Attention Overload*

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

September 9, 2026

Abstract

We introduce an Attention Overload Model (AOM) in which alternatives compete for attention, so each alternative's consideration probability weakly decreases as the choice problem expands. This nonparametric restriction has a search-capacity foundation. We identify exactly which preference rankings are compatible with observed choices and establish sharp bounds on latent attention. We then consider heterogeneous preferences in settings where alternatives are presented in a list, establishing nonparametric identification results for attention and the distribution of preferences. We also develop high-dimensional inference and finite-sample methods for these models. Using the travel-mode experiment of Wang and Zhu (2025), we illustrate the methods and find that recovered pairwise preference shares closely match subjects' self-reported rankings.

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

*We thank Chris Chambers, Emel Filiz-Ozbay, Whitney Newey, Erkut Ozbay, and seminar participants at Bristol, Cornell, LSE, MIT, Pittsburgh, Princeton, Rice, Rutgers, UC Davis, UCLA, and conference participants at the 2020 Risk, Uncertainty & Decision Conference, the 2020 Southern Economic Association Annual Meeting, the 2022 Cowles Foundation Econometrics Conference, and the 2023 Science of Decision Making Conference for their comments. We also thank Zexuan Wang and Dianzhuo Zhu for generously sharing their experimental data. The editor and four anonymous reviewers provided excellent comments and suggestions. Cattaneo gratefully acknowledges financial support from the National Science Foundation (SES-2241575), and the John Simon Guggenheim Memorial Foundation through a 2026 Guggenheim Fellowship.

†Department of Economics, Princeton University.

‡Jindal School of Management, University of Texas at Dallas.

§Department of Economics, UC San Diego.

¶Department of Economics, University of Maryland.

1 Introduction

This paper studies choice when alternatives compete for limited attention. We define an alternative's *attention frequency* as the probability that it enters the consideration set and impose *attention overload*: holding an alternative fixed, its attention frequency cannot increase when more rivals become available. If attention were deterministic, an item considered in a large supermarket would also be considered in a smaller store containing that item. The restriction is nonparametric and is motivated by evidence that larger choice sets can dilute item-level inspection when time and cognitive resources are limited (Reutskaja and Hogarth, 2009; Visschers, Hess, and Siegrist, 2010; Reutskaja, Nagel, Camerer, and Rangel, 2011; Geng, 2016; Thomas, Molter, and Krajbich, 2021). For example, Reutskaja, Nagel, Camerer, and Rangel (2011) show that the mean share of items viewed falls as the number of alternatives increases. Thomas, Molter, and Krajbich (2021) provide complementary eye-tracking evidence that the probability that an individual item is inspected decreases as the choice set expands. We investigate the implications of this restriction and characterize what observed choices reveal about underlying preference orderings and attention frequencies.

Our identification strategy exploits menu variation in settings where preferences remain stable: changes in inventory or eligibility, randomized display sets, search results, product assortments, or repeated cross-sections exposed to different subsets of a common universe.[1] Attention overload is especially well suited to environments in which limited time and cognitive resources cause alternatives to compete for inspection. Its directional content makes the restriction both economically interpretable and directly testable, allowing the data to distinguish attention dilution from settings in which menu expansion raises item-level consideration. In the appendix, we show that the restriction arises naturally from a familiar mechanism featuring heterogeneous search orders and cognitive capacities.

Existing attention models restrict attention behavior by imposing assumptions on the attention rule. We differ from this literature by placing attention frequency at the center of our analysis, an empirically relevant object (Reutskaja, Nagel, Camerer, and Rangel, 2011; Caplin, Dean, and

[1] We do not treat changes in prices or product characteristics as menu variation. Comparability can be maintained by conditioning on these characteristics or redefining alternatives as product–state pairs. Section 5 uses stable travel-mode categories under its stated assumptions on attribute profiles and preferences.

Martin, 2011; Thomas, Molter, and Krajbich, 2021). Although existing models do not impose restrictions directly on this object, their assumptions nevertheless have implications for how attention frequencies vary across menus. We show that these implications need not align with Attention Overload. In the Independent Attention model of Manzini and Mariotti (2014) and the Elimination-by-Aspects model of Aguiar (2017), attention frequencies are menu invariant, so these models correspond to a knife-edge case of Attention Overload. By contrast, the parametric model of Brady and Rehbeck (2016) and the nonparametric Random Attention Model (RAM) of Cattaneo, Ma, Masatlioglu, and Suleymanov (2020) can generate attention frequencies that increase as the menu expands, the opposite of the comparative static underlying Attention Overload. These models therefore do not encode the empirically motivated dilution force that is central to our framework. Because the models restrict different features of attention, they can rationalize different choice data, reveal different preferences, and address different empirical questions. Section 2.4 makes these distinctions explicit.

Motivated by these theoretical developments, a growing empirical literature establishes both the relevance and the testability of random attention. Experiments that observe consideration directly document widespread stochastic limited consideration, with monotonicity among the more demanding implications (Ellis, Filiz-Ozbay, and Ozbay, 2025; Chadd, 2023; Khan, Krajbich, Raymond, Sundaresan, and Veiga, 2026). Wang and Zhu (2025) find that RAM accepts the preference ranking implied by subjects' self-reports in a travel-mode experiment while rejecting rankings that place an objectively dominated option first; a health-preference experiment likewise finds largely consistent revealed preferences under RAM and RUM (Wang, Sicsic, and Chauvin, 2026). Choice overload experiments and field evidence also connect larger menus to limited or costly search and reduced purchase frequency (Dean, Ravindran, and Stoye, 2025; de Lara and Dean, 2024; Natan, 2024). Frame- and covariate-based studies reject full-attention random utility in some settings, support latent consideration, and show that accounting for heterogeneous choice sets can materially affect recovered preferences and welfare conclusions (Aguiar, Boccardi, Kashaev, and Kim, 2023; Abaluck and Adams, 2021; Barseghyan, Coughlin, Molinari, and Teitelbaum, 2021; Barseghyan, Molinari, and Thirkettle, 2021). Together, they show that limited consideration is empirically consequential.[2]

[2]Section SA-1 provides a fuller discussion of this evidence and broader comparisons with related models.

Our first contribution is a revealed-preference and revealed-attention theory for AOM under homogeneous preferences. We show that a candidate preference represents the observed choice rule if and only if $\succ$-Regularity holds, and we give a direct construction of a rationalizing attention rule. $\succ$-Regularity relaxes classic regularity of stochastic choice in a way that captures the attention-dilution: inferior alternatives may become more likely to be chosen as the menu expands. The characterization sharply identifies all compatible preferences. It also yields computationally attractive implications based on binary comparisons and a mixed-integer formulation that represents the identified set implicitly without enumerating all $|X|!$ rankings. We derive sharp component-wise bounds on attention frequencies, which is the main interest of this paper. Thus, the model recovers an economically meaningful feature of latent attention without parametrizing the consideration-set distribution.

Our second contribution is a new treatment of heterogeneous preferences. We introduce a list-based attention-overload model with heterogeneous preferences (H-LAO) for environments in which the researcher observes a presentation order, such as ranked search results or advertisements. H-LAO permits an unrestricted marginal distribution over strict preferences and imposes Sequential Path Independence (SPI) on a *list-based* attention rule. SPI is a cross-menu restriction requiring that the probability of reaching the next item on a list equal the probability of reaching the current item times a continuation probability. We provide a novel construction of the stochastic choice rule that would prevail under full attention. H-LAO is then sharply characterized by a mild choice overload condition and the standard Block–Marschak inequalities applied to this recovered full-attention rule. We provide sufficient conditions for point identification, partial-identification results when these conditions are not met, and illustrate both with an example.[3]

Our third contribution is econometric. For homogeneous AOM, we develop estimation and inference for compatible preferences and attention bounds when the number of menus and inequalities is large. The methods use high-dimensional Gaussian approximation, feasible correlated-Gaussian and universal critical values, and test inversion; binary screening reduces both statistical and computational burden. The approximation error depends logarithmically on the number of comparisons and on the smallest effective menu sample size, rather than on a fixed-dimensional asymptotic ap-

[3]The Supplemental Appendix discusses limited menu data, dependence between stopping and preferences, and violations of the maintained support condition.

proximation. We also give a common finite-sample projection method for both models. For H-LAO, exact projection gives confidence sets for preference distributions under the benchmark and both sensitivity specifications. Explicit pairwise intervals remain valid without dividing by weak reach, and dependence-robust preference-event bounds are obtained by linear programming. A studentized inversion of the undivided pairwise moment improves precision when its standard error is estimable, while a fully linear projection procedure covers the model without Sequential Path Independence.

We also contribute to the empirical literature by providing evidence from the independently collected travel-mode experiment of Wang and Zhu (2025). Its complete menu manipulation, dominated outside option, repeated choices, information-search records, and ex post rankings allow us to compare RAM and AOM, apply H-LAO, and provide an external comparison of recovered pairwise preference shares with separately elicited reports that are not used to estimate H-LAO. Subject-clustered specification diagnostics do not reject H-LAO, and simultaneous confidence intervals for all six recovered–reported pairwise differences contain zero. Due to space constraints, Sections SA-5 and SA-6 of the Supplemental Appendix report implementation benchmarks and simulation evidence on size, coverage, power, robustness, and computational performance across complete and incomplete menu designs.

The paper lies at the intersection of revealed preference, discrete choice, limited consideration, and econometrics. See Matzkin (2007), Matzkin (2013), and Molinari (2020) for reviews of nonparametric and partial identification, and de Clippel and Rozen (2024) and Strzalecki (2025) for broader treatments of boundedly rational and stochastic choice. Our distinctive behavioral primitive is the marginal attention received by each alternative. Restricting this object across menus yields revealed-preference content and sharp attention-frequency bounds under homogeneous preferences; with an observed list and explicit data and support conditions, it also permits recovery of reach and the full-attention stochastic choice rule under heterogeneous preferences. On the econometric side, we develop high-dimensional inference for the resulting choice-probability inequalities and finite-sample projection methods for partially identified preference and attention objects, including procedures valid under weak reach.

The rest of the paper proceeds as follows. Section 2 develops the homogeneous-preference theory and compares AOM with other random-attention models. Section 3 develops H-LAO. Section

4 presents estimation and inference. Section 5 reports the empirical application. The Supplemental Appendix contains additional literature, proofs, alternative inference procedures, simulation evidence, and computational details. We also provide an `R` package and replication code.

2 Choice under Attention Overload

This section studies homogeneous preferences and monotonicity of attention frequency; Section 3 allows preferences to be heterogeneous. Let X be a finite universe of alternatives and $\mathcal{D}$ a collection of nonempty menus. We permit incomplete menu data, so $\mathcal{D}$ may be a strict subset of $2^X \setminus \{\emptyset\}$. A *choice rule* is a map $\pi : X \times \mathcal{D} \to [0,1]$ such that $\pi(a|S) = 0$ for $a \notin S$ and $\sum_{a \in S} \pi(a|S) = 1$. We interpret $\pi(a|S)$ as the population probability of choosing a from S and assume it is identified, or consistently estimable, for observed menus.

An important feature of our model is that consideration sets can be random. An *attention rule* is a map $\mu : 2^X \times \mathcal{D} \to [0,1]$ such that $\mu(T|S) = 0$ for every $T \not\subseteq S$ and $\sum_{T \subseteq S} \mu(T|S) = 1$ for every $S \in \mathcal{D}$. The value $\mu(T|S)$ is the probability of considering exactly $T \subseteq S$ when the menu is S. This formulation includes deterministic attention rules; for example, $\mu(S|S) = 1$ represents full attention. Our discussion on attention overload leads to the following novel theoretical construct. Defined with respect to an attention rule μ, we summarize the attention received by an individual alternative by adding the probabilities of all consideration sets containing it.

Definition 1 (Attention Frequency)**.** Given μ, the *attention frequency* map $\phi_\mu : X \times \mathcal{D} \to [0,1]$ is $\phi_\mu(a|S) := \sum_{T \subseteq S:\, a \in T} \mu(T|S)$.

$\phi_\mu(a|S)$ is the probability that a attracts attention in S. Whenever μ is clear from context, we omit the subscript. Attention frequency is binary under a deterministic attention rule but may take any value in $[0,1]$ under stochastic attention. When decision makers face an abundance of options, the alternatives compete for attention. We capture this competition by requiring an alternative's attention frequency to weakly decrease when the menu expands. We call this property *Attention Overload*, the paper's central nonparametric restriction.

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

Attention overload is intuitively motivated by competition among alternatives and is supported by the empirical evidence discussed in the Introduction. As choice sets become larger, two empirical patterns are observed: (i) aggregate search effort increases, and (ii) decision avoidance becomes more prevalent. We now clarify how attention overload relates to these observations.

First, research on consumer search (Reutskaja and Hogarth, 2009; Caplin, Dean, and Martin, 2011; Thomas, Molter, and Krajbich, 2021) shows that individuals may exert greater search effort as menu size increases and, consequently, inspect more alternatives on average. At first glance, this pattern may appear to contradict attention overload. An increase in aggregate search effort, however, does not imply that any particular alternative becomes more likely to be considered.

To illustrate this distinction, suppose that from a menu of size N, a decision maker first draws a search quota $q \leq N$ from an arbitrary distribution with the mean k_N, and conditional on q, inspects a uniformly selected subset of that size. Then, the attention frequency of every alternative is $\frac{k_N}{N}$.[4] Hence, Attention Overload is equivalent to, $\frac{k_N}{N} \geq \frac{k_M}{M}$ for two nested menus with sizes $N < M$. Thus, total expected search k_N may increase with menu size while Attention Overload continues to hold. Note that there is no restriction in the distributional assumption on search quota other than the mean. What matters is that the average number of inspected alternatives does not increase proportionally faster than the size of the menu.[5]

Second, the literature also documents that individuals may become more likely to avoid making a choice when presented with a larger set of options (Iyengar and Lepper, 2000). This phenomenon is commonly referred to as *choice overload*. At first glance, attention overload and choice overload may appear to go hand in hand. The two concepts, however, are logically distinct.

We use the example above with two alternatives ($N = 2$) to illustrate this point. In our framework, the empty consideration set can be interpreted as the DM not considering any alternatives (decision avoidance). We will keep the attention frequency the same, while varying the probability of empty consideration set depending on the distributional assumption on search quota. Consider two distribution assumptions. In the first one, the search quota is either 0 or 2 with equal proba-

[4] Conditional on a search quota q, the probability that any given alternative is inspected is $\binom{N-1}{q-1}/\binom{N}{q} = \frac{q}{N}$. Averaging over the distribution of q gives k_N/N.

[5] Interestingly, this is precisely the pattern documented by Thomas, Molter, and Krajbich (2021), who show that participants inspect more alternatives as choice-set size increases (k_N increasing in N), while each individual alternative becomes less likely to be inspected and the fraction of the menu that is inspected declines ($\frac{k_N}{N}$ is decreasing).

bility, and in the second one, the search quota is exactly 1. Note that, in both cases, the average number of products considered is 1 and the attention frequency of each alternative will be the same, $\frac{1}{2}$. However, the probability of decision avoidance differs: It is $\frac{1}{2}$ or 0, respectively. Attention Overload therefore does not determine the probability of decision avoidance.

Finally, before formally introducing the model, we provide a microfoundation for Attention Overload based on salience ordering and search quota. Suppose each decision maker draws a menu-independent salience ordering and a quota, and considers top alternatives in that more salient up to her quota capacity. If an item is considered in a larger menu, eliminating other items can only move it earlier in the search order, so it remains reached in every smaller menu containing it. This item-level monotonicity holds even under heterogeneous quota and salience rankings. The Supplemental Appendix formally states and proves this search foundation. Section 2.4 compares Assumption 1 with other attention restrictions.

Given the discussion above, for simplicity, we assume $\mu(\emptyset|S) = 0$. To model choice overload alongside attention overload, one can instead allow the consideration set to be empty and additionally require $\mu(\emptyset|S)$ to weakly increase as the menu expands.[6] Given a nonempty realized consideration set, the decision maker chooses its best alternative according to a strict preference ordering $\succ$. Combining this behavior with attention overload gives the following representation.

Definition 2 (Attention Overload Representation)**.** An *Attention Overload Model* (AOM) is a pair $(\succ, \mu)$ consisting of a preference ordering $\succ$ over X and an attention rule μ satisfying attention overload (Assumption 1). The AOM $(\succ, \mu)$ represents a choice rule π if

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

for all $a \in S$ and $S \in \mathcal{D}$. We call π *AOM-compatible* if some AOM represents it. Analogously, $\succ$ represents π if there exists an attention rule μ such that the AOM $(\succ, \mu)$ represents π.

AOM applies to structurally comparable menus: the same label a represents the same relevant alternative, and the population preference ordering remains stable, as S changes. Natural sources

[6]The characterization below continues to hold if one allows $\mu(\emptyset|S) \neq 0$ and interprets it as the choice probability of the outside option, as in the literature (e.g., Manzini, Mariotti, and Ülkü, 2024; Brady and Rehbeck, 2016; Aguiar, 2017), because its proof places no restriction on $\mu(\emptyset|S)$. Separately, Section SA-2.8 studies attention overload conditional on engagement.

of such variation include inventory or eligibility changes, randomized displays, and repeated cross sections drawn from the same population. When prices or characteristics vary, comparability can be preserved by conditioning on them or redefining alternatives as product-state pairs. Applications with endogenous menu selection similarly require comparable preference and attention populations across menus. Under these conditions, cross-menu choice variation isolates the modeled attention response; the econometric analysis conditions on the realized menu assignments.

The unknown components of the model are the attention rule μ and preference ordering $\succ$; standard choice data identify only π. The analysis first characterizes all compatible preferences and then studies what can be learned about attention frequencies without identifying the probability assigned to every consideration set.

2.1 Behavioral Implications and Revealed Preferences

We begin with characterization and preference elicitation. For a candidate preference $\succ$, the central question is whether some attention rule satisfying attention overload can reproduce the observed choice pattern. Directly searching over attention rules is cumbersome when the universe is large. Our first result replaces that search with testable inequalities and thereby characterizes the complete set of compatible preferences. To motivate, we consider a few behavioral implications of the AOM.

Assume that $(\succ, \mu)$ represents π, and let $U_{\succeq}(a) = \{b \in X : b \succeq a\}$ denote the weak upper contour set of a. Classical regularity requires that removing alternatives can only help each alternative that remains: $\pi(a|T) \geq \pi(a|S)$ whenever $T, S \in \mathcal{D}$ and $a \in T \subseteq S$. This is too strong when attention is scarce. Shrinking the menu can free attention that flows to alternatives that a would not have beaten, so a's own choice probability need not rise. Attention overload instead requires the smaller menu to generate enough choice probability on a or alternatives preferred to it. The following observable condition therefore bounds the choice share of a's upper contour set rather than only a's own share.

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

To establish necessity, observe that choice of a requires attention to a, while attention overload gives $\phi(a|T) \geq \phi(a|S)$ for observed menus $T, S \in \mathcal{D}$ with $T \subseteq S$. Moreover, whenever a is considered

in T, the chosen alternative belongs to $U_{\succeq}(a)$. Hence

$$\pi(U_{\succeq}(a)|T) \geq \phi(a|T) \geq \phi(a|S) \geq \pi(a|S),$$

which is $\succ$-Regularity.

$\succ$-Regularity coincides with classical regularity when a is the best item in T, permits an individual lower-ranked item to violate classical regularity because choices of better items provide compensation, and is vacuous when a is worst in T. Given a candidate $\succ$, the condition is directly testable from choice probabilities. The following result shows that it is also sufficient, so testing $\succ$-Regularity is equivalent to testing whether $\succ$ can rationalize π under attention overload without constructing an attention rule.

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

We proved necessity above; the supplemental appendix proves sufficiency by directly constructing a rationalizing attention rule. Thus, $\mathcal{P}_\pi := \{\succ\colon \pi \text{ satisfies } \succ\text{-Regularity}\}$ is exactly the identified set of compatible preferences. The characterization also yields revealed preferences. Alternative b is revealed preferred to a if $b \succ a$ for every preference representing π. Theorem 1 motivates the following binary relation based directly on $\succ$-Regularity:

$$b\mathrm{P}_\pi a \quad \text{if } b \succ a \text{ for all preference orderings such that } \pi \text{ satisfies } \succ\text{-Regularity.}$$

We remark that the relation P_π does not require constructing an attention rule for each candidate preference. Checking $\succ$-Regularity uses only observed choice probabilities, a fact we exploit for inference in Section 4.

Corollary 1 (Revealed Preference)**.** Let π be AOM-compatible. Then, b is revealed preferred to a if and only if $b\mathrm{P}_\pi a$.

Although Theorem 1 and Corollary 1 avoid constructing the underlying attention rule for revealed-preference analysis, they still require checking every possible preference ordering, which

becomes computationally expensive as $|X|$ grows. Two approaches substantially reduce this computational burden.

The first approach admits a mixed-integer linear formulation that avoids explicitly enumerating preference orderings. For distinct $a, b \in X$, let $z_{ba} = 1$ mean that $b \succ a$, and let $\mathcal{Z}_\pi$ contain all binary vectors $\mathbf{z} = (z_{ba})_{b \neq a}$ satisfying

$$\begin{aligned} z_{ab} + z_{ba} &= 1 && \text{for all distinct } a, b, \\ z_{ab} + z_{bc} - 1 &\leq z_{ac} && \text{for all distinct } a, b, c, \\ \pi(a|S) &\leq \pi(a|T) + \sum_{b \in T \setminus \{a\}} \pi(b|T) z_{ba} && \text{for all } a \in T \subseteq S \text{ with } T, S \in \mathcal{D}. \end{aligned}$$

Corollary 2 (Mixed-Integer Characterization)**.** The map taking a strict ordering $\succ$ to its binary incidence vector $\mathbf{z}$ is a bijection between $\mathcal{P}_\pi$ and $\mathcal{Z}_\pi$. Consequently, π is AOM-compatible if and only if $\mathcal{Z}_\pi$ is nonempty. If π is AOM-compatible, then $b \mathrm{P}_\pi a$ if and only if $\min_{z \in \mathcal{Z}_\pi} z_{ba} = 1$.

The corollary replaces a search over $|X|!$ candidate rankings with a mixed-integer linear feasibility problem. It uses $|X|(|X| - 1)$ binary variables, cubically many ordering constraints, and one linear inequality for each observed $a \in T \subseteq S$. Binary screens can first fix individual z_{ba} coordinates, while nonbinary violations add restrictions that eliminate incompatible rankings. Pairwise revealed preference is then obtained by testing whether the opposite comparison remains feasible. For perspective, with ten alternatives, the formulation uses 90 directed binary variables, only 45 of which are independent after imposing pairwise completeness, instead of enumerating $10! = 3{,}628{,}800$ rankings. The Supplemental Appendix proves the formulation and discusses its implementation; Example 1 below illustrates the mechanics.

Our second approach provides an optimization-free screening device when the object of interest is a single pairwise comparison. As we will show below, regularity violations involving binary menus provide a simple and computationally attractive device for revealed preference analysis, although they do not exhaust the identifying power of attention overload. Alternatively, the result can also be used as an additional screening step if the goal is to obtain the entire set of compatible preferences $\mathcal{P}_\pi$.

Specifically, if $a, b \in S$ and $\pi(a|S) > \pi(a|\{a, b\})$, then every representation must rank b above

a. Otherwise $a \succ b$, so a is best in the binary menu and $\phi(a|\{a,b\}) = \pi(a|\{a,b\})$. But then

$$\phi(a|\{a,b\}) = \pi(a|\{a,b\}) < \pi(a|S) \leq \phi(a|S),$$

contradicting attention overload. The next proposition formalizes this observation.

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

Proposition 1 gives an easy-to-implement revealed-preference test without requiring a complete representation. With a nonbinary smaller menu T, the conclusion is weaker: a violation $\pi(a|S) > \pi(a|T)$ implies only that some alternative in T is preferred to a. Indeed, if a were best in T, then

$$\phi(a|T) = \pi(a|T) < \pi(a|S) \leq \phi(a|S),$$

again contradicting attention overload.

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

Both propositions use regularity violations, and Proposition 2 specializes to Proposition 1 when T is binary. The following example uses these propositions to screen preference orderings before applying Theorem 1; regularity violations alone do not exhaust the identifying power of attention overload.

Example 1. Consider the following choice data:

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

There are $4! = 24$ candidate preference orderings. First, applying Proposition 1 from $\{a,b,c\}$ to $\{a,b\}$ rules out every ordering with $b \succ a$, leaving 12 candidates. Two further violations involve

nonbinary smaller menus. From $\{a,b,c,d\}$ to $\{a,b,c\}$, c violates regularity, so Proposition 2 rules out the four candidates in which c is preferred to both a and b. Analogously, the violation for d from $\{a,b,c,d\}$ to $\{b,c,d\}$ rules out the four candidates in which d is preferred to both b and c. Four orderings remain: $a \succ b \succ c \succ d$, $a \succ b \succ d \succ c$, $a \succ c \succ b \succ d$, and $a \succ c \succ d \succ b$. Finally, checking $\succ$-Regularity requires both $b \succ d$ and $c \succ d$, leaving only $a \succ b \succ c \succ d$ and $a \succ c \succ b \succ d$. ▲

The example therefore has two AOM-compatible preferences, $a \succ_1 b \succ_1 c \succ_1 d$ and $a \succ_2 c \succ_2 b \succ_2 d$. The revealed-preference relation P_π is complete except for the comparison between b and c. Starting from 24 candidates, the binary screen eliminates 12, the nonbinary violations eliminate another 8, and the full characterization leaves exactly these two. Example 1 thus shows how the results can be applied sequentially to reduce the computational burden.

Corollary 2 reaches the same conclusion without enumerating the 24 rankings. For these data, every $\mathbf{z} \in \mathcal{Z}_\pi$ satisfies (i) $z_{ab} = z_{ac} = z_{ad} = z_{bd} = z_{cd} = 1$, and (ii) $z_{bc} + z_{cb} = 1$. Thus, $\mathcal{Z}_\pi$ contains exactly two incidence vectors, corresponding to $\succ_1$ and $\succ_2$. Minimizing each coordinate over $\mathcal{Z}_\pi$ gives one for the five common comparisons above, whereas

$$\min_{z \in \mathcal{Z}_\pi} z_{bc} = \min_{z \in \mathcal{Z}_\pi} z_{cb} = 0.$$

The mixed-integer formulation therefore reveals every comparison except that between b and c, exactly as the direct characterization does.

2.2 Attentive at Binaries

Revealed-preference analysis can be strengthened by limiting extreme inattention at binary menus. Because competition for attention is weaker in binary menus, one may expect decision makers to be more attentive when only two alternatives are available. We formalize this additional restriction by bounding the probability of considering only one of the two available alternatives.

$$\text{Fix } \eta \in [0.5, 1], \quad \text{and } \eta \geq \max\Big\{\mu(\{a\}|\{a,b\}), \mu(\{b\}|\{a,b\})\Big\} \text{ for every } \{a,b\} \in \mathcal{D}. \tag{1}$$

This condition limits extreme inattention without imposing full attention. A larger η weakens the restriction, and $\eta = 1$ is vacuous. Even at $\eta = 0.5$, the full binary menu may receive zero attention: both singleton consideration sets can have probability 0.5.

Condition (1) generates additional revealed preferences. For every observed binary menu, define $a \, \mathrm{P}_B \, b$ whenever $\pi(a|\{a,b\}) > \eta$. The choice probability of a then cannot be generated entirely by singleton consideration because $\mu(\{a\}|\{a,b\}) \leq \eta$. Hence $\mu(\{a,b\}|\{a,b\}) > 0$, and choosing a when both alternatives are considered reveals $a \succ b$.

The parameter η can be interpreted as the analyst's degree of caution in drawing welfare conclusions. At $\eta = 1$, binary choice alone yields no strict comparison, so $\mathrm{P}_B = \emptyset$. The familiar threshold $\eta = 0.5$ (Marschak, 1959; Fishburn, 1998) gives the strongest binary revealed-preference relation in this class, although it need not identify the complete ordering.

The necessity and sufficiency of this additional ordering restriction are formalized next.

Corollary 3 (Attentive-at-Binaries Characterization)**.** Fix $\eta \in [0.5, 1]$. A strict ordering $\succ$ belongs to an AOM representation satisfying (1) if and only if π satisfies $\succ$-Regularity and $\succ$ extends P_B.

Thus, P_B can be used with Propositions 1 and 2 to screen the candidate orderings further. The result applies to an incomplete domain: it restricts only the binary menus that belong to $\mathcal{D}$. We revisit Example 1 to illustrate.

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

2.3 Revealed Attention

Attention overload restricts marginal attention across menus but generally does not identify the probability assigned to each consideration set. The same choice rule can therefore admit multiple attention rules with different attention frequencies. This section derives observable lower and upper bounds on those frequencies and characterizes the corresponding joint identified set. These objects

can be useful, for example, when a firm studies product visibility or a policymaker asks whether consumers attend to higher-quality options.

We first formalize the joint identified set. For a candidate preference $\succ$, let $\mathcal{M}_\pi(\succ)$ contain all attention rules μ such that (i) the choice equations in Definition 2 hold for every observed menu and (ii) $\phi_\mu(a|R) \leq \phi_\mu(a|S)$ for every $S, R \in \mathcal{D}$ with $a \in S \subseteq R$. For fixed $\succ$, these are finitely many linear equalities and inequalities in μ.

Proposition 3 (Joint Identified Set)**.** The sharp identified set for the preference and attention rule is $\mathcal{I}_{\mu,\succ}(\pi) = \bigcup_{\succ \in \mathcal{P}_\pi} \left(\mathcal{M}_\pi(\succ) \times \{\succ\}\right)$. Moreover, $\mathcal{M}_\pi(\succ)$ is a nonempty polytope if and only if $\succ \in \mathcal{P}_\pi$. Sharp bounds on any linear functional of μ are obtained by minimizing and maximizing that functional over $\mathcal{M}_\pi(\succ)$ for each $\succ \in \mathcal{P}_\pi$, and then taking the extrema across preferences.

This proposition gives an operational joint characterization. Theorem 2 below provides closed-form solutions for the important marginal-attention coordinates, avoiding the need to solve the linear programs separately.

Consider bounding ϕ from below first. For any observed superset $R \in \mathcal{D}$ of S, attention overload implies that $\pi(a|R) \leq \phi(a|R) \leq \phi(a|S)$. Therefore, for any $S \in \mathcal{D}$, $\phi(a|S) \geq \max_{R \in \mathcal{D}:\, R \supseteq S} \pi(a|R)$. This lower bound on attention uses only the observed choice rule and does not require knowledge of the underlying attention rule. The preference-dependent upper bound is obtained analogously from the observed subsets $T \in \mathcal{D}$ of S that contain a. These observations give the following theorem.

Theorem 2 (Revealed Attention)**.** Let $(\succ, \mu)$ be an AOM representing π. Then, for every $S \in \mathcal{D}$ and $a \in S$,

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

The theorem is conditional on a compatible preference $\succ$. Its lower bound is common to every $\succ \in \mathcal{P}_\pi$, while its upper bound can vary with $\succ$. This distinction matters when preferences are only partially identified and is made precise below.

Three extreme cases help interpret Theorem 2. If both endpoints are one, a receives full attention in S; if both are zero, a is revealed unattended; and if the interval is exactly $[0, 1]$, the data reveal nothing about its attention frequency. These are the only possibilities under deterministic

choice, the setting studied by Lleras, Masatlioglu, Nakajima, and Ozbay (2017). The theorem supplies a common characterization for deterministic and stochastic choice.

Stochastic choice also permits partial revealed attention: the lower endpoint can be strictly positive or the upper endpoint strictly below one without the two coinciding. We illustrate this possibility by revisiting Example 1.

Example 1 (continued). For the menu $\{a, b, c, d\}$, Theorem 2 gives $\phi(a|\{a, b, c, d\}) = 0.05$, places both $\phi(b|\{a, b, c, d\})$ and $\phi(c|\{a, b, c, d\})$ in $[0.1, 0.25]$, and places $\phi(d|\{a, b, c, d\})$ in $[0.75, 1]$, where the extrema range only over the menus observed in the example. ▲

Attention can be point identified for selected alternatives. For example, suppose in addition that $R \in \mathcal{D}$, $R \supseteq \{a, b, c, d\}$, and $\pi(c|R) = 0.25$. The preference-free lower bound on $\phi(c|\{a, b, c, d\})$ is then 0.25, equal to its upper bound, so this attention frequency is point identified.

The lower and upper endpoints can be viewed as pessimistic and optimistic evaluations of attention. A remaining question is whether the coordinatewise pessimistic evaluations can be attained jointly by one attention rule satisfying attention overload. The next result answers affirmatively; the proof of Theorem 1 also constructs the analogous optimistic rule.

Theorem 3 (Pessimistic Evaluation for Attention)**.** Let $(\succ, \mu)$ be an AOM representing π. Then there exists a pessimistic attention rule μ^* such that $(\succ, \mu^*)$ is also an AOM representing π. That is, for every $S \in \mathcal{D}$ and $a \in S$, $\phi_{\mu^*}(a|S) = \max_{R \in \mathcal{D}:\, R \supseteq S} \pi(a|R)$.

This theorem concerns the pessimistic evaluation; the proof of Theorem 1 also constructs the analogous optimistic attention rule. Consequently, for each fixed $\succ \in \mathcal{P}_\pi$, the two bounds in Theorem 2 are sharp componentwise. When preferences are not point identified, the projection of the joint set in Proposition 3 onto attention frequencies is $\mathcal{I}_\phi(\pi) = \{\phi_\mu : (\mu, \succ) \in \mathcal{I}_{\mu,\succ}(\pi)\}$. Projecting this set onto a single coordinate $\phi(a|S)$ gives the common lower endpoint $\max_{R \in \mathcal{D}:\, R \supseteq S} \pi(a|R)$ and the upper endpoint $\max_{\succ \in \mathcal{P}_\pi} \min_{T \in \mathcal{D}:\, a \in T \subseteq S} \pi(U_\succeq(a)|T)$. These projected bounds do not imply that every vector formed by independently selecting values inside the coordinatewise intervals belongs to $\mathcal{I}_\phi(\pi)$, nor do they identify the complete attention rule μ. Our sharpness claim concerns the identified preference set and the stated componentwise attention functionals.

2.4 Comparison to Other Random Attention Models

Random-attention models differ in the object they discipline. Independent attention and elimination by aspects impose parametric structure on marginal or category-level attention (Tversky, 1972; Manzini and Mariotti, 2014; Aguiar, 2017). Logit attention specifies probabilities for consideration sets (Brady and Rehbeck, 2016). RAM instead imposes the nonparametric set-level restriction

$$\mu(D|S) \leq \mu(D|S \setminus \{b\}) \qquad \text{when } b \notin D \subseteq S.$$

AOM places no monotonicity restriction on the probability of a particular consideration set. It requires only that the attention frequency $\phi(a|S) = \sum_{D:\, a \in D} \mu(D|S)$ not increase with the menu. Set-level RAM monotonicity and marginal attention overload therefore neither imply nor are implied by one another, and they have distinct implications for choice behavior. The following example illustrates their different revealed-preference implications. Consider

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

Under RAM, the fall in b's choice probability when c is removed reveals $b \succ c$, while the analogous fall for c when b is removed reveals $c \succ b$.[7] Under AOM, the same choice patterns reveal $a \succ b$ and $a \succ c$. Thus, RAM ranks the alternative whose choice probability falls above the removed alternative, whereas AOM reveals that the alternative whose choice probability falls is outranked by an alternative remaining in the smaller menu. The data therefore violate RAM through a revealed-preference cycle but admit either $a \succ b \succ c$ or $a \succ c \succ b$ under AOM. Section SA-1 also gives a four-alternative choice rule that passes RAM but fails AOM, establishing observable non-nesting in the other direction.

Two extreme cases further illustrate the distinction between attention models at the attention

[7] These conclusions follow from Lemma 1 of Cattaneo, Ma, Masatlioglu, and Suleymanov (2020).

level. In a larger menu S, Ann considers every item, while Ben considers only a. For a smaller menu $a \in T \subset S$, attention overload requires Ann to retain full item-level attention and requires Ben to keep considering a, but it permits Ben to add items that were previously ignored. RAM does the reverse in these examples: it leaves Ann's behavior in T unrestricted but forces Ben's consideration set to remain $\{a\}$. Table 1 also reports the corresponding predictions of common parametric models.

	Ann	Ben
	Larger set S	
	Full attention	Limited attention
	$\phi(x\|S) = 1\ \forall x \in S$	$\phi(a\|S) = 1$ and $\phi(x\|S) = 0\ \forall x \neq a$
	Predictions for a smaller set $a \in T \subset S$	
Manzini and Mariotti (2014)	$\phi(x\|T) = 1\ \forall x \in T$	$\phi(a\|T) = 1$ and $\phi(x\|T) = 0\ \forall x \neq a$
Tversky (1972); Aguiar (2017)	$\phi(x\|T) = 1\ \forall x \in T$	$\phi(a\|T) = 1$ and $\phi(x\|T) = 0\ \forall x \neq a$
Brady and Rehbeck (2016)	No restriction	$\phi(a\|T) = 1$ and $\phi(x\|T) = 0\ \forall x \neq a$
Cattaneo, Ma, Masatlioglu, and Suleymanov (2020)	No restriction	$\phi(a\|T) = 1$ and $\phi(x\|T) = 0\ \forall x \neq a$
Demirkan and Kimya (2020)	No restriction	No restriction
Attention Overload Model (AOM)	$\phi(x\|T) = 1\ \forall x \in T$	$\phi(a\|T) = 1$; otherwise unrestricted

Table 1: Predictions for a smaller menu T, given Ann's full attention and Ben's singleton consideration in the larger menu S.

The independent-attention and elimination-by-aspects models hold item or category attention parameters fixed across menus. They therefore preserve Ann's full attention but cannot let Ben begin considering previously ignored items merely because the menu shrinks. Menu-dependent independent attention (Demirkan and Kimya, 2020) can allow either direction and does not by itself impose attention overload. Logit attention and RAM allow Ann to change arbitrarily in the smaller menu, while their set-level structure keeps Ben at $\{a\}$ in this example.

The different primitives also explain the different roles played by regularity. In Cattaneo, Ma, Masatlioglu, and Suleymanov (2020), violations of classical choice regularity are informative about preferences under RAM's set-level monotonicity. AOM does not simply require the opposite data pattern. Its characterization uses the preference-indexed $\succ$-Regularity condition: choice of a in a larger menu is bounded by choice from the entire upper contour set of a in a smaller menu. Classical regularity is recovered only when a is best in that smaller menu; lower-ranked alternatives may violate it. Conversely, a classical regularity violation at a binary menu reveals that the other binary

alternative must be preferred to a under AOM. These implications arise from marginal competition for attention, not from relabeling the RAM restriction.

The appropriate model is application-dependent. RAM is natural when removing unavailable alternatives should not reduce the probability of drawing a fixed surviving consideration set; AOM is natural when adding rivals dilutes item-level visibility. When both restrictions are plausible they can be imposed jointly. When they conflict, auxiliary attention measures or their distinct choice implications can guide specification testing.

3 Heterogeneous Preferences

The baseline AOM attributes stochastic choice to heterogeneous attention under a common preference. This section allows both preferences and attention to vary across a population. The objective is to learn the distribution of preferences from aggregate choice probabilities while retaining the attention-overload mechanism.

Without further structure, suppose there exists a probability distribution τ over rational preferences and an attention rule μ satisfying attention overload such that

$$\pi(a|S) = \sum_{T:\, a \in T \subseteq S} \tau(\{\succ \in \mathcal{P} : a \text{ is } \succ\text{-best in } T\}) \cdot \mu(T|S)$$

for all $a \in S$ and $S \in \mathcal{D}$. Here τ captures preference heterogeneity and μ attention heterogeneity, with the two independent conditional on the environment being studied.

An observed list supplies the additional structure needed to separate preference and attention heterogeneity. Many choice environments present alternatives sequentially, allowing the list to discipline the path of attention without restricting which preference orderings may occur. This structure distinguishes combinations of preference and attention heterogeneity that the unrestricted mixture alone would leave observationally equivalent. In addition, as is standard in the literature, we introduce an outside option o^*, chosen if and only if the consideration set is empty. Therefore, we expand the choice rule so that $\pi(\cdot|S)$ is defined as a probability distribution over $S \cup \{o^*\}$. For convenience, we maintain the support condition that the outside option is not chosen with

probability one, so that $\pi(S|S) > 0$ for every nonempty S.

Throughout this section, the structural H-LAO model is defined on the full menu domain $2^X \setminus \{\emptyset\}$, while $\mathcal{D} \subseteq 2^X \setminus \{\emptyset\}$ denotes the collection of menus for which choice probabilities are observed. Thus, the underlying choice and attention processes are defined for every nonempty menu, even when the researcher observes only their restriction to $\mathcal{D}$. This separation between the domain of the primitives and observability will make the partial-identification result more convenient to state.

3.1 List-Based Attention Overload

Consumers often encounter an observed order: search results, ranked advertisements, product recommendations, ballots, or alternatives on a screen. In real-world applications, the presentation order can plausibly carry informational value. Here, instead of explicitly modeling how the list is generated, we take the observed list as the economically relevant state variable and impose structure on how attention is allocated conditional on that list. Ordered search has a long history in decision theory (Rubinstein and Salant, 2006; Horan, 2010; Guney, 2014; Yildiz, 2016; Aguiar, Boccardi, and Dean, 2016; Kovach and Ülkü, 2020; Ishii, Kovach, and Ülkü, 2021; Tserenjigmid, 2021; Yegane, 2022; Koshevoy and Savaglio, 2023; Manzini, Mariotti, and Ülkü, 2024; Caplin, Dean, and Martin, 2011; Caplin and Dean, 2011). We condition on a presentation order observed by the researcher and common within the population whose aggregate choices define π. When different groups face different lists, the analysis applies within list strata and therefore accommodates multiple observed presentation regimes.

As a running example, consumers browse search results from top to bottom and stop at heterogeneous positions. Once a consumer stops, her consideration set consists of the available items encountered up to that point. This prefix structure makes the mapping from attention frequencies to the distribution of stopping points one-to-one. Formally, let $\{a_1, a_2, \ldots, a_{|X|}\}$ denote the observed list, where $j < k$ means that a_j appears earlier than a_k. Each menu $S \subseteq X$ inherits this order and is written $\{a_{s_1}, \ldots, a_{s_{|S|}}\}$.

Our new attention rule imposes a sequential structure on consideration set formation. Any consideration set must respect the underlying order: if a_k is included, then every feasible alternative

preceding a_k must also be included. In other words, once an alternative is considered, all earlier alternatives in the list are necessarily considered as well. Consequently, given S, the admissible consideration sets take the following form, and every other subset receives zero attention probability:

$$\emptyset,\ \{a_{s_1}\},\ \{a_{s_1}, a_{s_2}\},\ \{a_{s_1}, a_{s_2}, a_{s_3}\},\ \ldots$$

For $j = 1, \ldots, |S|$, define the jth prefix and suffix of S by $S_{:j} := \{a_{s_1}, \ldots, a_{s_j}\}$ and $S_{j:} := \{a_{s_j}, \ldots, a_{s_{|S|}}\}$, respectively. Thus, $S_{:j} \subseteq S_{:k}$ if and only if $j \leq k$, whereas $S_{j:} \subseteq S_{k:}$ if and only if $j \geq k$. We call an attention rule *list-based* if, for every nonempty $S \subseteq X$, it assigns positive probability only to $\emptyset$ and the prefixes in S. The key consequence of the list structure is a one-to-one mapping between attention frequencies and attention rules. Formally, a map $\phi : X \times (2^X \setminus \{\emptyset\}) \to [0, 1]$ is a *list-based attention frequency* if, for every $S = \{a_{s_1}, \ldots, a_{s_{|S|}}\}$, $1 \geq \phi(a_{s_1}|S) \geq \cdots \geq \phi(a_{s_{|S|}}|S) \geq 0$. Every list-based attention frequency induces the unique list-based attention rule

$$\mu(S_{:j}|S) = \phi(a_{s_j}|S) - \phi(a_{s_{j+1}}|S), \qquad \mu(\emptyset|S) = 1 - \phi(a_{s_1}|S),$$

where the terminal convention is $\phi(a_{s_{|S|+1}}|S) = 0$. Without the prefix structure in attention rule, the same attention frequencies may be generated by many different distributions over consideration sets.

The list-based attention frequency remains highly permissive. By itself, it does not imply attention overload: when the menu expands, the attention frequency of a given alternative may still increase. We therefore retain attention overload as a cross-menu restriction by imposing two assumptions on the attention frequencies.

Assumption 2 (Sequential Path Independence (SPI))**.** For every nonempty $S = \{a_{s_1}, \ldots, a_{s_{|S|}}\} \subseteq X$ and every $j \neq 1$, $\phi(a_{s_j}|S) = \phi(a_{s_{j-1}}|S)\phi(a_{s_j}|S_{j:})$.

The first assumption, SPI, is a cross-menu condition on sequential search. The probability of reaching the next item equals the probability of reaching the current item times a continuation probability that depends only on the remaining alternatives. This is a recursive multiplicative restriction on the attention-frequency function.

On a suffix-closed domain,[8] let $q(S) = \phi(a_{s_1}|S)$ denote first-position reach. Iterating SPI gives, for every menu S and position j,

$$\phi(a_{s_j}|S) = \prod_{k=1}^{j} q(S_{k:}). \tag{2}$$

Conversely, any assignment $q : \mathcal{D} \to [0,1]$ on a suffix-closed domain defines through (2) a list-based attention frequency satisfying SPI. On the full-menu domain, the model therefore retains $2^{|X|} - 1$ unrestricted first-position probabilities. Arbitrary q does not by itself impose attention overload, which remains the separate cross-menu restriction above. The following examples show that familiar stopping processes satisfy SPI.

Example 2 (Constant Stopping Attention)**.** This example gives a parsimonious model with a single parameter describing the attention rule. At each stage, search continues with probability $\alpha \in [0,1]$, so the stopping hazard is $1-\alpha$ and $q(S) = \alpha$ for every menu S. Then $\phi(a_{s_j}|S) = \alpha^j$. For $j < |S|$, $\mu(S_{:j}|S) = (1-\alpha)\alpha^j$, while $\mu(S|S) = \alpha^{|S|}$. Notably, $\phi(a_{s_j}|S)$ depends only on the position j, not on the size of S. ▲

Example 3 (Rank-Dependent Attention)**.** In this example, continuation depends on the number of alternatives remaining, capturing the idea that a lengthy remaining list may deter consumers from proceeding. Let $q(R) = h(|R|)$ for some weakly decreasing function $h : \{1, \ldots, |X|\} \to [0,1]$. Then, $\phi(a_{s_j}|S) = \prod_{m=|S|-j+1}^{|S|} h(m)$. Because h is weakly decreasing, the attention frequency is weakly decreasing as the menu expands. There are at most $|X|$ parameters in this class. ▲

Example 4 (Alternative-Dependent Attention)**.** Unlike the previous two examples, continuation here depends on the specific alternative rather than only on its position. Let $\beta(a) \in [0,1]$ be the item-specific conditional factor for continuing far enough to inspect a after reaching its predecessor, and set $q(S) = \beta(a_{s_1})$. Then $\phi(a_{s_j}|S) = \prod_{k \le j} \beta(a_{s_k})$. For $j < |S|$, $\mu(S_{:j}|S) = \big(1 - \beta(a_{s_{j+1}})\big) \prod_{k \le j} \beta(a_{s_k})$, while $\mu(S|S) = \prod_{k \le |S|} \beta(a_{s_k})$. This class uses only $|X|$ parameters. ▲

Finally, we need one more condition to guarantee attention overload.

Assumption 3 (Weak Attention Overload)**.** For $a = a_{s_1} = a_{t_1} \in S \subseteq T$, $\phi(a|S) \ge \phi(a|T)$.

[8] A domain $\mathcal{D}$ is suffix-closed (prefix-closed) if the suffixes (prefixes) of S belong to $\mathcal{D}$ whenever $S \in \mathcal{D}$.

Because the structural menu domain is full, Sequential Path Independence and Weak Attention Overload jointly imply Attention Overload.

3.2 List-Based Attention Overload with Heterogeneous Preferences

We are now ready to define the heterogeneous model. As in RUM, any strict preference ordering may occur. Let $\mathcal{P}$ denote the set of strict preference orderings over X.

Definition 3 (List-Based Attention Overload with Heterogeneous Preferences)**.** A choice rule π has a *List-Based Attention Overload with Heterogeneous Preferences* representation (H-LAO) if there exists a list-based attention frequency ϕ satisfying Assumptions 2 and 3, and a distribution τ on $\mathcal{P}$, such that

$$\pi(a|S) = \sum_{\substack{1\leq j\leq |S| \\ a\in S_{:j}}} \tau(\{\succ\in\mathcal{P} : a \text{ is } \succ\text{-best in } S_{:j}\}) \underbrace{\left(\phi(a_{s_j}|S) - \phi(a_{s_{j+1}}|S)\right)}_{\mu(S_{:j}|S)},$$

and $\pi(o^*|S) = 1 - \phi(a_{s_1}|S)$, for every nonempty $S \subseteq X$ and every $a \in S$.

Benchmark H-LAO permits an unrestricted marginal distribution τ over $\mathcal{P}$ and heterogeneous stopping positions, while maintaining conditional independence between preference and stopping type. RUM is the special case in which every consumer reaches the end of the list. Limited attention allows H-LAO to accommodate behavior inconsistent with RUM without tying admissible preferences to positions on the list. We later relax independence and derive a sharp identified set under arbitrary menu-specific dependence between preferences and stopping, followed by a second sharp set that also removes Sequential Path Independence.

We now discuss related models in the literature. The closest list-based benchmark is the Random Depth Model of Ishii, Kovach, and Ülkü (2021), which uses exogenous variation in list order for a fixed menu; H-LAO instead holds the observed order fixed and exploits variation across menus to study attention overload. Subsequent work by Masatlioglu and Vu (2024) orders heterogeneous consideration sets by attentiveness and identifies their composition and frequency. Its primitive is expansion of consideration sets across attention types, whereas H-LAO disciplines cross-menu reach along an observed list and permits an unrestricted marginal preference distribution. The

set-monotone and stable RAUM of Kashaev and Aguiar (2022) more generally combines random utility with RAM-style random attention and consequently delivers partial identification. H-LAO's stronger sequential and cross-menu structure recovers attention and the full-attention stochastic choice rule under positive reach. Other approaches use parametric consideration restrictions or enriched data such as mixture choice (Dardanoni, Manzini, Mariotti, and Tyson, 2020; Dardanoni, Manzini, Mariotti, Petri, and Tyson, 2023). We retain a broader comparison in the Supplemental Appendix.

3.3 Behavioral Characterization

The characterization relies on recovering the attention frequencies and using them to reconstruct the choice probabilities under full attention. First, we define, for every S, $\varphi(a_{s_1}|S) = 1 - \pi(o^*|S)$. Intuitively, $\varphi(a_{s_1}|S)$ equals the probability that the decision-maker attends to at least one alternative in S, since by assumption they choose o^* if and only if they do not consider any alternative. This allows us to recursively define, for all $S = \{a_{s_1}, a_{s_2}, \ldots, a_{s_{|S|}}\}$ and $j \neq 1$,

$$\varphi(a_{s_j}|S) = \varphi(a_{s_{j-1}}|S)\varphi(a_{s_j}|S_{j:}).$$

If the model is correctly specified, φ coincides with the underlying ϕ, and hence all attention frequencies are recovered from the choice data.[9] The recursion ensures that φ satisfies Assumption 2, but attention overload remains an additional testable cross-menu restriction.

Iterating the recursion gives

$$\varphi(a_{s_j}|S) = \prod_{k=1}^{j} \left[1 - \pi(o^*|S_{k:})\right]. \tag{3}$$

Every factor lies in $[0, 1]$, so $1 \geq \varphi(a_{s_1}|S) \geq \cdots \geq \varphi(a_{s_{|S|}}|S) \geq 0$ holds by construction. Under the maintained support condition, every factor in (3) is strictly positive whenever the corresponding suffix is observed. In particular, under full menu data, terminal reach is strictly positive, so the recursion defining f is well-defined.

To guarantee that the recovered attention frequency satisfies the required cross-menu restriction,

[9] We discuss the implication of incomplete observed menus in the next section.

it is enough to impose the following observable condition on outside-option choice. This type of monotonicity is commonly used as a measure of choice overload (see Chernev, Böckenholt, and Goodman, 2015 for a review).

Axiom 2 (Weak Choice Overload)**.** For $a_{s_1} = a_{t_1} \in S \subseteq T$, $\pi(o^*|T) \geq \pi(o^*|S)$.

The remaining testability question concerns the full-attention preference component. To calculate it, we define an auxiliary stochastic choice rule f recursively. The base case is $f(x|\{x\}) := 1$. For $S = \{a_{s_1}, \ldots, a_{s_{|S|}}\}$ with $|S| \geq 2$ and $k = 2, \ldots, |S|$, let[10]

$$f(a_{s_k}|S) := \frac{1}{\varphi(a_{s_{|S|}}|S)} \left(\pi(a_{s_k}|S) - \sum_{j=k}^{|S|-1} \big(\varphi(a_{s_j}|S) - \varphi(a_{s_{j+1}}|S)\big) f(a_{s_k}|S_{:j}) \right).$$

Finally, define $f(a_{s_1}|S) = 1 - \sum_{k=2}^{|S|} f(a_{s_k}|S)$. Thus, f adds up to one by construction. When π is generated by the model, $f(a|S)$ is the probability that a is chosen under full attention and therefore records the stochastic-choice implications of τ. If the model holds, f must be rationalizable by at least one distribution τ over strict preferences. Under the finite-universe, full-menu, adding-up, and strict-ordering conditions used here, the Block–Marschak characterization of Falmagne (1978) says that this is equivalent to nonnegativity of the Block–Marschak polynomials derived from f.

Axiom 3 (Non-negativity of BM)**.** For all T and all $x \in T$, $\sum_{S:\ S \supseteq T} (-1)^{|S \setminus T|} f(x|S) \geq 0$.

Notably, this condition also ensures that f is non-negative. We are then ready to state the characterization theorem.

Theorem 4 (Characterization)**.** Let $\mathcal{D} = 2^X \setminus \{\emptyset\}$. Suppose that $\pi(R|R) > 0$ for every nonempty $R \subseteq X$. The choice rule π has a H-LAO representation if and only if π satisfies Axiom 2 and 3.

Theorem 4 is an if-and-only-if, and hence sharp, behavioral characterization of H-LAO under full menu data. It separates limited attention from preference heterogeneity: the construction of φ, together with Weak Choice Overload, guarantees Sequential Path Independence and Attention

[10] The recursion defining f requires the denominator $\varphi(a_{s_{|S|}}|S)$ to be nonzero. This is ensured by our maintained support condition $\pi(S|S) > 0$ for every nonempty S. We discuss the corresponding result when this condition is relaxed in the Supplementary Appendix.

Overload, while BM non-negativity tests whether the implied full-attention choice rule is generated by a distribution of preferences. The next result states what is point identified under those conditions.

Theorem 5 (Identification)**.** Let $\mathcal{D} = 2^X \setminus \{\emptyset\}$ and (ϕ, τ) be a H-LAO representation of π. Then $\phi = \varphi$ and $\tau(\{\succ \in \mathcal{P} : a \text{ is } \succ\text{-best in } T\}) = f(a|T)$.

Thus, full-menu data point identify the attention frequency ϕ and the full-attention stochastic choice rule f. They need not uniquely identify the complete distribution τ over rankings. Once f has been recovered, however, the remaining conclusions are standard consequences of finite random-utility theory (Falmagne, 1978; Turansick, 2022). In particular, the sharp compatible set is

$$\mathcal{T}(f) = \left\{ \tau \in \Delta(\mathcal{P}) : f(a|S) = \sum_{\succ \in \mathcal{P}} \mathbb{1}\{a \text{ is } \succ\text{-best in } S\}\tau(\succ) \text{ for every } a \in S \subseteq X \right\},$$

and probabilities of preference events are bounded by minimizing and maximizing their τ-mass over this polytope. Falmagne's Mobius inversion also recovers the probability of each lower contour set and hence every alternative's rank distribution. For example,

$$\tau(\{\succ : a \succ b\}) = f(a|\{a,b\}), \qquad \tau(\{\succ : a \text{ is top ranked}\}) = f(a|X).$$

The H-LAO contribution is the recovery of f from limited-attention choice data; the Supplemental Appendix records the standard random-utility formulas needed to translate the recovered rule into preference-event information.

Example 5 (Recovering attention and preference types)**.** Let the observed list be a, b, c. Suppose continuation at each stage is geometric with probability 0.8, and let τ assign probabilities 0.5, 0.3, and 0.2 to $a \succ b \succ c$, $b \succ c \succ a$, and $c \succ a \succ b$, respectively. The resulting choice probabilities are reported below, rounded to two decimal places; subsequent calculations use the unrounded probabilities.

π [f]	a	b	c	o^*
$\{a\}$	0.80 [1.00]	–	–	0.20 [0.00]
$\{b\}$	–	0.80 [1.00]	–	0.20 [0.00]
$\{c\}$	–	–	0.80 [1.00]	0.20 [0.00]
$\{a, b\}$	0.61 [0.70]	0.19 [0.30]	–	0.20 [0.00]
$\{a, c\}$	0.48 [0.50]	–	0.32 [0.50]	0.20 [0.00]
$\{b, c\}$	–	0.67 [0.80]	0.13 [0.20]	0.20 [0.00]
$\{a, b, c\}$	0.51 [0.50]	0.19 [0.30]	0.10 [0.20]	0.20 [0.00]

Outside-option choices recover first-position attention of 0.8 in every menu. Sequential Path Independence then gives reach probabilities 0.8, $(0.8)^2$, and $(0.8)^3$ along the three-item list. Substituting these values into the recursion recovers the full attention choice probabilities illustrated above in square brackets. Note that $f(a|X) < \pi(a|X)$ while $f(b|X) > \pi(b|X)$. In this three-alternative example, $\mathcal{T}(f)$ is a singleton and recovers the three stated preference types. With larger universes, $\mathcal{T}(f)$ need not be a singleton even though f and the standard lower-contour and rank distributions remain identified. ▲

3.4 Identification under Limited Data

In the previous section, we consider identification under full domain. Here, we outline the sufficient conditions for identification under limited data, and provide partial identification results. First, recall that a domain $\mathcal{D}$ is suffix-closed (prefix-closed) if the suffixes (prefixes) of S belong to $\mathcal{D}$ whenever $S \in \mathcal{D}$.

Proposition 4 (Sufficient Condition for Identification: Attention and Full-attention choice)**.** Let $\mathcal{D}$ be suffix-closed and (ϕ, τ) be a H-LAO representation of π. Then, $\phi(a|S)$ is point identified for every $S \in \mathcal{D}$ and $a \in S$. If, in addition, $\mathcal{D}$ is also prefix-closed, $\tau(\{\succ \in \mathcal{P} : x \text{ is } \succ\text{-best in } S\})$ is point identified for every $S \in \mathcal{D}$ and $x \in S$.

Thus suffix data identify attention frequencies, whereas recovery of the full-attention choice probabilities additionally requires the relevant smaller-prefix problems. The latter condition can be checked recursively and requires observation of the relevant contiguous list intervals.

For later use, define the attention-recoverable domain

$$\mathcal{D}_\phi = \{S \in \mathcal{D} : S_{j:} \in \mathcal{D} \text{ for every } j = 1, \ldots, |S|\}.$$

Let $\mathcal{D}_f$ be the smallest collection of menus that contains every singleton and has the following recursive property: a menu $S \in \mathcal{D}_\phi$ with $|S| \geq 2$ belongs to $\mathcal{D}_f$ whenever $\varphi(a_{s_{|S|}}|S) > 0$ and every proper prefix $S_{:j}$ belongs to $\mathcal{D}_f$. Thus $\varphi(a|S)$ is recovered for $S \in \mathcal{D}_\phi$, while $f(a|S)$ is recovered for $S \in \mathcal{D}_f$.

We now provide an additional identification result for binary comparisons. Such comparisons may be particularly relevant to policymakers seeking to determine the proportion of the population that prefers one policy option to another. This information can help assess the relative support for competing policies and inform decisions between alternative interventions. Suppose a precedes b and the menus $\{a, b\}$ and $\{b\}$ are observed. Then $[1 - \pi(o^*|\{a, b\})]\,[1 - \pi(o^*|\{b\})]$ is the probability of reaching b in the binary menu. Then, one can show that

$$\tau(\{\succ : b \succ a\}) = \frac{\pi(b|\{a, b\})}{[1 - \pi(o^*|\{a, b\})]\,[1 - \pi(o^*|\{b\})]}$$

Proposition 5 (Sufficient Condition for Identification: Pairwise Preferences)**.** Let $\mathcal{D}$ contain $\{a, b\}$, $\{a\}$ and $\{b\}$, and (ϕ, τ) be a H-LAO representation of π. Then, $\tau(\{\succ \in \mathcal{P} : b \succ a\})$ is point identified.

Hence singleton and binary menus identify all pairwise shares. This conclusion uses only binary and singleton menus rather than suffix and prefix closure. In general, when the aforementioned domain conditions are not satisfied, attention or preference need not be point identified. Nevertheless, useful partial-identification bounds remain available.

Proposition 6 (Partial Identification of Attention)**.** Let (ϕ, τ) be an H-LAO representation of π, let $S \subseteq X$ be nonempty, and let $a \in S$. Then

$$\max_{R \in \mathcal{D}_\phi : R \supseteq S} \varphi(a|R) \leq \phi(a|S) \leq \min_{T \in \mathcal{D}_\phi : a \in T \subseteq S} \varphi(a|T)$$

A bound is omitted if its index set is empty.

The proposition bounds the attention frequency of a even when the target menu S is unob-

served.[11] If $S \in \mathcal{D}$ and the suffixes required to recover $\varphi(a|S)$ are observed, both bounds include S and collapse to the identified value. For example, suppose that every menu in Example 5 is observed except $\{a, c\}$. Then, the lower bound is given by $\varphi(c|\{a,b,c\}) = (1-\pi(o^*|\{a,b,c\}))(1-\pi(o^*|\{b,c\}))(1-\pi(o^*|\{c\})) = 0.8^3$, and the upper bound is given by $\varphi(c|\{c\}) = (1-\pi(o^*|\{c\})) = 0.8$, whereas the actual $\phi(c|\{a,c\})$ is 0.8^2. This example shows that neither bound requires data on $\{a, c\}$.

We now discuss the identification of the distribution of types under limited data. We exploit two distinct sources of identifying power. First, because f is random-utility rationalizable, it satisfies the classic regularity so that $f(a|T) \geq f(a|S) \geq f(a|R)$ for $a \in T \subseteq S \subseteq R$. Second, the list structure can strengthen the lower bound. When S is a prefix of R, define

$$\Psi(a, S, R) = \sum_{\substack{1 \leq j \leq |R|: \\ a \in R_{:j} \subsetneq S}} f(a|R_{:j})\mu(R_{:j}|R).$$

Write $\ell(S) := a_{s_{|S|}}$ for the last alternative of S. For $a \in S$, define the eligible-superset collection

$$\mathcal{R}_\Psi(a, S) = \Big\{ R \in \mathcal{D}_\phi : S \text{ is a prefix of } R,\ \varphi(\ell(S)|R) > 0, R_{:j} \in \mathcal{D}_f \text{ whenever } a \in R_{:j} \subsetneq S \Big\}.$$

With this defined, we are ready to state the proposition.

Proposition 7 (Partial Identification of τ)**.** Let (ϕ, τ) be an H-LAO representation of π, let $S \subseteq X$ be nonempty, and let $a \in S$. Write $\theta(a, S) := \tau(\{\succ : a \text{ is } \succ\text{-best in } S\})$. Then

$$\max\left\{ \max_{R \in \mathcal{R}_\Psi(a,S)} \frac{\pi(a|R) - \Psi(a,S,R)}{\varphi(\ell(S)|R)},\ \max_{R \in \mathcal{D}_f : R \supseteq S} f(a|R) \right\} \leq \theta(a, S) \leq \min_{T \in \mathcal{D}_f : a \in T \subseteq S} f(a|T).$$

Each extremum is omitted if its index set is empty.

To utilize this result, suppose in Example 5 every menu is observed except $\{a, b\}$. Then, the first lower bound for $\tau(\{\succ : a \text{ is } \succ\text{-best in } \{a,b\}\})$ can be calculated as

$$\frac{\pi(a|\{a,b,c\}) - (\varphi(a|\{a,b,c\}) - \varphi(b|\{a,b,c\}))}{\varphi(b|\{a,b,c\})} = \frac{0.5056 - (0.8 - 0.8^2)}{0.8^2} = 0.540$$

[11] Although S need not belong to the observed domain $\mathcal{D}$, the structural H-LAO attention frequency is defined on the full menu domain.

whereas the second lower bound is not applicable since $f(a|\{a,b,c\})$ is not well-defined if $\{a,b\}$ is missing from $\mathcal{D}$. On the other hand, suppose in Example 5 every menu is observed except $\{a,c\}$. Then, the first lower bound for $\tau(\{\succ\colon a \text{ is } \succ\text{-best in } \{a,c\}\})$ does not apply since there is not a set R so that $\Psi(a,\{a,c\},R)$ is defined. Nonetheless, the second lower bound is well-defined since $f(a|\{a,b,c\})$ is calculated as 0.5 from π given the dataset.

The Supplemental Appendix reports two nested sensitivity envelopes around benchmark H-LAO. The set $\mathcal{T}_{\mathcal{D}}^{\text{dep}}(\pi)$ retains recovered prefix masses but permits arbitrary menu-specific dependence between preferences and stopping. The set $\mathcal{T}_{\mathcal{D}}^{\text{noPI}}(\pi)$ additionally treats the prefix masses as latent and drops Sequential Path Independence. Both are sharp for their stated repeated-cross-section specifications, and $\mathcal{T}_{\mathcal{D}}(\pi) \subseteq \mathcal{T}_{\mathcal{D}}^{\text{dep}}(\pi) \subseteq \mathcal{T}_{\mathcal{D}}^{\text{noPI}}(\pi)$ whenever the three sets are defined. This nesting makes the identifying content of preference–stopping independence and Sequential Path Independence transparent. All definitions, proofs, multiple-list results, and projection procedures appear in the Supplemental Appendix.

4 Econometric Methods

This section develops estimation and inference for both models. We begin with homogeneous AOM, using Theorem 1, Propositions 1 and 2, and Theorem 2 to test candidate preferences and form confidence bounds for attention. We then turn to H-LAO and construct confidence sets for preference distributions and attention that remain valid under weak reach.

Related econometric work identifies preferences in the presence of costly search using rich covariate variation and derivative restrictions (Abaluck, Compiani, and Zhang, 2026). Our procedures instead take menu and list variation as the source of identification and conduct simultaneous inference directly on the resulting choice-probability restrictions. Section SA-1 compares additional approaches to latent consideration and preference heterogeneity.

4.1 Inference for Homogeneous AOM

For homogeneous AOM, we observe choices for n units indexed by $i = 1,\ldots,n$. Unit i faces menu $Y_i \in \mathcal{D}$ and chooses $y_i \in Y_i$, where $\mathcal{D} \subseteq 2^X \setminus \{\emptyset\}$ is the collection of observed menus.

Assumption 4 (Choice Data)**.** The data consist of a random sample of menus and observed choices $\{(Y_i, y_i) : 1 \leq i \leq n\}$ with $Y_i \in \mathcal{D}$ and $\mathbb{P}[y_i = a | Y_i = S] = \pi(a|S)$.

We conduct inference conditionally on the realized menu assignments $(Y_1, \ldots, Y_n)$, or equivalently on the menu-specific sample sizes and observation indices. Conditional on these assignments, choices from different menus are independent and $N_S = \sum_{i=1}^{n} \mathbb{1}(Y_i = S)$, the sample size for the choice problem S, is fixed. All probability statements below use this conditional interpretation; integrating over the menu assignments gives the corresponding unconditional conclusions. We restrict $\mathcal{D}$ to menus with $N_S \geq 2$ and consider sequences for which the relevant menu-specific sample sizes diverge. Throughout this section, every logarithmic factor is understood to be truncated below at one, that is, $\max\{\log(\cdot), 1\}$.

A given preference ordering $\succ$ is compatible with our AOM if and only if $\succ$-Regularity holds. For inference, define the nonredundant comparison index

$$\mathcal{I}_{\succ}(\mathcal{D}) = \{(a, S, T) : T, S \in \mathcal{D},\ T \subset S,\ a \in T,\ U_{\succeq}(a) \cap T \neq T\}.$$

The last condition is equivalent to requiring that a not be the $\succ$-worst alternative in T, as otherwise $\pi(U_{\succeq}(a)|T) = 1$ and the corresponding restriction $\pi(a|S) \leq 1$ is automatically satisfied. We therefore collect the informative inequality constraints in the following hypothesis:

$$\mathsf{H}_0 : \max_{(a,S,T) \in \mathcal{I}_{\succ}(\mathcal{D})} D(a|S,T) \leq 0, \quad \text{where } D(a|S,T) = \pi(a|S) - \pi(U_{\succeq}(a)|T). \tag{4}$$

To construct a test statistic for the hypotheses in (4), we can replace the unknown choice probabilities by their estimates. Specifically, for any event $A \subseteq S$, let $\widehat{\pi}(A|S)$ be an estimated choice probability, with $\sigma^2(A|S)$ and $\widehat{\sigma}^2(A|S)$ denoting its true and estimated variances:

$$\widehat{\pi}(A|S) = \frac{1}{N_S} \sum_{i=1}^{n} \mathbb{1}(Y_i = S, y_i \in A),$$
$$\sigma^2(A|S) = \frac{\pi(A|S)(1 - \pi(A|S))}{N_S}, \quad \widehat{\sigma}^2(A|S) = \frac{\widehat{\pi}(A|S)(1 - \widehat{\pi}(A|S))}{N_S}.$$

We write $\widehat{\pi}(a|S)$, $\sigma^2(a|S)$, and $\widehat{\sigma}^2(a|S)$ as shorthand for the singleton event $A = \{a\}$.

Now define the test statistic

$$\mathsf{T}(\succ) = \max\left\{\max_{(a,S,T)\in\mathcal{I}_\succ(\mathcal{D})} \widehat{D}(a|S,T)/\widehat{\sigma}(a|S,T)\ ,\ 0\right\},$$

where $\widehat{\sigma}(a|S,T)$ is the standard error of $\widehat{D}(a|S,T)$, so $\widehat{\sigma}^2(a|S,T) = \widehat{\sigma}^2(a|S) + \widehat{\sigma}^2(U_\succeq(a)|T)$. Truncating the maximum at zero prevents rejection when every estimated difference is nonpositive. The notation $\mathsf{T}(\succ)$ emphasizes that the inequalities and the statistic depend on the candidate ordering.

The standardized procedure is used for comparisons with positive oracle variance. We interpret $0/0$ as zero and a nonzero numerator divided by zero as signed infinity, leaving the other comparisons active. Gaussian simulation sets coordinates with zero estimated standard errors to zero. Under the conditions below, the probability of a zero estimated standard error vanishes. The plug-in covariance matrix used for Gaussian simulation is computed from the menu-level multinomial covariance matrices and is therefore positive semidefinite.

The validity argument has two steps: a uniform Gaussian coupling and a distributional comparison for the feasible critical value. For one comparison, $\widehat{D}(a|S,T)/\sigma(a|S,T)$ is approximately Gaussian with unit variance and mean $D(a|S,T)/\sigma(a|S,T) \le 0$ under the null. The challenge is to approximate all comparisons simultaneously when their number may be large and their standard errors are estimated. The following theorem supplies that coupling.

Theorem 6 (Uniform Gaussian Coupling)**.** Assume Assumption 4 holds and all oracle variances entering the comparisons are positive. Let $\mathfrak{c}_\succ = |\mathcal{I}_\succ(\mathcal{D})|$ be the number of nonredundant comparisons in (4), and

$$\mathfrak{c}_\sigma = \min_{(a,S,T)\in\mathcal{I}_\succ(\mathcal{D})} \left\{ \min\left\{\frac{N_S}{\log(N_S)}, \frac{N_T}{\log(N_T)}\right\} \cdot \max\left\{\sigma(a|S),\ \sigma(U_\succeq(a)|T)\right\}\right\},$$

Suppose $\mathfrak{c}_\sigma \to \infty$ and $\log(\mathfrak{c}_\succ)/\mathfrak{c}_\sigma \to 0$. Then, on a possibly enlarged probability space, there exist mean-zero Gaussian random variables $\{\mathsf{Z}_\succ(a|S,T) : (a,S,T) \in \mathcal{I}_\succ(\mathcal{D})\}$ such that (i) their covariance structure matches that of the centered and oracle-standardized comparisons indexed by $\mathcal{I}_\succ(\mathcal{D})$; and (ii)

$$\max_{(a,S,T)\in\mathcal{I}_\succ(\mathcal{D})} \left|\frac{\widehat{D}(a|S,T) - D(a|S,T)}{\widehat{\sigma}(a|S,T)} - \mathsf{Z}_\succ(a|S,T)\right| = O_{\mathbb{P}}\left(\frac{\log(\mathfrak{c}_\succ)}{\mathfrak{c}_\sigma}\right).$$

The coupling error depends on the number of comparisons only logarithmically, so the approximation permits many inequalities relative to the effective menu sample sizes and provides an explicit sample-size/dimension benchmark. Because the universal constants are not calibrated, the Supplemental Appendix complements the theorem with simulations documenting finite-sample performance.

The definition of $\mathfrak{c}_\sigma$ keeps each comparison's menu sample sizes paired with its own standard errors before taking the minimum over comparisons. It therefore avoids multiplying the smallest sample size from one comparison by the least favorable standard error from an unrelated comparison. Intuitively, $\mathfrak{c}_\sigma$ measures the smallest effective sample size on a square-root scale, up to logarithmic factors. When menu sample sizes are comparable and the relevant choice indicators have variances of order $\pi(1-\pi)$, it simplifies to

$$\mathfrak{c}_\sigma \propto \min_{T,S\in\mathcal{D}} \min\left\{\frac{\sqrt{\pi(1-\pi)N_S}}{\log(N_S)},\ \frac{\sqrt{\pi(1-\pi)N_T}}{\log(N_T)}\right\} = \min_{S\in\mathcal{D}} \frac{\sqrt{\pi(1-\pi)N_S}}{\log(N_S)}.$$

Since $\pi(1-\pi)N_S$ is related to the expected number of times an alternative is selected, a Gaussian approximation to a multinomial distribution typically requires this effective sample size to diverge (even in a low and fixed dimensional setting).

There is consequently no explicit requirement that $|X|$ be $o(n)$ or $o(\sqrt{n})$. The formal restriction operates through $\log(\mathfrak{c}_\succ)$ and the smallest effective menu sample size. With complete menu data, the number of menus and inequalities grows combinatorially in $|X|$, making adequate per-menu sampling increasingly demanding. For example, if n observations are allocated approximately evenly across the $2^{|X|}-1$ nonempty menus, then $N_S \approx n/(2^{|X|}-1) \approx n/2^{|X|}$. Thus, even the basic requirement $N_S \to \infty$ entails $2^{|X|} = o(n)$, so $|X|$ can grow only roughly logarithmically with n; the theorem's additional logarithmic terms tighten this benchmark. This feature motivates our methods for incomplete domains and binary screening: the six-alternative simulation has 664 nonredundant inequalities under the complete domain but as few as 15 under the reported sparse design. Binary screening avoids forming all subset relations, after which Corollary 2 represents the surviving rankings through a mixed-integer feasibility problem instead of enumerating them. Section SA-5 of the Supplemental Appendix gives the implementation hierarchy, optimization dimensions, solver certificates, and reproducible scalability benchmarks for both AOM and H-LAO.

Theorem 6 attains the nearly optimal inverse-root-effective-sample-size scale. The proof represents each menu-level choice indicator as a function of a scalar uniform variable, making the centered comparisons an empirical process indexed by functions of total variation at most two. A bounded-variation strong coupling (Cattaneo, Feng, and Underwood, 2024, Lemma SA26), followed by control of the estimated standard errors, gives the result; see also Cattaneo and Yu (2025, Remark 1). Direct distributional approximations provide another route (see Section 5 of Chernozhukov, Chetverikov, and Koike, 2023). Our theorem does not require a lower bound on the minimum eigenvalue of the covariance matrix and is therefore closer in scope to Chernozhukov, Chetverikov, Kato, and Koike (2022); applying their general result directly would yield an error of order $\mathfrak{c}_\sigma^{-1/2}$ up to logarithmic factors.

The next step is to construct critical values with a feasible Gaussian approximation. Conditional on the data, we simulate an auxiliary Gaussian vector independently of the sample, with covariance equal to the estimated correlation matrix of the standardized comparisons. Let $\widehat{\mathsf{Z}}_\succ(a|S,T)$ denote its coordinates (see the Supplemental Appendix for the precise formula), and define

$$\widehat{\mathsf{T}}^{\mathsf{G}}(\succ) = \max\left\{\max_{(a,S,T)\in\mathcal{I}_\succ(\mathcal{D})} \widehat{\mathsf{Z}}_\succ(a|S,T)\ ,\ 0\right\}.$$

Define $\widehat{\mathrm{cv}}(\alpha,\succ) = \inf\{t : \mathbb{P}[\widehat{\mathsf{T}}^{\mathsf{G}}(\succ) \le t|\text{Data}] \ge 1-\alpha\}$. The next theorem bounds the test's size distortion.

Theorem 7 (Preference Elicitation with Theorem 1)**.** Assume the conditions of Theorem 6 hold, fix $\alpha \in (0, 1/2)$, and suppose the remainder displayed below converges to zero. Then, under the null hypothesis H_0 in (4),

$$\mathbb{P}\big[\mathsf{T}(\succ) > \widehat{\mathrm{cv}}(\alpha,\succ)\big] \le \alpha + c\sqrt{\frac{\log^{\frac{5}{2}}(\mathfrak{c}_\succ)\log^{\frac{1}{2}}(\mathfrak{c}_\sigma)}{\mathfrak{c}_\sigma}}$$

for some constant c.

Standard test inversion gives the asymptotically valid confidence set $\mathsf{CS}(1-\alpha) = \{\succ\colon \mathsf{T}(\succ) \le \widehat{\mathrm{cv}}(\alpha,\succ)\}$. Condition (1) can be incorporated by adding the probability comparisons implied by the η-constrained binary relation.

The size bound above uses the fully general Gaussian comparison inequality in Proposition 2.1 of Chernozhukov, Chetverikov, Kato, and Koike (2022) and therefore delivers a conservative fourth-root error remainder without imposing nonsingularity. When the oracle correlation matrix is bounded away from singularity, Theorem 2.3 of Lopes (2022) gives nearly linear covariance-error dependence; the Supplemental Appendix verifies this condition for the important binary-screening design under balanced standard errors. It also gives a universal critical value and a finite-sample projection procedure based on simultaneous menu-probability bands, neither of which requires covariance comparison, as well as an alpha-split hybrid of the universal and correlated-Gaussian critical values. Generalized moment selection can improve power by reducing the influence of inequalities that are far from binding. The all-inequalities zero-centered critical value (labeled `LF` in the software) is the procedure covered directly by Theorem 7. The simulations report this baseline and the GMS refinement of Andrews and Soares (2010) from the same Gaussian draws. We retain the correlated-Gaussian result as the default because it uses the dependence structure and is generally more powerful.

Pairwise revealed preference can be assessed by checking whether, for example, $a \succ b$ for every $\succ \in \mathsf{CS}(1-\alpha)$. This approach uses the full identifying power of Theorem 1 but can be computationally costly. Proposition 1 supplies a simpler screen when binary menus are observed or pairwise comparisons are the primary target. As Example 1 illustrates, applying this screen first can sharply reduce the number of orderings that must be tested against $\succ$-Regularity.

We fix two alternatives, say a and b, and let $\mathcal{D}_{ab}$ be the collection of choice problems containing both a and b, excluding the binary comparison; that is, $\mathcal{D}_{ab} = \{S \in \mathcal{D} : \ S \supsetneq \{a,b\}\}$. We also recall the simplified notation $\pi(a|b) = \pi(a|\{a,b\})$ and $N_{ab} = N_{\{a,b\}}$ for binary comparisons. Then, we may deduce the preference ordering $b \succ a$ if we are able to reject $\mathsf{H}_0 : \max_{S\in\mathcal{D}_{ab}} D(a|S,b) \leq 0$, where we define $D(a|S,b) = \pi(a|S) - \pi(a|b)$. The corresponding test statistic is

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

where $\widehat{D}(a|S,b) = \widehat{\pi}(a|S) - \widehat{\pi}(a|b)$, and $\widehat{\sigma}(a|S,b)$ is its standard error. We employ the same technique to construct a critical value. Specifically, let $\widehat{\mathsf{Z}}_{ab}(S)$ for $S \in \mathcal{D}_{ab}$ be a collection of Gaussian random variables with zero mean, unit variance, and covariance matching the estimated correlation

matrix of $\widehat{D}(a|S,b)$. The critical value is $\widehat{\mathrm{cv}}(\alpha, ab) = \inf\{t : \mathbb{P}[\widehat{\mathsf{T}}^{\mathsf{G}}(ab) \leq t|\text{Data}] \geq 1-\alpha\}$ with $\widehat{\mathsf{T}}^{\mathsf{G}}(ab) = \max\{\max_{S\in\mathcal{D}_{ab}} \widehat{\mathsf{Z}}_{ab}(S), 0\}$. The next result follows from Theorem 7.

Corollary 4 (Preference Elicitation with Proposition 1)**.** Assume the conditions of Theorem 7 hold for the binary-screening comparisons. Define $\mathfrak{c}_\sigma$ as

$$\mathfrak{c}_\sigma = \min_{S\in\mathcal{D}_{ab}} \left\{ \min\left\{ \frac{N_S}{\log(N_S)}, \frac{N_{ab}}{\log(N_{ab})} \right\} \cdot \max\Big\{ \sigma(a|S), \sigma(a|b) \Big\} \right\}.$$

Then, under the null hypothesis that $\mathsf{H}_0 : \max_{S\in\mathcal{D}_{ab}} D(a|S,b) \leq 0$, the size distortion satisfies

$$\mathbb{P}\big[\mathsf{T}(ab) > \widehat{\mathrm{cv}}(\alpha, ab)\big] \leq \alpha + c\sqrt{\frac{\log^{\frac{5}{2}}(|\mathcal{D}_{ab}|)\log^{\frac{1}{2}}(\mathfrak{c}_\sigma)}{\mathfrak{c}_\sigma}}$$

for some constant c. If, in addition, there is a constant $C < \infty$ such that

$$C^{-1} \leq \frac{\sigma(a|S)}{\sigma(a|b)} \leq C \qquad \text{for every } S \in \mathcal{D}_{ab},$$

and $\log^{5/2}(|\mathcal{D}_{ab}|\mathfrak{c}_\sigma)/\mathfrak{c}_\sigma \to 0$, then the fixed-pair size distortion satisfies the sharper bound

$$\mathbb{P}\big[\mathsf{T}(ab) > \widehat{\mathrm{cv}}(\alpha, ab)\big] \leq \alpha + c(1+C^2)\frac{\log^{5/2}(|\mathcal{D}_{ab}|\mathfrak{c}_\sigma)}{\mathfrak{c}_\sigma}.$$

The comparisons $\widehat{D}(a|S,b)$ share only the binary-menu estimate $\widehat{\pi}(a|b)$; the estimates $\widehat{\pi}(a|S)$ are independent across $S \in \mathcal{D}_{ab}$. If $\sigma(a|S) \asymp \sigma(a|b)$ uniformly over $\mathcal{D}_{ab}$, this diagonal-plus-rank-one structure keeps the minimum eigenvalue of the correlation matrix bounded away from zero. The nondegenerate Gaussian comparison therefore removes the outer square root in the fixed-pair error rate, as formalized in Corollary 4. The Supplemental Appendix provides the covariance calculation and rate derivation. A screen that stacks multiple ordered pairs remains covered by the singularity-robust bound unless nondegeneracy of its joint correlation matrix is established separately.

We use the lower attention bound $\phi(a|S) \geq \max_{R\in\mathcal{D}:\, R\supseteq S} \pi(a|R)$ from Theorem 2 to construct a one-sided confidence bound; the upper bound is handled analogously. Maximizing noisy estimates creates upward selection bias and can overstate the lower bound. We therefore construct $\underline{\widehat{\phi}}(a|S)$ with coverage approaching $1-\alpha$.

Our construction is based on computing the maximum over a collection of adjusted empirical choice probabilities. This is a finite intersection-bound problem. As emphasized by Chernozhukov, Lee, and Rosen (2013), directly taking a supremum or infimum of noisy estimates creates selection bias and can make the estimated identified set too small. Our correction follows the same general principle, but specializes it to a finite collection of menu-specific multinomial probabilities. This yields a pivotal Gaussian maximum because the menu subsamples are conditionally independent; their framework also accommodates continua of inequalities, nonparametric first stages, and adaptive inequality selection. To be precise, we define

$$\widehat{\underline{\phi}}(a|S) = \max_{R \supseteq S, R \in \mathcal{D}} \Big\{ \widehat{\pi}(a|R) - \mathrm{cv}(\alpha, \underline{\phi}(a|S)) \cdot \widehat{\sigma}(a|R) \Big\}.$$

Here, $\mathrm{cv}(\alpha, \underline{\phi}(a|S))$ is the $(1-\alpha)$-quantile of $\max_{R \supseteq S, R \in \mathcal{D}} \mathsf{Z}_{a,S}(R)$, where $\mathsf{Z}_{a,S}(R)$ are independent standard Gaussian random variables. As such, the critical value is pivotal. If any required estimated standard error is zero, we conservatively set $\widehat{\underline{\phi}}(a|S) = 0$. Otherwise, the construction uses the normal approximation $\widehat{\pi}(a|R) \overset{\text{a}}{\sim} \mathrm{Normal}(\pi(a|R),\ \sigma^2(a|R))$. Conditional on the menu assignments, the estimated choice probabilities are mutually independent because they are constructed from different subsamples, and

$$\begin{aligned}
&\mathbb{P}\Big[\widehat{\pi}(a|R) - \mathrm{cv}(\alpha, \underline{\phi}(a|S)) \cdot \widehat{\sigma}(a|R) \le \pi(a|R),\ \forall R \supseteq S,\ R \in \mathcal{D}\Big] \\
&\quad = \mathbb{P}\Big[\max_{R \supseteq S, R \in \mathcal{D}} \frac{\widehat{\pi}(a|R) - \pi(a|R)}{\widehat{\sigma}(a|R)} \le \mathrm{cv}(\alpha, \underline{\phi}(a|S))\Big] \\
&\quad \approx \mathbb{P}\Big[\max_{R \supseteq S, R \in \mathcal{D}} \mathsf{Z}_{a,S}(R) \le \mathrm{cv}(\alpha, \underline{\phi}(a|S))\Big] = 1 - \alpha.
\end{aligned}$$

where $\approx$ denotes a large-sample approximation. Thus, with probability approaching $1-\alpha$, every adjusted estimate lies below its population choice probability, and their maximum lies below the population attention bound. The next theorem makes this argument precise.

Theorem 8 (Attention Frequency Elicitation with Theorem 2)**.** Let π be AOM-compatible, assume Assumption 4 holds, and suppose $0 < \pi(a|R) < 1$ for every $R \in \mathcal{D}$ with $R \supseteq S$ entering the construction. Define

$$\mathfrak{c}_{\supseteq S} = \big|\{R \in \mathcal{D} : R \supseteq S\}\big|$$

as the number of supersets of S, and

$$\mathfrak{c}_\sigma = \min_{R \supseteq S, R \in \mathcal{D}} \frac{\sqrt{N_R \pi(a|R)(1-\pi(a|R))}}{\log(N_R)}.$$

Suppose $\mathfrak{c}_\sigma \to \infty$ and $\log(\mathfrak{c}_{\supseteq S})/\mathfrak{c}_\sigma \to 0$. Then,

$$\mathbb{P}\big[\phi(a|S) \geq \underline{\widehat{\phi}}(a|S)\big] \geq 1 - \alpha - c\frac{\log^{\frac{3}{2}}(\mathfrak{c}_{\supseteq S})\log(\mathfrak{c}_\sigma)}{\mathfrak{c}_\sigma}$$

for some constant c.

4.2 Estimation and Inference for Heterogeneous H-LAO

H-LAO (partially) identifies the preferences distribution rather than selecting one candidate ordering. Estimation and statistical inference again start from the empirical choice rule. Replacing π by $\widehat{\pi}$ in (3) gives plug-in estimates of all reach probabilities and prefix masses on $\mathcal{D}_\phi$. On menus in $\mathcal{D}_f$, the recursion also gives a plug-in estimate of the full-attention rule f, after which preference-event bounds are obtained by linear programming over $\mathcal{T}(f)$. On suffix-closed domains, the direct polytope $\mathcal{T}_{\mathcal{D}}(\widehat{\pi})$ instead uses undivided observed-choice equations. It is the appropriate formulation near zero reach or when the proper-prefix support needed for $\mathcal{D}_f$ is unavailable.

We extend Assumption 4 by allowing the observed outcome set to be $\mathcal{Y}(S) = S \cup \{o^*\}$. Let

$$d_{\mathcal{D}} = \sum_{S \in \mathcal{D}} |\mathcal{Y}(S)|, \qquad \varepsilon_S(\alpha) = \sqrt{\frac{\log(2d_{\mathcal{D}}/\alpha)}{2N_S}},$$

and form simultaneous cell bands

$$\underline{p}(a|S) = \max\{0, \widehat{\pi}(a|S) - \varepsilon_S(\alpha)\}, \qquad \overline{p}(a|S) = \min\{1, \widehat{\pi}(a|S) + \varepsilon_S(\alpha)\}.$$

Hoeffding's inequality and a union bound imply that all true menu-outcome probabilities belong to these bands with probability at least $1-\alpha$, conditionally on the menu assignments. Let $\mathcal{C}^{\mathrm{H}}_{1-\alpha}$ be the set of choice rules satisfying the bands and the menu-specific adding-up restrictions. To distinguish the data-generating object from the choice-rule arguments over which the projection

is computed, write $p_0 = \pi$ for the true choice rule under the maintained model and use p for a generic candidate choice rule. More generally, let $\mathcal{C}_{1-\alpha}$ be any confidence region for p_0 satisfying $\mathbb{P}[p_0 \in \mathcal{C}_{1-\alpha}] \geq 1-\alpha$; the Hoeffding region is a finite-sample example.

The population identified sets below are defined in Propositions SA-4–SA-6 of the Supplemental Appendix. For a suffix-closed $\mathcal{D}$, define

$$\widehat{\mathcal{T}}^{\text{ind}}_{1-\alpha} = \bigcup_{p \in \mathcal{C}_{1-\alpha}} \mathcal{T}_{\mathcal{D}}(p), \qquad \widehat{\mathcal{T}}^{\text{dep}}_{1-\alpha} = \bigcup_{p \in \mathcal{C}_{1-\alpha}} \mathcal{T}^{\text{dep}}_{\mathcal{D}}(p).$$

On any observed-menu domain, also define

$$\widehat{\mathcal{T}}^{\text{noPI}}_{1-\alpha} = \bigcup_{p \in \mathcal{C}_{1-\alpha}} \mathcal{T}^{\text{noPI}}_{\mathcal{D}}(p).$$

For the first two sets, the identified set is taken to be empty when recovered attention is invalid.

Theorem 9 (Projection Inference for H-LAO)**.** Suppose $\mathbb{P}[p_0 \in \mathcal{C}_{1-\alpha}] \geq 1-\alpha$. If p_0 has an independent H-LAO representation with preference distribution τ_0, then

$$\mathbb{P}\left[\tau_0 \in \widehat{\mathcal{T}}^{\text{ind}}_{1-\alpha}\right] \geq 1-\alpha.$$

The analogous conclusion holds for $\widehat{\mathcal{T}}^{\text{dep}}_{1-\alpha}$ when preferences and stopping may be dependent, and for $\widehat{\mathcal{T}}^{\text{noPI}}_{1-\alpha}$ on any observed-menu domain when Sequential Path Independence is also removed. Projecting the corresponding confidence set gives simultaneous coverage for any collection of preference-event probabilities. All conclusions hold in finite samples when $\mathcal{C}_{1-\alpha} = \mathcal{C}^{\text{H}}_{1-\alpha}$.

The reason is direct: on the primitive-region coverage event, $p = p_0$ is included in the union and sharp identification places τ_0 in the corresponding population polytope. Applying the same union construction to the reach maps recovered from each candidate p gives simultaneous confidence sets for attention frequencies under Sequential Path Independence. Without Sequential Path Independence, one instead projects the feasible latent prefix masses. Exact independent projection preserves preference–stopping independence but leads to polynomial feasibility constraints when p varies.

Pairwise inference has a particularly transparent weak-reach implementation. If a precedes b,

H-LAO implies the undivided moment

$$\pi(b|\{a,b\}) = r_b(a,b)\tau(\{\succ\colon b \succ a\}),$$

where $r_b(a,b) = [1-\pi(o^*|\{a,b\})][1-\pi(o^*|\{b\})]$. Let $[\underline{r}_b(a,b), \overline{r}_b(a,b)]$ be the product bounds induced by the outside-option bands. A simultaneous confidence set for the pairwise share is

$$\widehat{C}^{\mathrm{H}}_{ba} = \left\{\theta \in [0,1] : [\underline{r}_b(a,b)\theta, \overline{r}_b(a,b)\theta] \cap [\underline{p}(b|\{a,b\}), \overline{p}(b|\{a,b\})] \neq \emptyset\right\}.$$

When $\overline{r}_b(a,b) > 0$, its endpoints are

$$\max\left\{0, \frac{\underline{p}(b|\{a,b\})}{\overline{r}_b(a,b)}\right\} \qquad \text{and} \qquad \min\left\{1, \frac{\overline{p}(b|\{a,b\})}{\underline{r}_b(a,b)}\right\},$$

where the second ratio is $+\infty$ if $\underline{r}_b(a,b) = 0$. If the upper reach bound is zero and the inside-choice band contains zero, the confidence set is $[0,1]$. Thus, the procedure remains valid at zero reach instead of dividing by a weak denominator.

Corollary 5 (Finite-Sample Pairwise Inference under Weak Reach)**.** Suppose the independent H-LAO model holds. Then, simultaneously for every pair supported by the observed singleton and binary menus,

$$\mathbb{P}\left[\tau_0(\{\succ\colon b \succ a\}) \in \widehat{C}^{\mathrm{H}}_{ba} \text{ for every such pair } (a,b)\right] \geq 1-\alpha.$$

Empty intervals provide a finite-sample specification check.

The preceding interval protects uniformly against a weak or zero denominator but can be conservative. A sharper alternative directly studentizes the same undivided moment. Write

$$\widehat{g}_{ba}(\theta) = \widehat{\pi}(b|\{a,b\}) - \theta\left[1-\widehat{\pi}(o^*|\{a,b\})\right]\left[1-\widehat{\pi}(o^*|\{b\})\right]$$

and let $\widehat{V}_{ba}(\theta)$ be its plug-in delta-method variance, computed from the multinomial covariance

matrix for $\{a, b\}$ and the binomial variance for $\{b\}$. For a fixed collection $\mathcal{B}$ of ordered pairs, define

$$\widehat{C}^{\mathrm{S}}_{ba} = \left\{ \theta \in [0,1] : \widehat{g}_{ba}(\theta)^2 \leq \left[\Phi^{-1}\left(1 - \frac{\alpha}{2|\mathcal{B}|}\right) \right]^2 \widehat{V}_{ba}(\theta) \right\}.$$

If $\widehat{V}_{ba}(\theta) = 0$ for every $\theta \in [0,1]$, set $\widehat{C}^{\mathrm{S}}_{ba} = [0,1]$ by convention.

Corollary 6 (Studentized Pairwise Inference)**.** Suppose the independent H-LAO model holds and $\mathcal{B}$ is fixed. For every $(b, a) \in \mathcal{B}$, suppose the binary- and singleton-menu sample sizes diverge, the variance at the true share is positive and consistently estimated, and the Lyapunov and product-remainder conditions in the Supplemental Appendix hold. Then

$$\liminf_{\min_{(b,a)\in\mathcal{B}}(N_{ab} \wedge N_b) \to \infty} \mathbb{P}\left[\tau_0(\{\succ : b \succ a\}) \in \widehat{C}^{\mathrm{S}}_{ba} \text{ for every } (b,a) \in \mathcal{B} \right] \geq 1 - \alpha.$$

Because the null is imposed before studentization, this Anderson–Rubin/Fieller construction never divides the estimator by reach and remains informative under weak positive reach whenever the moment variance is estimable. The finite-sample projection set $\widehat{C}^{\mathrm{H}}_{ba}$ remains the appropriate safeguard at exact or nearly degenerate zero-reach boundaries. The Supplemental Appendix gives the closed-form variance and proof. The formal result and simulations use Bonferroni calibration; the appendix separately describes a possible covariance-aware extension.

For general preference events under dependence, simultaneous product bounds on reach can be combined with the linear constraints defining $\mathcal{T}^{\mathrm{dep}}_{\mathcal{D}}(p)$. Minimizing and maximizing the event probability over this enlarged feasible set gives finite-sample-valid outer confidence bounds by linear programming. Under the model without Sequential Path Independence, prefix masses are latent variables rather than products of outside-choice probabilities. Combining their adding-up and attention-overload restrictions directly with the primitive cell bands and the constraints defining $\mathcal{T}^{\mathrm{noPI}}_{\mathcal{D}}(p)$ gives a fully linear, finite-sample projection procedure on any observed-menu domain. The Supplemental Appendix gives the exact projection results, the pairwise proofs, and both LP formulations. Correlated-Gaussian or multiplier-bootstrap regions for the primitive probabilities can replace the Hoeffding bands to improve power while preserving the same projection logic.

5 Empirical Application

We apply our theory and methods to the laboratory travel-mode experiment of Wang and Zhu (2025). The application is useful for three reasons. First, menus are manipulated over a common four-alternative universe and a dominated "Default" is always available. Second, the experiment records information acquisition before choice, providing an observable measure of search and engagement with individual modes. Third, subjects report their preference rankings after completing the choice tasks, allowing an external comparison with the preference shares recovered from choice data alone.

5.1 Data and Implementation

The experiment contains 39 subjects, each completing 48 tasks, for 1,872 subject–task observations. In every task the inside menu is a subset of $\{A, B, C, D\}$, while a strictly dominated Default remains available. All 16 inside-menu subsets occur in the data. Twenty-one subjects face the complete 16-menu design. The other 18 face the 11 menus containing at least two inside alternatives. The workbook records menu availability, price and travel-time attributes, whether each attribute was displayed, the order of information acquisition, final choice, and an ex post ranking. Each of the 39 subjects reports one ex post ranking; 37 place all four inside alternatives above Default, while two place Default above an inside alternative.

Table 2 summarizes menu variation and information acquisition. Search intensity rises with menu size: the mean number of displayed attributes increases from 0.36 on singleton menus to 1.44 on the four-alternative menu, while the fraction of tasks with no displayed attribute falls from 0.75 to 0.54. Subjects can choose without displaying an attribute because they are told the ordinal attribute pattern and may rely on prior knowledge. A display therefore provides direct evidence of costly information acquisition about an alternative. Because a few Default choices follow an inside-option display, we interpret displays as recording pre-choice search rather than the final H-LAO consideration set. Conversely, failure to display is not evidence of inattention because subjects can rely on prior knowledge. Default choice outside the empty menu is uncommon, ranging from 0.016 on singleton menus to 0.043 on the grand menu.

Table 2: Menu Size, Information Acquisition, and Default Choice

Inside-menu size	Observations	Mean displays	No-display share	Default-choice share
0	63	0.000	1.000	1.000
1	252	0.357	0.754	0.016
2	648	0.832	0.628	0.031
3	630	1.106	0.579	0.037
4	279	1.441	0.538	0.043

Notes: The inside menu contains travel modes A–D. Default is always available and is the only alternative when the inside menu is empty. A display is one opened price or travel-time attribute. Statistics are computed from the 1,872 subject–task observations supplied by Wang and Zhu (2025).

Prices and travel times vary across tasks. Alternative A is generally fastest and most expensive, B is slowest and least expensive, and C and D are intermediate. Across observations in which each mode is available, mean prices for A–D are 80.6, 24.2, 40.5, and 53.6, while mean travel times are 115.1, 392.8, 310.6, and 156.9. The workbook contains 93 exact four-mode attribute profiles, and no exact profile occurs under more than one menu. Thus exact profile matching cannot remove attribute variation. Our analysis maintains that the distribution of price and travel-time profiles is independent of menu composition. Under this assumption, integrating over the profiles yields the menu-invariant marginal distribution of rankings over stable travel-mode categories required by H-LAO and incorporates attribute variation into the preference heterogeneity allowed by the model.

We use two analysis samples. The primary full-domain sample contains the 21 subjects who face every menu; it contributes 945 observations on the 15 nonempty inside menus. Following the original RAM exercise of Wang and Zhu (2025), the dominance-consistent sample contains the 37 subjects whose ex post reports rank all four inside alternatives above Default. We retain their sample definition for the RAM replication, the matched RAM–AOM comparison, and the broader H-LAO analyses. For the pooled analyses, we maintain that assignment to the complete and reduced-menu designs is as good as random and that design version does not affect behavior conditional on the displayed menu; menu-specific choice and search frequencies therefore identify probabilities for a common population. The criteria serve different purposes: the samples overlap in 20 subjects because one complete-design subject is excluded from the dominance-consistent sample.

We code A–D as the inside alternatives, Default as the outside option, and A–B–C–D as the

observed list order. Default is always available, objectively dominated, and ranked below every inside mode by 37 of the 39 subjects. These features make it a natural empirical counterpart to H-LAO's outside option. Accordingly, its choice represents no effective inside consideration at the final-choice stage, while pre-choice displays record earlier information acquisition. The empty inside menu is used to verify coding but omitted from H-LAO estimation because it selects Default mechanically. For the complete-design cohort, the 15 nonempty menus recover list reach and prefix-stopping probabilities under Sequential Path Independence. We then compute pairwise preference shares, the full-attention choice rule when positive reach permits it, Block–Marschak diagnostics, and preference-event bounds under independence and preference–stopping dependence. Within the dominance-consistent sample, the common-menu analysis restricts attention to the 11 menus shared across the two experimental designs and uses the direct observed-domain polytope and its no-Sequential-Path-Independence extension rather than extrapolating to unobserved singleton menus.

The homogeneous analysis applies the AOM preference and attention procedures to the same menu-choice frequencies and compares the compatible AOM rankings with the RAM rankings reported by Wang and Zhu (2025). Following their original RAM analysis, this exercise adopts a common ranking over the travel-mode categories. H-LAO then allows the marginal distribution of these rankings to be heterogeneous. We report a benchmark that treats task-level observations as independent, but subjects are the independent sampling units because each contributes 48 choices. Our preferred analysis therefore permits arbitrary dependence among a subject's choices across tasks and menus. It conducts inference at the subject level and uses conservative finite-sample bounds for sparse outcomes. For the H-LAO pairwise shares, we construct simultaneous intervals by allowing the menu-specific choice probabilities to vary jointly within confidence bands and calculating the preference shares consistent with those bands and the model. The primitive region uses covariance-aware Gaussian bands for regular row-i.i.d. cells and cluster-multiplier bands for regular clustered cells, with exact-binomial and cluster-Hoeffding fallbacks, respectively, for nonregular cells. When both components are present, they split the common error budget. All simulated critical values in the application use 4,999 draws. Section SA-4.10 of the Supplemental Appendix gives the construction. The complete cohort has $G = 21$ subjects and the dominance-consistent

Table 3: Replication of the RAM Ranking Exercise

		LF		GMS	
Candidate rankings	Sampling	Nonrejected	Min. p-value	Nonrejected	Min. p-value
Nine rational	Row-i.i.d.	9	0.792	9	0.424
Nine rational	Clustered	9	0.186	9	0.066
Two irrational	Row-i.i.d.	0	0.000	0	0.000
Two irrational	Clustered	0	0.000	0	0.000

Notes: The sample contains the 37 subjects with rational ex post reports. The nine rational candidate rankings are the distinct reported rankings of A–D with Default placed last. The two irrational candidate rankings place Default first or above one inside mode. Attentive-at-binaries is set to 2/3, as in Wang and Zhu (2025). Nonrejection counts and minimum p-values are computed within each candidate-ranking group at level 0.05.

sample has $G = 37$; at these cluster counts, the simulations in Section SA-6 show substantially better coverage for the projection intervals than for the shorter studentized alternative.

Observed information acquisition provides valuable additional evidence. For mode a and menu S, define the observable search-and-choice proxy as the fraction of tasks in which the subject either displays at least one attribute of a or chooses a. The proxy combines direct information search with final choice and therefore captures engagement that either measure alone may miss. We use its variation across menus as evidence complementary to the model-based attention estimates. Because displays record pre-choice search, the proxy is a measure of the decision process rather than a literal observation of the final structural consideration set.

5.2 Results

For both RAM and AOM, we report the least-favorable (LF) Gaussian-maximum procedure and its generalized moment selection (GMS) refinement described in Section 4.

We first reproduce the laboratory RAM exercise of Wang and Zhu (2025). Following their design, we remove the two subjects with irrational reports, set attentive-at-binaries to 2/3, and test the nine distinct rational reported rankings together with the two irrational rankings. Table 3 reproduces their result under row-i.i.d. sampling: all nine rational rankings are retained and both irrational rankings are rejected. The conclusion is unchanged after clustering by subject at the 5% level (the smallest GMS p-value among the rational rankings falls from 0.424 to 0.066).

Table 4: AOM Ranking Inference

		LF		GMS	
Sample	Sampling	Nonrejected	Min. p-value	Nonrejected	Min. p-value
Complete	Row-i.i.d.	24	0.987	24	0.562
Complete	Clustered	24	0.829	24	0.260
Dominance-consistent	Row-i.i.d.	24	0.794	24	0.225
Dominance-consistent	Clustered	24	0.159	16	0.024

Notes: "Complete" uses the 21 subjects who face every menu; "Dominance-consistent" uses the same 37 subjects as Table 3. Each sample tests all 24 rankings of A–D that place the dominated Default last. "Nonrejected" is the number of candidate rankings not rejected at the 5% level. GMS uses generalized moment selection; LF uses the least-favorable Gaussian maximum. Clustered inference treats subjects as independent and uses subject multiplier draws. The minimum p-value is taken across the 24 candidate rankings.

Table 4 reports the AOM ranking tests. We treat Default as a fifth, always-available alternative and test all 24 rankings that place this dominated option last. In the complete cohort every ranking survives both GMS and least-favorable calibration under both sampling assumptions. On the 37-subject dominance-consistent sample used in Table 3, row-i.i.d. inference again retains all 24 rankings, while subject-clustered GMS eliminates eight and the least-favorable procedure retains all 24. The nine rational reported rankings in Table 3 are included among the 24 AOM candidates. In the same-subject comparison, clustered RAM retains all nine reported rankings, while clustered AOM GMS retains eight of those nine and narrows the full set from 24 to 16 candidates. Thus, the data are consistent with AOM, and the subject-clustered GMS procedure is informative about homogeneous preferences. The broader complete-cohort set reflects its smaller sample, while the least-favorable procedure provides the deliberately conservative benchmark.

Table 5 reports the direct H-LAO specification diagnostics using least-favorable calibration. The largest recovered-attention discrepancy is 0.078 in the complete cohort and 0.057 in the dominance-consistent sample; the corresponding largest Block–Marschak discrepancies are 0.175 and 0.072. None is statistically significant under either the row-i.i.d. benchmark or subject-clustered inference. The data are therefore jointly consistent with H-LAO's attention-overload and recovered random-utility restrictions, including under our preferred subject-clustered inference.

Table 6 compares the H-LAO estimates with the ex post reports in both samples. In the matched 20-subject complete-design validation sample, the mean absolute difference across the six

Table 5: H-LAO Specification Diagnostics

			Row-i.i.d.		Clustered	
Sample	Restriction	Max. discrepancy	Lower	p-value	Lower	p-value
Complete	Attention overload	0.078	0.000	0.451	0.000	0.510
Complete	Block–Marschak	0.175	-0.102	0.940	-0.034	0.890
Complete	Omnibus	0.175	0.000	0.451	0.000	0.510
Dominance-consistent	Attention overload	0.057	0.000	0.798	0.000	0.745
Dominance-consistent	Block–Marschak	0.072	-0.053	0.915	-0.016	0.274
Dominance-consistent	Omnibus	0.072	0.000	0.798	0.000	0.274

Notes: "Complete" uses the 21 subjects who face every menu; "Dominance-consistent" uses the 37 subjects retained in the original RAM analysis and combines the two experimental designs. A discrepancy is written so that a positive population value violates the indicated restriction. "Lower" is the simultaneous 95% one-sided lower confidence endpoint. The table uses least-favorable (LF) calibration: all nondegenerate inequalities enter the zero-centered maximum, without GMS recentering. The row-i.i.d. columns treat task-level observations as independent; the clustered columns use subject influence vectors and multiplier calibration. The p-values are calibrated over the joint family of diagnostic inequalities using the same Gaussian or subject-cluster multiplier maxima as the confidence endpoints. Full-attention probability restrictions are included in the omnibus test but omitted as a separate row because their largest estimated violation is zero.

pairwise shares is 0.038 and the largest difference is 0.074. In the dominance-consistent sample, the corresponding differences are 0.039 and 0.099. Every reported share lies within its clustered H-LAO interval, and all simultaneous difference intervals contain zero under both sampling schemes. Thus, the point estimates are close, and equality is not rejected by the simultaneous clustered intervals. The comparison uses information not employed in estimating H-LAO: H-LAO uses only menu availability, outside-option choice, inside choice, and the maintained list structure. Inclusion in the validation samples requires a usable dominance-consistent report, but the reported rankings are not used to estimate H-LAO.

Table 7 uses the 11 menus observed for every subject in the dominance-consistent sample and drops Sequential Path Independence. The simultaneous outer confidence bounds remain informative despite this weaker specification: every reversal of the baseline ordering $A \succ B \succ C \succ D$ has a strictly positive clustered lower endpoint. The upper endpoints equal one, so the common-menu design establishes the presence of each preference reversal without tightly measuring its population share.

The search records provide separate descriptive evidence for the dominance-consistent sample. Table 8 reports larger-menu minus smaller-menu differences for 76 nested-menu comparisons. Attribute display alone is mixed: 43 differences are positive, 33 are negative, and none is signifi-

Table 6: Revealed and Self-Reported Pairwise Preference Shares

	Point estimates		Row-i.i.d.		Clustered	
	H-LAO	Reported	Projection CI	Difference CI	Projection CI	Difference CI
Complete-design validation sample						
$B \succ A$	0.874	0.800	[0.657, 1.000]	[-0.206, 0.354]	[0.587, 1.000]	[-0.099, 0.247]
$D \succ B$	0.400	0.400	[0.196, 0.795]	[-0.333, 0.333]	[0.106, 1.000]	[-0.220, 0.220]
$D \succ A$	0.814	0.800	[0.633, 1.000]	[-0.257, 0.284]	[0.644, 1.000]	[-0.193, 0.220]
$C \succ B$	0.576	0.600	[0.360, 1.000]	[-0.360, 0.313]	[0.287, 1.000]	[-0.227, 0.179]
$C \succ A$	0.847	0.900	[0.678, 1.000]	[-0.274, 0.169]	[0.629, 1.000]	[-0.183, 0.078]
$D \succ C$	0.267	0.200	[0.082, 0.594]	[-0.213, 0.346]	[0.013, 1.000]	[-0.084, 0.217]
Dominance-consistent sample						
$B \succ A$	0.760	0.784	[0.598, 1.000]	[-0.238, 0.190]	[0.519, 1.000]	[-0.161, 0.113]
$D \succ B$	0.419	0.432	[0.251, 0.745]	[-0.266, 0.240]	[0.196, 1.000]	[-0.162, 0.136]
$D \succ A$	0.803	0.811	[0.643, 1.000]	[-0.216, 0.199]	[0.652, 1.000]	[-0.144, 0.128]
$C \succ B$	0.536	0.541	[0.355, 0.923]	[-0.262, 0.252]	[0.308, 1.000]	[-0.160, 0.150]
$C \succ A$	0.779	0.865	[0.625, 1.000]	[-0.277, 0.105]	[0.576, 1.000]	[-0.216, 0.044]
$D \succ C$	0.369	0.270	[0.221, 0.639]	[-0.127, 0.326]	[0.155, 1.000]	[-0.052, 0.250]

Notes: For the complete-design validation panel, both H-LAO estimates and reported shares use the 20 complete-design subjects with usable dominance-consistent reports. The 21-subject complete cohort remains the sample for the main H-LAO analysis. For the dominance-consistent sample, both estimates use the 37 subjects retained in the original RAM analysis. Difference is H-LAO minus Reported. The H-LAO intervals allow the menu-specific choice probabilities to vary jointly within simultaneous confidence bands and report all pairwise shares consistent with those bands and the model. The row-i.i.d. benchmark uses correlated-Gaussian bands for cells with at least five successes, at least five failures, and positive estimated variance, and exact binomial bands otherwise. Clustered inference treats subjects as the independent units, accounts for dependence among observations from the same subject, and uses multiplier bands when at least five clusters contribute successes, at least five contribute failures, and the estimated variance is positive; other cells use cluster-Hoeffding bands. When both components occur, each receives half of the five-percent familywise error budget. The difference intervals use a delta-method approximation, appropriate when reach is bounded away from zero, with Gaussian multiplier calibration jointly over the six comparisons. The clustered intervals account for covariance between choices and reports from the same subjects; the row-i.i.d. benchmark treats these components as independent. Because these intervals concern equality rather than one-sided inequalities, the GMS–LF distinction does not apply.

cant after simultaneous calibration. Strikingly, the broader search-and-choice proxy—displayed or chosen—declines in 70 of the 76 comparisons, with a mean difference of -0.255; 36 declines are significant under subject-clustered inference, whereas no increase is significant. This widespread pattern provides independent descriptive evidence consistent with attention dilution. Because the proxy measures search and choice engagement rather than final structural consideration, we do not use it as a formal model test.

Four features of the design guide interpretation. First, price and travel-time profiles vary across tasks and cannot be held fixed by exact matching. The maintained menu-independent

Table 7: Preference-Share Confidence Bounds without Sequential Path Independence

	95% outer confidence bounds	
Preference event	Row-i.i.d.	Subject-clustered
$B \succ A$	[0.595, 1.000]	[0.521, 1.000]
$C \succ A$	[0.623, 1.000]	[0.578, 1.000]
$D \succ A$	[0.640, 1.000]	[0.653, 1.000]
$C \succ B$	[0.352, 1.000]	[0.310, 1.000]
$D \succ B$	[0.248, 1.000]	[0.198, 1.000]
$D \succ C$	[0.220, 1.000]	[0.157, 1.000]

Notes: Both columns use the same dominance-consistent sample of 37 subjects observed on the 11 menus with at least two inside alternatives. The entries are simultaneous 95% outer confidence bounds. They allow the menu-specific choice probabilities to vary jointly within their simultaneous confidence region and report the most conservative preference bounds consistent with the model without Sequential Path Independence. A strictly positive lower endpoint establishes the presence, but not the precise population prevalence, of the corresponding preference reversal. The row-i.i.d. benchmark treats task-level observations as independent. The subject-clustered bounds account for dependence among observations from the same subject and use conservative finite-sample bounds for sparse outcomes.

Table 8: Observed Information Search across Nested Menus

	Differences			Row-i.i.d.		Clustered	
Attention proxy	Mean	Negative	Positive	Sig. negative	Sig. positive	Sig. negative	Sig. positive
Displayed	0.009	33	43	0	0	0	0
Displayed or chosen	-0.255	70	6	38	0	36	0

Notes: The calculations use the 37-subject dominance-consistent sample. Each difference is the observed attention-proxy rate in the larger menu minus its rate in the smaller menu; negative differences therefore agree with attention dilution. Counts of significant negative and positive differences use simultaneous 95% two-sided multiplier intervals across all 76 nested-menu comparisons. "Displayed" records whether at least one attribute of the mode was opened; "Displayed or chosen" records whether the mode was displayed or selected as the final choice.

profile distribution incorporates this variation into the stable marginal distribution of travel-mode-category rankings used by H-LAO; homogeneous AOM provides the corresponding common-ranking benchmark. Variation in the underlying composition of observed categories can nevertheless weaken category-level random-utility implications (Liao, Saito, and Sandroni, 2026). Second, the complete menu domain is available for 21 subjects. The common-menu no-Sequential-Path-Independence analysis uses the broader dominance-consistent sample and therefore changes both menu support and subject composition; we interpret it as a sensitivity analysis rather than a direct precision comparison. Third, subjects rather than task-level rows are the independent sampling units. We therefore base our preferred inferential conclusions on subject-clustered procedures and report row-

i.i.d. results as benchmarks. Fourth, attribute displays measure costly pre-choice information search. We report displays separately and use displayed-or-chosen as a complementary descriptive measure of mode-level engagement.

The experiment's rare combination of full menu manipulation, a dominated outside option, repeated choices, observed information acquisition, and separately elicited rankings makes it unusually informative for comparing RAM, AOM, and H-LAO and for examining the preference and attention objects recovered by our methods. We use the application as an empirical illustration for stable travel-mode categories; because price and travel-time profiles vary across tasks, it is not a design-based test conditional on fixed profiles.

6 Conclusion

This paper introduces attention overload, a nonparametric restriction requiring an alternative's attention frequency to weakly decrease as its menu expands. Mixtures of menu-independent search priorities and inspection capacities provide a behavioral foundation for the restriction. Despite leaving the consideration-set distribution unrestricted, attention overload yields a sharp behavioral characterization, revealed-preference implications, and informative bounds on latent attention frequencies. The constructive characterization also supports econometric procedures that remain valid when the number of menus and inequalities is large, while its mixed-integer formulation represents the compatible-preference set implicitly and answers feasibility and pairwise-revelation queries without enumerating all rankings.

We also develop H-LAO, a list-based extension that accommodates heterogeneous preferences and attention. Outside-option choices reveal how far consumers proceed down a list. When the relevant menus are observed, Sequential Path Independence allows us to recover attention frequencies; with a positive probability of reaching the end of the list, additional menu variation recovers the choices that would prevail under full attention. Our identification analysis also characterizes what remains learnable when some menus are missing or consumers do not reach the end of a list. Under the stated support conditions, singleton and binary data can be enough to identify pairwise preference shares. The Supplemental Appendix develops confidence procedures for complete and

incomplete menu data, including cases in which reach is small or zero, and linear-programming methods that allow preferences and stopping behavior to be dependent or relax Sequential Path Independence.

Independent empirical studies establish the relevance of random limited attention and show that its restrictions are empirically testable. Attention overload turns the item-level dilution mechanism into an application-specific comparative static whose validity can be assessed directly from choice data. The Supplemental Appendix reports simulations evaluating the finite-sample behavior and computational demands of the proposed procedures. In the travel-mode application, subject-clustered diagnostics do not reject H-LAO, its pairwise preference shares closely match subjects' separately elicited rankings, and all simultaneous confidence intervals for the differences contain zero. The displayed-or-chosen search-and-choice proxy declines in 70 of 76 nested-menu comparisons, with 36 significant declines and no significant increases under subject-clustered simultaneous inference. Together, the homogeneous and heterogeneous models show how economically interpretable cross-menu attention restrictions can recover preference and attention information from standard choice data without committing to a parametric consideration-set model.

References

Abaluck, J., and A. Adams (2021): "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses," *Quarterly Journal of Economics*, 136(3), 1611–1663.

Abaluck, J., G. Compiani, and F. Zhang (2026): "A Method to Estimate Discrete Choice Models That Is Robust to Consumer Search," *Journal of Political Economy*, 134(7), 1967–2022.

Aguiar, V. H. (2017): "Random Categorization and Bounded Rationality," *Economics Letters*, 159, 46–52.

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

Aguiar, V. H., M. J. Boccardi, N. Kashaev, and J. Kim (2023): "Random utility and limited consideration," *Quantitative Economics*, 14(1), 71–116.

Andrews, D. W., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

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

Barseghyan, L., F. Molinari, and M. Thirkettle (2021): "Discrete Choice under Risk with Limited Consideration," *American Economic Review*, 111(6), 1972–2006.

Brady, R. L., and J. Rehbeck (2016): "Menu-Dependent Stochastic Feasibility," *Econometrica*, 84(3), 1203–1223.

Caplin, A., and M. Dean (2011): "Search, Choice, and Revealed Preference," *Theoretical Economics*, 6(1), 19–48.

Caplin, A., M. Dean, and D. Martin (2011): "Search and Satisficing," *American Economic Review*, 101(7), 2899–2922.

Cattaneo, M. D., Y. Feng, and W. G. Underwood (2024): "Uniform Inference for Kernel Density Estimators with Dyadic Data," *Journal of the American Statistical Association*, 119(548), 2695–2708.

Cattaneo, M. D., X. Ma, Y. Masatlioglu, and E. Suleymanov (2020): "A Random Attention Model," *Journal of Political Economy*, 128(7), 2796–2836.

Cattaneo, M. D., and R. R. Yu (2025): "Strong Approximations for Empirical Processes Indexed by Lipschitz Functions," *The Annals of Statistics*, 53(3), 1203–1229.

Chadd, I. (2023): "Random Network Consideration: Theory and Experiment," *Journal of Economic Behavior & Organization*, 211, 251–269.

Chernev, A., U. Böckenholt, and J. Goodman (2015): "Choice overload: A conceptual review and meta-analysis," *Journal of Consumer Psychology*, 25(2), 333–358.

Chernozhukov, V., D. Chetverikov, K. Kato, and Y. Koike (2022): "Improved Central Limit Theorem and Bootstrap Approximations in High Dimensions," *Annals of Statistics*, 50(5), 2562–2586.

Chernozhukov, V., D. Chetverikov, and Y. Koike (2023): "Nearly Optimal Central Limit Theorem and Bootstrap Approximations in High Dimensions," *The Annals of Applied Probability*, 33(3), 2374–2425.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

Dardanoni, V., P. Manzini, M. Mariotti, H. Petri, and C. J. Tyson (2023): "Mixture Choice Data: Revealing Preferences and Cognition," *Journal of Political Economy*, 131(3), 687–715.

Dardanoni, V., P. Manzini, M. Mariotti, and C. J. Tyson (2020): "Inferring Cognitive Heterogeneity From Aggregate Choices," *Econometrica*, 88(3), 1269–1296.

de Clippel, G., and K. Rozen (2024): "Bounded Rationality in Choice Theory: A Survey," *Journal of Economic Literature*, 62(3), 995–1039.

de Lara, L., and M. Dean (2024): "Rational Choice Overload," Working paper.

Dean, M., D. Ravindran, and J. Stoye (2025): "A Better Test of Choice Overload," Working paper.

Demirkan, Y., and M. Kimya (2020): "Hazard Rate, Stochastic Choice and Consideration Sets," *Journal of Mathematical Economics*, 87, 142–150.

Ellis, K., E. Filiz-Ozbay, and E. Y. Ozbay (2025): "Consistency and Heterogeneity in Consideration and Choice," Working paper.

Falmagne, J.-C. (1978): "A Representation Theorem for Finite Random Scale Systems," *Journal of Mathematical Psychology*, 18(1), 52–72.

Fishburn, P. C. (1998): "Stochastic Utility," in *Handbook of Utility Theory*, ed. by S. Barbera, P. Hammond, and C. Seidl, pp. 273–318. Kluwer Dordrecht.

Geng, S. (2016): "Decision Time, Consideration Time, and Status Quo Bias," *Economic Inquiry*, 54(1), 433–449.

Guney, B. (2014): "A theory of iterative choice in lists," *Journal of Mathematical Economics*, 53, 26–32.

Horan, S. (2010): "Sequential search and choice from lists," *Working Paper*.

Ishii, Y., M. Kovach, and L. Ülkü (2021): "A model of stochastic choice from lists," *Journal of Mathematical Economics*, 96, 102509.

Iyengar, S. S., and M. R. Lepper (2000): "When choice is demotivating: Can one desire too much of a good thing?," *Journal of Personality and Social Psychology*, 79(6), 995–1006.

Kashaev, N., and V. H. Aguiar (2022): "A random attention and utility model," *Journal of Economic Theory*, 204, 105487.

Khan, I., I. Krajbich, C. Raymond, S. Sundaresan, and A. Veiga (2026): "Measuring Consideration: The Attentional Origins of Stochastic Choice," Working paper.

Koshevoy, G., and E. Savaglio (2023): "On rational choice from lists of sets," *Journal of Mathematical Economics*, 109, 102891.

Kovach, M., and L. Ülkü (2020): "Satisficing with a variable threshold," *Journal of Mathematical Economics*, 87, 67–76.

Liao, Y., K. Saito, and A. Sandroni (2026): "Random Utility with Aggregation," Working paper, arXiv:2506.00372, revised 2026.

Lleras, J. S., Y. Masatlioglu, D. Nakajima, and E. Y. Ozbay (2017): "When More Is Less: Limited Consideration," *Journal of Economic Theory*, 170, 70–85.

Lopes, M. E. (2022): "Central Limit Theorem and Bootstrap Approximation in High Dimensions: Near $1/\sqrt{n}$ Rates via Implicit Smoothing," *The Annals of Statistics*, 50(5), 2492–2513.

Manzini, P., and M. Mariotti (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82(3), 1153–1176.

Manzini, P., M. Mariotti, and L. Ülkü (2024): "A model of approval with an application to list design," *Journal of Economic Theory*, 217, 105821.

Marschak, J. (1959): "Binary choice constraints and random utility indicators," in *Theory and Decision Library*, p. 218–239. Springer.

Masatlioglu, Y., and T. P. Vu (2024): "Growing Attention," *Journal of Economic Theory*, 222, 105929.

Matzkin, R. L. (2007): "Nonparametric Identification," in *Handbook of Econometrics, Volume VIB*, ed. by J. Heckman, and E. Leamer, pp. 5307–5368. Elsevier Science B.V.

——— (2013): "Nonparametric Identification in Structural Economic Models," *Annual Review of Economics*, 5, 457–486.

Molinari, F. (2020): "Microeconometrics with Partial Identification," in *Handbook of Econometrics, Volume VIIA*, ed. by S. Durlauf, L. Hansen, J. Heckman, and R. Matzkin, pp. 355–486. Elsevier Science B.V.

Natan, O. R. (2024): "Choice Frictions in Large Assortments," *Marketing Science*, 44(3), 593–625.

Reutskaja, E., and R. M. Hogarth (2009): "Satisfaction in Choice as a Function of the Number of Alternatives: When "Goods Satiate"," *Psychology and Marketing*, 26(3), 197–203.

Reutskaja, E., R. Nagel, C. F. Camerer, and A. Rangel (2011): "Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study," *American Economic Review*, 101(2), 900–926.

Rubinstein, A., and Y. Salant (2006): "A model of choice from lists," *Theoretical Economics*, 1(1), 3–17.

Strzalecki, T. (2025): *Stochastic Choice Theory*, Econometric Society Monographs. Cambridge University Press.

Thomas, A. W., F. Molter, and I. Krajbich (2021): "Uncovering the Computational Mechanisms Underlying Many-Alternative Choice," *eLife*, 10, e57012.

Tserenjigmid, G. (2021): "The order-dependent Luce model," *Management Science*, 67(11), 6915–6933.

Turansick, C. (2022): "Identification in the Random Utility Model," *Journal of Economic Theory*, p. 105489.

Tversky, A. (1972): "Elimination by aspects: A theory of choice," *Psychological review*, 79(4), 281.

Visschers, V. H., R. Hess, and M. Siegrist (2010): "Health Motivation and Product Design Determine Consumers Visual Attention to Nutrition Information on Food Products," *Public Health Nutrition*, 13(7), 1099–1106.

Wang, Z., J. Sicsic, and P. Chauvin (2026): "From Random Utility to Random Attention: Validating the Random Attention Model in Health Preference Analysis," Working paper.

Wang, Z., and D. Zhu (2025): "Identification of Daily Travel Mode Preferences under Random Attention Allocation: Evidence from a Lab Experiment and Paris," Working paper.

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

Yildiz, K. (2016): "List-rationalizable choice," *Theoretical Economics*, 11(2), 587–599.