Choice and Attention across Time

Xi Zhi Lim*

This version: 2024/03/02
Most recent and public: http://s.xzlim.com/sequence

Abstract

I study how past and future choices are linked in the framework of attention. Attention cannot be observed but past choices are necessarily considered in future decisions. This link connects two types of rationality violations, counterfactual and realized, where the former results from inattention and the latter fully pins down preferences. Results show that the necessary traces of limited attention lie within choice sequences because they enable and compel a decision maker to correct their "mistakes". The framework accommodates different attention structures and extends to framing, introducing choice sequences as an important channel to formulate, identify, and scrutinize limited attention.

Keywords: Choice sequences, limited attention, limited consideration, framing effects
JEL: D01, D11

*Shanghai Jiao Tong University (email: xzlim@sjtu.edu.cn). I am grateful to co-advisors Mark Dean and Pietro Ortoleva for guidance and to committee members Hassan Afrouzi, Navin Kartik, and Qingmin Liu for feedback. I thank Kyle Chauvin, Xiaoyu Cheng, Paul Cheung, Soo Hong Chew, David Dillenberger, Evan Friedman, Shaowei Ke, Matthew Kovach, Jian Li, Senran Lin, Shuo Liu, Yusufcan Masatlioglu, Bin Miao, Daisuke Nakajima, Norio Takeoka, Rui Tang, Fan Wang, Chen Zhao, seminar participants at AMES 2022, CMES 2021, Columbia, MEET 2022, NYU Shanghai, Peking University, RUD 2023, SHUFE, SWUFE (CCBEF), and VEAEBES for feedback. This work appears in the 3rd chapter of my PhD dissertation at Columbia University (2020).

1 Introduction

The generalization of a standard theory to explain “non-standard” behavior benefits from richer data; intuitively, data can compensate for the added flexibility. Observing reference points, endowments, or status quo reconciles behaviors that contradict a single, consistent preference.[1] Observing menu preferences helps test the hypothesis of temptation and enables studies of self-control and addiction.[2] Observing compound lottery and multi-dimensional risk brings new insights to traditional risk preference anomalies like the Allais paradox.[3] These datasets are appreciated because they are useful and innovative, even if they change the ways data have to be collected. Can a dataset that is useful but *ubiquitous* also receive consideration? Stochastic choice is one example,[4] choice sequences could be another.

The growing literature on limited consideration has thus far *zoomed in* on a choice problem to study an attention mechanism, investigating the possibility of a search process that considers a subset of alternatives or a rule of thumb that eliminates alternatives from final decisions. This paper *zooms out* and studies how the evolution of choices can hint at the role of attention even if no assumptions are imposed on attention structures.

Limited consideration occurs when a decision maker (DM) fails to consider every alternative in every choice set. It results in seemingly irrational decisions even if the DM has a standard and consistent preference, thereby capturing a straightforward form of bounded rationality that has received substantial attention.[5] But its simplicity and intuitiveness is not without cost. Attention being inherently hard to observe can leave us with multiplicity of estimated preferences, which burdens our

[1]For example, Kahneman and Tversky (1979); Munro and Sugden (2003); Sugden (2003); Masatlioglu and Ok (2005); Ortoleva (2010); Masatlioglu and Nakajima (2013); Masatlioglu and Ok (2014); Dean et al. (2017); Kovach (2020); Ellis and Masatlioglu (2022).

[2]For example, Gul and Pesendorfer (2001, 2007); Dekel et al. (2009); Noor (2011); Dillenberger and Sadowski (2012); Ahn et al. (2019); Freeman (2021).

[3]For example, Segal (1990); Dillenberger (2010); Lanzani (2022); Chew et al. (2022); Halevy and Ozdenoren (2022); Zhang (2023); Ke and Zhang (2023).

[4]For example, Gul and Pesendorfer (2006); Manzini and Mariotti (2014); Gul et al. (2014); Fudenberg et al. (2015); Brady and Rehbeck (2016); Aguiar (2017); Echenique et al. (2018); Cattaneo et al. (2020); Kovach and Tserenjigmid (2022); Kovach and Suleymanov (2023); Kibris et al. (2024).

[5]Earliest work traceable to Wright and Barbour (1977)'s discussion of consideration set, related applications in marketing and finance that include Hauser and Wernerfelt (1990); Roberts and Lattin (1991), and related choice theories that include Manzini and Mariotti (2007); Masatlioglu et al. (2012); Cherepanov et al. (2013).

analysis of economic consequences and welfare. Moreover, the model specification that x receives attention in choice problem A but not in choice problem B is almost impossible to test with a within-subject design; once the DM experiences A and pays attention to x, would she immediately forget x when asked to choose from B?

This paper addresses these issues by linking past and future choices, allowing an analyst to exploit the wealth of information contained in the natural evolution of choices. The innovation lies in the primitive—a dataset of choice sequences, which departs from the standard "one-shot" setting where the DM makes one real choice.

This is accompanied by an intuitive assumption: past choices should be automatically considered when the DM makes future decisions. Then, decisions from the same choice set that vary with experience hint at a mechanism of evolving attention. Moreover, because choices must become more informative, a latter choice that contradicts an earlier choice reveals true preference. This contributes to necessary pleasantries: either behavior is standard, or its "non-standard" manifestation reveals unobserved parameters. It turns out that this simple assumption, captured using an axiomatic foundation, allows us to test the hypothesis of limited consideration using realized choices, confirm "mistakes", and fully identify preferences.

To illustrate, suppose you are unaware of the vacation destination Penang (an island in my country Malaysia) even though you can afford it, so an analyst who observes your choice of Hawaii may falsely jump to the conclusion that you prefer Hawaii over Penang by the theories of revealed preferences. However, when you attend a conference in Southeast Asia, you might consider and choose Penang for a drop-by vacation. This incident makes Penang then and forever an option you are aware of, and your future decisions of whether to return to Penang will more informatively convey your true preference between Penang and other destinations.

The underlying intuition applies broadly: A DM who uses the iPad may or may not have considered a Surface Go, but a DM who converted to an iPad from a Surface Go probably prefers iPad to Surface Go. A person who reads physical books may actually prefer e-readers, but one who left e-readers for physical books probably prefers physical books. A colleague who has not begun to referee papers could be a remarkable reviewer, but one who used to do so yet is no longer invited may not be the best reviewer.

To learn from this intuition, consider a DM with an *Attention Across Time* (AAT) representation. The DM has a subjective attention function that determines what she would normally consider from each choice set A, denoted by $\Gamma(A)$. Moreover, the DM's past experiences, h, identifies a collection of previously chosen alternatives that will continue to be considered, denoted by $\boldsymbol{c}(h)$. The DM's decision therefore solves

$$\tilde{c}(h)(A) = \max_{x \in \Gamma(A) \cup (\boldsymbol{c}(h) \cap A)} u(x).$$

Of course, past experiences h are historical choice problems, so elements of $\boldsymbol{c}(h)$ come from the same procedure, just at an earlier time. This contrasts the "standard" DM who has full attention $\Gamma(A) = A$, for whom the problem reduces to the familiar

$$\tilde{c}(h)(A) = \max_{x \in A} u(x).$$

The first inquiry helps us understand the behavioral content of this model; three key axioms that relate past and future choices underpin AAT behavior. *Weak Stability* (Axiom 1) imposes restriction on behavior over time: in contrast to full compliance with WARP, it allows for one-time switches between every pair of alternatives. *Past Dependence* (Axiom 2) limits the way past choices may affect future choices; specifically, if a recent experience affects the choice from the current choice set, then the new choice is limited to the recently chosen alternative. The third and last axiom captures the behavioral signature of attention. First, revealed preference is defined when observed choices suggest that the DM is aware of y when x is chosen. Specifically, when x is chosen over y after y was previously chosen, or when x is chosen from a choice set where y is chosen "by default". *Default Attention* (Axiom 3) posits that if x is revealed preferred to y, then y will never be chosen from a choice set where x is chosen by default. The axiom essentially says that the default choice from a choice set should receive attention in that choice set no matter the history, and therefore a subjectively inferior alternative should never be chosen. All three axioms are trivially satisfied in the conventional setting with only one period of choice.

Then, a series of straightforward observations forms the core of this paper: exploring what economists can learn from the wealth of information in choice sequences.

The first observations concern the identification of preferences. Preferences are pinned down. To illustrate the intuition, suppose a WARP violation occurs between the choice of x from $\{x, y\}$ and the choice of y from $\{x, y, z\}$. It turns out that if we present the "problematic" choice sets in an alternating order, then the DM is bound to consider both x and y in the second problem, resulting in a choice that reveals their preference.[6] This key empirical strategy leverages the fact that an experienced DM has considered more options and therefore their choices more informatively convey genuine preference. Moreover, it is possible to identify preferences using just one carefully designed sequence of choice sets. On the other hand, attention Γ cannot be pinned down even with the richer dataset of choice sequences, but the set of possible attention functions can be characterized using a maximal set that allows an analyst to overestimate or underestimate attention.

The second observations study the link between limited attention and rationality violations. To see its relevance, note that limited attention *can* result in "non-standard" behavior, but it is not immediately clear whether it *must*. An investigation into this link begins with a distinction between two kinds of rationality violations: those that occur on a *counterfactual* basis and those that will *realize*.[7] Counterfactual violations are typically observed in a between-subject design, where a population of subjects makes inconsistent decisions from randomly assigned choice problems; it means some subjects are predisposed to commit these violations. It turns out that the lack of counterfactual violations cannot rule out limited attention. A simple example illustrates a DM who never commits counterfactual violations but is inconsistent with full attention due to history-dependent behavior.

Unlike counterfactual violations, realized violations come from the continuous observation of one DM—when they are seen choosing x over y at some point and y over x in others. It turns out that full attention is ruled out if and only if the DM commits realized WARP violations. If no violations of this kind can be detected, then the DM is observationally identical to one with full attention. These observations suggest that realized violation provides the true test for limited attention even though it is inexorably obscured in studies that focus on one-shot decisions.

But where does the attention function Γ come from? Because AAT only imposes structure on attention across time, its silence on attention structures invites

[6] If genuine preference is $u(x) > u(y)$, then the sequence of choice sets $(\{x, y\}, \{x, y, z\})$ produces choices (x, x) and the sequence of choice sets $(\{x, y, z\}, \{x, y\})$ produces choices (y, x), both cases reveal that x is preferred to y. The opposite holds for $u(y) > u(x)$. Subsection 3.2 provides details and an illustration using Figure 3.1. Online Appendix B Example 12 describes a test using a population of DMs.

[7] Subsection 3.3 provides a formal definition and an illustration using Figure 2.1.

a complementary relationship with models that propose structures for Γ. Masatlioglu et al. (2012) propose a theory about the intrinsic nature of consideration sets, that the removal unconsidered alternatives should not affect what is considered. It can be shown that if the attention function in AAT satisfies this structure, then future attention, even though it is evolving as it adapts to past experiences, will continue to satisfy this structure. The same can be said for the models proposed by Manzini and Mariotti (2007) and Manzini and Mariotti (2012) where criteria like shortlisting and categorization are used to exclude alternatives from final consideration—future attention stems from revised shortlists and re-categorizations. As a consequence, even though accumulating experience improves decisions, the DMs do not fundamentally depart from their intrinsic attention structures.

A complete characterization of compatible models suggests that the ultimate content of AAT is the correction of WARP violations. One-shot models that have this feature are compatible with AAT, and it allows us to connect *attention across time* and *attention across choice sets*, providing a robust framework of limited consideration and proposing new ways to test and verify inattention.

Last, the framework is extended to incorporate framing, which enables an analysis of how different frames can affect a DM's current behavior and future attention. It begins with a general representation that captures framing in full generality. Different frames can draw the DM's attention to different alternatives even though the choice set is fixed. Successful frames induce lasting consideration for the future. Special cases of framing are then introduced and characterized, namely *ordered lists* where the DM considers alternatives from top to bottom but may stop at some point and *recommendations* where certain alternatives are made salient. In both cases, postulates are imposed on choice sequences, proposing new empirical directions to test whether and how framing works. The model formalizes a number of intuitive observations, such as the futile repetition of unsuccessful frames and the crucial role that genuine preference (or quality) plays in the facilitation of lasting consideration.

The findings of this paper are undoubtedly limited—there is more to learn from choice sequences—but they underscore a broader agenda of using richer data, in place of assumptions, to learn *from* individual behavior. If we believe that behavior is boundedly rational, then choice sequences emerge as an important dataset that allows an analyst to observe and study the correction of "mistakes". *Whether* corrections occur, *when* they occur, and *how* they occur each contributes significantly to a comprehensive examination of bounded rationality. Limited consideration is only one of many possible examples.

I proceed as follows: Related literature comes next. Section 2 introduces the primitive and the axiomatic foundation. Section 3 introduces AAT and basic results regarding identification and rationality violations. Section 4 analyzes different attention structures in the literature to investigate the link between attention across time and attention across choice sets. Section 5 extends the framework to incorporate frames and studies the short- and long-term effects of framing on choices and attention. Section 6 concludes. Appendix A contains main proofs. Online Appendix B contains omitted proofs and results.

1.1 Related Literature

Closest to AAT is the choice theory literature that studies attention using consideration sets, even though this literature has not considered the evolution of choices. Masatlioglu et al. (2012)'s *choice with limited attention* belongs to this category, Lleras et al. (2017) study consideration sets that preserve considered alternatives in subsets, and Geng (2022) introduces triggers and capacity. Other heuristics and choice procedures can also give rise to consideration sets, including Manzini and Mariotti (2007)'s *rational shortlist method*, Manzini and Mariotti (2012)'s *categorize-then-choose*, and Cherepanov et al. (2013)'s *rationalization*, Rideout (2021)'s *justification* and Geng and Özbay (2021)'s *shortlisting procedure with capacity*, where potentially superior alternatives are eliminated from final decisions.[8] These studies complement AAT's silence on attention structures; in turn, AAT suggests how choice sequences can serve as a new channel to study inattention.

Different methods have been deployed to infer preference under limited attention. Masatlioglu et al. (2012)'s CLA already has this feature, inferring attention partially identifies preference. Caplin and Dean (2011) and Caplin et al. (2011) study search processes that take place by observing tentative choices in a single choice problem. Manzini and Mariotti (2014) and Cattaneo et al. (2020) exploit the richness in stochastic choice to pin down preference; Kovach and Suleymanov

[8]Similar procedures that study the iterative scrutiny of alternatives can be found in Xu and Zhou (2007); Apesteguia and Ballester (2013); Yildiz (2016)

(2023) additionally consider reference-dependent distributions of consideration sets. Gossner et al. (2021) study how behavior may react to the exogenous manipulation of attention.

The general intuition that past behavior can influence future behavior is shared elsewhere. In static settings, Gilboa and Schmeidler (1995); Gilboa et al. (2002) study *case-based utility* where a case can be past experiences, Bordalo et al. (2020) study how a database of past experiences can affect the evaluation of alternatives through different perceived norms. Models of *habit formation* consider DMs who are affected by the past and attempt to shape future behavior (Gul and Pesendorfer, 2007; Rozen, 2010; Tserenjigmid, 2020; Hayashi and Takeoka, 2022).

The extension to frames relates to a general setting studied in Salant and Rubinstein (2008)'s *choice with frames*. Guney (2014) studies deterministic behavior under observable lists whereas Manzini et al. (2021); Tserenjigmid (2021); Ishii et al. (2021) study stochastic behavior/data. Cheung and Masatlioglu (2024) study observable recommendation and use stochastic choice data to reveal the influence of recommendation on both attention and utility. This paper adds to the literature new ways to test, analyze, and use frames—using choice sequences.

2 Setup

2.1 Primitive

Let X be a countable set of alternatives and let $\mathcal{A}$ be the set of all finite subsets of X with at least two elements.[9] Let X^{N} be the set of all infinite sequences of alternatives and let $\mathcal{A}^{\mathrm{N}}$ be the set of all infinite sequences of choice sets. The primitive is a choice function that assigns to each infinite sequence of choice sets an infinite sequence of choices,

$$c : \mathcal{A}^{\mathrm{N}} \to X^{\mathrm{N}},$$

where for every sequence of choice sets $(A_n) \in \mathcal{A}^{\mathrm{N}}$ and any natural number k, the corresponding choice $c\left((A_n)\right)_k$ is an element of the corresponding choice

[9]The exclusion of singleton choice sets is to avoid passive choices, since "choosing" something without a choice might not result in awareness/consideration. All results go through if we let $\mathcal{A}$ be the set of all finite and nonempty subsets of X.

set $(A_n)_k$.[10] The first choice (when $k = 1$) can be treated as the first decision since the DM enters the analyst's observation. I sometimes write A_k for $(A_n)_k$. Denote by $\hat{X}$ the set of alternatives that are *ever-chosen*, i.e., $\hat{X} := \{x \in X : c((A_n))_i = x$ for some (A_n) and $i\}$.[11] Like more standard primitives, this dataset can be extracted from a population of DMs each randomly assigned to a sequence.[12]

To focus on how past choices affect future choices (instead of the opposite), the scope of this paper is limited to situations in which two sequences of choice sets that are identical up to a certain point give the same choices up to that point. This property, henceforth *future independence*, rules out choices that are made with perfect foresight of future choice sets. Formally, for any $(A_n), (B_n) \in \mathcal{A}^{\mathrm{N}}$, if $A_k = B_k$ for all $k \leq K$, then $c((A_n))_k = c((B_n))_k$ for all $k \leq K$.

Then, c can be decomposed into an object more familiar to most economists: a collection of history-dependent choice functions. Formally, let $\mathcal{A}^{<\mathrm{N}}$ be the set of all finite sequences of choice sets, including the empty sequence denoted by $\emptyset$. For each history of choice sets $h \in \mathcal{A}^{<\mathrm{N}}$, denote by

$$\tilde{c}(h) : \mathcal{A} \to X$$

the *one-shot choice function* that assigns a choice to each upcoming choice set (right after h). Each $\tilde{c}(h)$ captures a cross-section of the original primitive c, illustrated in Figure 2.1. Due to future independence, $\left\{\tilde{c}(h) \,|\, h \in \mathcal{A}^{<\mathrm{N}}\right\}$ is fully and uniquely pinned down. Choice without (observable) history is captured by $\tilde{c}(\emptyset)$, and I refer to it as c_0 for convenience.

[10] Even though I work with infinitely long choice sequences, the main representation theorem only requires choice sequences to have at least length 4.

[11] There is at most one *never-chosen* alternative, since the set of all binary choice sets leaves at most one alternative never chosen.

[12] The concern that a DM can only be observed under one sequence of choice sets does not prohibit us from discussing what the DM would have chosen from other choice sequences. The same limitation applies to standard settings where we consider a DM who produces a choice for each of multiple choice sets. The solution is to use a between-subject design with random assignments. If a population of subjects is each randomly assigned to choice set A or choice set B, and aggregate data presents a WARP violation, then we know that some subjects are predisposed to commit WARP violations. Online Appendix B Example 12 describes this approach for choice sequences.

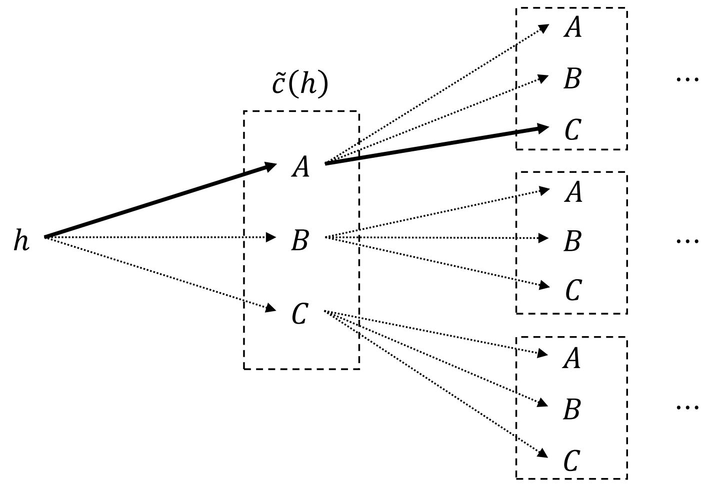

Figure 2.1: Each rectangle is made up of one-shot choice sets; therefore, a sequence of choice sets includes at most one choice set from each rectangle. The solid path represents (part of) a sequence of choice sets.

2.2 Axioms

Aligned with the motivation of this paper, all three axioms impose restrictions on choices across time; they are trivially satisfied if choice sequences were not considered. The first two axioms provide basic structure.

Axiom 1 (Weak Stability)**.** *For any $(A_n) \in \mathcal{A}^{\mathrm{N}}$ and $h < i < j$, if $c((A_n))_h = x$, $c((A_n))_i = y \neq x$, $x \in A_i$, and $y \in A_j$, then $c((A_n))_j \neq x$.*

Axiom 1 imposes a version of the infamous weak axiom of revealed preferences (WARP) with two key differences.[13] First, it is imposed *within* a sequence of choice problems $(A_n) \in \mathcal{A}^{\mathrm{N}}$. In particular, there is no restriction on how choices differ across sequences.[14] Second, it does so without demanding full compliance with WARP, but limits the instances of WARP violations. A conforming DM may switch between x and y, but she does not go back and forth between them.

To illustrate, suppose a DM first chooses x in the presence of y and then chooses y in the presence of x. The latter choice violates WARP, and it may be due to the

[13]There are many (roughly) equivalent definitions for WARP, I use "if x is chosen in y's presence, then y is never chosen in x's presence".

[14]Let $A = \{x, y, a\}$ and $B = \{x, y, b\}$. Suppose sequence one $(A, B, B, B, ...)$ produces $(x, b, b, b, ...)$ and sequence two $(B, A, A, A, ...)$ produces $(y, a, a, a...)$, then choosing x over y in A but the opposite in B is a WARP violation across sequence, but there is no violation of Axiom 1.

emerging consideration of y. Axiom 1 does not exclude this behavior, but it posits that from here on, the choice between x and y has finalized. In other words, the DM may "flip" but must not "flip-flop". A DM who never switches automatically satisfies this axiom.

Axiom 2 (Past Dependence)**.** *For any* $(A_1, ..., A_K) \in \mathcal{A}^{<N}$ *and* $B \in \mathcal{A}$*,*

$$\tilde{c}\left(\left(A_1, ..., A_K\right)\right)(B) \in \left\{\tilde{c}\left(\left(A_1, ..., A_{K-1}\right)\right)(B), \tilde{c}\left(\left(A_1, ..., A_{K-1}\right)\right)\left(A_K\right)\right\}.$$

Axiom 2 allows past choices to affect future choices. One way to understand this axiom is to first consider the removal of "$\tilde{c}\left(\left(A_1, ..., A_{K-1}\right)\right)\left(A_K\right)$". In that case, the DM's choice after history $(A_1, ..., A_K)$ does not depend on whether or not she had experienced A_K, i.e., choices are past *independent*. Axiom 2 weakens past independence by allowing for one type of departure: the next choice is exactly the choice it succeeded, $\tilde{c}\left(\left(A_1, ..., A_{K-1}\right)\right)\left(A_K\right)$. That is, after facing a history of choice sets $(A_1, ..., A_K)$, what a DM chooses from B is either what she would have chosen had she not experienced A_K or exactly what she just chose from A_K. The postulate is therefore a delimited weakening of past independence. First, even though past choices may affect future choices, it must do so in a period-by-period manner; this provides tractability and important testable predictions. Moreover, said effect is limited to "helping" the recently chosen alternative to be chosen again; other forms of past dependence remain prohibited.

The next and last axiom embodies the behavioral signature of attention. Recall that $c_0(A)$ denotes the choice from A without (observable) history. Consider a definition of revealed preference that captures the analyst's inference that x is better than y. This relationship is identified either when x is chosen over y when y was chosen in the past or when x is chosen from a choice set that y is chosen when there is no observable history. Formally, for distinct x and y, let xPy if at least one of the following is true for some $(A_n) \in \mathcal{A}^{\mathbb{N}}$: (1) $c\left(\left(A_n\right)\right)_j = x$, $y \in A_j$ and $c\left(\left(A_n\right)\right)_i = y$ such that $i < j$ or (2) $c\left(\left(A_n\right)\right)_j = x$ such that $c_0\left(A_j\right) = y$.

Axiom 3 (Default Attention)**.** *If* $c_0(A)\,Py$*, then* $y \notin \tilde{c}(h)(A)$ *for all* $h \in \mathcal{A}^{<N}$*.*

Axiom 3 restricts how a DM can depart from her default choice in a choice set. Specifically, it posits that if the default choice from A, denoted by $c_0(A)$, is identified to be better than y, then y is never going to be chosen from A (no matter the history). Intuitively, it captures the idea that certain alternatives will always receive attention when they appear in certain choice sets (regardless of the history), hence "default attention". These may be the most salient alternatives insofar as to attract attention: pizza is always in the consideration set for football night, even though the additional consideration of satay (Malaysian skewers) depends on whether the DM has learned of this dish. In general, since DM was not born into the analyst's observation, default attention may be interpreted as the attention structure formed in the (unobserved) past.

3 Model

3.1 Attention Across Time

We are ready for the main representation theorem. Denote by $\boldsymbol{c}\left(h\right)$ the set of (previously) chosen alternatives in the history h, i.e., $\boldsymbol{c}\left(\left(A_1, ..., A_K\right)\right) := \left\{\tilde{c}\left(\emptyset\right)\left(A_1\right)\right\} \cup \left\{\tilde{c}\left(\left(A_1\right)\right)\left(A_2\right)\right\} \cup \left\{\tilde{c}\left(\left(A_1, ..., A_{k-1}\right)\right)\left(A_k\right) : k = 3, ..., K\right\}$.

Definition 1. c admits an Attention Across Time (AAT) representation if there exist a *utility function* $u : X \to \mathbb{R}$ and an *attention function* $\Gamma : \mathcal{A} \to 2^X \setminus \{\emptyset\}$, where $\Gamma\left(A\right) \subseteq A$, such that

$$\tilde{c}\left(h\right)\left(A\right) = \underset{x \in \tilde{\Gamma}\left(h\right)\left(A\right)}{\arg\max}\, u\left(x\right)$$

where $\tilde{\Gamma}\left(h\right)\left(A\right) = \Gamma\left(A\right) \cup \left(\boldsymbol{c}\left(h\right) \cap A\right)$.

Theorem 1. *c satisfies Axioms 1, 2, and 3 if and only if it admits an Attention Across Time (AAT) representation.*

AAT suggests the following choice procedure: When a DM faces choice set B, she not only considers alternatives that she would always consider when she faces B but also the alternatives that she had chosen in the past. The former is *history-independent* and may capture what is salient (to her) in the underlying choice problem. The latter is *history-dependent* and receives her attention due to her past experiences. The intuition is straightforward—a DM may be unaware of certain alternatives, but she must be aware of the alternatives that she had chosen, which depend on her past experiences.

Attention structures The generality of Γ puts no restriction on choice behavior without history—every c_0 is consistent with every utility function under some Γ—but AAT imposes on what happens *next*.[15] For example, consider Masatlioglu et al. (2012)'s use of attention filter to explain why, although x is preferred to y, it is chosen from $\{x, y\}$ but not from $\{x, y, z\}$. A possible specification is $\Gamma(\{x, y\}) = \{x, y\}$, $\Gamma(\{x, y, z\}) = \{y, z\}$. AAT accommodates this specification, but it further imposes that if x is chosen from $\{x, y\}$ in the past, then x must be considered when the DM faces $\{x, y, z\}$ in the future, i.e., $x \notin \Gamma(\{x, y, z\})$ but $x \in \tilde{\Gamma}(\{x, y\})(\{x, y, z\})$. Section 4 provides a comprehensive examination of "attention across time" as related to "attention across choice sets".

Inferring preference An analyst knows definitively that the DM prefers a to b if she chose b in the past and chooses a over b in the future, since both a and b are within the consideration set of the latter choice problem. In fact, one way to elicit such preference is to first introduce a choice set under which the DM would choose b, and then ask the DM to choose from a choice set from which she would normally choose a. This simple observation helps pin down preferences even if attention cannot be directly observed. Subsection 3.2 investigates further and proposes empirical strategies to reveal preference.

Variety of experience Does having more experience help the DM make better decisions? The model adds details to this intuition. Better decisions ultimately come from an expanding $\boldsymbol{c}(h)$, but facing the same problem repeatedly does not contribute to the expansion of $\boldsymbol{c}(h)$, even though h becomes longer. Instead, the variety of past experiences is key to expanding $\boldsymbol{c}(h)$. To see this, note that if two identical choice sets are faced consecutively, the choice from the latter must coincide with the choice from the former, because the consideration set has not expanded. This means the second decision adds nothing to $\boldsymbol{c}(h)$. Similar arguments can be made if a choice set is repeated but not consecutively. Curiously, increased experience can result in better decisions that appear "irrational", Subsection 3.3 provides details.

History-dependent decisions Different alternatives can be chosen from the same choice set at different points in history, even if the DM is never indifferent and has

[15] For any $c_0 : \mathcal{A} \to X$, let $\Gamma(A) = \{c_0(A)\}$ for all A and use any utility function.

no noise in their decisions. Although this qualifies as a (trivial) WARP violation, their occurrence counter-intuitively suggests that the DM *used to* be "irrational," because she chose a sub-optimal option, but has since become more rational through the consideration of better options. Subsection 3.4 delves deeper into the significance of history-dependent behaviors.

Relevance of experience Certain types of past experiences are more useful than others, and this depends on what the future entails. To see this, suppose x is added to $\boldsymbol{c}(h)$ but there is no future choice set that contains x, then the increased awareness of x will not improve future decisions. The same limitation holds if future choice sets that contain x also contain superior alternatives that are themselves considered, since x will not be chosen anyway. Therefore, the relevance of experience holds greater long-term value than sheer quantity of experience. Section 5 investigates the effects of framing on future attention; it cautions against using frames to "help" a DM because short-term benefits can result in harmful long-term inattention.

Other theories Choice sequences can invite interest in other theories excluded in AAT. One possibility is learning, where a DM initially unsure of her preference discovers that she doesn't like something after consuming it, causing her to change her behavior in the future. The DM can violate Axiom 1 by switching back and forth. To see this, consider a DM in her Penang vacation where she first tries *petai*, realizes that it is not as good, and then tries *tempoyak* only to find out that it is worse, sticking to *petai* for the rest of her trip. Other DMs may be building a bundle to seek variety, or deliberately randomizing, or simply being stochastic; these behaviors violate Axiom 2 because the axiom requires the same alternative to be chosen when a choice set is repeated consecutively.[16]

3.2 Identification of Parameters

Next, a series of observations assert that standard economics problems that concern welfare and incentives, which rely on the identification of preferences, are possible to study even when the analyst cannot directly observe attention.

[16]Deliberately randomize: Agranov and Ortoleva (2017); Cerreia-Vioglio et al. (2019).

Uniqueness of preference

The first observation is uniqueness of preferences. In AAT, preference is unique for ever-chosen alternatives $\hat{X}$. On the contrary, analyzing past and future choices in isolation often results in a dilemma trying to conclude whether choices are due to genuine preferences or due to the lack of attention.

Proposition 1. *If c admits AAT representations $(u_1, \Gamma_1), (u_2, \Gamma_2)$ and $x, y \in \hat{X}$, then $u_1(x) > u_1(y)$ if and only if $u_2(x) > u_2(y)$.*

Convergence

To delve deeper, I address the question of *how* to identify preferences, beginning with an intuitive and useful testable prediction.

Consider the following test that compels the correction of "mistakes" and reveals preference. Suppose a DM is predisposed to commit a WARP violation by choosing x from $\{x, y\}$ but y from $\{x, y, z\}$. They correspond to the top two circles in Figure 3.1, implying that x receives consideration in $\{x, y\}$ and y receives consideration in $\{x, y, z\}$, which is not enough to determine whether x or y is preferred. However, if a second choice is elicited using alternating choice sets, then both x and y will receive consideration, one due to default attention inferred from first period behavior and another due to a recent experience. The second period choice thus reveals preference.

Moreover, since they maximize the same preference, second period's choices must coincide, either converging on x or converging on y. This *convergence* property is generalized to arbitrary choice sets and history in Proposition 2; it is an important testable prediction of AAT and the key indicator that the DM has a stable preference despite the influence of limited consideration. Online Appendix B Example 12 extends this test to a population of DMs with unknown preference parameters.

Proposition 2. *Suppose c admits an AAT representation, $\tilde{c}(h)(A) = x$, $\tilde{c}(h)(B) = y$, $\{x, y\} \subseteq A \cap B$, and $x \neq y$. Either $\tilde{c}((h, A))(B) = \tilde{c}((h, B))(A) = x$, which implies $u(x) > u(y)$, or $\tilde{c}((h, A))(B) = \tilde{c}((h, B))(A) = y$, which implies $u(y) > u(x)$.*[17]

[17] Notation (h, A) refers to the history that begins with history h and followed by choice set A.

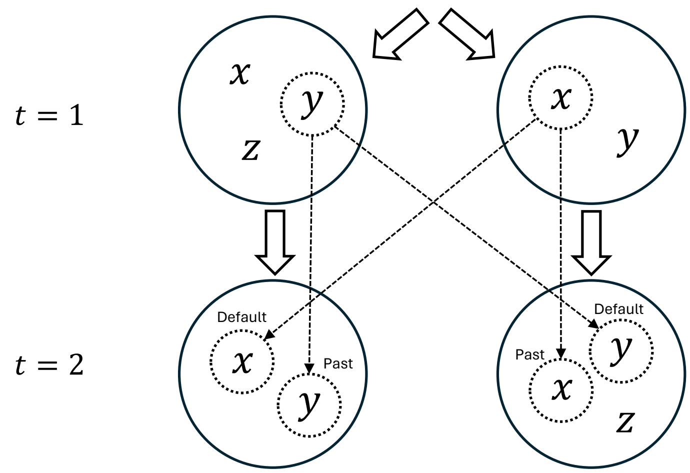

Figure 3.1: Big circles represent choice sets, small circles represent (inferred) consideration.

Convergence rules out other choice patterns that hold different interpretations, namely sticky choice (i.e., $\tilde{c}(\{x,y\})(\{x,y,z\}) = x$, $\tilde{c}(\{x,y,z\})(\{x,y\}) = y$), which may be explained by habit formation, and past independence (i.e., $\tilde{c}(\{x,y\})(\{x,y,z\}) = y$, $\tilde{c}(\{x,y,z\})(\{x,y\}) = x$), where it is as if the DM completely neglects past experiences.

Switches

However, we do not always have the liberty of designing clever tests. How can we infer preferences in general? Consider a formal definition of the event that allows us to reveal preference—switches. Note that they need not be WARP violations.

Definition 2. Given c and suppose $x \neq y$.

1. Let $x\mathbb{S}_{(A_n)}y$ if $c((A_n))_i = y$ and $c((A_n))_j = x$ with $y \in A_j$ for some $i < j$.
2. Let $x\mathbb{S}y$ if $x\mathbb{S}_{(A_n)}y$ for some $(A_n) \in \mathcal{A}^{\mathrm{N}}$.

Next, Proposition 3 makes a number of observations about switches. The first is already discussed: if we see a switch from choosing y to choosing x over y, then x is preferred. Moreover, this kind of evidence can always be found as long as both x and y are ever chosen. Also, because preference is unique, evidence can never be contradictory: if it is possible to reveal that x is preferred to y in one sequence, it is impossible to reveal the opposite in any sequence. Finally, if there are only finitely many alternatives, it is possible to collect all of these evidences using just one sequence of choice sets by asking the right questions in the right order.[18]

Proposition 3. *Suppose c admits an AAT representation (u, Γ).*

1. *If $x\mathbb{S}y$, then $u(x) > u(y)$.*
2. *The relation $\mathbb{S}$ on $\hat{X}$ is a strict total order.*[19]
3. *If X is finite, then there exists (A_n) such that $\mathbb{S}_{(A_n)}$ on $\hat{X}$ is a strict total order.*

Identification of Attention

It remains to address whether attention function Γ is also unique, and the answer is negative.

For the behavior that results from full attention, it makes no difference whether the DM has considered the alternatives that she did not chose; they are inferior anyway. But the same intuition applies more generally, since we can always alter Γ by adding or removing inferior alternatives without changing the model's prediction. Therefore, the set of possible model specifications is characterized by a maximal set: Given c, consider $\Gamma^+ : \mathcal{A} \to 2^X \setminus \{\emptyset\}$ where

$$\Gamma^+(A) := \left\{ x \in A : c_0(A)\,\mathbb{S}x \text{ or } x = c_0(A) \text{ or } x \notin \hat{X} \right\}.$$

Proposition 4. *If c admits an AAT representation, then it also admits an AAT representation with attention function Γ if and only if $c_0(A) \in \Gamma(A) \subseteq \Gamma^+(A)$ for all A.*

In practice, this means an analyst in doubt has the freedom to different model specifications that would not alter predictions, including overestimating the size of attention function by taking the union of possible candidates of Γ or by taking a conservative approach using intersections (Online Appendix B Corollary 3).

[18] It is impossible to construct a c-independent sequence that fully identifies preferences. Online Appendix B Example 13 provides a counterexample, Corollary 2 outlines the best a c-independent sequence can do.

[19] A strict total order is a binary relation that is asymmetric (if $x\mathbb{S}y$, then not $y\mathbb{S}x$), transitive (if $x\mathbb{S}y$ and $y\mathbb{S}z$, then $x\mathbb{S}z$), and connected (if $x \neq y$, then $x\mathbb{S}y$ or $y\mathbb{S}x$).

3.3 Counterfactual Violations

Next, I argue that our existing tests for limited attention cannot possibly be complete without taking choice sequences into account. The key idea lies in the distinction between two kinds of WARP violations that are manifested differently: counterfactual and realized.

For every history h, $\tilde{c}(h) : \mathcal{A} \to X$ is a one-shot choice function that captures, at a given point in time, what the DM would choose for each upcoming choice set. When $\tilde{c}(h)$ violates WARP, for instance

$$\tilde{c}(h)(\{x,y\}) = x \;\&\; \tilde{c}(h)(\{x,y,z\}) = y,$$

it is a *counterfactual* violation. This is typically observed in a between-subject design where each DM is randomly assigned to one of two choice sets and aggregate behavior suggests someone is “non-standard”. It differs from WARP violations that have *realized* as we observe the same DM over time, obscured in studies that focus on one-shot decisions, for instance

$$\tilde{c}(\emptyset)(\{x,y\}) = x \;\&\; \tilde{c}((\{x,y\}))(\{x,y,z\}) = y.$$

Consider Figure 2.1 and suppose $x, y \in A \cap C$. Suppose in the largest rectangle x is chosen from A and y is chosen from C, then this is a counterfactual WARP violation at history h. In contrast, if along the solid path x is chosen from A and y is chosen from C, then this is a *realize*d WARP violation.

Perhaps unexpected at first, it can be shown that even though full attention rules out counterfactual WARP violations, the absence of counterfactual WARP violations does not rule out limited attention.

To see this, Example 1 asserts that full attention rules out counterfactual WARP violations. However, Example 2 shows that a DM who satisfies WARP initially may violate WARP in the future, so the conventional setting that only considers one-shot decisions can fall short of detecting all instances of limited attention. Curious at first, it also highlights that WARP compliance can worsen with increased experience, even though decisions have definitely improved. Perhaps surprisingly, a DM who *never* commits counterfactual WARP violations can be (non-trivially) affected by limited attention, illustrated in Example 3.

Example 1. Consider an AAT representation where $\Gamma(A) = A$ for all A, so $\tilde{\Gamma}(h)(A) = A$ for all h, which means the DM considers everything all the time and therefore by utility maximization, she never commits counterfactual WARP violations.

Example 2. Consider $X = \{x, y, z, z'\}$ and $\Gamma(A) = \{z\}$ if $z \in A$, $\Gamma(A) = \{x\}$ if $z \notin A$ and $x \in A$, $\Gamma(\{y, z'\}) = \{y\}$, and $u(x) > u(y) > u(z) > u(z')$. Notice that c_0 can be explained by the maximization of preference ranking $z \succ_0 x \succ_0 y \succ_0 z'$, thereby complies with WARP. But with history $h = (\{y, z'\})$, choices $\tilde{c}(h)(\{x, y, z, z'\}) = y$ and $\tilde{c}(h)(\{x, y\}) = x$ form a counterfactual WARP violation.

Example 3. Consider $X = \{x, y, z\}$ and $\Gamma(A) = \{x\}$ if $x \in A$, $\Gamma(\{y, z\}) = \{y\}$, and $u(z) > u(y) > u(x)$. Notice that $c_0(\{x, y\}) = x$ but $\tilde{c}(\{y, z\})(\{x, y\}) = y$, i.e., the choice from $\{x, y\}$ varies with history, so the DM is inconsistent with full attention. However, c_0 can be explained by the maximization of preference ranking $x \succ_0 y \succ_0 z$, and it can be shown that the DM never commits counterfactual WARP violations.[20]

Online Appendix B Corollary 4 shows that a counterfactual WARP violation is present if and only if there is a sufficiently large difference between initial behavior and true preference.

3.4 Necessary Departures

If even the lack of counterfactual violations cannot rule out full attention, can anything rule out full attention? It turns out that the necessary traces of limited attention lie in realized violations and history-dependent choices.

Recall that AAT admits the special case where choice behavior is consistent with full attention, i.e., $\Gamma(A) = A$ for all A. In this case, every decision appears to be independent of history, and no WARP violations will be detected, be it counterfactual

[20] A WARP violation requires a better alternative to receive attention in some but not all choice sets, which is impossible when only the worst alternative of each choice set receives attention initially. Formally, a WARP violation at $\tilde{c}(h)$ requires $a, b \in X$ and $A, B \in \mathcal{A}$ such that $\{a, b\} \subseteq A \cap B$, $u(a) > u(b)$, $a \in \tilde{\Gamma}(h)(A)$ and $a \notin \tilde{\Gamma}(h)(B)$. But this is not possible. If $a \in \Gamma(A)$, then it is not possible that $b \in A$ since only the worst outcome is considered by default. If $a \notin \Gamma(A)$, then $a \in \tilde{\Gamma}(h)(A)$ implies $a \in \boldsymbol{c}(h)$, which means $a \in \tilde{\Gamma}(h)(B)$, a contradiction.

or realized. To see this, notice that for all h and A,

$$\tilde{c}(h)(A) = \underset{x \in \Gamma(A) \cup (\boldsymbol{c}(h) \cap A)}{\arg\max} u(x) = \underset{x \in A \cup (\boldsymbol{c}(h) \cap A)}{\arg\max} u(x) = \underset{x \in A}{\arg\max}\, u(x).$$

Of course, just because behavior *can* be represented this way does not mean the DM actually has full attention; perhaps she got lucky by only paying attention to the best things. Aware of this distinction, we avoid calling this behavior full attention and instead say that it admits a *standard utility representation.*

It turns out that in AAT, behavior must admit a standard utility representation unless it involves *both* realized WARP violations and history dependence, making them the quintessential markings of limited consideration.

In order to make this argument formal, consider two new definitions that result from strengthening Axiom 1 and Axiom 2.

Definition 3.

1. *Full Stability:* for every $(A_n) \in \mathcal{A}^{\mathrm{N}}$, if $c((A_n))_i = x$, $\{x, y\} \subseteq (A_n)_i \cap (A_n)_j$, and $x \neq y$, then $c((A_n))_j \neq y$.

2. *Past Independence:* for every $(A_1, ..., A_K) \in \mathcal{A}^{<\mathrm{N}}$ and $B \in \mathcal{A}$, $\tilde{c}((A_1, ..., A_K))(B) = \tilde{c}((A_1, ..., A_{K-1}))(B)$.

Full Stability captures WARP within sequence. Past Independence captures choices that do vary with past experiences. The former strengthens Axiom 1 and the latter strengthens Axiom 2.

Theorem 2. *Suppose c admits an AAT representation. The following are equivalent:*

1. *c satisfies Full Stability*

2. *c satisfies Past Independence*

3. *c admits a standard utility representation*

Theorem 2 summarizes the observations we discussed with one additional finding: Full Stability and Past Independence are linked under AAT, even though they are in general non-nested and could be interpreted as different aspects of "rational" behaviors. In particular, Past Independence has remarkable implications: it unites one-shot choices and choice sequences; if we return to Figure 2.1, it means the same alternative has to be chosen from A no matter when and where A appears. If we believe in Past Independence, then choice sequences are redundant. But bounded rationality often includes the prospect of correcting "mistakes", so Past Independence will likely fail and choice sequences become important.

Combining Theorem 2 and the examples in Subsection 3.3 makes the case for studying choice sequences: The lack of counterfactual violations neither rules out limited attention nor rules out realized violations, but the lack of realized violations implies consistency with standard utility representation (Theorem 2) and therefore rules out counterfactual violations.

4 Attention Structures

AAT puts no restriction on attention structures, i.e., Γ can be anything, but an extensive literature suggests that certain attention structures make more sense than others. For instance, Masatlioglu et al. (2012)'s *choice with limited attention* (CLA) proposes that dropping an alternative that does not receive consideration should not alter the consideration set, and hence a particular structure is required on Γ. Manzini and Mariotti (2007)'s *rational shortlist method* (RSM) involves a criterion that removes alternatives from final consideration and the universal application of the same criterion imposes structure on Γ.

It is easy to just plug in these structures into Γ, but an immediate concern arises: Will the accumulation of experiences, which can alter consideration sets, cause a DM to depart from a particular attention structure? Surprisingly, the answer is probably not.

Subsection 4.1 considers a "vertical merger" between AAT and CLA, Subsection 4.2 does the same for RSM, and Subsection 4.3 does the same for Manzini and Mariotti (2012)'s *categorize-then-choose* (CTC). These models turn out to be highly compatible with AAT; they introduce structures on attention *across choice sets*, whereas AAT introduces structure on attention *across time*. Their complementarity provides a robust framework and proposes new ways to verify inattention. Subsection 4.4 concludes with a complete characterization of compatible models.

4.1 Attention Filter

Due to Masatlioglu et al. (2012):

Definition 4. A mapping $\hat{\Gamma} : \mathcal{A} \to \mathcal{A}$ is an *attention filter* if $\hat{\Gamma}(A) \subseteq A$ and $y \notin \hat{\Gamma}(A)$ implies $\hat{\Gamma}(A \setminus \{y\}) = \hat{\Gamma}(A)$

Definition 5. A choice function $\hat{c} : \mathcal{A} \to X$ is a *choice with limited attention* (CLA) if there exist $\hat{u} : X \to \mathbb{R}$ and an *attention filter* $\hat{\Gamma} : \mathcal{A} \to \mathcal{A}$ such that

$$\hat{c}(A) = \arg\max_{x \in \bar{\Gamma}(x)} \hat{u}(x).$$

It turns out that in AAT, a DM who has an attention filter will always have an (possibly different) attention filter, even though her consideration sets have changed as she accumulates experiences. Proposition 5 formalizes this statement and Example 4 illustrates.

Proposition 5. *If c admits an AAT representation (u, Γ) where Γ is an attention filter, then for any history $h \in \mathcal{A}^{<N}$,*

1. $\tilde{\Gamma}(h) : \mathcal{A} \to \mathcal{A}$ *is an attention filter,*
2. $\tilde{c}(h) : \mathcal{A} \to X$ *is a CLA.*

Example 4 (Secret Menu)**.** A fast food chain offers four items, cheeseburger a, hamburger b, Flying Dutchman d, and Animal Fries e. A customer is initially unaware of the latter two. If both a and b are unavailable, the chain will recommend e, bringing it to the consumer's attention. And if e is also unavailable, then the store recommends d. The attention function is therefore $\Gamma(A) = \{a, b\} \cap A$ if $\{a, b\} \cap A \neq \emptyset$, $\Gamma(\{d, e\}) = \Gamma(\{e\}) = \{e\}$, and $\Gamma(\{d\}) = \{d\}$, which satisfies the property of an attention filter. Suppose $u(d) > u(e) > u(b) > u(a)$. Consider the history $h = (\{d, e\})$ from which e is considered and chosen, the consumer has since discovered e and includes it in her future consideration sets, i.e., $\tilde{\Gamma}(h)(A) = \Gamma(A) \cup \{e\}$ if $e \in A$ and $\tilde{\Gamma}(h)(A) = \Gamma(A)$ otherwise. Although $\tilde{\Gamma}(h)$ differs from Γ, it is still an attention filter because dropping d from a choice set does not change the consideration set.

One way to characterize these behaviors is to impose CLA's original axioms on c_0; they amount to putting (testable) restrictions on counterfactual choices.[21] One

[21] Masatlioglu et al. (2012) proposes an axiom called *WARP with Limited Attention*.

might wonder if it is possible to, instead, impose restrictions on choices across time, and the answer is positive. Axiom 4 introduces the behavioral manifestation of CLA in choice sequences. It posits that if dropping y results in x no longer chosen—for which CLA would infer x is preferred to y—then the DM should never switch from choosing x to choosing y over x.

Axiom 4. *If $c_0(T) = x$ and $c_0(T \setminus \{y\}) \neq x$, then not $y \mathbb{S} x$.*

Proposition 6. *c satisfies Axioms 1, 2, 3 and 4 if and only if it admits an Attention Across Time (AAT) representation (u, Γ) where Γ is an attention filter.*

This robust framework strengthens our identification of parameters. Results from Section 3 carry over, so preferences are pinned down even though CLA by itself is insufficient. And in the instances where CLA infers x is preferred to y, then the "direct evidence" where the DM switches from choosing y to choosing x over y can be found in choice sequences.[22] On the other hand, CLA narrows down the permissible set of attention functions to those that are supported by the theory of attention filters, mitigating the issue of multiplicity and providing important interpretations.

4.2 Shortlisting

Due to Manzini and Mariotti (2007):

Definition 6. A choice function $\hat{c} : \mathcal{A} \to X$ is a *rational shortlist methods* (RSM) if there exist asymmetric binary relations P_1 and P_2 on X such that

$$\hat{c}(A) = \max(\max(A, P_1), P_2).$$

Here, $\max(A, S) := \{x \in A | \text{ not } ySx \, \forall y \in A\}$. The model describes a choice procedure that involves sequentially making a choice, where the DM first creates a shortlist using a rationale, P_1, and then makes a final decision using P_2. The shortlist can therefore be viewed as a consideration set with certain features, defined in Definition 7. Like the case for an attention filter, Proposition 7 suggests that a DM

[22] If $c_0(T) = x$, $c_0(T \setminus \{y\}) \neq x$, and $y \in \hat{X}$, then not $y\mathbb{S}x$, so $x\mathbb{S}y$ due to Proposition 3 (2). The reverse is not always true: if the DM has $u(x) > u(y)$, always considers everything, and y is not the worst alternative, then $x\mathbb{S}y$ but choices never violate counterfactual WARP.

who uses a shortlist will always use a shortlist, even if accumulating experience has changed the DM's behavior.

Definition 7. A mapping $\hat{\Gamma} : \mathcal{A} \to \mathcal{A}$ is a *shortlist* if there exists an asymmetric binary relation S on X such that $\hat{\Gamma}(A) = \max(A, S)$ for all A.

Proposition 7. *If c admits an AAT representation (u, Γ) such that Γ is a shortlist, then for any history $h \in \mathcal{A}^{<N}$,*

1. *$\tilde{\Gamma}(h) : \mathcal{A} \to \mathcal{A}$ is a shortlist,*
2. *$\tilde{c}(h) : \mathcal{A} \to X$ is an RSM.*

The merger between RSM and AAT captures an intuitive process: The DM shortlists alternatives before making final decisions but revises her rationales as she accumulates experience. Specifically, when a DM gains experience with an alternative, the original rationale P_1 is revised so that nothing eliminates said alternative, thereby guaranteeing its consideration in the future and resulting in better decisions. Example 5 illustrates.

Example 5 (Shortlisting Suppliers)**.** A firm can choose from a set of suppliers, but a Malaysian supplier a and a Thailand supplier b are removed from consideration when a China supplier d is available. So Γ is a shortlist derived from the rationale bP_1a and bP_1d. Suppose $u(a) > u(b) > u(d)$. During COVID-19 lockdowns, d is temporarily unavailable, resulting in the choice set $\{a, b\}$ from which a is chosen. After this experience ($h = (\{a, d\})$), the firm cancels bP_1a but maintains bP_1d; so $\tilde{\Gamma}(h)$ is still a shortlist even though $\tilde{\Gamma}(h) \neq \Gamma$.

As a consequence, the DM's decisions not only become increasingly informative of her true preferences but also explain whether her past choices were in fact influenced by shortlisting. To see this, consider a sequence of observations where x was initially chosen over y, but after y became chosen in an incidental choice problem, future comparisons resolve in favor of y. This confirms that the initial choice of x was driven by a rationale that eliminated y instead of the reflection of genuine preference.

4.3 Dominated Categories

Related to the intuition of shortlisting is when an entire category of options is removed from consideration when another category (not necessarily better) is present. Due to Manzini and Mariotti (2012):

Definition 8. A choice function $\hat{c} : \mathcal{A} \to X$ is *a categorize-then-choose* (CTC) if there exist asymmetric binary relations $\succ_s$ on $2^X \setminus \{\emptyset\}$ and $\succ^*$ on X such that $\hat{c}(A) = \max\left(\max^s(A, \succ^s), \succ^*\right)$.

Here, $\max^s(A, \succ^s) := \{x \in A | \nexists R', R'' \subseteq A : R'' \succ^s R' \text{ and } x \in R'\}$ and $\max(A, \succ^*)$ is defined in Subsection 4.2. The model describes a choice procedure where alternatives belonging to dominated categories are eliminated using a shading relation $\succ^s$ before final decision is made using $\succ^*$. The authors describe the first stage as coarse maximization, using categories. Definition 9 provides a formal definition.

Definition 9. A mapping $\hat{\Gamma} : \mathcal{A} \to \mathcal{A}$ is a *coarse-max* if there exists an asymmetric binary relation S on $2^X \setminus \{\emptyset\}$ such that $\hat{\Gamma}(A) = \max^s(A, S)$ for all A.

Similar to before, Proposition 8 suggests that a DM who coarse-max will continue to coarse-max even with accumulating experience. However, future coarse-max becomes "finer" as they involve smaller (dominated) categories, resulting in larger consideration sets and better decisions due to increased experience. Example 6 illustrates the idea.

Proposition 8. *If c admits an AAT representation (u, Γ) such that Γ is a coarse-max, then for any history $h \in \mathcal{A}^{<N}$,*

1. *$\tilde{\Gamma}(h) : \mathcal{A} \to \mathcal{A}$ is a coarse-max,*
2. *$\tilde{c}(h) : \mathcal{A} \to X$ is a CTC.*

Example 6 (Favorite Restaurants)**.** Consider the example from Manzini and Mariotti (2012) where the availability of {Italian restaurants} shades {Mexican restaurants}, excluding the latter category from final decisions. Imagine a history in which a sole Mexican restaurant was open during Ferragosto, visited by the consumer, and was subsequently re-categorized as a special Mexican restaurant (perhaps a favorite) exempted from shading. The other Mexican restaurants continue to be shaded by Italian restaurants. In the future, the consumer's genuine preference will determine whether they return to this Mexican restaurant.

4.4 Characterizing Compatibility

CLA, RSM, and CTC are different models that capture different behavior, but their compatibility with AAT contributes to a robust framework where AAT identifies *preferences* and these models provide insights to the intrinsic formation of *(in)attention*. Can other models be compatible? I now characterize a sufficient and necessary property.

Let X be a countable set of alternatives and let $\mathcal{A}$ be the set of all subsets of X. Let $\mathbb{C}_{All}$ be the collection of all (one-shot) choice functions $\hat{c} : \mathcal{A} \to X$ such that $\hat{c}(A) \in A$ for all $A \in \mathcal{A}$. A subset $\mathbb{C} \subseteq \mathbb{C}_{All}$, which may include some choice functions and exclude others, can be viewed as the universe of behaviors explained by a given choice model, or equivalently, those that satisfy some given axioms. Let $\mathbb{C}_{WARP}$ characterize the set of all choice functions that satisfy WARP (i.e., $\hat{c} \in \mathbb{C}_{WARP}$ if and only if $\hat{c}(T) = \hat{c}(S)$ whenever $\hat{c}(S) \in T \subseteq S$). By convention, $f(\mathcal{A}) := \{f(A) : A \in \mathcal{A}\}$.

Definition 10. $\mathbb{C}$ is *compatible with AAT* if for every $f \in \mathbb{C}$, there exists an AAT representation such that $c_0 = f$ and $\tilde{c}(h) \in \mathbb{C}$ for all $h \in \mathcal{A}^{<\mathrm{N}}$.

Definition 10 captures compatibility in the following sense. Suppose we observe a DM who suffers from limited attention and whose "non-standard" behavior f belongs to a choice model $\mathbb{C}$. If f is a particularly interesting or important behavior, then the ability for $\mathbb{C}$ to explain f is good news for $\mathbb{C}$. However, if the DM is also AAT, then her behavior *tomorrow*, f', could differ from f, perhaps because her experience today causes her to expand her awareness or consideration of alternatives. The question that compatibility asks is whether f' still belongs to $\mathbb{C}$, and whether this continues to hold when we observe $f'', f''', f'''', \ldots$ subsequently. If not, then $\mathbb{C}$ and AAT are not compatible. Note that for f' to belong to $\mathbb{C}$, there is a certain "advantage" for $\mathbb{C}$ to be large and include many choice functions; but that comes at a cost. An overly large $\mathbb{C}$ may include a certain odd behavior g where g or one of $g', g'', g''' \ldots$ fails to belong to $\mathbb{C}$, resulting again in incompatibility.

Compatibility is therefore non-trivial.[23] CLA, RSM, and CTC are compatible with AAT due to a common feature called WARP-convex.

Definition 11. Let $f, g, \kappa \in \mathbb{C}_{All}$.

[23] Online Appendix B Example 14 provides a $\mathbb{C}$ that is not compatible with AAT.

1. g is a *κ-cousin* of f if for some finite $T \subseteq f(\mathcal{A})$, $g(A) = \kappa(\{f(A)\} \cup [A \cap T])$.
2. $\mathbb{C}$ is *WARP-convex* if for all $f \in \mathbb{C}$, there exists $\kappa \in \mathbb{C}_{WARP}$ such that every κ-cousin of f is in $\mathbb{C}$.

Intuitively, a model (or a set of axioms) is WARP-convex if, whenever a WARP-violating choice function is predicted by the model, choice functions that can be derived by reconciling some violations with a WARP-complying choice function can also be explained by the model. It captures a model's tolerance to the correction of "mistakes".

A $\mathbb{C}$ that only contains WARP-conforming choice functions (even if $\mathbb{C} \neq \mathbb{C}_{WARP}$) is trivially WARP-convex, including expected utility, exponential discounting, and generalizations that preserve WARP. The notion is meaningful when we consider non-WARP models. It turns out that despite its non-triviality, all of the limited consideration models I looked at are WARP-convex, including RSM (Manzini and Mariotti, 2007), CLA (Masatlioglu et al., 2012), CTC (Manzini and Mariotti, 2012), *rationalization* (Cherepanov et al., 2013), and *overwhelming choice* (Lleras et al., 2017).

Theorem 3. *$\mathbb{C}$ is WARP-convex if and only if it is compatible with AAT.*

Theorem 3 has two components. First, it suggests that a (one-shot) choice model is compatible with AAT only if its WARP violations are *correctable*. This direction is intuitive; since AAT compels the correction of WARP violations in future choices, a compatible choice model must tolerate a DM who is in the process of these corrections.

Perhaps unexpected at first, the opposite is also true. A choice model is compatible with AAT *as long as* WARP violations are correctable, which highlights the fact that AAT does nothing more than correcting WARP violations in a specific way. If no WARP violation is present, then AAT would accommodate a choice behavior as is; but if violations are present, then AAT would religiously correct them without introducing new or different types of WARP violations.[24]

[24]When a model is compatible, it does not mean $\tilde{\Gamma}(h)$ will, unlike CLA, RSM, and CTC, satisfy the structures imposed by these models. To see this, Geng (2022)'s *limited consideration model with capacity-k* puts a cap on the size of consideration sets. So $\tilde{\Gamma}(h)(A)$ will certainly exceed this capacity when A is large and h is long. Curiously, their model is compatible with AAT, essentially because we can always find $\hat{\Gamma}(h) \neq \tilde{\Gamma}(h)$, by removing inferior alternatives from consideration, such that $\hat{\Gamma}(h)$ predicts the same choices while staying within capacity. The same observation applies to Geng and Özbay (2021)'s *shortlisting with capacity-k*.

5 Attention under Frames

The analysis readily presents an extended framework that incorporates *frames*. It allows us to study how frames affect attention (now and in the future) and, more fundamentally, whether frames work.[25] A representation for generic frames is introduced, and later specialized to the cases where frames are *lists* (the DM searches from top to bottom but may stop at any point) and *recommendations* (the DM considers recommended alternatives and possibly more).

5.1 Generic Frames

I assume that $\bar{\mathcal{A}}$ is a collection of choice problems with the typical element $\bar{A} = (A, F) \in \bar{\mathcal{A}}$. The primitive is a function that assigns to each infinite sequence of choice problems (choice sets with frames) an infinite sequence of choices, $c : \bar{\mathcal{A}}^{\mathrm{N}} \to X^{\mathrm{N}}$, and other assumptions are analogous to Section 2.[26] For now, a frame is "generic" in the sense that it can represent any observable difference in the presentation of alternatives. Choices are therefore captured by a collection of history dependent one-shot choice functions,

$$\tilde{c}(h) : \bar{\mathcal{A}} \to X,$$

where $h \in \bar{\mathcal{A}}^{<\mathrm{N}}$ is a finite sequence of choice sets each presented under some frame. Importantly, the DM can make different choices for the same choice set A when it appears under different frames, i.e., $\tilde{c}(h)(A, F) \neq \tilde{c}(h)(A, F')$.

It turns out that the original set of axioms in Section 2, after cosmetic modifications, suffices for an AAT representation with frames.[27]

[25] There is a general interest in identifying the effects of frames, see for example Goldin and Reck (2020).

[26] Formally, let X be a countable set of alternatives and let $\mathcal{A}$ be the set of all finite subsets of X with at least two elements. Let $\bar{\mathcal{A}}$ be a collection of choice problems that satisfies $\{A : (A, F) \in \bar{\mathcal{A}}\} = \mathcal{A}$, that is, every choice set appears at least once, even if not every frame appears in the dataset. I abuse notation by writing "$x \in \bar{A}$" when I mean $x \in A$ where $\bar{A} = (A, F)$. The primitive $c : \bar{\mathcal{A}}^{\mathrm{N}} \to X^{\mathrm{N}}$ satisfies $c\left(\left(\bar{A}_n\right)\right)_k \in \left(\bar{A}_n\right)_k$ for each $\left(\bar{A}_n\right) \in \bar{\mathcal{A}}^{\mathrm{N}}$ and $k \in \mathbb{N}$. I continue to assume future independence. One-shot choice functions $\tilde{c}$, choice without history c_0, and the revealed preference relation P are defined in the same way as in Section 2. I use $\bar{A}_k$ for $\left(\bar{A}_n\right)_k$ and use $f(A, F)$ for $f((A, F))$ where f can be c_0, $\tilde{c}(h)$, Γ, or $\tilde{\Gamma}(h)$.

[27]

Axiom 5. *For any $\left(\bar{A}_n\right) \in \bar{\mathcal{A}}^{N}$ and $h < i < j$, if $c\left(\left(\bar{A}_n\right)\right)_h = x$, $c\left(\left(\bar{A}_n\right)\right)_i = y \neq x$, $x \in \bar{A}_i$, and*

Definition 12. c admits an Attention Across Time with Frames (AAT-F) representation if there exist a *utility function* $u : X \to \mathbb{R}$ and an *attention function* $\Gamma : \bar{\mathcal{A}} \to 2^X \setminus \{\emptyset\}$, where $\Gamma(A, F) \subseteq A$, such that

$$\tilde{c}(h)(A, F) = \underset{x \in \tilde{\Gamma}(h)(A,F)}{\arg\max}\; u(x)$$

where $\tilde{\Gamma}(h)(A, F) = \Gamma(A, F) \cup (\boldsymbol{c}(h) \cap A)$.

Theorem 4. *c satisfies Axioms 5, 6, and 7 if and only if it admits an Attention Across Time with Frames (AATF) representation.*

As before, the DM pays attention to a (weak) subset of alternatives $\Gamma(A, F) \subseteq A$ and considers historically chosen alternatives $\boldsymbol{c}(h)$ if they are available. But unlike before, different frames can induce different consideration sets for same choice set A, i.e., $\Gamma(A, F') \neq \Gamma(A, F'')$, thereby resulting in different choices. Because these choices will remain in lasting consideration, frames have short-term and long-term effects.

5.2 Effective Frames

A number of basic observations capture the effect of framing in this framework.

First, Example 7 suggests that a frame can successfully draw the DM's attention to a target alternative now and in the future.

Example 7. Suppose the history is h and a certain frame F is introduced for choice set A, with the intention of alerting the DM to a target alternative x. If the DM ends up choosing x, i.e., $\tilde{c}(h)(A, F) = x$, then the frame is successful and the DM will always consider x in the future, i.e., $x \in \tilde{\Gamma}(h')(A', F')$ if $x \in A'$, where h' is any history that begins with h and (A, F).

However, Example 8 suggests that repeating an unsuccessful frame will neither alter behavior now nor affect attention in the future.

$y \in \bar{A}_j$, then $c\left(\left(\bar{A}_n\right)\right)_j \neq x$.

Axiom 6. *For any $\bar{B} \in \bar{\mathcal{A}}$, $\tilde{c}\left(\left(\bar{A}_1, ..., \bar{A}_K\right)\right)\left(\bar{B}\right) \in \left\{\tilde{c}\left(\left(\bar{A}_1, ..., \bar{A}_{K-1}\right)\right)\left(\bar{B}\right), \tilde{c}\left(\left(\bar{A}_1, ..., \bar{A}_{K-1}\right)\right)\left(\bar{A}_K\right)\right\}$.*

Axiom 7. *If $c_0\left(\bar{A}\right) Py$, then $y \notin \tilde{c}(h)\left(\bar{A}\right)$ for all $h \in \bar{A}$.*

Example 8. In Example 7, if the frame was not successful, i.e., $\tilde{c}(h)(A, F) \neq x$, then repeating the same frame will be futile. To see this, suppose $\tilde{c}(h)(A, F) = y$ and let h' be the history h followed by (A, F). Since additional consideration is paid only to the newly chosen alternative y, which was already receiving consideration, the DM's consideration set when she encounters (A, F) for the second time is the same as the first time, i.e., $\tilde{\Gamma}(h')(A, F) = \tilde{\Gamma}(h)(A, F)$, which results in the same choice $\tilde{c}(h')(A, F) = y$. Since the target alternative x is still not chosen, it will not be added into future considerations.

These observations capture some intuitive aspects of framing. In particular, if the target alternative is inferior, then using a frame to elevate it will not produce meaningful results; the DM will simply consider it and choose something else. Because lasting consideration can only result from the DM choosing the target alternative at least once, the model suggests that effective framing still depends substantially on the relative quality of the target alternative in the choice set (or the consideration set). For example, if a frame can cause more appealing alternatives to *not* receive consideration, then it may help the target alternative to be chosen now and remain in consideration in the future.

Can we help a DM by alerting them to superior alternatives? The model cautions this endeavor. Section 3 highlights that the quality of future decisions depends on the complementary between past experiences and future problems. For instance, alerting the DM to a better alternative tends to be effective (since it is better, the DM will choose it) and improves current utility; but if this alternative is unavailable in future, then the intervention could lead to negative long-term consequences where an always-available option never receives consideration, as in Example 9.

Example 9 (Hidden Talent)**.** A stand-up comedy show needs an emergency substitute, and only fringe performers a (very talented, an economist) and b (talented, a full-time performer) are available for last-minute arrangements. The show is only aware of b and would have chosen her, but a friend brings a to the show's consideration resulting in the hiring of a. However, a later returns to her full-time job and is no longer available in the future. Facing the normal selection of performers, the show does not consider b, even though b is better than most of the chosen performers. The show's overall utility could be better had it chosen b earlier and kept her in lasting consideration.

5.3 Ordered Lists

Suppose each frame is an ordered list, that is, for each $\bar{A} = (A, F) \in \bar{\mathcal{A}}$, F is a complete, transitive, and antisymmetric binary relation (a linear order) on A, where yFx ($y \neq x$) is interpreted as "y is listed above x".

Do DMs search from the top just because alternatives are ordered this way? Consider Axiom 8, which says that if x is chosen from $\bar{A} = (A, F)$ when there is no (observable) history, in which it is listed below y (according to F), then a switch from x to y cannot occur, i.e., not $y\mathbb{S}x$.[28] Intuitively, the consideration of x *should* have implied the consideration of y, assuming that the ordered list works, and choosing x means x is better than y; therefore the DM has no reason to switch from choosing x to choosing y over x.

Axiom 8. *If* $yFc_0\left(\bar{A}\right)$ *for some* $\bar{A} = (A, F)$, *then not* $y\mathbb{S}c_0\left(\bar{A}\right)$.

It turns out that this postulate fully characterizes the expected behavior for a top to bottom search.

Proposition 9. *c satisfies Axioms 5, 6, 7 and 8 if and only if it admits an AATF representation where $x \in \Gamma(A, F)$ and yFx implies $y \in \Gamma(A, F)$.*

Example 10 (Baby Diapers)**.** An academic goes online to purchase baby diapers for a newborn. The search result presents 100 brands in a list and the academic searches from top to bottom but does not consider everything, stopping at a certain point (say, 15 brands). It is not clear if the academic ultimately purchases the most preferred diaper, but if item number 13, brand x, is purchased, then it is preferred over the previous 12 items. Moreover, when the academic returns to top up on diapers, brand x may be listed beyond the point the academic searches, but already aware of this brand the academic considers it anyway.

5.4 Recommendations

Consider one last application where a frame is a set of recommended options. For each $\bar{A} = (A, F) \in \bar{\mathcal{A}}$, $F(A) \subseteq A$ is a set of alternatives that are highlighted or

[28] Similar to Section 3, we say $y\mathbb{S}x$ if there exists $\left(\bar{A}_n\right) \in \bar{\mathcal{A}}^{\mathbb{N}}$ such that $c\left(\left(\bar{A}_n\right)\right)_i = x$ and $c\left(\left(\bar{A}_n\right)\right)_j = y \neq x$ with $x \in \bar{A}_j$ for some $i < j$.

made salient to the DM. For each choice set A, different recommendations F, F' can be made, and they potentially result in different decisions.

Whether or not recommendations work, and whether they work as intended, is neither assumed nor observed; we infer using choice sequences. Axiom 9 says that if y is a recommended option in choice problem $\bar{A}$, but the DM chooses x, then a switch from x to y cannot occur.

Axiom 9. *If $y \in F(A)$ for some $\bar{A} = (A, F)$, then not $y \mathbb{S} c_0 \left(\bar{A}\right)$.*

Proposition 10. *c satisfies Axioms 5, 6, 7 and 9 if and only if it admits an AATF representation where $x \in F(A)$ implies $x \in \Gamma(A, F)$.*

Although consideration sets cannot be directly observed, Axiom 9 provides a test for whether recommendations work. If the axiom fails, that means there is a choice problem $\bar{A}$ from which the decision maker fails to consider a recommended option. On the contrary, if the axiom is satisfied, then behavior is consistent with consideration sets that include all recommended options.

Example 11 (Unsought Advice)**.** Out of two libraries on campus, Lehman a and Butler b, a professor recommends a to a PhD student, expecting that the student will at least consider a (the student has a consistent preference and knows which one is better as long as it is considered). Surprisingly, the student chooses b, leading the professor to conclude that the student prefers b over a. A couple of months later, the student stops going to b and switches to a, indicating to the professor that their initial recommendation was, in fact, disregarded.

6 Conclusion

This paper introduces a framework that studies how past experiences can lead to the consideration of previously chosen alternatives in future decisions. The intuition is captured by a model called *Attention Across Time* (AAT), which allows an analyst to fully pin down preferences even if attention is not directly observed, paving the way to sharper welfare analysis in the presence of limited attention. A wide range of implications is drawn, including empirical techniques to reveal preferences, the separation of counterfactual and realized violations, more robust tests of limited consideration, compatibility with different attention structures, and an extension for studying the short- and long-term effects of frames. The findings center around one key message: that the wealth of information in choice sequences contributes meaningfully to the examination of boundedly rational behavior, where limited consideration is one of many possible examples. Should choice sequences receive lasting consideration?

References

Agranov, M. and P. Ortoleva (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy 125*(1), 40–68.

Aguiar, V. (2017). Random categorization and bounded rationality. *Economics Letters 159*, 46–52.

Ahn, D. S., R. Iijima, Y. Le Yaouanq, and T. Sarver (2019). Behavioural characterizations of naivete for time-inconsistent preferences. *The Review of Economic Studies 86*(6), 2319–2355.

Apesteguia, J. and M. A. Ballester (2013). Choice by sequential procedures. *Games and Economic Behavior 77*(1), 90–99.

Bordalo, P., N. Gennaioli, and A. Shleifer (2020). Memory, attention, and choice. *The Quarterly Journal of Economics 135*(3), 1399–1442.

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., X. Ma, Y. Masatlioglu, and E. Suleymanov (2020). A random attention model. *Journal of Political Economy 128*(7), 2796–2836.

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2019). Deliberately stochastic. *American Economic Review 109*(7), 2425–2445.

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

Cheung, P. and Y. Masatlioglu (2024). Disentangling attention and utility channels in recommendations. Working paper.

Chew, S. H., G. Kader, and W. Wang (2022). Source recursive expected utility on rich mixture sets. Working paper.

Dean, M., Ö. Kıbrıs, and Y. Masatlioglu (2017). Limited attention and status quo bias. *Journal of Economic Theory 169*, 93–127.

Dekel, E., B. L. Lipman, and A. Rustichini (2009). Temptation-driven preferences. *The Review of Economic Studies 76*(3), 937–971.

Dillenberger, D. (2010). Preferences for one-shot resolution of uncertainty and allais-type behavior. *Econometrica 78*(6), 1973–2004.

Dillenberger, D. and P. Sadowski (2012). Ashamed to be selfish. *Theoretical Economics 7*(1), 99–124.

Echenique, F., K. Saito, and G. Tserenjigmid (2018). The perception-adjusted luce model. *Mathematical Social Sciences 93*, 67–76.

Ellis, A. and Y. Masatlioglu (2022). Choice with endogenous categorization. *The Review of Economic Studies 89*(1), 240–278.

Freeman, D. J. (2021). Revealing naïveté and sophistication from procrastination and preproperation. *American Economic Journal: Microeconomics 13*(2), 402–38.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015). Stochastic choice and revealed perturbed utility. *Econometrica 83*(6), 2371–2409.

Geng, S. (2022). Limited consideration model with a trigger or a capacity. *Journal of Mathematical Economics 101*, 102692.

Geng, S. and E. Y. Özbay (2021). Shortlisting procedure with a limited capacity. *Journal of Mathematical Economics 94*, 102447.

Gilboa, I. and D. Schmeidler (1995). Case-based decision theory. *The Quarterly Journal of Economics 110*(3), 605–639.

Gilboa, I., D. Schmeidler, and P. P. Wakker (2002). Utility in case-based decision theory. *Journal of Economic Theory 105*(2), 483–502.

Goldin, J. and D. Reck (2020). Revealed-preference analysis with framing effects. *Journal of Political Economy 128*(7), 2759–2795.

Gossner, O., J. Steiner, and C. Stewart (2021). Attention please! *Econometrica 89*(4), 1717–1751.

Gul, F., P. Natenzon, and W. Pesendorfer (2014). Random choice as behavioral optimization. *Econometrica 82*(5), 1873–1912.

Gul, F. and W. Pesendorfer (2001). Temptation and self-control. *Econometrica 69*(6), 1403–1435.

Gul, F. and W. Pesendorfer (2006). Random expected utility. *Econometrica 74*(1), 121–146.

Gul, F. and W. Pesendorfer (2007). Harmful addiction. *The Review of Economic Studies 74*(1), 147–172.

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

Halevy, Y. and E. Ozdenoren (2022). Uncertainty and compound lotteries: Calibration. *Economic Theory 74*(2), 373–395.

Hauser, J. R. and B. Wernerfelt (1990). An evaluation cost model of consideration sets. *Journal of Consumer Research 16*(4), 393–408.

Hayashi, T. and N. Takeoka (2022). Habit formation, self-deception, and self-control. *Economic Theory 74*(2), 547–592.

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

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of choice under risk. *Econometrica 47*(2), 263–292.

Ke, S. and M. Zhang (2023). Multidimensional choices under uncertainty. Working paper.

Kibris, Ã., Y. Masatlioglu, and E. Suleymanov (2024). A random reference model. *American Economic Journal: Microeconomics 16*(1), 155–209.

Kovach, M. (2020). Twisting the truth: Foundations of wishful thinking. *Theoretical Economics 15*(3), 989–1022.

Kovach, M. and E. Suleymanov (2023). Reference dependence and random attention. *Journal of Economic Behavior & Organization 215*, 421–441.

Kovach, M. and G. Tserenjigmid (2022). Behavioral foundations of nested stochastic choice and nested logit. *Journal of Political Economy 130*(9), 2411–2461.

Lanzani, G. (2022). Correlation made simple: Applications to salience and regret theory. *The Quarterly Journal of Economics 137*(2), 959–987.

Lleras, J. S., Y. Masatlioglu, D. Nakajima, and E. Y. Ozbay (2017). When more is less: Limited consideration. *Journal of Economic Theory 170*, 70–85.

Manzini, P. and M. Mariotti (2007). Sequentially rationalizable choice. *American Economic Review 97*(5), 1824–1839.

Manzini, P. and M. Mariotti (2012). Categorize then choose: Boundedly rational choice and welfare. *Journal of the European Economic Association 10*(5), 1141–1165.

Manzini, P. and M. Mariotti (2014). Stochastic choice and consideration sets. *Econometrica 82*(3), 1153–1176.

Manzini, P., M. Mariotti, and L. Ulku (2021). Sequential approval: A model of "likes", paper downloads and other forms of click behaviour. Working paper.

Masatlioglu, Y. and D. Nakajima (2013). Choice by iterative search. *Theoretical Economics 8*(3), 701–728.

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

Masatlioglu, Y. and E. A. Ok (2005). Rational choice with status quo bias. *Journal of Economic Theory 121*(1), 1–29.

Masatlioglu, Y. and E. A. Ok (2014). A canonical model of choice with initial endowments. *The Review of Economic Studies 81*(2), 851–883.

Munro, A. and R. Sugden (2003). On the theory of reference-dependent preferences. *Journal of Economic Behavior & Organization 50*(4), 407–428.

Noor, J. (2011). Temptation and revealed preference. *Econometrica 79*(2), 601–644.

Ortoleva, P. (2010). Status quo bias, multiple priors and uncertainty aversion. *Games and Economic Behavior 69*(2), 411–424.

Rideout, S. (2021). Choosing for the right reasons. Working paper.

Roberts, J. H. and J. M. Lattin (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research 28*(4), 429–440.

Rozen, K. (2010). Foundations of intrinsic habit formation. *Econometrica 78*(4), 1341–1373.

Salant, Y. and A. Rubinstein (2008). (A, f): Choice with Frames. *The Review of Economic Studies 75*(4), 1287–1296.

Segal, U. (1990). Two-stage lotteries without the reduction axiom. *Econometrica 58*(2), 349–377.

Sugden, R. (2003). Reference-dependent subjective expected utility. *Journal of Economic Theory 111*(2), 172–191.

Tserenjigmid, G. (2020). On the characterization of linear habit formation. *Economic Theory 70*(1), 49–93.

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

Wright, P. and F. Barbour (1977). Phased decision strategies: Sequels to an initial screening. In M. K. Starr and M. Zeleny (Eds.), *Multiple Criteria Decision Making*, pp. 91–109. Amsterdam: North-Holland Publishing Co.

Xu, Y. and L. Zhou (2007). Rationalizability of choice functions by game trees. *Journal of Economic theory 134*(1), 548–556.

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

Zhang, M. (2023). Procedural expected utility. Working paper.

A Appendix: Proofs

Simplifying notations: (i) A: a choice set. (ii) $\boldsymbol{A}$: a sequence of choice sets (of any length). (iii) $[\boldsymbol{A}]_{t=k}^{l}$: the subsequence of $\boldsymbol{A}$ including only elements in positions k through l. (iv) $A \in \boldsymbol{A}$: a choice set A that is in the sequence of choice sets $\boldsymbol{A}$. (v) ABC: the sequence of choice sets that starts with choice set A, followed by choice set B, and ends with choice set C. (vi) x: an alternative. (vii) Ax: alternative x is chosen from the choice set A, which is the only choice set in the sequence. (viii) $A^{\ni y}x$: alternative x is chosen from the choice set A, which is the only choice set in the sequence, and alternative y is in A. (ix) $AxByC^{\ni x}z$: the sequence of choice sets ABC from which x, y, z are chosen respectively, and alternative x is also in C. (x) $\boldsymbol{A}By$: a sequence of choice sets $\boldsymbol{A}$, followed by the choice set B from which alternative y is chosen. (xi) $\boldsymbol{A}xzBy$: a sequence of choice sets $\boldsymbol{A}$, from which alternative x is chosen from some choice set $A \in \boldsymbol{A}$ and alternative z is chosen from some choice set $C \in \boldsymbol{A}$ (in no particular order), followed by choice set B from which alternative y is chosen.

Implications of axioms: Axiom 1, Axiom 2, and Axiom 3 imply the following conditions. Online Appendix B Subsection B.1.1 provides detailed proofs. Once Theorem 1 is proven, these are also implied by an AAT representation.

Condition A.1. If $\boldsymbol{A}Bx$, either Bx or $\boldsymbol{A}x$.

Condition A.2. If $\boldsymbol{A}x$, then Bx for some B.

Condition A.3. If $AxBx$ and $x \neq y$, then not $ByAy$.

Condition A.4. If Ax and $\boldsymbol{B}Ay$ where $x \neq y$, then $AxCyAy$ for some C.

Condition A.5. If $\boldsymbol{A}x$ and $\boldsymbol{B}y$ where $x \neq y$, then either $\boldsymbol{C}yE^{\ni y}x$ or $\boldsymbol{D}xF^{\ni x}y$.

Condition A.6. Suppose $x \neq y$. (1) If $\boldsymbol{A}yB^{\ni y}x$, then not $\boldsymbol{C}xD^{\ni x}y$. (2) If $\boldsymbol{A}xyBx$, then not $\boldsymbol{C}xyBy$. (3) If $\boldsymbol{A}xyB^{\ni y}x$, then not $\boldsymbol{A}xyD^{\ni x}y$.

Condition A.7. If xPy, then not yPx.

Condition A.8. If $\boldsymbol{c}(\boldsymbol{A}) = \boldsymbol{c}(\boldsymbol{B})$, then $\tilde{c}(\boldsymbol{A})(D) = \tilde{c}(\boldsymbol{B})(D)$ for all D.

A.1 Proof of Theorem 1

Fix c. As is common, necessity of axioms (**if**) is straightforward, so I will focus on showing sufficiency of axioms (**only if**). The plan goes as follows. We start by constructing $\succ$ in stage 1, the true underlying preference that the anticipated utility function represents. Stage 2 shows that the constructed $\succ$ has the desirable properties to be represented by a utility function. In stage 3, we construct Γ and show that $(\succ, \Gamma)$ explains choices. Note that by Condition A.2, $\hat{X} = \{x \in X : x = \tilde{c}(\emptyset)(A) \, \exists A \in \mathcal{A}\}$. Moreover, $|X \backslash \hat{X}| \leq 1$ because if $z \in X \backslash \hat{X}$, then $\{z, x\}\, x$ for all $x \neq z$, which means $x \in \hat{X}$ for all $x \neq z$.

Stage 1, construction of $\succ$ Consider any pair $x, y \in X$ such that $x \neq y$. If $x \in \hat{X}$ and $y \in X \backslash \hat{X}$, set $x \succ_P y$. If $x, y \in \hat{X}$, suppose WLOG that $\{x, y\}\, x$.

1. If there exists A such that $Ay\, \{x, y\}\, x$, we set $x \succ_S y$.
2. If there exists A such that $Ay\, \{x, y\}\, y$, we set $y \succ_D x$.

Claim. Either $x \succ_S y$ or $y \succ_D x$ and not both.

Proof. The existence of Ay is guaranteed by $y \in \hat{X}$ and Condition A.2, so either $x \succ_S y$ or $y \succ_D x$. Suppose for contradiction both, so for some A we have $Ay\, \{x, y\}\, x$ and for some B we have $By\, \{x, y\}\, y$, but this violates Condition A.8. □

Stage 2, properties of $\succ$ By the claim, $\succ_S \cup \succ_D$, a subset of $\hat{X} \times \hat{X}$, is connected and antisymmetric. By construction, $\succ_P$, a subset of $\hat{X} \times X \backslash \hat{X}$, is connected by clearly antisymmetric. Hence the relation $\succ := \succ_S \cup \succ_D \cup \succ_P$, a subset of $X \times X$, is connected and antisymmetric.

Claim. $\succ$ is transitive.

Proof. Take any $x, y, z \in X$. If one of x, y, z is in $X \backslash \hat{X}$ (at most one due to $|X \backslash \hat{X}| \leq 1$), say WLOG $x, y \in \hat{X}$ and $z \in X \backslash \hat{X}$, then $x \succ z$ and $y \succ z$ means no violation of transitivity is possible. Now suppose $x, y, z \in \hat{X}$. Suppose for contradiction $x \succ y$, $y \succ z$, and $z \succ x$. Since $x, y, z \in \hat{X}$, each of these $\succ$'s is either $\succ_S$ or $\succ_D$, which implies xPy, yPz, and zPx (definition of P given in Section 2 prior to Axiom 3). Suppose WLOG $\{x, y, z\}\, x$. Since $y, z \in \hat{X}$, Condition A.2 guarantees the existence of Ay and Bz. Either $AyBz$ or $BzAy$ (or both). To see this, suppose not $AyBz$, then by Condition A.1 we have $AyBy$, but Condition A.3 implies not $BzAz$, then by Condition A.1 we have $BzAy$. Suppose WLOG $AyBz$, consider $AyBz\{x,y,z\}\alpha$. If $\alpha = x$, then xPz (due to a switch). If $\alpha = y$, then yPx (due to the default of x). If $\alpha = z$, then zPy (due to a switch). So there is bound to be a contradiction of Condition A.7. $\square$

Stage 3, model explains choice Since $\succ$ on X is a strict total order and X is countable, let $u : X \to \mathbb{R}$ be real-valued function such that $u(x) > u(y)$ if and only if $x \succ y$. Moreover, construct the attention function $\Gamma : \mathcal{A} \to \mathcal{A}$ by $\Gamma(A) := \{\tilde{c}(\emptyset)(A)\}$, for all $A \in \mathcal{A}$. We check that this model specification explains choices. Throughout, we use c_{model} to label the choice function *given by the model*, and from it $\tilde{c}_{\text{model}}$ the one-shot choice functions.

Claim. $\tilde{c}(\emptyset)(A) = \arg\max_{x\in\tilde{\Gamma}(\emptyset)(A)} u(x)$.

Proof. Due to $\Gamma(A) = \{\tilde{c}(\emptyset)(A)\}$ and $\tilde{\Gamma}(\emptyset)(A) = \Gamma(A)$. $\square$

We now show that (u, Γ) explains the entire c. Take any sequence of choice sets $(A_n) \in \mathcal{A}^{\mathbf{N}}$, and suppose for contradiction that, for some i,

$$c((A_n))_i \neq \arg\max_{x\in\tilde{\Gamma}((A_1,\dots,A_{i-1}))(A_i)} u(x).$$

Let $\boldsymbol{i}$ be the set of all such i's; they correspond to the set of all choice sets in (A_n) from which the actual choice is not the same as the model prediction. Denote the minimum element of $\boldsymbol{i}$ by $i^* := \min \boldsymbol{i}$, which is well-defined. The earlier claim implies $i^* \neq 1$. Consider $i^* \geq 2$. For notational convenience, denote the *choice* and the *model prediction* by, respectively,

$$c^R := c((A_n))_{i^*} \text{ and } c^P := c_{\text{model}}((A_n))_{i^*} = \arg\max_{x\in\tilde{\Gamma}\left((A_1,\dots,A_{i^*-1})\right)(A_{i^*})} u(x).$$

Claim. $\left\{c^P, c^R\right\} \subseteq \tilde{\Gamma}((A_1,\dots,A_{i^*-1}))(A_{i^*})$.

Proof. By definition, $c^P \in \tilde{\Gamma}((A_1,\dots,A_{i^*-1}))(A_{i^*})$. Also, Condition A.1 implies either $c^R = \tilde{c}(\emptyset)(A_{i^*})$ or $c^R \in \boldsymbol{c}((A_1,\dots,A_{i^*-1}))$, and since i^* is the first instance of disagreement, we have $c^R \in \tilde{\Gamma}((A_1,\dots,A_{i^*-1}))(A_{i^*})$. We continue with $\left\{c^P, c^R\right\} \subseteq \tilde{\Gamma}((A_1,\dots,A_{i^*-1}))(A_{i^*})$. $\square$

Claim. Either $c^P \succ_S c^R$ or $c^P \succ_D c^R$, which implies $c^P P c^R$.

Proof. By definition, $c^R \in \hat{X}$. Since $c^P \in \tilde{\Gamma}\left(\left(A_1, ..., A_{i^*-1}\right)\right)\left(A_{i^*}\right)$, and since i^* is the first instance of disagreement, either $c^P \in \Gamma\left(A_{i^*}\right)$ which by the construction of Γ implies $c^P = \tilde{c}\left(\emptyset\right)\left(A_{i^*}\right)$ or $c^P \in \boldsymbol{c}\left(\left(A_1, ..., A_{i^*-1}\right)\right)$, each would imply $c^P \in \hat{X}$. Since the model predicts c^P to be chosen over c^R even though c^R was paid attention to, we have $u\left(c^P\right) > u\left(c^R\right)$, which imply $c^P \succ_S c^R$ or $c^P \succ_D c^R$ from the construction of u. Then $c^P P c^R$ follows from the definition of P. □

Claim. $c^R P c^P$.

Proof. Since i^* is the first instance of disagreement, $c^P \in \tilde{\Gamma}\left(\left(A_1, ..., A_{i^*-1}\right)\right)\left(A_{i^*}\right)$ implies either $c^P = \tilde{c}\left(\emptyset\right)\left(A_{i^*}\right)$ or $c^P \in \boldsymbol{c}\left(\left(A_1, ..., A_{i^*-1}\right)\right)$. The formal and c^R chosen from A_{i^*} implies $c^R P c^P$. The latter and c^R chosen over c^P after c^P was previously chosen implies $c^R P c^P$. □

But $c^P P c^R$ and $c^R P c^P$ contradict Condition A.7. We showed that if model prediction and actual choices mismatch for the first time in a sequence, that necessarily leads to a contradiction; this means no mismatches can ever happen.

A.2 Proof of Theorem 2

A standard utility representation with parameter u paired with $\Gamma\left(A\right) = A$ for all A will imply the other two, which is also illustrated in the main text.

Past Independence implies standard utility representation: Applying Past Independence iteratively gives $\tilde{c}\left(h\right)\left(A\right) = \tilde{c}\left(\emptyset\right)\left(A\right)$ for all h and A. Suppose $\tilde{c}\left(\emptyset\right)\left(A\right)$ and $\tilde{c}\left(\emptyset\right)\left(B\right)$ violate WARP, then the sequence $\left(A, B, A, B, ...\right)$ will violate Axiom 1, so $\tilde{c}\left(\emptyset\right)$ satisfies WARP and it is standard that there exists $u : X \rightarrow \mathbb{R}$ such that $\tilde{c}\left(\emptyset\right)\left(A\right) = \arg\max_{x \in A} u\left(x\right)$. So $\tilde{c}\left(h\right)\left(A\right) = \arg\max_{x \in A} u\left(x\right)$ for all h and A.

Full Stability implies Past Independence: Past Independence is equivalent to $\tilde{c}\left(h\right)\left(A\right) = \tilde{c}\left(\emptyset\right)\left(A\right)$ for all h and A (sufficiency of Past Independence uses iterative argument, necessity of Path Independence is straightforward). Suppose Past Independence is not satisfied, so there exists A such that $\tilde{c}\left(\emptyset\right)\left(A\right) = x$ and $\tilde{c}\left(h\right)\left(A\right) = y \neq x$. Due to AAT, this means $u\left(y\right) > u\left(x\right)$ and $y \in \boldsymbol{c}\left(h\right)$. The latter invokes Condition A.2 to guarantee existence of B such that $\tilde{c}(\emptyset)(B) = y$. Consider the sequence of choice set $(A, B, \{x, y\}, ...)$. In AAT, x is chosen from A, y is chosen from B (because $u(y) > u(z)$ for all $z \in \Gamma(B) \cup \{x\}$), and y is chosen from $\{x, y\}$ (because $y \in \tilde{\Gamma}((A, B))(\{x, y\})$ and $u(y) > u(x)$). But x chosen over y (from A) and then y chosen over x (from $\{x, y\}$) within a sequence violates Full Stability.

A.3 Proof of Theorem 3

Only if: Take any $f \in \mathbb{C}$. Since $\mathbb{C}$ is WARP-convex, consider the $\kappa \in \mathbb{C}_{WARP}$ such that every κ-cousin of f is in $\mathbb{C}$. Construct the AAT representation (u, Γ) where u represents κ in standard utility maximization and $\Gamma(A) = \{f(A)\}$. It is clear that $c_0 = f$. Now consider any history $h \in \mathcal{A}^{<\mathrm{N}}$, we show that $\tilde{c}(h) \in \mathbb{C}$. Consider any $A \in \mathcal{A}$. By AAT, $\tilde{\Gamma}(h)(A) = \Gamma(A) \cup [A \cap \boldsymbol{c}(h)] = \{f(A)\} \cup [A \cap \boldsymbol{c}(h)]$, so $\tilde{c}(h)(A) = \arg\max_{x \in \{f(A)\} \cup [A \cap \boldsymbol{c}(h)]} u(x) = \kappa(\{f(A)\} \cup [A \cap \boldsymbol{c}(h)])$, where the second equality is due to the fact that u represents κ. By setting $T = \boldsymbol{c}(h)$, we conclude that $\tilde{c}(h)$ is a κ-cousin of f, and therefore $\tilde{c}(h) \in \mathbb{C}$.

If: Suppose $\mathbb{C}$ is compatible with AAT but not WARP-convex, so for some $f \in \mathbb{C}$, there is no $\kappa \in \mathbb{C}_{WARP}$ such that every κ-cousin of f is in $\mathbb{C}$. Consider this f and suppose the AAT representation (u, Γ) is such that $c_0 = f$ and $\tilde{c}(h) \in \mathbb{C}$ for all $h \in \mathcal{A}^{<\mathrm{N}}$. Since u represents some $\kappa \in \mathbb{C}_{WARP}$ in standard utility maximization, consider the κ-cousin of f that is not in $\mathbb{C}$, exists because $\mathbb{C}$ is not WARP-convex, and call it g, which is associated with some finite $T \subseteq f(\mathcal{A})$. Now we construct a history $h \in \mathcal{A}^{<\mathrm{N}}$ such that $\boldsymbol{c}(h) = T$. Consider the alternatives $x_1, ..., x_n$ such that $\{x_1, ..., x_n\} = T$ and $u(x_i) < u(x_{i+1})$ for each i. For each i, since $x_i \in T \subseteq f(\mathcal{A})$, there exists A_i such that $\tilde{c}(\emptyset)(A_i) = x_i$, which means $x_i \in \Gamma(A_i)$ and $x_i = \arg\max_{z \in \Gamma(A_i)} u(z)$. Consider the history $h = (A_1, ..., A_n)$. Note that $\tilde{c}(\emptyset)(A_1) = x_1$. Then, inductively, for each $i = 2, ..., n$, $\tilde{\Gamma}((A_1, ..., A_{i-1}))(A_i) = \Gamma(A_i) \cup \{x_1, ..., x_{i-1}\}$ and $u(x_k) < u(x_i)$ for all $k < i$, so $\tilde{c}((A_1, ..., A_{i-1}))(A_i) = x_i$. So $\boldsymbol{c}(h) = T$, and therefore $\tilde{c}(h)(A) = \arg\max_{z \in \Gamma(A) \cup [A \cap \boldsymbol{c}(h)]} u(z) = \arg\max_{z \in \{f(A)\} \cup [A \cap T]} u(z) = g(A)$ for all $A \in \mathcal{A}$, so $\tilde{c}(h) = g$. It is impossible that $\tilde{c}(h) \in \mathbb{C}$ but $g \notin \mathbb{C}$.

A.4 Proof of Theorem 4

The proof is identical to the proof of Theorem 1, with these differences: **(1)** Whenever the proof of Theorem 1, including its preceding lemmas and up to stage 3, uses a choice set $A \in \mathcal{A}$, switch it to a choice problem $(A, F) \in \bar{\mathcal{A}}$. If there are multiple of them, $(A, F), (A, F')$, it is WLOG to pick one. This is where the assumption that $\bar{\mathcal{A}}$ includes every choice set $A \in \mathcal{A}$ at least once becomes relevant. Note that the definition of $\hat{X}$ now takes into account frames; that is, if there is any choice problem (A, F) such that when it appears in any sequences the alternative x is chosen, then $x \in \hat{X}$. **(2)** Then at stage 3, where a complete and transitive $\succ$ on X has already been established, proceed with building a utility function $u : X \to \mathbb{R}$ as usual. Now construct the attention function by $\Gamma(A, F) := \{\tilde{c}(\emptyset)(A, F)\}$. **(3)** The rest of the proof shows that choices coincide with model prediction. When it says, "for all choice set $A \in \mathcal{A}$", change it to "for all choice problem $(A, F) \in \bar{\mathcal{A}}$".

Supplemental Materials

(Online)

B Online Appendix: Omitted Proofs and Results

Simplifying notations are given in Appendix A.

B.1 Omitted Proofs

B.1.1 Proofs of Conditions

Lemma 1. *If c satisfies Axiom 2, then it satisfies Condition A.1.*

Proof. Take $\boldsymbol{A}Bx$, and suppose not Bx. Let K be the length of $\boldsymbol{A}$. By Axiom 2, either $[\boldsymbol{A}]_{t=1}^{K-1}\,[\boldsymbol{A}]_{t=K}^{K}\,x$ or $[\boldsymbol{A}]_{t=1}^{K-1}\,Bx$ (or both). If it is the former, we are done since $\boldsymbol{A}x$. Suppose it is the latter, then by Axiom 2 again we have either $[\boldsymbol{A}]_{t=1}^{K-2}\,[\boldsymbol{A}]_{t=K-1}^{K-1}\,x$ or $[\boldsymbol{A}]_{t=K}^{K-2}\,Bx$ (or both). Again, if it is the former, we are done, otherwise we keep moving backward until we find q such that $1 \leq q < K$ and $[\boldsymbol{A}]_{t=1}^{q}\,[\boldsymbol{A}]_{t=q+1}^{q+1}\,x$. If this process does not end when $q = 1$, then $[\boldsymbol{A}]_{t=1}^{1}\,x$ by Axiom 2, so $\boldsymbol{A}x$. □

Lemma 2. *If c satisfies Axiom 2, then it satisfies Condition A.2.*

Proof. Say $\boldsymbol{A}x$, and in particular x is chosen from the K-th element, i.e., $[\boldsymbol{A}]_{t=1}^{K-1}\,[\boldsymbol{A}]_{t=K}^{K}\,x$. By Condition A.1, either $[\boldsymbol{A}]_{t=K}^{K}\,x$ or $[\boldsymbol{A}]_{t=1}^{K-1}\,x$. If the former, let $B = [\boldsymbol{A}]_{t=K}^{K}$ and we are done. If the latter, by Condition A.1 again, either $[\boldsymbol{A}]_{t=K-1}^{K-1}\,x$ or $[\boldsymbol{A}]_{t=1}^{K-2}\,x$. If the former, let $B = [\boldsymbol{A}]_{t=K-1}^{K-1}$ and we are done. Otherwise we keep going backward until we find q such that $1 \leq q < K$ and $[\boldsymbol{A}]_{t=q}^{q}\,x$. If this process does not end by $q = 2$, then it must be that $[\boldsymbol{A}]_{t=1}^{1}\,x$, so Bx where $B = [\boldsymbol{A}]_{t=1}^{1}$. □

Lemma 3. *If c satisfies Axiom 3, then it satisfies Condition A.3.*

Proof. Suppose for contradiction $AxBx$ and $ByAy$. So $\{x, y\} \subseteq A \cap B$. By definition, By and $AxBx$ jointly imply xPy. Then Ax, $ByAy$, and xPy jointly violate Axiom 3. □

Lemma 4. *If c satisfies Axiom 2 and Axiom 3, then it satisfies Condition A.4*

Proof. Suppose Ax and $\boldsymbol{B}Ay$, then yPx by the definition of P. Moreover Condition A.2 implies Cy for some y. Now consider $AxC\alpha A\beta$. Condition A.1 and Cy imply $\alpha \in \{x, y\}$, but $\alpha = x$, Cy, and yPx jointly contradict Axiom 3, so $\alpha = y$. Condition A.1 and Ax imply $\beta \in \{x, y\}$. If $\beta = x$, then xPy, but Ax, $\boldsymbol{B}Ay$, and xPy jointly contradict Axiom 3. So $AxCyAy$. □

Lemma 5. *If c satisfies Axiom 2 and Axiom 3, then it satisfies Condition A.5.*

Proof. Suppose $\boldsymbol{A}x$ and $\boldsymbol{B}y$, then by Condition A.2 there exist Ax and By. Suppose WLOG $\{x, y\}\,x$. Consider $ByA\alpha$. Condition A.1 implies $\alpha \in \{x, y\}$. Suppose $\alpha = x$, then consider $ByAx\,\{x, y\}\,\beta$. No matter β we are done. Suppose $\alpha = y$, then Condition A.1 and Condition A.3 imply $AxBy$. Now consider $AxBy\,\{x, y\}\,\beta$. No matter β we are done. $\square$

Lemma 6. *If c satisfies Axiom 1 and Axiom 3, then it satisfies Condition A.6.*

Proof. (2) and (3) are immediate given (1), which I now show. Suppose for contradiction $\boldsymbol{A}yB^{\ni y}x$ and $\boldsymbol{C}xD^{\ni x}y$, and suppose WLOG $\{x, y\}\,x$. By Axiom 1, $\boldsymbol{C}xD^{\ni x}y\,\{x, y\}\,y$. But this along with $\{x, y\}\,x$ and xPy (by the definition of P due to $\boldsymbol{A}yB^{\ni y}x$) jointly violate Axiom 3. $\square$

Lemma 7. *If c satisfies Axiom 1, Axiom 2, and Axiom 3, then it satisfies Condition A.7.*

Proof. Suppose xPy and yPx. By the definition of P, either (i) $\boldsymbol{A}yB^{\ni y}x$ and $\boldsymbol{C}xD^{\ni x}y$, (ii) $\boldsymbol{A}yB^{\ni y}x$ and $[Cx$ and $\boldsymbol{D}Cy]$, or (iii) $[Ay$ and $\boldsymbol{B}Ax]$ and $[Cx$ and $\boldsymbol{D}Cy]$ (a fourth case is WLOG the first case and is omitted). Case (i) contradicts Condition A.6. Due to Condition A.4, (ii) and (iii) also contradict Condition A.6. $\square$

Lemma 8. *If c satisfies Axiom 1, Axiom 2, and Axiom 3, then it satisfies Condition A.8.*

Proof. Suppose Dz. Let $c_1 = \tilde{c}(\boldsymbol{A})(D)$ and $c_2 = \tilde{c}(\boldsymbol{B})(D)$, and suppose for contradiction $c_1 \neq c_2$. Say $c_1, c_2 \in \boldsymbol{c}(\boldsymbol{A})$, then Condition A.6 (3) is violated. Instead, suppose WLOG $c_1 \notin \boldsymbol{c}(\boldsymbol{A})$, so $c_1 = z$ by Condition A.1 and $c_2 \neq z$, which means c_2Pz and, by Condition A.1, $c_2 \in \boldsymbol{c}(\boldsymbol{A})$. So $\boldsymbol{A}c_2D^{\ni c_2}z$, which means zPc_2. But c_2Pz and zPc_2 jointly contradict Condition A.7. $\square$

B.1.2 Proof of Proposition 1

Suppose $x, y \in \hat{X}$ and suppose for contradiction c admits AAT representations with specifications (u_1, Γ_1) and (u_2, Γ_2) but $u_1(x) > u_1(y)$ and $u_2(x) < u_2(y)$. Since $x, y \in \hat{X}$, Condition A.5 guarantees either $\boldsymbol{C}yE^{\ni y}x$, which contradicts $u_2(y) > u_2(x)$, or $\boldsymbol{D}xF^{\ni x}y$, which contradicts $u_1(x) > u_1(y)$.

B.1.3 Proof of Proposition 2

Suppose $\boldsymbol{C}Ax$ and $\boldsymbol{C}By$. Suppose $u(x) > u(y)$. Since $\tilde{\Gamma}(\boldsymbol{C}A)(B) = \tilde{\Gamma}(\boldsymbol{C})(B) \cup \{x\}$ and $u(x) > u(y) > u(z)$ for all $z \in \tilde{\Gamma}(\boldsymbol{C})(B) \setminus \{x, y\}$, so $\tilde{c}(\boldsymbol{C}A)(B) = x$. Since $\tilde{\Gamma}(\boldsymbol{C}B)(A) = \tilde{\Gamma}(\boldsymbol{C})(A) \cup \{y\}$ and $u(x) > u(z)$ for all $z \in \tilde{\Gamma}(\boldsymbol{C})(A) \setminus \{x\}$ and $u(x) > u(y)$, so $\tilde{c}(\boldsymbol{C}B)(A) = x$. We showed convergence on x when $u(x) > u(y)$. If $u(y) > u(x)$ instead, analogous arguments yield convergence on y.

B.1.4 Proof of Proposition 3

(1) Suppose $x\mathbb{S}_{(A_n)}y$, which by definition means $\boldsymbol{A}yB^{\ni y}x$ exists. Due to AAT, $y \in \boldsymbol{c}(\boldsymbol{A}) \subseteq \tilde{\Gamma}(\boldsymbol{A})(B)$, and so $\tilde{c}(\boldsymbol{A})(B) = x$ implies $u(x) > u(y)$.

(2) Since $x, y \in \hat{X}$, Condition A.5 guarantees either $\boldsymbol{C}yE^{\ni y}x$, which means $x\mathbb{S}y$, or $\boldsymbol{D}xF^{\ni x}y$, which means $y\mathbb{S}x$. Moreover, $x\mathbb{S}y$ and $u(y) > u(x)$ jointly contradict AAT, so $x\mathbb{S}y$ if and only if $u(x) > u(y)$ for all $x, y \in \hat{X}$, hence $\mathbb{S}$ is also asymmetric and transitive on $\hat{X}$.

(3) Suppose AAT representation (u, Γ) that represents c. Since $\hat{X}$ is finite, enumerate the alternatives in $\hat{X}$ by $u(\cdot)$ so that $\{x_1, ..., x_n\} = \hat{X}$ and $u(x_i) < u(x_{i+1})$ for all i. For each i, Condition A.2 guarantees existence of A_i such that $c_0(A_i) = x_i$. Now construct the sequence that begins with $(A_1, ..., A_n)$, followed by the finite sequence of all possible binary choice problems $(B_1, ..., B_k)$ (in any order), with arbitrary completion of what happens next (since (A_n) must be an infinite sequence). Note that $c_0(A_1) = x_1$. For $j \in \{2, ..., n\}$, if $\tilde{\Gamma}((A_1, ..., A_{j-1}))(A_j) \subseteq \Gamma(A_j) \cup \{x_1, ..., x_{j-1}\}$, then $\tilde{c}((A_1, ..., A_{j-1}))(A_j) = \arg\max_{x \in \tilde{\Gamma}((A_1, ..., A_{j-1}))(A_j)} u(x) = x_j$. By induction, this gives $\boldsymbol{c}((A_1, ..., A_n)) = \hat{X}$. Then, in the $(B_1, ..., B_k)$ phase, either $x\mathbb{S}_{(A_n)}y$ or $y\mathbb{S}_{(A_n)}x$ for all $x, y \in \hat{X}$ and $x \neq y$. Since $x\mathbb{S}_{(A_n)}y$ if and only if $u(x) > u(y)$ for all $x, y \in \hat{X}$, $\mathbb{S}_{(A_n)}$ is also asymmetric and transitive on $\hat{X}$.

B.1.5 Proof of Proposition 4

Corollary 1. *If c admits an AAT representation (Γ, u) and $\Gamma^* \subseteq \Gamma$ such that $c_0(A) \in \Gamma^*(A)$ for all A, then c also admits an AAT representation (Γ^*, u).*

Proof. Denote the resulting choices from (Γ^*, u) by c^*, $\tilde{c}^*$, and $\boldsymbol{c}^*$. Consider $h = \emptyset$. For A and $y \in \Gamma^* (A)$, since $y \in \Gamma (A)$, so $u (\tilde{c} (\emptyset) (A)) > u (y)$ if $\tilde{c} (\emptyset) (A) \neq y$. Then $\tilde{c} (\emptyset) (A) \in \Gamma^* (A)$ implies $\tilde{c}^* (\emptyset) (A) = \tilde{c} (\emptyset) (A)$. Then, analogous arguments apply to $h \neq \emptyset$ inductively. For any history h, suppose $\boldsymbol{c} (h) = \boldsymbol{c}^* (h)$ (by induction). For any A,

$$\tilde{\Gamma}^* (h) (A) = \Gamma^* (A) \cup [A \cap \boldsymbol{c}^* (h)] \subseteq \Gamma (A) \cup [A \cap \boldsymbol{c} (h)] = \tilde{\Gamma} (h) (A) .$$

So for any $y \in \tilde{\Gamma}^* (h) (A)$, since $y \in \tilde{\Gamma} (h) (A)$, so $u (\tilde{c} (h) (A)) > u (y)$ if $\tilde{c} (h) (A) \neq y$. By AAT, either $\tilde{c} (h) (A) = \tilde{c} (\emptyset) (A) \in \Gamma^* (A)$ or $\tilde{c} (h) (A) \in \boldsymbol{c} (h) \cap A = \boldsymbol{c}^* (h) \cap A$, so $\tilde{c} (h) (A) \in \tilde{\Gamma}^* (h) (A)$, so $\tilde{c}^* (h) (A) = \tilde{c} (h) (A)$. $\square$

Only if: Fix c. Suppose c admits an AAT representation (Γ, u). Fix any $A \in \mathcal{A}$. Suppose for contradiction $y \in \Gamma (A) \setminus \Gamma^+ (A)$, so $y \mathbb{S} c_0 (A)$ by definition of Γ^+, then by Proposition 3 (1) we have $u (y) > u (c_0 (A))$. In this case, (Γ, u) gives $\arg\max_{x \in \tilde{\Gamma}(\emptyset)(A)} u (x) \neq c_0 (A)$ a contradiction that (Γ, u) represents c. So $\Gamma (A) \setminus \Gamma^+ (A) = \emptyset$, or $\Gamma (A) \subseteq \Gamma^+ (A)$. It is straightforward that $c_0 (A) \in \Gamma (A)$, otherwise $\arg\max_{x \in \tilde{\Gamma}(\emptyset)(A)} u (x) \neq c_0 (A)$, a contradiction that (Γ, u) represents c.

If: Fix c. Suppose c admits an AAT representation (Γ, u). If $z \in X \setminus \hat{X}$, suppose WLOG that $u (x) > u (z)$ for all $x \in \hat{X}$. By Corollary 1, (Γ^*, u) where $\Gamma^* (A) = \{c_0 (A)\}$ also represent c. Now I show that (Γ^+, u) also represent c, and then the proof is complete by invoking Corollary 1. Suppose for contradiction there exist (A_n) and integer k such that the model predictions of (Γ^*, u) and (Γ^+, u) disagree. Let i^* be the integer that represents the first disagreement in (A_n), let $h = (A_1, ..., A_{i^*-1})$, and denote the model predictions by, respectively,

$$a := \underset{x \in \tilde{\Gamma}^*(h)(A_{i^*})}{\arg\max} \ u (x) \ \text{ and } \ b := \underset{x \in \tilde{\Gamma}^+(h)(A_{i^*})}{\arg\max} \ u (x) .$$

Since this is the first disagreement in (A_n), $\boldsymbol{c}^* (h) = \boldsymbol{c}^+ (h)$. Furthermore, since $\Gamma^* (A_{i^*}) = \{c_0 (A_{i^*})\} \subseteq \Gamma^+ (A_{i^*})$ by construction, $\tilde{\Gamma}^* (h) (A_{i^*}) \subseteq \tilde{\Gamma}^+ (h) (A_{i^*})$. So the disagreement is caused by $b \in \Gamma^+ (A_{i^*}) \setminus \{c_0 (A_{i^*})\}$. Because $u (b) > u (a)$, $b \in \hat{X}$. So by the definition of Γ^+ we have $c_0 (A_{i^*}) \mathbb{S} b$. By Proposition 3 (1) this implies $u (c_0 (A_{i^*})) > u (b)$, so it is impossible for (Γ^+, u) to predict b.

B.1.6 Proof of Proposition 5

Given (u, Γ) where Γ is an attention filter. Consider any h. If $x \notin \tilde{\Gamma}(h)(B)$, then $x \notin \Gamma(B)$ and $x \notin \boldsymbol{c}(h)$. Note that $x \notin \Gamma(B)$ also implies $\Gamma(B \backslash \{x\}) = \Gamma(B)$ since Γ is an attention filter. So $\tilde{\Gamma}(h)(B \backslash \{x\}) = \Gamma(B \backslash \{x\}) \cup [(B \backslash \{x\}) \cap \boldsymbol{c}(h)] = \Gamma(B) \cup [B \cap \boldsymbol{c}(h)] = \tilde{\Gamma}(h)(B)$. We established (1), and (2) is implied by definition.

B.1.7 Proof of Proposition 6

If: Since c admits an AAT representation, it satisfies Axiom 1, Axiom 2, and Axiom 3 (Theorem 1). Suppose Γ is an attention filter. If $\tilde{c}(\emptyset)(T) = x$ and $\tilde{c}(\emptyset)(T \backslash \{y\}) \neq x$ where $x \neq y$, it must be that $\Gamma(T) \neq \Gamma(T \backslash \{y\})$, so $y \in \Gamma(T)$ since Γ is an attention filter. Then, since $\tilde{c}(\emptyset)(T) = x$, $u(x) > u(y)$. Since $y \mathbb{S} x$ would have implied $u(y) > u(x)$ due to Proposition 3 (1), not $y \mathbb{S} x$. Hence Axiom 4.

Only if: Since c satisfies Axiom 1, Axiom 2 and Axiom 3, by Theorem 1 it admits an AAT representation (Theorem 1). Consider another AAT representation where the attention function is Γ^+, guaranteed by Proposition 4. We show that Γ^+ is an attention filter. Suppose for contradiction there exists A and y such that $y \in A \backslash \Gamma^+(A)$ and $\Gamma^+(A) \neq \Gamma^+(A \backslash \{y\})$. By the definition of Γ^+, $y \notin \Gamma^+(A)$ implies $y \in \hat{X}$ and $\neg c_0(A) \mathbb{S} y$, and Proposition 3 (2) implies $y \mathbb{S} c_0(A)$. Next we argue that $c_0(A) \neq c_0(A \backslash \{y\})$. Suppose for contradiction $c_0(A) = c_0(A \backslash \{y\})$, then $y \mathbb{S} c_0(A)$ and the definition of Γ^+ would imply $\Gamma^+(A) = \Gamma^+(A \backslash \{y\})$, a contradiction. So $y \mathbb{S} c_0(A)$ and $c_0(A) \neq c_0(A \backslash \{y\})$, but they jointly contradict Axiom 4.

B.1.8 Proof of Proposition 7

Given (u, Γ) where Γ is a shortlist, let S be the corresponding rationale. Consider any h. Define a new rationale $S^* \subseteq X \times X$ by $x S^* y$ if $x S y$ and $y \notin \boldsymbol{c}(h)$. Then let $\tilde{\Gamma}^*$ be the shortlist given by S^*. Finally we check that $\tilde{\Gamma}^* = \tilde{\Gamma}(h)$. Fix any $A \in \mathcal{A}$. If $y \in \tilde{\Gamma}(h)(A)$, then either $y \in \Gamma(A)$ (so $\neg x S y$ for all $x \in X$) or $y \in \boldsymbol{c}(h)$ (so $\neg x S^* y$ for all $x \in X$), so $y \in \tilde{\Gamma}^*(A)$. If $y \in \tilde{\Gamma}^*(A)$, then $\neg x S^* y$ for all $x \in A$, which by definition of S^* can be due to (i) $\neg x S y$ for all $x \in A$, which means $y \in \Gamma(A)$, or (ii) $y \in \boldsymbol{c}(h)$. In either case, $y \in \tilde{\Gamma}(h)(A)$. We established (1), and (2) is implied by definition.

B.1.9 Proof of Proposition 8

Given (u, Γ) where Γ is a coarse-max, let S be the corresponding rationale. Consider any h. Define a new rationale $S^* \subseteq \left[2^X \backslash \{\emptyset\}\right] \times \left[2^X \backslash \{\emptyset\}\right]$ with the following rules. If $R'SR''$, then let $R'S^*R''$ if $R'' \cap \boldsymbol{c}(h) = \emptyset$ and let $R'S^*\left[R'' \backslash \boldsymbol{c}(h)\right]$ if $R'' \cap \boldsymbol{c}(h) \neq \emptyset$. The operation removes $\boldsymbol{c}(h)$ from categories that would be dominated, while forming new categories $R'' \backslash \boldsymbol{c}(h)$ that are dominated. Note that as a result, if $R'S^*R''$, then $R'' \cap \boldsymbol{c}(h) = \emptyset$. Then let $\tilde{\Gamma}^*$ be the coarse-max defined by S^*. Finally we check that $\tilde{\Gamma}^* = \tilde{\Gamma}(h)$. Fix any $A \in \mathcal{A}$. If $y \in \tilde{\Gamma}(h)(A)$, then either $y \in \Gamma(A)$ (so $\neg R'SR''$ for all $R', R'' \subseteq A$ and $y \in R''$) or $y \in \boldsymbol{c}(h)$ (so $\neg R'S^*R''$ for all $R'' \ni y$), so $y \in \tilde{\Gamma}^*(A)$. If $y \in \tilde{\Gamma}^*(A)$, then $\neg R'S^*R''$ for all $R', R'' \subseteq A$ and $y \in R''$, which by definition of S^* can be due to (i) $\neg R'SR''$ for all $R', R'' \subseteq A$ and $y \in R''$, which means $y \in \Gamma(A)$, or (ii) $y \in \boldsymbol{c}(h)$. In either case, $y \in \tilde{\Gamma}(h)(A)$. We established (1), and (2) is implied by definition.

B.1.10 Proof of Proposition 9

If: Suppose $yFc_0\left(\bar{A}\right)$ for some $\bar{A} = (A, F) \in \bar{\mathcal{A}}$ and, for contradiction, $y\mathbb{S}c_0\left(\bar{A}\right)$. So $u(y) > u\left(c_0\left(\bar{A}\right)\right)$ by an adaption of Proposition 3 (1). But

$$c_0\left(\bar{A}\right) \in \Gamma\left(\bar{A}\right) \text{ and } yFc_0\left(\bar{A}\right) \tag{B.1}$$

imply $y \in \Gamma\left(\bar{A}\right)$, so the choice without history from $\bar{A}$ cannot be $c_0\left(\bar{A}\right)$ since y brings greater utility and is considered, a contradiction.

Only if: The existence of an AATF representation (u^*, Γ^*) is given by Theorem 4. If $\hat{X} = X$ (there is no never chosen alternatives), let $u := u^*$. If $X \backslash \hat{X} = \{z\}$ (there is at most one never chosen alternatives because all binary choice sets are in $\mathcal{A}$), let $u := u^*$ and then modify it by setting $u(z) := \min u(a) - 1$. Given $\bar{A} = (A, F) \in \bar{\mathcal{A}}$, let

$$\Gamma(A, F) := \left\{c_0\left(\bar{A}\right)\right\} \cup \left\{y \in A : yFc_0\left(\bar{A}\right)\right\}. \tag{B.2}$$

For each $y \neq c_0\left(\bar{A}\right)$ included into consideration, there are two possibilities. If $y \notin \hat{X}$, then $u(y) < u(z)$ for all $z \in X$, so including y into consideration when $c_0\left(\bar{A}\right)$ is also considered does not affect choice. If $y \in \hat{X}$, an adaption of Proposition 3 (1) guarantees that either $y\mathbb{S}c_0\left(\bar{A}\right)$ or $c_0\left(\bar{A}\right)\mathbb{S}y$, but Axiom 8 rules out the former, so

$c_0\left(\bar{A}\right)\mathbb{S}y$, which implies $u\left(c_0\left(\bar{A}\right)\right) > u\left(y\right)$, hence including y into consideration when $c_0\left(\bar{A}\right)$ is also considered does not affect choice. By construction, Γ satisfies the desired properties.

B.1.11 Proof of Proposition 10

If: The same as the proof of Proposition 9 except that Equation B.1 is replaced by $y \in F\left(A\right)$. **Only if:** The same as the proof of Proposition 9 except that Equation B.2 is replaced by $\Gamma\left(A, F\right) := \left\{c_0\left(\bar{A}\right)\right\} \cup \left\{y \in F\left(A\right)\right\}$ and Axiom 8 is replaced by Axiom 9.

B.2 Omitted Examples and Results

Example 12. This example discusses the empirical test of the convergence property introduced in Subsection 3.2. Consider a population of DMs, each with a (deterministic) AAT representation. For choice sets $\left\{\left\{x, y, z\right\}, \left\{x, y\right\}\right\}$, there are 189 sets of relevant $\left(u, \Gamma\right)$ parameters (due to 9 variations in preferences and 21 variations of preferences). Instead of brute-forcing it, let's categorize the DMs using what they would choose in the first period, which creates a partition of these 189 sets using observable differences. Let P_{ij} be the fraction that would have chosen i from $\left\{x, y, z\right\}$ and j from $\left\{x, y\right\}$ when they encounter these choice sets in the first period; $\sum_{i\in\{x,y,z\}, j\in\{x,y\}} P_{ij} = 1$. Within the P_{xy} fraction, let λ_{xy} be the fraction that prefers x to y and $\left(1 - \lambda_{xy}\right)$ the fraction that prefers y to x; preference for z is irrelevant since DMs in this group will never choose z in the problems we consider (see Proposition 2). Define λ_{yx} similarly. For the P_{zx} fraction, let λ_{zx} be the fraction that prefers x to z and $\left(1 - \lambda_{zx}\right)$ the fraction that prefers z to x. For the P_{zy} fraction, let λ_{zy} be the fraction that prefers z to y and $\left(1 - \lambda_{zy}\right)$ the fraction that prefers y to z.

Now consider a random assignment of the population of DMs into two groups (treatments). In group A, DMs first choose from $\left\{x, y, z\right\}$ (first period) and then choose from $\left\{x, y\right\}$ (second period). In group B, $\left\{x, y\right\}$ first and $\left\{x, y, z\right\}$ second. Suppose for now $P_{zx} = P_{zy} = 0$, for instance when z is a dominated alternative / decoy. Note that if every DM has full attention, then $P_{xy} = P_{yx} = 0$, but we do not assume this.

In group A, for first period's choice (from $\left\{x, y, z\right\}$), $P_{xx} + P_{xy}$ fraction chooses x and $P_{yy} + P_{yx}$ fraction chooses y, with relative fraction of x to y being $R_A^{t=1} = \frac{P_{xx}+P_{xy}}{P_{yy}+P_{yx}}$.

Similarly, in group B, for first period's choice (from $\{x,y\}$), $R_B^{t=1} = \frac{P_{xx}+P_{yx}}{P_{yy}+P_{xy}}$. If $P_{xy} = P_{yx} = 0$, then $R_A^{t=1} = R_B^{t=1}$, so $R_A^{t=1} \neq R_B^{t=1}$ means some DMs have limited attention.

Now we consider second period's choices. In group A, for the second period choice (from $(\{x,y\}$), AAT predicts that the P_{xx} fraction and the P_{yy} fraction will continue to choose x and y respectively, but the P_{xy} fraction and the P_{yx} fraction now consider both x and y and choose according to their preferences; for example, for the P_{xy} fraction, λ_{xy} fraction will choose x and $(1-\lambda_{xy})$ fraction will choose y (see Proposition 2). The same holds for group B. As a result, relative fractions of x to y for group A and group B both equal

$$R_A^{t=2} = R_B^{t=2} = \frac{P_{xx} + P_{xy}\lambda_{xy} + P_{yx}\lambda_{yx}}{P_{yy} + P_{xy}\left(1-\lambda_{xy}\right) + P_{yx}\left(1-\lambda_{yx}\right)}. \tag{B.3}$$

So $R_A^{t=2} = R_B^{t=2} = R^{t=2}$, i.e., convergence, even if $P_{xy}, P_{yx} > 0$.

If we assume that the P_{xx} fraction and the P_{yy} fraction genuinely prefer x and y respectively, then $R^{t=2}$ reveals the overall fraction of DMs who genuinely prefers x relative to the overall fraction of DMs who genuinely prefers y. In general, it is possible that a DM belonging to the P_{xx} fraction actually prefers y to x but never considers y.

Now suppose $P_{zx}, P_{zy} > 0$. If we recalculate $R_A^{t=2}$ and $R_B^{t=2}$, they become

$$R_A^{t=2} = \frac{... + P_{zx}}{... + P_{zy}}, \quad R_B^{t=2} = \frac{... + P_{zx}\lambda_{zx}}{... + P_{zy}\left(1-\lambda_{zy}\right)}, \tag{B.4}$$

where $...$ follow Equation B.3, so in general $R_A^{t=2} \neq R_B^{t=2}$. This occurs due to distributional differences in preferences parameters, and it would occur even with a population of fully standard DMs (i.e., $P_{xy} = P_{yx} = 0$ and $\lambda_{zx} = (1-\lambda_{zy}) = 0$). But this richer dataset provides a solution! In group A, $P_{zx}\lambda_{zx}$ fraction will exhibit choices (z,x), and $P_{zy}\left(1-\lambda_{zy}\right)$ fraction will exhibit (z,y). In group B, $P_{zx}\left(1-\lambda_{zx}\right)$ fraction exhibits (x,z) and $P_{zy}\left(\lambda_{zy}\right)$ fraction exhibits (y,z). So data allows us to pin down $P_{zx}, P_{zy}, \lambda_{zx}, \lambda_{zy}$, hence we can check if Equation B.4 holds!

Example 13. Let $X = \{x,y,z\}$. Consider (u_1, Γ_1) where $u_1(x) > u_1(y)$, $\{A : y \in \Gamma_1(A)\} = \{\{x,y,z\}\}$, $\Gamma_1(\{x,y,z\}) = \{y\}$, $\Gamma_1(\{x,y\}) = \{x\}$, then $x, y \in \hat{X}$ and any sequence (A_n) such that $x\mathbb{S}_{(A_n)}y$ must present $\{x,y,z\}$ before $\{x,y\}$. Now consider (u_2, Γ_2) where $u_2(x) > u_2(y)$, $\{A : y \in \Gamma_2(A)\} = \{\{x,y\}\}$, $\Gamma_2(\{x,y\}) =$

$\{y\}$, $\Gamma_2(\{x,y,z\}) = \{x\}$, then $x, y \in \hat{X}$ and any sequence (B_n) that reveals $x\mathbb{S}_{(B_n)}y$ must present $\{x,y\}$ before $\{x,y,z\}$.

Example 14. The following $\mathbb{C}$ is not compatible with AAT: Let $X = \{1,2,3,4\}$. Suppose $\mathbb{C}$ consists of all WARP-conforming choice functions (to make this example non-trivial) and, additionally, a single choice function f where $f(\{1,2\}) = 1$, $f(\{2,3\}) = 2$, $f(\{3,4\}) = 3$, $f(\{4,1\}) = 4$, $f(\{1,2,3\}) = 3$, $f(\{2,3,4\}) = 4$, $f(\{3,4,1\}) = 1$, and $f(\{4,1,2\}) = 2$ (the other choice sets are irrelevant). No other choice functions are in $\mathbb{C}$. Suppose for contradiction that $\mathbb{C}$ is compatible with AAT. To accommodate f, by the symmetry of f, suppose without loss of generality $u(1) > u(2) > u(3) > u(4)$. After history $h = (\{2,3\})$, $2 \in \tilde{\Gamma}(h)(A)$ if $2 \in A$, but this does not change the choices $\tilde{f}(h)(\{4,1,2\}) = 2$ and $\tilde{f}(h)(\{1,2\}) = 1$, so $\tilde{f}(h)$ violates WARP and $\tilde{f}(h) \neq f$, hence $\tilde{f}(h) \notin \mathbb{C}$, a contradiction that $\mathbb{C}$ is compatible with AAT.

Corollary 2. *Suppose X is finite. There exists a sequence of choice sets $(A_n) \in \mathcal{A}^{\mathbb{N}}$ such that for any c that admits an AAT representation, there exists a subset of alternatives $\bar{X}_c \subseteq X$ such that $\mathbb{S}_{(A_n)}$ on $\bar{X}_c$ is a strict total order and $\left|X \backslash \bar{X}_c\right| \leq 1$.*

Proof. We prove by construction. Consider any sequence (A_n) that begins with the finite sequence of all possible binary choice problems $(B_1, \ldots, B_k)$ (in any order) and then repeats itself, with arbitrary completion of what happens next (since (A_n) must be an infinite sequence). Now consider any c. Note that all but at most one alternative would have been chosen in the first iteration of $(B_1, \ldots, B_k)$ (if z is not chosen, then all other alternatives have been chosen, so z is the only alternative that has not been chosen), denote this set by $\bar{X}_c$. Then, during the repetition of $(B_1, \ldots, B_k)$, either $x\mathbb{S}_{(A_n)}y$ or $y\mathbb{S}_{(A_n)}x$ for all $x, y \in \bar{X}_c$ and $x \neq y$. Moreover, since c admits an AAT representation, due to Proposition 1 and $\bar{X}_c \subseteq \hat{X}$, we have $x\mathbb{S}_{(A_n)}y$ only if $u(x) > u(y)$. Hence $\mathbb{S}_{(A_n)}$ is also asymmetric and transitive. $\square$

Corollary 3. *If c admits AAT representations (Γ, u) and (Γ', u), then it also admits AAT representations $(\Gamma \cup \Gamma', u)$ and $(\Gamma \cap \Gamma', u)$.*

Proof. The case for intersection is shown in Corollary 1; note that an intersection between Γ and Γ' guarantees that $\Gamma^* := \Gamma \cap \Gamma'$ satisfies $\Gamma^* \subseteq \Gamma$ and $c_0(A) \in \Gamma^*(A)$ for all A. For the case for union, the "only if" of Proposition 4 guarantees that $c_0(A) \in \Gamma(A) \subseteq \Gamma^+(A)$ and $c_0(A) \in \Gamma'(A) \subseteq \Gamma^+(A)$, and the "if" part subsequently guarantees (Γ^*, u^*) where $\Gamma^* = \Gamma \cup \Gamma'$ is an AAT representation for the same behavior c. If $u \neq u^*$, then by Proposition 1, $X \backslash \hat{X} = \{z\}$ and $u(z) > u(x)$ for some $x \in \hat{X}$ (the proof of Proposition 4 constructs u^* using $u(x) > u(z)$). But since $(\Gamma, u), (\Gamma', u)$ both represent c, if $c_0(A) = x$, then $z \notin \Gamma(A) \cup \Gamma'(A)$, so (Γ^*, u) also represent c. $\square$

Corollary 4. *Suppose c admits an AAT representation (u, Γ) and suppose c_0 is explained by some strict total order $(\succ_0, X)$ (i.e., $c_0(A) \succ_0 z$ for all $z \in A \backslash \{c_0(A)\}$). There exists history h such that $\tilde{c}(h)$ violates counterfactual WARP if and only if there are $x, y, z \in \hat{X}$ such that $z \succ_0 x \succ_0 y$ but $u(x) > u(y) > u(z)$.*

Proof. **If:** This is given by Example 2, with $c_0(\{y, z'\}) = y$ replaced by $c_0(A) = y$ for some A, which is guaranteed by $y \in \hat{X}$ and Condition A.2. **Only if:** Suppose c_0 satisfies counterfactual WARP and $\tilde{c}(h)$ fails counterfactual WARP for some $h \in \mathcal{A}^{<\mathrm{N}}$. So there exist $x, y \in X$ and $A, B \in \mathcal{A}$ such that $\{x, y\} \subseteq A \cap B$, $\tilde{c}(h)(A) = x$, and $\tilde{c}(h)(B) = y$. Suppose WLOG $\tilde{c}(h)(\{x, y\}) = x$. We will separately demonstrate $u(x) > u(y)$, $u(y) > u(c_0(B))$, $c_0(B) \succ_0 x$, and $x \succ_0 y$, which completes the proof. Note that $c_0(\{x, y\}) = x$; otherwise, $c_0(\{x, y\}) = y$ implies $y \in \Gamma(\{x, y\}) \subseteq \tilde{\Gamma}(h)(\{x, y\})$, so $\tilde{c}(h)(\{x, y\}) = x$ means $u(x) > u(y)$ and $x \in \boldsymbol{c}(h)$, but then $\tilde{c}(h)(B) = y$ is a contradiction. This means $x \succ_0 y$. Also, since c_0 satisfies WARP and $x \in B$, $c_0(B) \neq y$, so the only way we have $\tilde{c}(h)(B) = y$ is $u(y) > u(c_0(B))$ and $y \in \boldsymbol{c}(h)$, which also means $c_0(B) \neq x$. Because c_0 satisfies WARP, $c_0(B) \succ_0 x$ and $c_0(B) \succ_0 y$. Finally, $y \in \boldsymbol{c}(h)$ and $\tilde{c}(h)(\{x, y\}) = x$ imply $u(x) > u(y)$. $\square$