functions Using Revealed Preference Inequalities." *J. Econometrics* 179 (2): 112–27.

Brady, Richard L., and John Rehbeck. 2016. "Menu-Dependent Stochastic Feasibility." *Econometrica* 84 (3): 1203–23.

Bugni, Federico A. 2016. "Comparison of Inferential Methods in Partially Identified Models in Terms of Error in Coverage Probability." *Econometric Theory* 32 (1): 187–242.

Bugni, Federico A., Ivan A. Canay, and Xiaoxia Shi. 2015. "Specification Tests for Partially Identified Models Defined by Moment Inequalities." *J. Econometrics* 185 (1): 259–82.

Canay, Ivan A. 2010. "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity." *J. Econometrics* 156 (2): 408–25.

Canay, Ivan A., and Azeem M. Shaikh. 2017. "Practical and Theoretical Advances for Inference in Partially Identified Models." In *Advances in Economics and Econometrics*, vol. 2, *Eleventh World Congress*, edited by Bo Honoré, Ariel Pakes, Monika Piazzesi, and Larry Samuelson, 271–306. Cambridge: Cambridge Univ. Press.

Dardanoni, Valentino, Paola Manzini, Marco Mariotti, and Christopher J. Tyson. 2020. "Inferring Cognitive Heterogeneity from Aggregate Choices." *Econometrica* 88 (3): 1269–96.

Dean, Mark, Özgür Kıbrıs, and Yusufcan Masatlioglu. 2017. "Limited Attention and Status Quo Bias." *J. Econ. Theory* 169:93–127.

Deb, Rahul, Yuichi Kitamura, John K. H. Quah, and Jörg Stoye. 2018. "Revealed Price Preference: Theory and Stochastic Testing." Working Paper no. 2087, Cowles Found. Res. Econ., New Haven, CT.

Echenique, Federico, and Kota Saito. 2019. "General Luce Model." *Econ. Theory* 68:811–26.

Echenique, Federico, Kota Saito, and Gerelt Tserenjigmid. 2018. "The Perception-Adjusted Luce Model." *Math. Soc. Sci.* 93:67–76.

Fishburn, Peter C. 1973. "Binary Choice Probabilities: On the Varieties of Stochastic Transitivity." *J. Math. Psychology* 10 (4): 327–52.

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki. 2015. "Stochastic Choice and Revealed Perturbed Utility." *Econometrica* 83 (6): 2371–409.

Gaundry, Marc J. I., and Marcel G. Dagenais. 1979. "The Dogit Model." *Transportation Res. B* 13 (2): 105–11.

Goeree, Michelle Sovinsky. 2008. "Limited Information and Advertising in the US Personal Computer Industry." *Econometrica* 76 (5): 1017–74.

Gul, Faruk, Paulo Natenzon, and Wolfgang Pesendorfer. 2014. "Random Choice as Behavioral Optimization." *Econometrica* 82 (5): 1873–912.

Hauser, John R., and Birger Wernerfelt. 1990. "An Evaluation Cost Model of Consideration Sets." *J. Consumer Res.* 16 (4): 393–408.
Hausman, Jerry A., and Whitney K. Newey. 2017. "Nonparametric Welfare Analysis." *Ann. Rev. Econ.* 9:521–46.
Ho, Kate, and Adam M. Rosen. 2017. "Partial Identification in Applied Research: Benefits and Challenges." In *Advances in Economics and Econometrics*, vol. 2, *Eleventh World Congress*, edited by Bo Honoré, Ariel Pakes, Monika Piazzesi, and Larry Samuelson, 307–59. Cambridge: Cambridge Univ. Press.
Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino. 2017. "Advertising, Consumer Awareness, and Choice: Evidence from the US Banking Industry." *RAND J. Econ.* 48 (3): 611–46.
Horan, Sean. 2019. "Random Consideration and Choice: A Case Study of 'Default' Options." *Math. Soc. Sci.* 102:73–84.
Houthakker, Hendrik S. 1950. "Revealed Preference and the Utility Function." *Economica* 17 (66): 159–74.
Huber, Joel, John W. Payne, and Christopher Puto. 1982. "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *J. Consumer Res.* 9 (1): 90–98.
Imbens, Guido W., and Charles F. Manski. 2004. "Confidence Intervals for Partially Identified Parameters." *Econometrica* 72 (6): 1845–57.
Kaido, Hiroaki, Francesca Molinari, and Joerg Stoye. 2019. "Confidence Intervals for Projections of Partially Identified Parameters." *Econometrica* 87 (4): 1397–432.
Kawaguchi, Kohei. 2017. "Testing Rationality without Restricting Heterogeneity." *J. Econometrics* 197 (1): 153–71.
Kawaguchi, Kohei, Kosuke Uetake, and Yasutora Watanabe. 2016. "Identifying Consumer Attention: A Product-Availability Approach." SSRN Working Paper no. 2529294, Soc. Sci. Res. Network.
Kitamura, Yuichi, and Jörg Stoye. 2018. "Nonparametric Analysis of Random Utility Models." *Econometrica* 86 (6): 1883–909.
Lleras, Juan Sebastian, Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay. 2017. "When More Is Less: Limited Consideration." *J. Econ. Theory* 170:70–85.
Manzini, Paola, and Marco Mariotti. 2014. "Stochastic Choice and Consideration Sets." *Econometrica* 82 (3): 1153–76.
Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay. 2012. "Revealed Attention." *A.E.R.* 102 (5): 2183–205.
Matzkin, Rosa L. 2013. "Nonparametric Identification in Structural Economic Models." *Ann. Rev. Econ.* 5:457–86.
Molinari, Francesca. 2020. "Microeconometrics with Partial Identification." In *Handbook of Econometrics*, vol. 7A, edited by Steven Durlauf, Lars Hansen, James Heckman, and Rosa Matzkin, forthcoming. Amsterdam: North-Holland.
Reutskaja, Elena, Rosemarie Nagel, Colin F. Camerer, and Antonio Rangel. 2011. "Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study." *A.E.R.* 101 (2): 900–926.
Rieskamp, Jörg, Jerome R. Busemeyer, and Barbara A. Mellers. 2006. "Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice." *J. Econ. Literature* 44 (3): 631–61.
Suppes, Patrick, and R. D. Luce. 1965. "Preference, Utility, and Subjective Probability." In *Handbook of Mathematical Psychology*, vol. 3, edited by Duncan R. Luce, Robert R. Bush, and Eugene Galanter, 249–410. New York: Wiley.
Tversky, Amos. 1969. "Intransitivity of Preferences." *Psychological Rev.* 76 (1): 31–48.

———. 1972. "Elimination by Aspects: A Theory of Choice." *Psychological Rev.* 79 (4): 281.

van Nierop, Erjen, Bart Bronnenberg, Richard Paap, Michel Wedel, and Philip Hans Franses. 2010. "Retrieving Unobserved Consideration Sets from Household Panel Data." *J. Marketing Res.* 47 (1): 63–74.