Choice and Attention across Time

Xi Zhi "RC" Lim*

This version: 2026/2/2 BRIC
Most recent public version: http://s.xzlim.com/sequence

Abstract

Dynamic Limited Consideration (DLC) provides a theoretical foundation that captures how bounded rationality "mistakes" are corrected over time. When a future choice contradicts a past choice from the same menu, this observed correction provides evidence for bounded rationality and uniquely identifies preferences. Conventional, static WARP violations are sufficient but not necessary for limited consideration, whereas dynamic WARP violations are both sufficient and necessary. A stochastic extension uniquely identifies the distribution of preferences in a heterogeneous population where agents differ in both preference and attention. Well-known attention structures such as *attention filters* and *shortlisting* are self-preserving under DLC.

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

*Antai College of Economics and Management, 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 RUD 2023, Asia-Pacific ESA 2023, AMES 2022, MEET 2022, CMES 2021, VEAEBES2021, Columbia, NYU Shanghai, Peking University, SHUFE, and SWUFE for feedback. Part of this work appears in the 3rd chapter of my PhD dissertation at Columbia University (2020).

Table of content will be removed in final production.

Contents

1 Introduction **4**

2 Primitive **8**

3 Model **9**

3.1 Dynamic Limited Consideration 9
3.2 Identification of Parameters 12
3.3 Static and Dynamic Violations 15
3.4 Stability and Narrow Bracketing 17
3.5 Characterization 18

4 Stochastic Choice **19**

4.1 Identification of Heterogeneous Preferences 20
4.2 Identification Algorithm 21

5 Attention Structures **23**

5.1 Attention Filter 23
5.2 Shortlisting 25
5.3 Dominated Categories 26
5.4 Characterizing Compatibility 27

6 Related Literature **29**

7 Conclusion **30**

A Appendix: Proofs **31**

A.1 Proof of Theorem 1 32
A.2 Proof of Theorem 2 32
A.3 Proof of Theorem 3 35
A.4 Proof of Theorem 4 38

References **39**

B Online Appendix: Omitted Proofs and Results **43**
B.1 Omitted Proofs 43
B.1.1 Proofs of Conditions 43
B.1.2 Proof of Proposition 1 44
B.1.3 Proof of Proposition 2 45
B.1.4 Proof of Proposition 3 45
B.1.5 Proof of Proposition 4 45
B.1.6 Proof of Proposition 5 47
B.1.7 Proof of Corollary 1 47
B.1.8 Proof of Proposition 6 47
B.1.9 Proof of Proposition 7 47
B.1.10 Proof of Proposition 8 48
B.1.11 Proof of Proposition 9 48
B.2 Omitted Examples and Results 48
B.2.1 Other Examples 48
B.2.2 Other Results 49

Table of content will be removed in final production.

1 Introduction

Theories of limited consideration in economics and marketing, where a boundedly rational decision maker (DM) exhibits contradictory choices because they fail to consider every alternative despite having a stable underlying preference, pose three broad challenges rooted in the unobservability of mental consideration.[1]

First, essential for welfare analysis, the identification of preference parameters often requires the analyst to impose specific assumptions on menu-dependent consideration sets, which effectively restrict the types of "mistakes" that a DM can exhibit. Second, when it comes to testing the theories, the model specification that a DM considers x in menu A but not in menu B is difficult to test—once the DM faces A and chooses x, the DM may inevitably consider x when facing B simply due to recent experience. Third, the manifestation of limited consideration in the long-run is unclear: should a DM narrowly-bracket every problem and therefore repeat her "mistakes," or should she learn from her "mistakes" and gravitate towards rational behavior?

However, the correction of "mistakes," if observed, present an opportunity to address these questions. This paper extends the conventional static framework to a dynamic setting, where the analyst observes the *evolution of choices*, and axiomatically characterizes a model of Dynamic Limited Consideration (DLC). The main feature of the model is the retention of past choices in future consideration sets. Intuitively, even though a DM may not consider all options in all menus, she should at least be able to determine whether a newly-considered option is better than a previously chosen one. Thus, DLC provides a theoretical foundation for the correction of bounded rationality "mistakes," which in turn allows the analyst to uniquely identify preferences and test the hypothesis of limited consideration dynamically.

In a DLC, the DM has a utility function u and a subjective attention function Γ. The set $\Gamma(A)$, possibly a strict subset of the menu A, captures what she would normally consider when she encounters A. Moreover, the DM's past experiences h, a sequence of previously encountered menus, endogenously determine a collection of previously chosen alternatives that will continue to be considered, denoted by

[1] 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).

$\boldsymbol{c}(h)$. The DM's decision solves

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

The set $\boldsymbol{c}(h)$ comes from the same procedure, just at an earlier time. If a DM is "standard" and has full attention $\Gamma(A) = A$, then $\Gamma(A) \cup (\boldsymbol{c}(h) \cap A) = A$, and choices are independent of h as in static frameworks:

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

Under DLC, observed time-variation in choice implies an expansion of the DM's consideration set, which leads to choices that are more aligned with genuine preference. Therefore, a latter choice that contradicts an earlier choice is both an elicitation of genuine preference and an evidence that the past choice was made under limited consideration.

For example, suppose you have failed to consider Penang (in Malaysia) as a vacation destination even though you can afford it, and an analyst who observes your choice of Hawaii is at the brink of jumping to conclusion that you prefer Hawaii over Penang by the theories of revealed preference. However, when you later attend a conference in Southeast Asia and choose Penang for a drop-by vacation, since you must have then considered Penang, your future decisions of whether to return to Penang should more informatively convey your true preference between Penang and other destinations. This intuition applies broadly: A consumer who reads physical books might prefer e-readers, but one who left e-readers for physical books probably prefers physical books. A firm that advertises on Google may or may not have considered advertising on TikTok, but one that has switched from TikTok to Google probably prefers Google. A colleague who has not begun to review papers could be a remarkable reviewer who is simply not known, but one who is not re-invited may not be the best reviewer.

Identifying Preference The evolution of choices under DLC has a temporal structure that uniquely identifies preference even in the absence of additional assumptions on menu-dependent consideration sets $\Gamma(\cdot)$ (Proposition 1). To see this, suppose a DM is predisposed to violate WARP by choosing x from $\{x, y\}$ and y from

$\{x, y, z\}$. If the "problematic" choice sets are presented one after another, then the DM is bound to consider both x and y in the latter problem regardless of the order, either due to default attention $\Gamma(A)$ or due to the experience from the earlier problem $\boldsymbol{c}(h)$. Figure 1.1 provides an illustration. Thus, the choice from the latter problem always reflects the DM's genuine preference (formally, Proposition 2). Intuitively, an experienced DM has considered more options and their long-run choices, which are better aligned with their genuine preference, provide identification for preference parameters. Section 4 turns to stochastic choice data and shows the distribution of preferences can be identified from a heterogeneous population where agents differ in both preference and attention parameters.

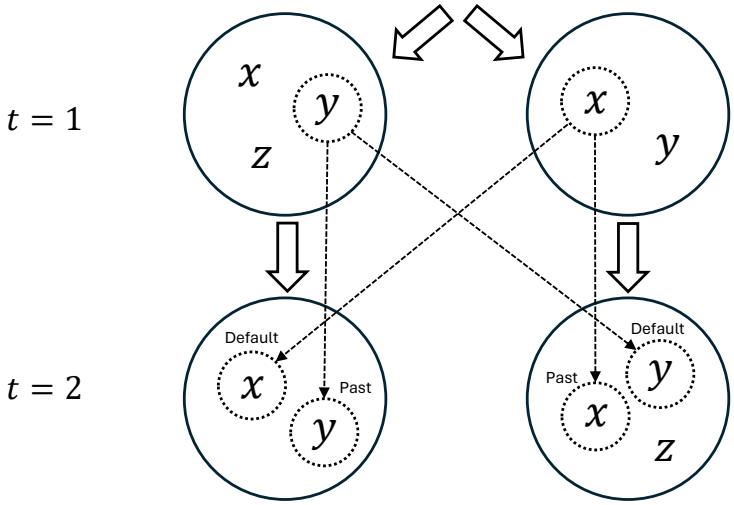


Figure 1.1: Big circles represent choice sets, small circles represent (inferred) consideration. Second period consideration sets are bound to include both x and y if they constitute a WARP violation in the first period. Subsection 3.2 provides details.

Counterfactual vs Realized "Mistakes" The dynamic framework allows for two forms of rationality (WARP) violations, static and dynamic. *Counterfactual violations* (static) occur when, given a history h, a DM will choose x from $\{x, y\}$ but y from $\{x, y, z\}$. Differently, *realized violations* (dynamic) occur when, in a single sequence of problems, the DM chooses x over y in the past but y over x in the future, an observation only possible in a dynamic setting. It turns out that the absence of

counterfactual violations does not in general imply consistency with full consideration (Example 2). However, under DLC, a DM commits realized WARP violations if and only if full consideration is ruled out (Theorem 1). Therefore, realized WARP violations are robust indicators for limited consideration even though they cannot be detected from one-shot choices. DLC also links history-dependent decisions and realized WARP violations, even though these behavioral properties do not generally imply one another (Theorem 1).

Axiomatic Characterization Because attention and memory are not observed, Theorem 2 axiomatically characterizes the observed evolution of choice behavior that is consistent with DLC. *Weak Stability* (Axiom 1) imposes restriction on behavior over time. It generalizes WARP by allowing 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 axiom captures the behavioral signature of attention and memory. First, revealed preference is defined when observed choices suggest that the DM has considered 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 would have been chosen. *Default Attention* (Axiom 3) then posits that if x is revealed preferred to y, then y is never chosen from a choice set where x would have been chosen.

Self-preserving Attention Structures Static models in the limited consideration literature offer different restrictions on attention functions Γ, which are unrestricted in DLC. Masatlioglu et al. (2012) study, in a static setting, consideration sets that remain the same when unconsidered alternatives are removed. It turns out that a DM who aligns with this attention structure will always maintain this structure even if attention evolves with experiences (Proposition 6). The same holds for Manzini and Mariotti (2007)'s shortlisting and Manzini and Mariotti (2012)'s categorization, where evolving attention is manifested as revised shortlists and as recategorizations (Proposition 8, Proposition 9). These observations suggest that an inattentive DM preserves an intrinsic system of (in)attention even if experiences alter the exact consideration sets she uses at different times. Theorem 4 provides a complete characterization of compatible static models.

Section 2 introduces the primitive. Section 3 introduces DLC, proves identification, relates different static and dynamic rationality violations, and offers an axiomatic characterization. Section 4 extends the framework to stochastic choice and focuses on identification of heterogeneous preference parameters in population data. Section 5 incorporates well-known attention structures in the literature into DLC. Section 6 discusses related literature and Section 7 concludes. Appendix A contains main proofs. Online Appendix B contains omitted proofs and results.

2 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.[2] Given a positive integer $N \geq 2$, let X^N be the set of all sequences of alternatives of length N and let $\mathcal{A}^N$ be the set of all sequences of choice sets of length N. The primitive is a choice function that assigns to each sequence of choice sets a sequence of choices,

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

where for every sequence of choice sets $(A_n) \in \mathcal{A}^N$ and any natural number $k \leq N$, the choice $c((A_n))_k$ is an element of the choice set $(A_n)_k$. I sometimes write A_k instead of $(A_n)_k$. A stochastic version of this primitive is considered in Section 4.

I assume that two sequences of choice sets that are identical up to a certain point yield the same choices up to that point. Formally, *future independence* requires that for any $(A_n), (B_n) \in \mathcal{A}^N$, 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 is equivalent to a collection of history-dependent choice functions. Formally, let $\mathcal{A}^{<N}$ be the set of all sequences of choice sets with length less than N, including the empty sequence denoted by $\emptyset$. For each (historical) sequence of choice sets $h \in \mathcal{A}^{<N}$, denote by

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

[2]The exclusion of singleton choice sets is to avoid passive choices, since "choosing" something without a choice might not involve actual consideration. All results go through if I let $\mathcal{A}$ be the set of all finite and nonempty subsets of 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, $\left\{\tilde{c}(h) \,|h \in \mathcal{A}^{<N}\right\}$ is fully and uniquely pinned down. The notation (h, B) denotes a history that begins with history h and is followed by choice set B.

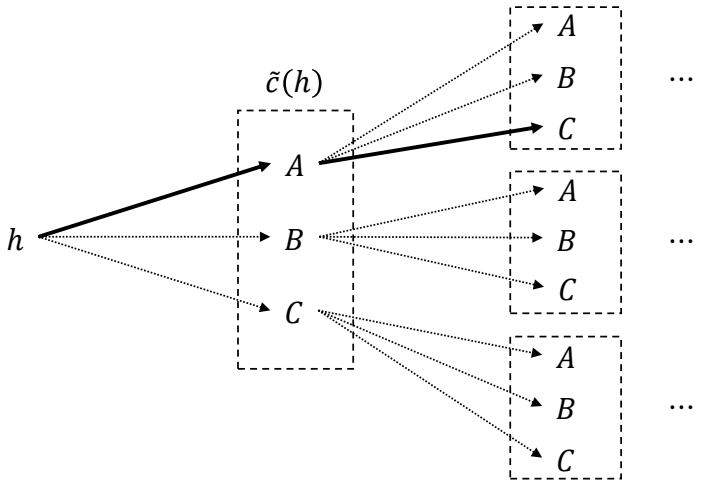


Figure 2.1: Each rectangle is made up of one-shot choice sets; therefore, a sequence of choice sets includes at most one choice set from each rectangle. The solid path represents (part of) a sequence of choice sets.

3 Model

3.1 Dynamic Limited Consideration

Denote by $\boldsymbol{c}(h)$ the set of previously chosen alternatives in the history $h \in \mathcal{A}^{<N}$, i.e., $\boldsymbol{c}((A_1, ..., A_K)) := \{\tilde{c}(\emptyset)(A_1), \tilde{c}((A_1))(A_2), \tilde{c}((A_1, A_2))(A_3), ...\}$.

Definition 1. c admits a Dynamic Limited Consideration (DLC) 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 $\Gamma(A) \subseteq A$, such that

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

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

DLC captures the following choice procedure: When a DM faces choice set A, she not only considers alternatives that she would always consider when she faces A, given by the attention set $\Gamma(A)$, but also the alternatives that she had chosen in the past. The former is *history-independent* and may capture what is intrinsically 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: even if a boundedly rational DM fails to consider every alternative in every menu, her past choices had already received her consideration and should remain so. Subsection 3.5 provides an axiomatic foundation for DLC.

Because the attention function Γ is fully general, DLC puts no restriction on choice behavior without history.[3] However, once there is a history, DLC governs what must happen next. For example, Masatlioglu et al. (2012) consider the use of an attention filter to explain why x is preferred to y but y is chosen from $\{x, y, z\}$; a possible specification is $\Gamma(\{x, y\}) = \{x, y\}$, $\Gamma(\{x, y, z\}) = \{y, z\}$. DLC accommodates this specification, but after the choice of x from $\{x, y\}$, DLC requires that x is considered when the DM subsequently faces $\{x, y, z\}$, i.e., $x \notin \Gamma(\{x, y, z\})$ but $x \in \tilde{\Gamma}(\tilde{c}(\{x, y\}))(\{x, y, z\})$. As a consequence, the choice $\tilde{c}(\tilde{c}(\{x, y\}))(\{x, y, z\}) = x$ differs from $\tilde{c}(\emptyset)(\{x, y, z\}) = y$ even though the underlying menu is the same. This illustration captures the central idea of DLC: that a DM's choice from a given menu can be history-dependent as past experiences may prompt consideration of a superior alternative that is typically excluded from the intrinsic consideration set for the menu.

An immediate implication is that experience can improve decisions, but only if there is sufficient variety. Better decisions ultimately come from an expanding $\boldsymbol{c}(h)$, but a longer h does not contribute to this goal if it mostly contains identical experiences. Notice that facing two identical choice sets one after another results in identical choices, so the second experience adds nothing to $\boldsymbol{c}(h)$. Similar arguments can be made if a choice set is repeated but not consecutively. Instead, the variety of past experiences is key to expanding $\boldsymbol{c}(h)$.

Moreover, the relevance of past experiences is more helpful than sheer quantity. To see this, suppose x is added to $\boldsymbol{c}(h)$ but there is no future choice set that contains x, then the increased consideration of x will not improve future decisions. Similar limitation holds if future choice sets that contain x also contain superior alterna-

[3] Every $\tilde{c}(\emptyset) : \mathcal{A} \to X$ can be explained by setting $\Gamma(A) = \{\tilde{c}(\emptyset)(A)\}$ for all A.

tives that are themselves considered. Therefore, the relevance of experience holds greater long-term value than sheer quantity of experience.

Equivalent formulations with imperfect memory The modeling feature that a DM may end up considering everything in every choice set may seem implausible, but it is behaviorally equivalent to related formulations of DLC where the DM, upon learning what is the best option in a choice set, only considers the best option itself, resulting in an "automatic" choice that may better align with our intuitions regarding cognitive constraints. Another behaviorally equivalent formulation involves a DM who compares the choice made during the most recent encounter with the same menu against alternatives that have received attention since then.

Formally, consider a general model where the DM (at least) keeps track of the best considered alternative for a menu. A third subjective parameter $\tilde{S} : \mathcal{A}^{<N} \times \mathcal{A} \to \mathcal{A}$, where $\tilde{S}(h, A)$ is any subset of $\Gamma(A) \cup (\boldsymbol{c}(h) \cap A)$ that contains its u-maximizer, replace $\tilde{\Gamma}(h)(A)$ in Definition 1. It can be shown that the new model is behaviorally equivalent to DLC despite having a more general formulation.[4] Moreover, automatic choice and rolling consideration are intuitive specializations of the new model and are also behaviorally equivalent to DLC.[5]

Other theories The evolution of choices can also be examined under other theories, for example learning and randomization, and a natural question is whether they provide alternative explanation for the same behavior under DLC. To address this formally, the axiomatic characterization in Subsection 3.5 provides the sufficient and necessary conditions for a DLC representation that many of these alternative theories violate. One example is learning, where a DM initially unsure of her preference discovers that she doesn't like something after consuming it, causing her to make different choices in the future (Ferreira and Gravel, 2024). This

[4]The new formulation is more general than DLC because $\tilde{S}(h, A) = \tilde{\Gamma}(h)(A)$ for all (h, A) gives DLC as a special case.

[5]**Special Case 1**: automatic choice after the best alternative $x^*(A) := \arg\max_{x \in A} u(x)$ is found, i.e., $\tilde{S}(h, A) = \{x^*(A)\}$ if $x^*(A) \in \tilde{\Gamma}(h')(A)$ for some truncation h' of h and $\tilde{S}(h, A) = \tilde{\Gamma}(h)(A)$ otherwise. **Special Case 2**: automatic choice after the DM has considered all options in A, $\tilde{S}(h, A) = \{x^*(A)\}$ if $\tilde{\Gamma}(h')(A) = A$ for some truncation h' of h and $\tilde{S}(h, A) = \tilde{\Gamma}(h)(A)$ otherwise. **Special Case 3** (rolling consideration): given (h, A), if A is in h, then let y^* be the most recent choice from A and let $T \subseteq \boldsymbol{c}(h)$ be the choices made between that experience and now, then $\tilde{S}(h, A) = \{y\} \cup (T \cap A)$; otherwise, let $\tilde{S}(h, A) = \tilde{\Gamma}(h)(A) = \Gamma(A) \cup (\boldsymbol{c}(h) \cap A)$.

behavior violates 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 seeking variety, facing random utility, or deliberately randomizing (Agranov and Ortoleva, 2017; Cerreia-Vioglio et al., 2019). These behaviors violate Axiom 2, which requires the same alternative to be chosen when the same choice set is repeated consecutively.

3.2 Identification of Parameters

A key takeaway from this study is that standard economic 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, is constrained to deterministic choice, and do not make assumptions about intrinsic attention sets Γ.

Identification of Preference

The first observation is that even for fully rational behavior, a general framework of limited attention does not permit the identification of preferences.

Problem 1. Consider $X = \{x, y, z\}$, $u(x) > u(y) > u(z)$, $\Gamma(A) = A$ for all A. The resulting choice function $c_{(u,\Gamma)}$ is the same as $c_{(u',\Gamma')}$ where $u(z) > u(x) > u(y)$ and $\Gamma(A) = A \setminus \{z\}$ for all A.

Given c, denote by $\hat{X}$ the set of alternatives that are *ever-chosen*, i.e., $\hat{X} := \{x \in X : c((A_n))_i = x \text{ for some } (A_n) \text{ and } i\}$. There is at most one *never-chosen* alternative, i.e., $|X \setminus \hat{X}| \leq 1$, and there is exactly one never-chosen alternative for fully rational behavior.[6] It can be shown that preferences are identified except for the one never-chosen alternative if it exists.

Proposition 1. *If c admits DLC 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 $u_2(x) > u_2(y)$.*

The key intuition for preference identification is captured by a simple design that compels the correction of "mistakes."

[6]There is at most one *never-chosen* alternative because the set of all binary choice sets leaves at most one alternative never chosen. The formal definition of fully rational behavior can be found in Subsection 3.4.

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 1.1, indicating 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, under DLC, both x and y will receive consideration if the two choice problems are encountered sequentially. For example in the left sequence of problems in Figure 1.1, x receives consideration due to default attention inferred from first period's behavior and y receives attention due to a recent experience of choosing y. Second period's choice thus reveals preference.

More fundamentally, regardless of the order of choice problems, choices converge to the same alternative (either x or y) because they maximize the same preference. This *convergence* property is generalized to arbitrary choice sets and histories in Proposition 2, and it indicates to the analyst that a DM has stable preference despite the influence of limited consideration.

Proposition 2. *Suppose c admits a DLC 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 $u(y) > u(x)$.*

Proposition 2 rules out the remaining two choice patterns, *sticky choice* and *narrow-bracketing*.[7]

In more general datasets, the *switch* from choosing y to choosing x over y reveals preference. For any distinct $x, y \in X$, 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 $i < j$, and let $x\mathbb{S}y$ if $x\mathbb{S}_{(A_n)}y$ for some $(A_n) \in \mathcal{A}^N$.

Proposition 3. *Suppose c admits a DLC 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.*[8]
3. *If X is finite and N is sufficiently large, then there exists (A_n) such that $\mathbb{S}_{(A_n)}$ on $\hat{X}$ is a strict total order.*[9]

[7] **Sticky choice**: $\tilde{c}((h, A))(B) = x$ and $\tilde{c}((h, B))(A) = y$, which may be explained by habit formation. **Narrow-bracketing**: $\tilde{c}((h, A))(B) = y$ and $\tilde{c}((h, B))(A) = x$, where the DM's choice is history-independent.

[8] 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 $y\mathbb{S}x$).

[9] It is impossible to construct a c-independent sequence that fully identifies preferences. Online

The first part of Proposition 3 has already been discussed: observing a switch from y to x indicates that x is preferred. The second part asserts that such evidence is always observable provided that x and y are ever-chosen, and evidence is never contradictory (i.e., if it is revealed in one sequence that x is preferred to y, then no sequence can reveal the opposite). Finally, the third part shows that if the set of alternatives is finite, then it is possible to collect all these evidences using a single finite sequence of choice sets; intuitively, this is achieved by asking the right questions in the right order, and the details are relegated to the proof.

Identification of Attention

Because Γ can differ by including or excluding inferior alternatives without changing model prediction, the set of possible model specifications is characterized by a maximal set. Given c, let $\Gamma^{+}:\mathcal{A}\to 2^{X}\setminus\{\emptyset\}$ be

$$\Gamma^{+}(A):=\left\{x\in A:\tilde{c}(\emptyset)(A)\,\mathbb{S}x \text{ or } x=\tilde{c}(\emptyset)(A) \text{ or } x\notin\hat{X}\right\}. \tag{3.1}$$

Proposition 4. *If c admits a DLC representation, then it also admits a DLC representation with attention function Γ if and only if $\tilde{c}(\emptyset)(A)\in\Gamma(A)\subseteq\Gamma^{+}(A)$ for all A.*

In practice, 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 (formally in Online Appendix B Corollary 4).

Note that the maximal set $\Gamma^{+}(\cdot)$, which is closest to the full attention benchmark where everything is considered, will improve Proposition 1 to the point where preference is unique even for the never-chosen alternative, which I state now and use later in Section 4 (stochastic choice).

Proposition 5. *If c admits DLC representations $(u_1,\Gamma^{+}),(u_2,\Gamma^{+})$, then $u_1(x)>u_1(y)$ if and only if $u_2(x)>u_2(y)$.*

Appendix B Example 6 provides a counterexample, Corollary 3 outlines the best a c-independent sequence can do.

3.3 Static and Dynamic Violations

Next, I argue that our existing tests for limited attention in static settings can fall short of detecting all instances of limited attention. The key idea lies in the distinction between two different manifestations of WARP violations that the primitive c allows for: counterfactual (static) and realized (dynamic).

Recall that the one-shot choice function $\tilde{c}(h) : \mathcal{A} \to X$ captures what the DM will choose, given a history h, from each upcoming choice set. When $\tilde{c}(h)$ violates WARP, for instance

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

I call it a *counterfactual* violation.[10] This corresponds to the WARP-violating choices within the left rectangular of Figure 3.1. A different kind of WARP violation is a stream of decisions where a past choice and a future choice form a violation, for instance

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

which I call a *realized* violation. This corresponds to the WARP-violating choices along the solid path in Figure 3.1.

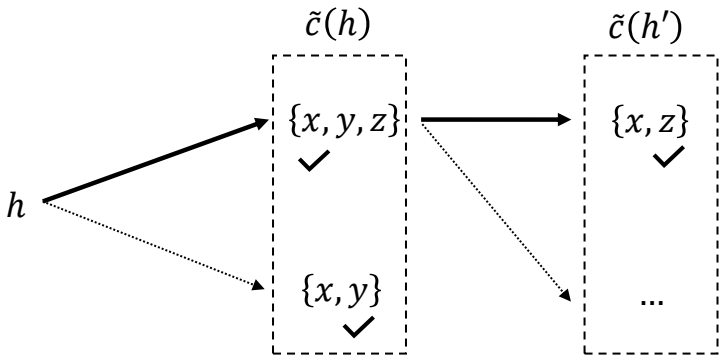


Figure 3.1: This figure follows the structure of Figure 2.1 but provides choices to illustrate two kinds of WARP violations: *counterfactual* (within the left rectangular) and *realized* (along the solid path).

[10] In experiments, 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."

In general, a c can be violation-free (e.g., standard, history-independent utility maximization), only exhibits realized WARP violations (e.g., models of habit formation), only exhibits counterfactual WARP violations (e.g., context effect with narrow-bracketing), or exhibits both kinds of violations.

Perhaps unexpectedly at first, the absence of counterfactual WARP violations does not rule out limited attention even though they have conventionally served as the motivation for theories of limited attention. The DM in Example 1 satisfies WARP initially, but violates WARP in the future. Interestingly, even though WARP compliance has worsened, decision quality (in terms of maximizing genuine preference) has improved.

Example 1. 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 $\tilde{c}(\emptyset)$ can be explained by the maximization of preference ranking $z \succ_0 x \succ_0 y \succ_0 z'$, hence it satisfies WARP. But after 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.

In the extreme, Example 2 shows that a DM who *never* commits counterfactual WARP violations (at any history) is non-trivially affected by limited attention because choices are history-dependent, which must rule out full attention.

Example 2. 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 $\tilde{c}(\emptyset)(\{x, y\}) = x$ 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, $\tilde{c}(\emptyset)$ can be explained by the maximization of preference ranking $x \succ_0 y \succ_0 z$, and it can be shown that the DM can never commit counterfactual WARP violations (for any history).[11]

A sufficient and necessary condition for counterfactual WARP violations is provided in Online Appendix B Corollary 5.

[11] 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.

3.4 Stability and Narrow Bracketing

To further investigate the relationship between counterfactual and realized WARP violations, note that DLC admits, as special cases, behaviors that are consistent with full attention, i.e., when $\Gamma(A) = A$ for all A. In that case, every decision is independent of history and no WARP violations will be detected,

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

Those c's are said to admit a *standard utility representation*.

The next result shows that a non-trivial DLC behavior (i.e., one that does not admit a standard utility representation) must exhibit realized WARP violations and history-dependent behavior. Formally, Definition 2 introduces two relatively demanding properties of choice behavior that will later be suitably relaxed into Axiom 1 and Axiom 2 in the axiomatic framework for DLC. Stability captures WARP within every sequence of choices (i.e., the absence of realized WARP violations), and Narrow Bracketing captures choices that do not vary with past experiences.

Definition 2.

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

2. Narrow Bracketing: for every $h \in \mathcal{A}^{<N}$ and $A, B \in \mathcal{A}$, $\tilde{c}((h, B))(A) = \tilde{c}(h)(A)$.

Stability and Narrow Bracketing are in general non-nested and could be interpreted as different aspects of "rational" behaviors, but Theorem 1 shows that, under DLC, they imply one another and their failures are necessary for limited attention.

Theorem 1. *Suppose $N \geq 3$ and c admits a DLC representation. The following are equivalent:*

1. *c satisfies Stability.*

2. *c satisfies Narrow Bracketing.*

3. *c admits a standard utility representation.*

To summarize the relationship between counterfactual and realized WARP violations: the absence of counterfactual violations neither rules out limited attention nor rules out realized violations (Example 2), but the absence of realized violations implies consistency with standard utility representation (Theorem 1) and therefore rules out counterfactual violations.

It may be of interest in itself that Narrow Bracketing constrains dynamic choices in a way that makes them no richer than static choices. This simplifying assumption appears to be particularly unappealing in studies of boundedly rationality where agents make "mistakes" but might be able to correct them over time.

3.5 Characterization

Choice behavior admits a DLC representation if and only if they satisfy three axioms that are silent on one-shot choices (i.e., if $N = 1$, then the axioms are trivial) but puts restriction on how choices evolve over time. The first two axioms generalize Stability and Narrow Bracketing, respectively.

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

Axiom 1 imposes a version of the well-known weak axiom of revealed preference (WARP) with two key differences.[12] First, it is imposed within a sequence of choice problems $(A_n) \in \mathcal{A}^N$, so there is no restriction on how choices may differ (or violate WARP) across sequences.[13] Second, instead of demanding full compliance with WARP, it limits the instances of WARP violations. A conforming DM may switch from choosing x to choosing y over x, but she does not go back and forth between them.

Axiom 2 (Past Dependence). *For any* $h \in \mathcal{A}^{<N-1}$ *and* $A, B \in \mathcal{A}$, $\tilde{c}((h, B))(A) \in \{\tilde{c}(h)(A), \tilde{c}(h)(B)\}$.

Axiom 2 captures a particular kind of history-dependent behaviors. Consider first the "standard" behavior where experiencing choice set B does not affect the

[12] 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."

[13] 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.

choice in A, then $\tilde{c}\left((h,B)\right)(A) = \tilde{c}\left(h\right)(A)$. The axiom allows for one type of departure where the choice from A is exactly the choice from B.

The next and last axiom embodies the behavioral signature of attention. Recall that $\tilde{c}\left(\emptyset\right)(A)$ 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 after y was chosen in the past or when x is chosen from a choice set where y would have been chosen in the absence of observable history. Formally, for distinct x and y, let xPy if at least one of the following is true for some $(A_n) \in \mathcal{A}^N$: (1) $c\left((A_n)\right)_i = y$ and $c\left((A_n)\right)_j = x$ where $y \in A_j$ such that $i < j$ or (2) $c\left((A_n)\right)_j = x$ such that $\tilde{c}\left(\emptyset\right)(A_j) = y$.

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

Intuitively, Axiom 3 posits that if the original choice of A, $\tilde{c}\left(\emptyset\right)(A)$, is already better than y, then y is never going to be chosen from A regardless of past experiences. This behavior suggests that every choice set contains alternatives that always receives consideration, for instant the most salient alternatives. For example, *pizza* is always considered for football night, whereas the consideration of *satay* depends on past experiences.

Theorem 2. *Suppose $N \geq 4$. The choice function c satisfies Axioms 1, 2, and 3 if and only if it admits a DLC representation.*

4 Stochastic Choice

Consider a stochastic extension of DLC where a population of DMs with unknown (heterogeneous) preference and attention parameters produce aggregate data. That is, instead of observing a choice function c that assigns a choice sequence to each sequence of choice sets, the analyst observes a *distribution* of choice sequences for each sequence of choice sets. Similar to population data in static settings, the interpretation is that an analyst can observe a DM's sequence of choices from one sequence of choice sets, but she does not know how this DM will behave when facing a different sequence of choice sets.

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. Given a finite $N \geq 2$, the primitive is

$\rho : \mathcal{A}^N \to \Delta\left(X^N\right)$, where for each (x_n) in the support of $\rho(A_n)$, we have $x_k \in A_k$ for all k. Per convention, the probability that (x_n) is chosen from (A_n) is denoted by $\rho(A_n)(x_n)$.

Let $\boldsymbol{u}$ be the set of all utility functions and let $\boldsymbol{\Gamma}$ be the set of all attention functions. Given $(u, \Gamma) \in \boldsymbol{u} \times \boldsymbol{\Gamma}$, denote by $c_{(u,\Gamma)} : \mathcal{A}^N \to X^N$ the choice function induced by DLC with parameter (u, Γ) (Definition 1). A (finite) population of DLC decision makers is captured using a discrete distribution of utility function - attention function pairs that is denoted by $\mu \in \Delta(\boldsymbol{u} \times \boldsymbol{\Gamma})$. Given μ, the fraction of DMs with parameter (u, Γ) is denoted by $\mu(u, \Gamma)$.

Definition 3. ρ is consistent with a DLC population $\mu \in \Delta(\boldsymbol{u} \times \boldsymbol{\Gamma})$ if for all $(A_n) \in \mathcal{A}^N$ and $(x_n) \in X^N$,

$$\rho(A_n)(x_n) = \sum_{(u,\Gamma): c_{(u,\Gamma)}(A_n) = (x_n)} \mu(u, \Gamma).$$

4.1 Identification of Heterogeneous Preferences

The next result, Theorem 3, asserts that a dynamic framework with sufficiently many observations can help identify preferences even if a population of DMs (1) experience limited consideration, (2) have heterogeneous preference and attention parameters, and (3) each DM only faces one sequence of choice sets.

It is well-known that utility functions are unique up to strictly increasing transformation. For any utility functions u and u', write $u' \sim u$ if u' and u rank alternatives in the same way and $u' \not\sim u$ otherwise.[14] Given a DLC population μ, for any ranking of alternatives consistent with u, the fraction of the population that shares the same ranking (but not necessarily the same utility function) and has attention function Γ is $\mu_{\sim u,\Gamma} := \sum_{u' \sim u} \mu(u', \Gamma)$, and the fraction that has the same ranking regardless of attention is $\mu_{\sim u} := \sum_{\Gamma \in \boldsymbol{\Gamma}} \mu_{\sim u,\Gamma}$.

The main challenge in identifying preferences is that different parameters (u, Γ) and (u', Γ'), where u, u' rank alternatives differently, can induce the same sequence of choices in some sequences of choices sets. Because attention functions are not unique even in the single-DM setting, I break ties by focusing on maximal attention functions Γ^+ (see Equation 3.1) since they are closest to the full consideration

[14] Formally, $u' \sim u$ if $u'(x) > u'(y) \Leftrightarrow u(x) > u(y)$ for all $x, y \in X$.

benchmark.[15] Formally, let $\boldsymbol{P}$ be the subset of $\boldsymbol{u} \times \boldsymbol{\Gamma}$ such that $(u, \Gamma) \in \boldsymbol{P}$ if and only if Γ satisfies Equation 3.1 (i.e., is maximal) for $c_{(u,\Gamma)}$. It follows from Proposition 4 that restricting attention to $\Delta(\boldsymbol{P})$ does not alter the set of admissible ρ's.[16]

Theorem 3. *Suppose X is finite. For sufficiently large $N < \infty$, if ρ is consistent with DLC populations $\mu, \mu' \in \Delta(\boldsymbol{P})$, then $\mu_{\sim u,\Gamma} = \mu'_{\sim u,\Gamma}$ for any $u \in \boldsymbol{u}$ and $\Gamma \in \boldsymbol{\Gamma}$.*

The proof is based on the observation that, given a target preference ranking, the "unluckiest" DMs who only pay attention to the worst alternatives (initially) are identified because they exhibit a one-of-a-kind choice pattern. Then, recursively, the second "unluckiest" DMs can be identified, and so on. Subsection 4.2 provides an illustration.

Corollary 1. *Suppose X is finite. For sufficiently large $N < \infty$, if ρ is consistent with DLC populations $\mu, \mu' \in \Delta(\boldsymbol{P})$, then $\mu_{\sim u} = \mu'_{\sim u}$ for any $u \in \boldsymbol{u}$.*

4.2 Identification Algorithm

To illustrate Theorem 3, this section focuses on $X = \{x, y, z\}$ and introduces an algorithm that uniquely identifies $\mu_{\sim \bar{u},\Gamma}$ for $\bar{u}(x) > \bar{u}(y) > \bar{u}(z)$.

Suppose ρ is consistent with a DLC population $\mu \in \Delta(\boldsymbol{P})$. The algorithm begins with the "smallest" attention function

$$\Gamma_1(\{x, y, z\}) = \{z\}, \Gamma_1(\{y, z\}) = \{z\}, \Gamma_1(\{x, y\}) = \{y\}, \Gamma_1(\{x, z\}) = \{x, z\}$$

where the DM of type $(\bar{u}, \Gamma_1)$ pays attention only to the worst alternatives initially (except for $\Gamma_1(\{x, z\}) = \{x, z\}$).[17] Consider the sequence of problems

$$(A_n) = (\{x, y, z\}, \{y, z\}, \{x, y\}, \{x, z\}, \{x, y\}, \{y, z\}).$$

[15] This restriction improves identification only if the population includes a DM (u, Γ) whose preference is not fully identified even in the single-DM setting. due to never-chosen alternatives.

[16] Suppose ρ is consistent with a DLC population $\mu \in \Delta(\boldsymbol{u} \times \boldsymbol{\Gamma})$. For every $(u, \Gamma) \in \text{supp}(\mu)$ such that in $c_{(u,\Gamma)}$, every alternative is ever-chosen ($\hat{X} = X$), replace (u, Γ) with (u, Γ^+) where Γ^+ satisfies Equation 3.1 for $c_{(u,\Gamma)}$. For every $(u, \Gamma) \in \text{supp}(\mu)$ such that in $c_{(u,\Gamma)}$, an alternatives is never-chosen ($\hat{X} = X \backslash \{z\}$), let u' has the same ordinal ranking as u except that z has the lowest utility, and replace (u, Γ) with (u', Γ^+) where Γ^+ satisfies Equation 3.1 for $c_{(u,\Gamma)}$. Denote the resulting population as $\mu' \in \Delta(\boldsymbol{P})$. Then, ρ is consistent with a DLC population μ'.

[17] It can be shown that (u, Γ), where $\Gamma(\{x, y, z\}) = \{z\}, \Gamma(\{y, z\}) = \{z\}, \Gamma(\{x, y\}) = \{y\}, \Gamma(\{x, z\}) = \{z\}$ and $u \sim \bar{u}$, is not in $\boldsymbol{P}$ because x is never-chosen in $c_{(u,\Gamma)}$, which requires $x \in \Gamma(\{x, y, z\})$ if Γ satisfies Equation 3.1 for $c_{(u,\Gamma)}$, a contradiction.

Suppose the analyst observes a fraction α of DMs choose $(x_n) = (z, z, y, x, x, y)$, that is, $\rho(A_n)(x_n) = \alpha$, then it must be that the types $(u, \Gamma_1) \in \boldsymbol{P}$ where $u \sim \bar{u}$ occur with aggregate probability α; that is,

$$\sum_{u \sim \bar{u}} \mu(u, \Gamma_1) = \alpha.$$

To see this, observe that final two choices require $u(x) > u(y)$ and $u(y) > u(z)$, respectively, because all of x, y, z must now be considered, so $u \not\sim \bar{u}$ cannot produce the observed behavior. To see that Γ_1 must be the attention function, I argue that only maximal attention function that can explain the first three choices in (x_n) is Γ_1. Given $u \sim \bar{u}$, suppose for contradiction (1) any of x, y is added into $\Gamma_1(\{x, y, z\}) = \{z\}$, then the first choice will be the addition instead of z; (2) if y is added to $\Gamma_1(\{y, z\}) = \{z\}$, the second choice will be y instead of z; (3) if x is added to $\Gamma_1(\{x, y\}) = \{y\}$, the third choice will be x instead of y. Hence it must be Γ_1. Intuitively, identification is possible because these "unluckiest" DMs who only pay attention to the worst alternatives (except for $\Gamma_1(\{x, z\}) = \{x, z\}$) exhibit one-of-a-kind behavior.

The sequence (A_n) is carefully constructed with two parts. In the first four positions, choice sets in $\mathcal{A}$ are organized so that the target DM gradually discovers better alternatives, and the last two positions are used to verify her preference.

Next, consider the "second unluckiest" DMs under $u \sim \bar{u}$, where

$$\Gamma_2(\{x, y, z\}) = \{z\}, \Gamma_2(\{y, z\}) = \{y, z\}, \Gamma_2(\{x, y\}) = \{y\}, \Gamma_2(\{x, z\}) = \{x, z\},$$

who choose $(y_n) = (z, y, y, x, x, y)$ from

$$(B_n) = (\{x, y, z\}, \{x, y\}, \{y, z\}, \{x, z\}, \{x, y\}, \{y, z\}).$$

It turns out that the only other types in $\boldsymbol{P}$ who choose (y_n) from (B_n) are $(u, \Gamma_1) \in \boldsymbol{P}$ such that $u \sim \bar{u}$.[18] So if $\rho(B_n)(y_n) = \beta$, then it must be that

$$\sum_{u \sim \bar{u}} \mu(u, \Gamma_2) = \beta - \alpha.$$

[18]Using a similar argument as before that the final two choices reveal $u \sim \bar{u}$ and the first three choices are inconsistent with any expansion of Γ_1 except for adding y to $\Gamma_1(\{y, z\})$, which yields Γ_2.

The process continues until $\sum_{u\sim\bar{u}}\mu\left(u,\Gamma_i\right)$ is pinned down for all Γ_i, recursively from the "unluckiest" DMs to the "luckiest" DMs.

5 Attention Structures

Because not all attention functions (Γ) are equally plausible, the literature has considered different attention structures (assumptions that identify a class of Γ's) in static settings. Most prominently, Masatlioglu et al. (2012) introduce *attention filters*, where the removal of an alternative that does not receive consideration from a menu does not alter the consideration set. 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 structural assumptions on Γ. Manzini and Mariotti (2012)'s *categorize-then-choose* (CTC) has similar features.

An immediate question is, if the accumulation of experience can alter attention, can it cause a DM to depart from a particular attention structure she once had?

To address this question formally, this section studies "vertical mergers" between DLC and static models. Results show that well-known static models are highly compatible with DLC in the sense that a DM who once had a CLA (resp. RSM, CTC) attention structure will always have a CLA (resp. RSM, CTC) attention structure even though consideration sets are necessarily evolving. In other words, these attention structures are self-preserving under DLC. Subsection 5.4 extends this observation by characterizing the universe of compatible static models and discusses its implications.

5.1 Attention Filter

Masatlioglu et al. (2012) introduce a form of limited consideration where the removal of unconsidered alternatives from a menu does not alter the DM's consideration set, as defined in Definition 4.

Definition 4. $\hat{\Gamma}$ is an *attention filter* if $\hat{\Gamma}\left(A\right)\subseteq A$ and $y\notin\hat{\Gamma}\left(A\right)$ implies $\hat{\Gamma}\left(A\backslash\left\{y\right\}\right)=\hat{\Gamma}\left(A\right)$.

In DLC, even though the accumulation of experience changes attention, a DM

who has an attention filter will always have an (possibly different) attention filter. Proposition 6 formalizes this statement and Example 3 illustrates.

Proposition 6. *If c admits a DLC representation (u, Γ) where Γ is an attention filter, then for any history $h \in \mathcal{A}^{<N}$, $\tilde{\Gamma}(h)$ is an attention filter and $\tilde{c}(h)$ is a CLA.*[19]

Example 3 (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. 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 DLC where Γ is an attention filter is to additionally impose CLA's original axioms on period zero static choices $\tilde{c}(\emptyset)$.[20] Differently, Axiom 4 introduces a behavioral manifestation of CLA in the evolution of choices: 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 $\tilde{c}(\emptyset)(T) = x$ and $\tilde{c}(\emptyset)(T \setminus \{y\}) \neq x$, then not $y \mathbb{S} x$.*

Proposition 7. *Suppose c admits a DLC representation. The following are equivalent:*

1. c satisfies Axiom 4.
2. c admits a DLC 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 when CLA infers x is preferred to y, the "direct evidence" where the DM switches from choosing y to choosing x over y can be found in choice

[19] In a CLA, $\hat{c}(A) = \arg\max_{x \in \hat{\Gamma}(A)} u(x)$ where $\hat{\Gamma}$ is an attention filter (Masatlioglu et al., 2012).
[20] Masatlioglu et al. (2012) proposes an axiom called *WARP with Limited Attention*.

sequences.[21] In return, CLA disciplines the permissible set of attention functions, which mitigates the issue of multiplicity and facilitates interpretations/applications.

5.2 Shortlisting

A different form of inattention is considered by Manzini and Mariotti (2007) where it is as if a DM first creates a shortlist using some rationale S and then makes a final decision using a different rationale that could be interpreted as genuine preference. The shortlist can therefore be viewed as an attention function with certain features, defined in Definition 5.

Definition 5. $\hat{\Gamma}$ is a *shortlist* if there exists an asymmetric binary relation S on X such that $\hat{\Gamma}(A) = \{x \in A | \text{ not } ySx \, \forall y \in A\}$.

It can be shown that a DM who uses a shortlist will always use some shortlists, even if accumulating experience changes the DM's behavior.

Proposition 8. *If c admits a DLC representation (u, Γ) such that Γ is a shortlist, then for any history $h \in \mathcal{A}^{<N}$, $\tilde{\Gamma}(h)$ is a shortlist and $\tilde{c}(h)$ is an RSM.*[22]

The merger between RSM and DLC is intuitive: The DM shortlists alternatives before making final decisions but revises her rationales as she accumulates experience. Upon gaining experience with an alternative, the original rationale S is revised so that nothing eliminates said alternative, thereby guaranteeing its consideration in the future and resulting in better decisions. Example 4 illustrates.

Example 4 (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 dSa and dSb. Suppose $u(a) > u(b) > u(d)$. During COVID-19, d is temporarily unavailable and a is chosen from $\{a, b\}$. After this experience ($h = (\{a, b\})$), the firm cancels dSa but maintains dSb; so $\tilde{\Gamma}(h)$ is still a shortlist even though $\tilde{\Gamma}(h) \neq \Gamma$.

[21] If $\tilde{c}(\emptyset)(T) = x$, $\tilde{c}(\emptyset)(T \setminus \{y\}) \neq x$, 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.

[22] In an RSM, $\hat{c}(A) = \max(\max(A, P_1), P_2)$ where P_1 and P_2 are asymmetric binary relations on X and $\max(A, S) = \{x \in A | \text{ not } ySx \, \forall y \in A\}$ (Manzini and Mariotti, 2007).

As the DM's decisions improve with experience, they also explain whether past choices were in fact influenced by shortlisting. If x was initially chosen over y, but future comparisons resolve in favor of y after y became chosen in an incidental problem, then it suggests to the analyst that the initial choice of x was driven by shortlisting instead of being the genuine reflection of preference.

5.3 Dominated Categories

Related to the intuition of shortlisting is when an entire category of options is removed from consideration when another category is present, introduced by Manzini and Mariotti (2012). The authors describe this process as coarse maximization. A set of alternatives R' is dominated (removed from consideration) when another set of alternatives R'' is available. Formally, $\max^s(A, S) := \{x \in A | \nexists R', R'' \subseteq A \text{ such that } R''SR' \text{ and } x \in R'\}$ where S is a subjective shading rationale. It imposes a particular structure on attention, defined in Definition 6.

Definition 6. $\hat{\Gamma}$ is a *coarse-max* if there exists an asymmetric binary relation S on $2^X \setminus \{\emptyset\}$ such that $\hat{\Gamma}(A) = \max^s(A, S)$.

Similar to before, Proposition 9 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 5 illustrates the idea.

Proposition 9. *If c admits a DLC representation (u, Γ) such that Γ is a coarse-max, then for any history $h \in \mathcal{A}^{<N}$, $\tilde{\Gamma}(h)$ is a coarse-max and $\tilde{c}(h)$ is a CTC.*[23]

Example 5 (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 exempted from

[23] In a CTC, $\hat{c}(A) \succ^* y$ for all $y \in \max^s(A, \succ^s) \setminus \{\hat{c}(A)\}$ where $\succ^s$ is an asymmetric binary relation on $2^X \setminus \{\emptyset\}$ and $\succ^*$ is an asymmetric and complete binary relation on X (Manzini and Mariotti, 2012).

shading. The other Mexican restaurants continue to be shaded by Italian restaurants. In the future, the consumer's genuine preference will determine whether she returns to this Mexican restaurant.

5.4 Characterizing Compatibility

Compatibility with DLC contributes to a robust framework where DLC identifies preferences and static, one-shot decision theoretic models provide insights to the intrinsic formation of (in)attention. I now characterize a sufficient and necessary property for compatibility.

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. A subset $\mathbb{C} \subseteq \mathbb{C}_{All}$, which may include some choice functions and exclude others, can be viewed as the collection of behaviors explained by a given model, or equivalently, those that satisfy a given set of axioms. Let $\mathbb{C}_{WARP}$ be the set of choice functions that satisfy WARP (i.e., $\hat{c} \in \mathbb{C}_{WARP}$ if and only if $\hat{c}(T) = \hat{c}(S)$ whenever $\hat{c}(S) \in T \subseteq S$). Per convention, $f(\mathcal{A}) := \{f(A) : A \in \mathcal{A}\}$.

Definition 7. $\mathbb{C}$ is *compatible with DLC* if for every $f \in \mathbb{C}$, there exists a DLC representation (u, Γ) such that $\tilde{c}(\emptyset) = f$ and $\tilde{c}(h) \in \mathbb{C}$ for all $h \in \mathcal{A}^{<\mathrm{N}}$.

Definition 7 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 DLC, 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 $f'', f''', f'''', ...$ subsequently. If not, then $\mathbb{C}$ and DLC 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 one of $g', g'', g'''...$ fails to belong to $\mathbb{C}$, resulting again in incompatibility.

Compatibility is therefore non-trivial.[24] CLA, RSM, and CTC are compatible with DLC due to a common feature called WARP-convex.

[24] Online Appendix B Example 7 provides a $\mathbb{C}$ that is not compatible with DLC.

Definition 8. Consider choice functions $f, g, \kappa \in \mathbb{C}_{All}$.

1. g is a κ-*cousin* of f if for some finite $T \subseteq f(\mathcal{A})$, $g(A) = \kappa(\{f(A)\} \cup [A \cap T])$ for all $A \in \mathcal{A}$.

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."

If $\mathbb{C}$ only contains WARP-conforming choice functions (even if $\mathbb{C} \neq \mathbb{C}_{WARP}$), then it is trivially WARP-convex; examples of such $\mathbb{C}$'s include expected utility, exponential discounting, and generalizations that preserve WARP. Therefore, The notion is meaningful for non-WARP models. Despite its non-triviality, prominent models related to limited consideration are WARP-convex, including CLA, RSM, CTC, *rationalization* (Cherepanov et al., 2013), and *overwhelming choice* (Lleras et al., 2017).

Theorem 4. *$\mathbb{C}$ is WARP-convex if and only if it is compatible with DLC.*

The "if" direction of Theorem 4 is intuitive: if a static, one-shot choice model is compatible with DLC, then its WARP violations are correctable. Intuitively, since DLC compels the gradual correction of WARP violations as experience accumulates, any compatible choice model must accommodate a DM who undergoes such corrections at every step.

But the opposite, which is less obvious, is also true: a choice model is compatible with DLC as long as WARP violations are correctable. This highlights the fact that DLC is silent on intrinsic attention structures and is focused on a parsimonious correction of WARP mistakes. If no WARP violations are present, then DLC accommodates behavior as-is; but if violations are present, then DLC religiously corrects them without introducing new or different types of WARP violations.[25]

[25]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 DLC, essentially because we

6 Related Literature

Closest to DLC is the choice theory literature that studies deterministic choice using consideration sets, even though this literature has (implicitly) focused on static settings. Masatlioglu et al. (2012)'s *choice with limited attention* (CLA) 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.[26] Although these studies consider static decisions, they complement DLC's silence on the intrinsic structure of attention sets; and in turn, DLC studies the dynamics of attention and leverages the evolution of choices to fully identify preferences without imposing additional assumptions on attention structures.

The literature has explored different ways to identify preference under the context of limited attention. Masatlioglu et al. (2012) partially identify preference by inferring attention. Caplin and Dean (2011) and Caplin et al. (2011) use tentative choices in a single choice problem to infer search processes. Manzini and Mariotti (2014) and Cattaneo et al. (2020) exploit the richness in stochastic choice to pin down preference; Kovach and Suleymanov (2023) additionally consider reference-dependent distributions of consideration sets and Masatlioglu and Vu (2024) consider a distribution of nested consideration sets. Gossner et al. (2021) use the exogenous manipulation of attention. Differently, the present study exploits the evolution of choices for identification.

Afield are studies that have not specifically looked at limited consideration but have studied the observed dynamics of choices. Gilboa and Schmeidler (1995); Gilboa et al. (2002) study case-based decisions, assuming that preferences under different possible past experiences (cases) are observed. Relatedly, studies of habit

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*.

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

formation take history-dependent preferences as given and focus on agents who understand how consumption affects future preferences (Gul and Pesendorfer, 2007; Rozen, 2010; Tserenjigmid, 2020; Hayashi and Takeoka, 2022). Frick et al. (2019) study dynamic random utility with history-dependent stochastic choice. Piermont et al. (2016) study endogenous subjective learning literature infers learning from preference over menus / consumption problems. Differently, Ferreira and Gravel (2024) look at just one sequence of choices to study preference discovery, and Cerigioni (2021) additionally observes "environments" to study "unconscious" decisions.

7 Conclusion

Models of bounded rationality often interpret behavior that deviates from standard benchmarks as "mistakes" that result from cognitive constrains. This paper characterizes a model called Dynamic Limited Consideration (DLC) to capture how past experiences should improve future decisions in the context of limited consideration. The richer setting of dynamic choice uniquely identifies preferences using the evolution of choices. This observation continues to hold in the stochastic choice setting where aggregate, population data comes from heterogeneous agents who differ in both preference and attention parameters. The paper then relates counterfactual (static) and realized (dynamic) WARP violations, and it shows that the latter is a stronger evidence for bounded rationality. DLC allows for any nonempty-valued initial attention function Γ, and well-known properties of Γ from static models (e.g., Masatlioglu et al. (2012), Manzini and Mariotti (2007), Manzini and Mariotti (2012)) are shown to be self-preserving under DLC. Finally, and of broader interest, this study demonstrates how dynamic choice can serve as a rich but natural dataset that allows the analyst to ask whether, when, and how corrections of "mistakes" occur, which provides a theoretical foundation for how bounded rationality plays out beyond static settings.

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 2 is proven, these are also implied by a DLC 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

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

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

Stability implies Narrow Bracketing: Narrow Bracketing is equivalent to $\tilde{c}(h)(A) = \tilde{c}(\emptyset)(A)$ for all h and A (sufficiency of Narrow Bracketing uses iterative argument, necessity of Path Independence is straightforward). Suppose Narrow Bracketing is not satisfied, so there exists A such that $\tilde{c}(\emptyset)(A) = x$ and $\tilde{c}(h)(A) = y \neq x$. Due to DLC, this means $u(y) > u(x)$ and $y \in \boldsymbol{c}(h)$. 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 DLC, 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 the first and third choice jointly violate Stability.

A.2 Proof of Theorem 2

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 Subsection 3.5 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. □

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 $\tilde{\Gamma}(\emptyset)(A) = \Gamma(A)$. □

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,

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

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,

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

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

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

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 $c^P \in \tilde{\Gamma}((A_1, ..., A_{i^*-1}))(A_{i^*})$, and since i^* is the first instance of disagreement, either $c^P \in \Gamma(A_{i^*})$ which by the construction of Γ implies $c^P = \tilde{c}(\emptyset)(A_{i^*})$ or $c^P \in \boldsymbol{c}((A_1, ..., A_{i^*-1}))$, 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, $c^P \in \tilde{\Gamma}((A_1, ..., A_{i^*-1}))(A_{i^*})$ implies either $c^P = \tilde{c}(\emptyset)(A_{i^*})$ or $c^P \in \boldsymbol{c}((A_1, ..., A_{i^*-1}))$. 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.3 Proof of Theorem 3

(A simpler illustration is provided in Subsection 4.2 under $X = \{x, y, z\}$.) Suppose ρ is consistent with a DLC population. The pair (u, Γ) is said to be *identified* if $\mu_{\sim u,\Gamma} = \mu'_{\sim u,\Gamma}$ for all $\mu, \mu' \in \Delta(\boldsymbol{P})$ such that ρ is consistent with DLC populations μ, μ'. Write $\max(Z, u) := \arg\max_{x \in Z} u(x)$, it selects the best alternative in Z according to u. Partition $\boldsymbol{P}$ into components $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ where $\boldsymbol{P}_1 = \{(u, \Gamma) \in \boldsymbol{P} : \cup_{A \in \mathcal{A}} \max(\Gamma(A), u) = X\}$ and $\boldsymbol{P}_2 = \boldsymbol{P} \backslash \boldsymbol{P}_1$.

Claim 1. If $(u, \Gamma) \in \boldsymbol{P}$ but $\cup_{A \in \mathcal{A}} \max(\Gamma(A), u) \neq X$ (i.e., $(u, \Gamma) \in \boldsymbol{P}_2$), then for some $z \in X$, (1) $\cup_{A \in \mathcal{A}} \max(\Gamma(A), u) = X \backslash \{z\}$, (2) z is never-chosen in $c_{(u,\Gamma)}$, and (3) $u(x) > u(z)$ for all $x \in X \backslash \{z\}$.

Proof. By the definition of $\boldsymbol{P}_2$, $\cup_{A \in \mathcal{A}} \max(\Gamma(A), u) \subsetneq X$. Observe that $x \in X \backslash \cup_{A \in \mathcal{A}} \max(\Gamma(A), u)$ is never-chosen in $c_{(u,\Gamma)}$ because they are never the best considered alternatives in $\tilde{c}_{(u,\Gamma)}(\emptyset) : \mathcal{A} \to X$ and can never become considered (or see Condition A.2, a necessary condition of DLC). Footnote 6 establishes that $c_{(u,\Gamma)}$ has exactly one never-chosen alternative, let it be z, and we have shown (1) and (2). Since $z \in X \backslash \hat{X}$ for $c_{(u,\Gamma)}$, because $(u, \Gamma) \in \boldsymbol{P}$ requires that Γ is maximal (see Equation 3.1), we have $z \in \Gamma(A)$ for all A that contains z. Suppose for contradiction $u(x) > u(z)$ for some $x \in X \backslash \{z\}$, then $\tilde{c}_{(u,\Gamma)}(\emptyset)(\{x, z\}) = z$, a contradiction of (2). Hence (3). □

Lemma 1. *Fix any $(u, \Gamma) \in \boldsymbol{P}_1$. If (u, Γ') is identified for all $(u, \Gamma') \in \boldsymbol{P}_1$ such that $u(\max(\Gamma(A), u)) \geq u(\max(\Gamma'(A), u))$ for all A with at least one strict inequality, then (u, Γ) is identified.*

Proof. Given $(u, \Gamma) \in \boldsymbol{P}_1$, enumerate the elements of X so that $u(x_i) < u(x_{i+1})$ for all $i = 1, ..., |X| - 1$. For each i, let $S_{x_i} := \{A \in \mathcal{A} : \max(\Gamma(A), u) = x_i\}$. Now construct a sequence of choice sets (A_n) using the following ***3-part algorithm***. In **Part 1**, list the choice sets in S_{x_1} in any order, followed by the choice sets in S_{x_2}, and so on. In **Part 2**, (re)list all binary choice sets in any order. In **Part 3**, (re)list any choice sets until (A_n) reaches length N. Since X is finite, $\mathcal{A}$ is finite, so Part

1 and Part 2 involve finitely many choice sets, and so for a sufficient large finite N, (A_n) with at least Part 1 and Part 2 covered can be constructed. Subsection 4.2 provides an example. Let (x_n) be a sequence of choices where in Part 1 of (A_n), x_i's are chosen from the choice sets in S_{x_i}, and in Part 2 and Part 3 of (A_n), choices maximize u. Observe that $c_{(u,\Gamma)}(A_n) = (x_n)$. To see this, in Part 1, whenever the DM with parameter (u,Γ) faces a choice set, she has not yet "discovered" any better alternatives, so she continues to choose $\max(\Gamma(A), u)$. And in Part 2 and Part 3, every alternative had already been chosen in Part 1 and would now receive consideration, so choices maximize u.

Let $\boldsymbol{Q} = \{(u',\Gamma') : u' \sim u\}$. Consider $(u',\Gamma') \in \boldsymbol{P}\backslash\boldsymbol{Q}$ such that $c_{(u',\Gamma')}(A_n) = (x_n)$. First suppose for contradiction $(u',\Gamma') \in \boldsymbol{P}_2$, then by Claim 1, some z is never-chosen by $c_{(u',\Gamma')}$, which contradicts the fact that z must be chosen from $A \in S_z$ when A appears in Part 1 of (A_n) (where the appearance of A is guaranteed because $(u,\Gamma) \in \boldsymbol{P}_1$ implies $S_{x_i} \neq \emptyset$ for all x_i). So $(u',\Gamma') \in \boldsymbol{P}_1$. Suppose for contradiction $u' \not\sim u$, either that choices in Part 1 of (A_n) are inconsistent with (x_n), or if they are consistent, then every alternative had already been chosen in Part 1, and $u' \not\sim u$ implies that for some x, y such that $u'(x) > u'(y)$ and $u(y) > u(x)$, when $\{x, y\}$ shows up in Part 2 of (A_n), the DM with (u',Γ') will choose x instead of y, which contradicts $c_{(u',\Gamma')}(A_n) = (x_n)$. So it must be that $u' \sim u$. Now we argue that $u(\max(\Gamma(A), u)) \geq u(\max(\Gamma'(A), u))$ for all A with at least one strict inequality. Suppose not, it cannot be equality everywhere because $(u',\Gamma') \notin \boldsymbol{Q}$, so let A^* be a choice set such that $u(\max(\Gamma(A^*), u)) < u(\max(\Gamma'(A^*), u))$ and let $x^* = \max(\Gamma(A^*), u)$. Since the DM with (u',Γ') considers something better than x^* in A^*, she does not choose x^* when A^* appears in Part 1 of (A_n), which contradicts $c_{(u',\Gamma')}(A_n) = (x_n)$.

To conclude, let $\tilde{\boldsymbol{P}}$ be the set of $(u',\Gamma') \in \boldsymbol{P}\backslash\boldsymbol{Q}$ such that $c_{(u',\Gamma')}(A_n) = (x_n)$; we have shown that if $(u',\Gamma') \in \tilde{\boldsymbol{P}}$, then $u' \sim u$ and $u(\max(\Gamma(A), u)) > u(\max(\Gamma'(A), u))$ for all A. The antecedent assumes that every $(u',\Gamma') \in \tilde{\boldsymbol{P}}$ is identified, that is, $\mu_{\sim u',\Gamma'}$ is unique. By definition, $\mu_{\sim u',\Gamma'} = \mu_{\sim u,\Gamma'}$ if $u' \sim u$, so we identify (u,Γ) using

$$\mu_{\sim u,\Gamma} = \rho(A_n)(x_n) - \sum_{(u',\Gamma')\in\tilde{\boldsymbol{P}}} \mu(u',\Gamma') = \rho(A_n)(x_n) - \sum_{(u,\Gamma')\in\tilde{\boldsymbol{P}}} \mu_{\sim u,\Gamma'}. \tag{A.1}$$

□

Lemma 2. *Fix any* $(u,\Gamma)\in\boldsymbol{P}_2$*. If* (u',Γ') *is identified for all* $(u',\Gamma')\in\boldsymbol{P}_1$ *and for all* $(u,\Gamma')\in\boldsymbol{P}_2$ *such that* $u\left(\max\left(\Gamma\left(A\right),u\right)\right)\geq u\left(\max\left(\Gamma'\left(A\right),u\right)\right)$ *for all* A *with at least one strict inequality, then* (u,Γ) *is identified.*

Proof. The proof is identical to the proof of Lemma 1 from the construction of (A_n) with the ***3-part algorithm***, to construction of (x_n), to the observation that $c_{(u,\Gamma)}\left(A_n\right)=(x_n)$.

Differently, when $(u',\Gamma')\in\boldsymbol{P}\backslash\boldsymbol{Q}$ such that $c_{(u',\Gamma')}\left(A_n\right)=(x_n)$ and $\boldsymbol{Q}=\{(u',\Gamma):u'\sim u\}$, it is now possible that $(u',\Gamma')\in\boldsymbol{P}_2$. Suppose so. Observe that because every alternative except the worst alternative z to u had been chosen in Part 1 of (A_n), and $c_{(u',\Gamma')}$ can have at most one never-chosen alternative due to Claim 1 (1) and (2), so z is the never-chosen alternative in $c_{(u',\Gamma')}$, and so z is also the worst alternative to u' due to Claim 1 (3). Furthermore, since Part 2 choices in $c_{(u',\Gamma')}\left(A_n\right)$ are consistent with (x_n), it must be that $u\left(x\right)>u\left(y\right)$ if and only if $u'\left(x\right)>u'\left(y\right)$ for all $x,y\in X\backslash\{z\}$, so $u'\sim u$. Following the proof of Lemma 1, we establish that $u\left(\max\left(\Gamma\left(A\right),u\right)\right)\geq u\left(\max\left(\Gamma'\left(A\right),u\right)\right)$ for all A with at least one strict inequality.

To conclude, let $\tilde{\boldsymbol{P}}$ be the set of $(u',\Gamma')\in\boldsymbol{P}\backslash\boldsymbol{Q}$ such that $c_{(u',\Gamma')}\left(A_n\right)=(x_n)$. If $(u',\Gamma')\in\tilde{\boldsymbol{P}}$, either $(u',\Gamma')\in\boldsymbol{P}_1$, which the antecedent assumes is identified, or $(u',\Gamma')\in\boldsymbol{P}_2$, for which we have shown $u'\sim u$ and $u\left(\max\left(\Gamma\left(A\right),u\right)\right)\geq u\left(\max\left(\Gamma'\left(A\right),u\right)\right)$ for all A with at least one strict inequality, and therefore the antecedent assumes (u',Γ') is identified. Let $\bar{\boldsymbol{u}}$ be the set of utility functions that assign integer utility values from 1 to $|X|$ (this is to avoid double counting in the following expression). We identify (u,Γ) using

$$\mu_{\sim u,\Gamma}=\rho\left(A_n\right)\left(x_n\right)-\sum_{(u',\Gamma')\in\tilde{\boldsymbol{P}}}\mu\left(u',\Gamma'\right)=\rho\left(A_n\right)\left(x_n\right)-\sum_{u'\in\bar{\boldsymbol{u}},(u',\Gamma')\in\tilde{\boldsymbol{P}}}\mu_{\sim u',\Gamma'}. \quad \text{(A.2)}$$

□

Finally, we show that every (u,Γ) can be identified recursively. Let $\min\left(Z,u\right):=\arg\min_{x\in Z}u\left(Z\right)$.

Start by fixing any u and consider the DMs with $(u,\Gamma)\in\boldsymbol{P}_1$ who are "unluckiest" in the sense that for all A but A^*, $\Gamma\left(A\right)=\min\left(A,u\right)$, and $\Gamma\left(A^*\right)=\min\left(A^*,u\right)\cup\min\left(A^*\backslash\min\left(A^*,u\right),u\right)$. Note that the only (u,Γ') such that $u\left(\max\left(\Gamma\left(A\right),u\right)\right)\geq u\left(\max\left(\Gamma'\left(A\right),u\right)\right)$ for all A with at least one strict inequality is such that $\Gamma'\left(A\right)=$

$\min(A,u)$ for all A, so the best alternative is never-chosen in $c_{(u,\Gamma')}$, and it can be shown that $(u,\Gamma') \notin \boldsymbol{P}$ (it will violate Claim 1 (3)), so $\mu_{\sim u,\Gamma'} = 0$. Invoke Lemma 1. Then, find the second unluckiest DMs by expanding Γ to include exactly one extra alternative for exactly one A. If the expansion results in $(u,\Gamma') \in \boldsymbol{P}_2$, we ignore it (this is possible because $\min(X,u)$ will at some point become never-chosen when the DM is sufficiently attentive). Otherwise, invoke Lemma 1. Proceed by gradually adding to Γ and invoking Lemma 1. Since there are only finitely many configurations of Γ's because X is finite, the process will end, upon which every $(u,\Gamma) \in \boldsymbol{P}_1$ has been identified.

Next, fix any u, and consider the DMs with $(u,\Gamma) \in \boldsymbol{P}_2$. These DMs never choose one alternative, and it has to be the worst alternative due to Claim 1. Consider the "unluckiest" among them, that is, they never choose $\{z\} := \min(X,u)$ but are "unluckiest" otherwise, that is, $\Gamma(A) = [\{z\} \cap A] \cup \min(A \setminus \{z\}, u)$ for all A. Note that the only $(u,\Gamma') \in \boldsymbol{P}$ such that $u(\max(\Gamma(A),u)) \geq u(\max(\Gamma'(A),u))$ for all A with at least one strict inequality is such that $\Gamma'(A) = \{z\}$ for some A, so $(u,\Gamma') \in \boldsymbol{P}_1$ ($(u,\Gamma') \in \boldsymbol{P}_2$ leads to a violation of Claim 1 (3)), so the antecedent of Lemma 2 is satisfied and we invoke it. Then, find the second unluckiest DMs by expanding Γ to include exactly one extra alternative for exactly one A. The expansion always results in $(u,\Gamma') \in \boldsymbol{P}_2$ because z is never-chosen. Note that if $(u,\Gamma'') \in \boldsymbol{P}$ is such that $u(\max(\Gamma'(A),u)) \geq u(\max(\Gamma''(A),u))$ for all A with at least one strict inequality, then either $(u,\Gamma'') \in \boldsymbol{P}_1$ or (u,Γ'') was identified earlier in this process, so the antecedent of Lemma 2 is satisfied and we invoke it. Proceed by gradually adding to Γ and invoking Lemma 2. Since there are only finitely many configurations of Γ's because X is finite, the process will end, upon which every $(u,\Gamma) \in \boldsymbol{P}_2$ has been identified.

A.4 Proof of Theorem 4

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 DLC representation (u,Γ) where u represents κ in standard utility maximization and $\Gamma(A) = \{f(A)\}$. It is clear that $\tilde{c}(\emptyset) = f$. 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 DLC, $\tilde{\Gamma}(h)(A) = \Gamma(A) \cup [A \cap \boldsymbol{c}(h)] = \{f(A)\} \cup [A \cap \boldsymbol{c}(h)]$, 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 DLC 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 DLC representation (u, Γ) is such that $\tilde{c}(\emptyset) = f$ 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 $g \notin \mathbb{C}$.

References

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

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

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.

Cerigioni, F. (2021). Dual decision processes: Retrieving preferences when some choices are automatic. *Journal of Political Economy 129*(6), 1667–1704.

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.

Ferreira, J. V. and N. Gravel (2024). Revealing preference discovery: A chronological choice framework. *Theory and Decision*, 1–32.

Frick, M., R. Iijima, and T. Strzalecki (2019). Dynamic random utility. *Econometrica 87*(6), 1941–2002.

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.

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

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

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.

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

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.

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

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

Piermont, E., N. Takeoka, and R. Teper (2016). Learning the krepsian state: Exploration through consumption. *Games and Economic Behavior 100*, 69–94.

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.

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

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.

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 3. *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 4. *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 5. *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 6. *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 7. *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. □

Lemma 8. *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. □

Lemma 9. *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. □

Lemma 10. *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. □

B.1.2 Proof of Proposition 1

Suppose $x, y \in \hat{X}$ and suppose for contradiction c admits DLC 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 DLC, $y \in \boldsymbol{c}(\boldsymbol{A}) \subseteq \tilde{\Gamma}(\boldsymbol{A})(B)$, 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 DLC, 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 DLC representation (u, Γ) that represents c. Since X is finite, $\hat{X}$ is finite, so 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 $\tilde{c}(\emptyset)(A_i) = x_i$. 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). Let $\tilde{N}$ be the length of the constructed sequence, which is finite. Suppose $N \geq \tilde{N}$. If $N > \tilde{N}$, embed arbitrary choice sets to the constructed sequence until it reaches length N. Note that $\tilde{c}(\emptyset)(A_1) = x_1$. 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 2. *If c admits a DLC representation (Γ, u) and $\Gamma^* \subseteq \Gamma$ such that $\tilde{c}(\emptyset)(A) \in \Gamma^*(A)$ for all A, then c also admits a DLC 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 DLC, 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 $\tilde{c}^*(h)(A) = \tilde{c}(h)(A)$. □

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

If: Fix c. Suppose c admits a DLC 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 2, (Γ^*, u) where $\Gamma^*(A) = \{\tilde{c}(\emptyset)(A)\}$ also represent c. Now I show that (Γ^+, u) also represent c, and then the proof is complete by invoking Corollary 2. 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,

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

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

B.1.6 Proof of Proposition 5

The relative ranking of $u(x)$ and $u(y)$ for $x, y \in \hat{X}$ is unique due to Proposition 1. If $\hat{X} = X$, we are done. Suppose $\{z\} = X \backslash \hat{X}$ (recall that there is at most one never-chosen alternative in any c, see Footnote 6). Then $z \in \Gamma^+(A) = \Gamma(A)$ for all A that contains z, and in particular $z \in \Gamma(\{z, x\})$ for all $x \in X \backslash \{z\}$. Since z is never-chosen, we must have $u(x) > u(z)$ for all $x \in X \backslash \{z\}$.

B.1.7 Proof of Corollary 1

From Theorem 3, $\mu_{\sim u} = \sum_{\Gamma \in \mathbf{\Gamma}} \mu_{\sim u, \Gamma} = \sum_{\Gamma \in \mathbf{\Gamma}} \mu'_{\sim u, \Gamma} = \mu'_{\sim u}$

B.1.8 Proof of Proposition 6

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 $\tilde{\Gamma}(h)(B \backslash \{x\}) = \Gamma(B \backslash \{x\}) \cup [(B \backslash \{x\}) \cap \boldsymbol{c}(h)] = \Gamma(B) \cup [B \cap \boldsymbol{c}(h)] = \tilde{\Gamma}(h)(B)$. CLA is by definition.

B.1.9 Proof of Proposition 7

(2) implies (1): 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 $\tilde{c}(\emptyset)(T) = x$, we have $u(x) > u(y)$. Since $y \mathbb{S} x$ would imply $u(y) > u(x)$ due to Proposition 3 (1), not $y \mathbb{S} x$. Hence Axiom 4.

(1) implies (2): Proposition 4 guarantees a DLC representation (u, Γ^+). 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 $\Gamma^+(A) \neq \Gamma^+(A \backslash \{y\})$. Note that this means $\tilde{c}(\emptyset)(A) = x \neq y$ and $x \in \hat{X}$. By the definition of Γ^+, $y \notin \Gamma^+(A)$ implies $y \in \hat{X}$ and $\neg x \mathbb{S} y$, and then Proposition 3 (2) gives $y \mathbb{S} x$. Next we argue that $\tilde{c}(\emptyset)(A \backslash \{y\}) \neq x$. Suppose for contradiction $\tilde{c}(\emptyset)(A \backslash \{y\}) = x$, then $y \mathbb{S} x$ implies $\Gamma^+(A) = \Gamma^+(A \backslash \{y\})$ by the definition of Γ^+, a contradiction. So $y \mathbb{S} x$ and $\tilde{c}(\emptyset)(A) = x \neq \tilde{c}(\emptyset)(A \backslash \{y\})$, but they contradict Axiom 4.

B.1.10 Proof of Proposition 8

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 xS^*y if xSy 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 $\neg xSy$ for all $x \in X$) or $y \in \boldsymbol{c}(h)$ (so $\neg xS^*y$ for all $x \in X$), so $y \in \tilde{\Gamma}^*(A)$. If $y \in \tilde{\Gamma}^*(A)$, then $\neg xS^*y$ for all $x \in A$, which by definition of S^* can be due to (i) $\neg xSy$ for all $x \in A$, which means $y \in \Gamma(A)$, or (ii) $y \in \boldsymbol{c}(h)$. In either case, $y \in \tilde{\Gamma}(h)(A)$. RSM is by definition.

B.1.11 Proof of Proposition 9

Given (u, Γ) where Γ is a coarse-max, let S be the corresponding rationale. Consider any h. Define a new rationale $S^* \subseteq \left[2^X \setminus \{\emptyset\}\right] \times \left[2^X \setminus \{\emptyset\}\right]$ with the following rules. If $R'SR''$, then let $R'S^*R''$ if $R'' \cap \boldsymbol{c}(h) = \emptyset$ and let $R'S^*\left[R'' \setminus \boldsymbol{c}(h)\right]$ if $R'' \cap \boldsymbol{c}(h) \neq \emptyset$. The operation removes $\boldsymbol{c}(h)$ from categories that would be dominated, while forming new categories $R'' \setminus \boldsymbol{c}(h)$ 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, $y \in \tilde{\Gamma}(h)(A)$. CTC is by definition.

B.2 Omitted Examples and Results

B.2.1 Other Examples

Example 6. 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 7. The following $\mathbb{C}$ is not compatible with DLC: 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 DLC. 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 DLC.

B.2.2 Other Results

Corollary 3. *Suppose X is finite. There exists a sequence of choice sets $(A_n) \in \mathcal{A}^N$ such that for any c that admits a DLC 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 $(B_1, ..., B_k)$ (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 $(B_1, ..., B_k)$ (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 $(B_1, ..., B_k)$, 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 a DLC 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. □

Corollary 4. *If c admits DLC representations (Γ, u) and (Γ', u), then it also admits DLC representations $(\Gamma \cup \Gamma', u)$ and $(\Gamma \cap \Gamma', u)$.*

Proof. The case for intersection is shown in Corollary 2; note that an intersection between Γ and Γ' guarantees that $\Gamma^* := \Gamma \cap \Gamma'$ satisfies $\Gamma^* \subseteq \Gamma$ and $\tilde{c}(\emptyset)(A) \in \Gamma^*(A)$ for all A. For the case for union, the "only if" of Proposition 4 guarantees that $\tilde{c}(\emptyset)(A) \in \Gamma(A) \subseteq \Gamma^+(A)$ and $\tilde{c}(\emptyset)(A) \in \Gamma'(A) \subseteq \Gamma^+(A)$, and the "if" part

subsequently guarantees (Γ^*, u^*) where $\Gamma^* = \Gamma \cup \Gamma'$ is a DLC 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 $\tilde{c}(\emptyset)(A) = x$, then $z \notin \Gamma(A) \cup \Gamma'(A)$, so (Γ^*, u) also represent c. □

Corollary 5. *Suppose c admits a DLC representation (u, Γ) and suppose $\tilde{c}(\emptyset)$ is explained by some strict total order $(\succ_0, X)$ (i.e., $\tilde{c}(\emptyset)(A) \succ_0 z$ for all $z \in A \backslash \{\tilde{c}(\emptyset)(A)\}$). 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 1, with $\tilde{c}(\emptyset)(\{y, z'\}) = y$ replaced by $\tilde{c}(\emptyset)(A) = y$ for some A, which is guaranteed by $y \in \hat{X}$ and Condition A.2. **Only if:** Suppose $\tilde{c}(\emptyset)$ 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 $u(x) > u(y)$, $u(y) > u(\tilde{c}(\emptyset)(B))$, $\tilde{c}(\emptyset)(B) \succ_0 x$, and $x \succ_0 y$, which completes the proof. Note that $\tilde{c}(\emptyset)(\{x, y\}) = x$; otherwise, $\tilde{c}(\emptyset)(\{x, y\}) = y$ 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 $\tilde{c}(\emptyset)$ satisfies WARP and $x \in B$, $\tilde{c}(\emptyset)(B) \neq y$, so the only way we have $\tilde{c}(h)(B) = y$ is $u(y) > u(\tilde{c}(\emptyset)(B))$ and $y \in \boldsymbol{c}(h)$, which also means $\tilde{c}(\emptyset)(B) \neq x$. Because $\tilde{c}(\emptyset)$ satisfies WARP, $\tilde{c}(\emptyset)(B) \succ_0 x$ and $\tilde{c}(\emptyset)(B) \succ_0 y$. Finally, $y \in \boldsymbol{c}(h)$ and $\tilde{c}(h)(\{x, y\}) = x$ imply $u(x) > u(y)$. □