When More is Less: Limited Consideration*

J. S. Lleras Y. Masatlioglu D. Nakajima E. Ozbay

October 16, 2010
Preliminary Draft

Abstract

There is well-established evidence that decision makers consistently fail to consider all available options. Instead, they restrict their attention to only a subset of it and then undertake a more detailed analysis of the reduced sets of alternatives. This systematic lack of consideration of available options can lead to a "too much choice" effect, where excess of options can be welfare-reducing for a DM. Building on this idea, we model individuals who might pay attention to only a subset of the choice problem presented to them. Within this smaller set, a decision maker is rational in the standard sense, and she chooses the maximal element with respect to her preference. We provide testable characterization results for choice behavior under different consideration structures. In addition, we show to which options the decision makers must pay attention to at each set, which elements are revealed preferred to which, and discuss welfare implications.

JEL Classification: D11, D81.
Keywords: Revealed Preferences, Awareness, Attention, Consideration Set, Preference Reversal

*We would like to thank David Ahn, Tilman Börgers, Emel Filiz-Ozbay, Stephan Lauermann, Paola Manzini, Marco Mariotti, Neslihan Uler, and participants of the Berkeley Theory Brunch for their helpful comments.

1 Introduction

In many choice problems, decision makers face a considerable number of alternatives. In addition, to having a large amount of choices, there has been an ongoing trend to increase even more the number of available options in many markets. For instance, nowadays a shopper in a supermarket needs to select from 285 varieties of cookies, 85 flavors and brands of juices, 230 different soups and 275 varieties of boxed cereal (Schwartz 2005). Also, Scheibehenne (2008) provides many examples of markets where we have had a "variety revolution", this trend of increasing the number of options offered to consumers: the number of types of products offered by companies that produce ice cream, potato chips, fast food, and orange juice have increased more than 10 fold in the past 50 years. And looking at internet markets, the number of options offered in markets like music, books, or movies is staggering: there are tens of thousands of available good in these markets. Classical rational choice theory would conclude this exuberance of choice has positive welfare consequences since increasing the number of available options in any market is unambiguously beneficial for consumers. However, alternatively to this classical idea of "more is better", there is recent evidence on a "Too Much Choice" (TMC) effect in many choice situations, where a large amount of options can have negative consequences in terms of welfare for decision makers.

Since it conflicts with the standard view of rational choice theory, the aforementioned Too Much Choice effect has also been labeled as "The Paradox of Choice" (see Schwartz (2005)). Schwartz (2005) claims that one of the reasons that motivate the Paradox of Choice is that more options can hurt consumers since the choice process may seem overwhelming: the overwhelming nature of some problems might lead DMs might fail to consider some objects that are available to them, which might even be preferred than the actual chosen element; or even to inaction. This idea of the possible negative effects of an excessive number of alternatives dates back to the seminal psychology contribution of Miller (1956), which studied the limited ability of decision makers to process bits of information. By designing experiments that test the cognitive capacity of subjects, Miller (1956) establishes that there is an upper bound on the amount of information that the average person can process in the short term (when suddenly presented with with new information.)

If DMs have a rather limited attention span, abundance of choice can create difficulties for the decision making process, which is something that has been extensively documented: decision makers often deal with a small number of alternatives which is far fewer than the total number of alternatives (Hauser and Wernerfelt 1990) (see also Chiang et al. (1998), Stigler (1961), Pessemier (1978)). For example, in financial economics, (see e.g. Huberman and Regev (2001)), it is known that investors make investing decision based on a limited number of all the available options. Similar examples can be found in job search (see Richards et al. (1975)), university choice (see e.g. Laroche et al. (1984), Rosen et al. (1998)), and airport choice (see Basar and Bhat (2004)). Hence, a common procedure is for decision makers to choose heuristics, or rules of thumb, to simplify the choice problem to one they can analyze (i.e. apply a maximization procedure). A classical example of such heuristic choices is to choose the second cheapest wine on the menu, which is arguably just a reduction of the dimensionality of the problem to facilitate a choice, and this rule of thumb is most likely orthogonal to any maximization procedure (see Tversky (1972), Tversky and Kahneman (1974), Martignon and Hoffrage (2002), or Katsikopoulos and Martignon (2006) for more examples, evidence, and analysis of heuristic choices.)

Complementary to the evidence of the recurrent use of heuristics to simplify complex choice problems, there have been several studies that have found that there is the TMC effect, where the excess of available options lead decision makers to be worse off than with fewer options. Iyengar and Lepper (2000) famous work, consider three different choice situations, on the field and on the lab, and showed that in the three studies subjects were better off under the situation of a smaller number choices rather than a bigger number of elements they could choose from. The first experiment involved a jam tasting and consumer welfare is measured in the likelihood of buying jam; the finding is that consumers were more likely to buy jam (redeem a coupon given to them for participating in the tasting) when they were in the groups that tasted less options. A lab experiment involved chocolate tasting, and consumers who faced less options said that they enjoyed more the tasting than the ones who faced more types of chocolate. Finally, Iyengar and Lepper (2000) also showed that the quality of extra credit essays, and the number of essays completed, increased as students were presented with fewer choices. Along the same lines, Iyengar et al. (2004), showed that people are more likely to enroll in company sponsored retirement plans when they have fewer available plans; and Chernev (2003) also showed that in a experiment with chocolate choices, subjects that were presented fewer options were more confident with their choice, when questioned about it by the experimenter (they were less likely to change the choice once suggested by the experimenter). Furthermore, Caplin et al. (2009) study a choice environment where a large amount of options can lead consumers to "make mistakes" (i.e. choose some element that is not their preferred one). By tracking choices over time, they conclude that people people tend to make more mistakes when there are more available options, and that these mistakes decrease by increasing the contemplation time. See Schwartz (2005) for more documented instances of the TMC effect.[1]

Given these results about limited cognitive capacity decision makers in different situations, the idea of the Paradox of Choice seems intuitive in terms of choice overload: how can

[1]Nonetheless, some authors, such a Scheibehenne et al. (2009) argue that the too much effect is small and unquantifiable at best under certain environments.

we know, by observing choices, that a consumer really chose her most preferred option out of a menu, when she is not able to pay attention to all available options?

In this paper we model, within the classical rational choice theory framework, a choice situation susceptible to a Too Much Choice effect. Based on this idea, and the empirical evidence of choice overload, we consider a choice situation where too many options can lead the DM to overlook some of the available option, which has intuitive negative welfare consequences. By using an an axiomatic choice theoretical framework we can analyze in which choice problems we, as analysts, can claim that "more is less", i.e. when more options can lead to lower consumer welfare, and what the DMs considered and the revealed preference at each possible menu.

We model a choice problem with two independent steps. First, agents consider some of the available options, this will be the DMs *consideration set*, and afterwards they make a choice. Usually axiomatic choice theory does not deal with the first step, by implicitly assuming that DMs sees everything that is available, focusing exclusively on the choice or maximization step. However, if we allow the possibility to not consider some elements, from observation we cannot disentangle each step. As analysts, we do not know what was observed and what was the procedure used for the making the choice separately. In this paper, we explicitly incorporate the consideration step of the decision-making process into the classical choice theory framework and explore how much of each process we can separately identify, conditional to some consistency conditions on the way observations work. Moreover, we use consideration of elements, or lack thereof, to identify a revealed preference (a la Samuelson (1938)) that come from some maximization procedure, and instances in the decision-making process that possibly came from heuristics.

Here, we require a minimal condition on the formation of consideration sets to capture the TMC effect: if the DM pay attention to some element in a large menu, then she will pay attention to that element on a smaller menu. If the number of available options is what affect the DM's consideration of products, for example it is reasonable to think that if the DM pay attention to a certain type of boxed cereal in a crowded supermarket shelf, then she will pay attention to that same cereal in a smaller neighborhood convenience store.[2]

Based on the consistency conditions for the formation of consideration sets, we provide axiomatic characterization, revealed preference results, and welfare implications for *Choice with Limited Consideration (CLC)*. By CLC we mean a procedure where the DM uses a

[2]In a companion paper, Masatlioglu et al. (2009), consider a choice environment, which is independent to ours, where DMs can be unaware of being unaware, which require a different condition on the formation of consideration sets. We extend the problem to satisfy both these conditions on section 4. Within a similar framework, we will see that the welfare implications of the two models are substantially different. In addition, we consider more restrictive consideration filter structures on section 5.

complete and transitive order to choose out of the consideration set, which can be smaller than the actual menu of options offered, i.e. a rational consumer that overlooks some of the available elements in the choice set. Formally, we will say that a choice function c is a CLC with a *consideration filter*, if there is a consideration filter Γ, which we define formally in section 2, and a utility function u,[3] such that the decision maker chooses the alternative with highest utility in the consideration set, i.e, for each A, the decision maker's maximization problem can be written as

$$\max \quad u(x) \quad \text{subject to } x \in \Gamma(A).$$

Within this framework, we can see that as consequence of allowing DMs to possibly overlook some available elements, we incur some limitations when we try to make revealed preference, structure of the consideration set, and welfare claims. How can we tell choice behavior that fails rationality when DM might overlook options? And to what extent can we make revealed preference and welfare assertions? In this paper we provide answers for these two questions within the classical choice theory framework. In this paper we characterize a model of choice where the DM might fail to observe elements in the choice set, following the lines of the "Too Much Choice" argument; thus as a consequence in our model if the choices can be characterized as a CLC with a consideration filter, whenever we observe a choice reversal we can conclude that at that instance the DM suffered from too much choice and didn't observe or analyze all the options, and thus the smaller menu is welfare enhancing.

The remainder of the paper will be structured as follows, section 2 formally defines and discusses the relevance consideration filters, section 3 characterizes the choice with limited consideration model for functions and linear orders and discusses the revealed preference, revealed attention and revealed inattention within our framework. Section 4 imposes an additional restriction on the consideration filter, from Masatlioglu et al. (2009), called attention filter, and adding these two conditions gives us a strong filter. For this type of filter we also provide a characterization result. Section 5 establishes characterization and welfare results for two independent special cases of the CLC with limited consideration: full attention on binary menus; and filters generated by a transitive order, which established a direct connection between our paper and Manzini and Mariotti (2007) Rational Shortlist Method. Section 6 discusses the related literature, and finally section 7 concludes.

[3] We consider in the body of paper the case of choice functions with a linear order $\succ$, which is more tractable. However, in the appendix we show that it is possible to extend the main characterization result to choice correspondences with a weak order $\succsim$.

2 Consideration Filters

The model as presented now has no empirical content. In other words, if there is no restriction on Γ, then any choice function can be interpreted as a CLC. Precisely, if Γ consists of only chosen element, then choice is rationalized by any order[4], and the problem is trivial. So to analyze this problem we have to impose some minimal condition on Γ.[5]

In many real-world markets, every product is competing with each other for the space in the consideration set of DMs, who have cognitive limitations. In these situations, if an alternative attracts attention when there are many others, then it must be considered when some of them become unavailable. For example, if a product is able to attract attention in a crowded supermarket shelf, the same product will be noticed in a smaller convenience store, where there are fewer alternatives. We call such consideration sets as attention filter, and we will formally define it later. Under this structure, we show that we deduce part of the DM's preference whenever her choices from a large set and a smaller set are inconsistent. Since the choice from the larger set must be considered in the smaller set, we can conclude that it is less preferred to the choice from the smaller set.

Consider some strategies to narrow down the set of available alternatives.

Example 1. A decision maker may pay attention to (be aware of) only

1. The three most advertised or safest cars in the market.
2. The cheapest car, the most fuel efficient car and the most advertised car in the market.

   This behavior is often called "all or nothing" or "extreme seeking" behavior. Gourville and Soman (2005) show that subjects, who presented with laptop computers that included a basic model, a fully-loaded model, and either one, two, or three intermediate models, increasingly chose one of the two extreme models as the overall assortment increased in size.
3. The top three job candidates in each field: theory, macro, and econometrics for hiring one assistant professor.
4. All products appearing in the first page of search result and/or sponsored links.

[4] We will only require that $x \sim y$ if $x, y \in c(A)$ for some A if we consider choice correspondences.

[5] For instance, in the companion paper, Masatlioglu et al. (2009), impose the minimal condition that $x \notin \Gamma(A) \Rightarrow \Gamma(A) = \Gamma(A \setminus x)$ to study the lack of attention, here we impose a similar but independent condition based on a different behavioral motivation: the effect of too many alternatives on lack of consideration of some available elements.

For many customers, web-search engines, such as Yahoo!, Google, MSN Search, and AOL, are now the primary method for finding products. Customers pay attention to products appearing in the top part of first page in search results.

5. The n cheapest options.

   In case of choosing a supplier for a particular product, Dulleck et al. (2008) show that consumers select a shortlist of suppliers by using the price variable only (for example ten cheapest suppliers) and they trade off reliability and price among these shortlisted suppliers.

6. All products appearing in the search result if the total number is 20 or less. Otherwise, the DM adds another keyword to narrow down her search.

7. All products of a particular brand, if there are none available choose another brand and consider only those products. This type of behavior is studied in the marketing literature as brand consideration (see Roberts and Lattin (1997)).

8. Consider all brands of cars that have n models or less.

The common property across all these examples: if an alternative is considered in a decision problem, it will also be considered when some alternatives are removed; we can think that alternatives are competing for the DMs attention. For example, consider the first example above. If a car is one of the three most advertised ones, then it will still be so when some of others are out of the market. Therefore we consider a particular type of consideration that is motivated by the vast research on that shows consumers are overwhelmed by the abundance of options. This lack of attention can be attributed to the size or complexity of the choice set (see e.g. (Schwartz 2005)), and evidence of this phenomenon goes back to the psychology literature to Miller (1956). We call this type of consideration structure a consideration filter. Before the definition, some notation is needed. Let X be an arbitrary non-empty finite set and $\mathcal{X}$ be the set of all non-empty subsets of X.

Definition. *A correspondence* $\Gamma : \mathcal{X} \to \mathcal{X}$ *is called an* ***Consideration Filter*** *if* $x \in S \subset T$ *and* $x \in \Gamma(T)$ *then* $x \in \Gamma(S)$.

Our goal here is to capture some systematic failures to consider options that have been considered in the literature, focusing on the "too much choice" phenomenon. Therefore we are modeling the observation process as a consideration filter, implicitly assuming that the reason why agents fail to consider options is the size or complexity of the opportunity set. This is in line with the Miller (1956) findings of the limited amount of information that DMs can process mentioned, and with the empirical evidence which shows the complexity of a decision process as a function of size of the menu such as Iyengar and Lepper (2000), Iyengar et al. (2004), etc. In general, these research shows when there is a larger set of options DMs seem to loose something in terms of welfare compared to the simpler, smaller, alternative (submenu).

Modeling consideration, and the advertising implications as a function of the size of the menu has been common on the marketing literature like Hauser and Wernerfelt (1990) and Shugan (1983), or Weinberger (2007). Hauser and Wernerfelt (1990) show that as both the number of options and the information about options increases, people consider fewer choices and process a smaller fraction of the overall information available regarding their choices. Consideration filter captures the idea that the more products are available, the more products a decision maker overlooks. For instance, if a decision maker overlooks a particular alternative in some set, and we think there is a reason (element, who captures the DMs attention) to overlook it. If a smaller set contains such a reason, a superset of it must contain it too. Therefore, the overlooked alternative in the small set should not attract attention in a larger set either.

The monotone cost of observation with respect to size was used since early as in Stigler (1961), as a rational search model; and more recently complexity of choice problems has been interpreted as a size of the choice set in choice theory work like Ergin (2003), Ergin and Sarver (2007), Tyson (2008), Dean and Caplin (2008), Dean (2008), and Masatlioglu and Nakajima (2010) or applied theory models like Eliaz and Spiegler (2007). Here we model the possible lack of attention or failure to consider as a function of size of the opportunity set with the consideration filter, but in general terms this approach of linking complexity of a problem and the size of the menu is common in different fields.

As previously mentioned, it is not novel to think that the size of the menu has an impact on the choices. For instance Tyson (2008) models the complexity of choices is a function of size of the menu in a very similar way. Tyson (2008) does it through the nestedness requirement for the system of preferences, mentioning that if it is possible to relate two elements x and y, on a big menu A, then the same relation will hold for any subset of A that contains x and y. We build this assumption of the relation of the size of the menu to the complexity of the problem through a contraction consistency property. In Tyson (2008) a DM might not be able to compare options on a bigger menu, and in our approach the DM might not consider options on a bigger menu, they are different things but nonetheless they share that same flavor of the complexity of the problem (the inability to compare, or to observe) is related to the size of the menu.

Even with this restriction on the consideration set of being the result of a consideration filter, our model "rationalizes", in the sense of a DM having a well-defined maximization process, several anomalies pointed in experimental literature such as the attraction effect, choice cycles and choosing pairwise dominated alternatives, and in the marketing literature such as advertising an irrelevant alternative and introducing overqualified products.

3 CLC: Characterization Results

A choice or plan assigns a unique chosen element to every non-empty feasible set. This choice can be represented by a choice function on $\mathcal{X}$, $c : \mathcal{X} \to X$, such that $c(S) \in S$ for every $S \in \mathcal{X}$. Let $\succ$ be a strict linear order on X. We denote the best element in S with respect to $\succ$ by $\max_{\succ} S$.

We propose a model to capture the idea of limited attention: our decision maker pays attention only a subset of all available alternatives and picks the best alternative among them. Her choice for a given feasible set S is the alternative $x \in S$ that she prefers (given her preference $\succ$) among all feasible alternatives to which she pays attention. Formally,

$$c(S) = \max_{\succ} \Gamma(S),$$

where $\emptyset \neq \Gamma(S) \subseteq S$ is the consideration set that consists of alternatives to which the decision maker pays attention under choice problem S. Then her choice will be the best element in $\Gamma(S)$ according to $\succ$.

If a decision maker is not aware of a particular product, it cannot be part of her consideration set. Therefore, awareness of the item is the necessary condition Lavidge and Steiner (1961). However, in some choice environments, individuals might be aware of a lot of products but they do not seriously consider all of them for purchase. Therefore, the consideration set can be smaller than the awareness set. Indeed, Jarvis and Wilcox (1977) shows that while the average number of known products may vary a lot for different class of products, the average size of the consideration set is three to eight products. In extreme cases, consumers actually have a consideration set of size one. In an empirical study of about 1,000 recent buyers of new cars, 22% of new-car buyers looked at only one brand (Lapersonne et al. (1995)). Even though the marketing literature has consideration models with different degrees of consideration or awareness, which would be important in a search model (see e.g. Wu and Rangaswamy (2003)), we abstract away from this notion which would make our model less clean and study the choice problem with only two options: either the DM considers something or not. This way we gain tractability and arguably do not loose too much in terms of behavior.

Now, as mentioned in section 2, without any structure on the consideration sets, the model has no empirical context since any choice could be rationalized this way by allowing the DM to only look at the element chosen. This would make the decision vacuously maximal with the respect to the observed options. Now we add the necessity of the consideration structure to be a consideration filter, we formally define the workhorse of this paper, the model of *Choice with Limited Consideration, or CLC.*

Definition. *A choice function c is a Choice with Limited Consideration with a consideration filter if there exists a strict order $\succ$ and a consideration set mapping Γ such that*

$$c(S) = \max_{\succ} \ \Gamma(S).$$

and Γ is a Consideration Filter.

Occasionally, we say that $(\Gamma, \succ)$ represents c. We also mention that $\succ$ represents c, which means that there exists some consideration set mapping Γ such that $(\Gamma, \succ)$ represents c. In the Appendix A.1 we show that it is possible to extend the main characterization result to correspondences and weak orders. However here we show all the characterization results only for functions for intuition and tractability.

Our main characterization result is concerned with finding necessary and sufficient conditions for the type of choice behavior is consistent with our model: How could one test whether choice data is consistent with CLC with a consideration filter? Surprisingly, it turns out that CLC with a consideration filter can be simply characterized through one observable property of choice, just like WARP in the classical choice setting.

We provide several examples which exhibit choice reversals and are consideration filters. Therefore, the postulate we provide should allow choice reversals. Before we state the test, recall the standard Weak Axiom of Revealed Preference (WARP) which does not allow any type of choice reversals. WARP requires that every set S has the "best" alternative x^* (for choice functions), that is, x^* must be chosen from a budget set T whenever x^* is available and the choice from T lies in S. Formally,

WARP. For any nonempty S, there exists $x^* \in S$ such that for any T including x^*,

$$c(T) = x^* \text{ whenever } c(T) \in S.$$

A rational decision maker making a choice with a consideration filter might exhibit a choice reversal as Aumann (2005) argues, which is incompatible with WARP, we need to relax WARP to characterize rational behavior when choices might be overlooked. Our Limited

Consideration WARP axiom allows choice reversals, but requires some consistency in the way they operate, consistent with the idea that the reason for a choice reversal is the lack of consideration of some elements in a menu. Recall that we are modeling the situation where we can conclude that "more is less", so choice reversals (in some consistent way) in our model will reveal a welfare enhancement, since the DM essentially failed to consider a better element in the large, and more complex, menu.

(A1) LC-WARP. For any nonempty S, there exists $x^* \in S$ such that for any T including x^*,

$$c(T) = x^* \text{ whenever } \quad \begin{array}{l} (i)\ c(T) \in S,\ \textit{and} \\ (ii)\ c(T') = x^* \text{ for some } T' \supset T \end{array}$$

Let us explain Limited Consideration WARP by comparing it with WARP. WARP states that every set S has the "best" item x^* in the sense that it is always chosen whenever the chosen element is within S (and x^* is available). LC-WARP requires that x^* must be chosen from T only when the chosen element from T is within S *and* x^* is chosen from some set larger than T.

To understand the axiom better, we provide an example which does not satisfy LC-WARP. Consider the following choice pattern:

$$c(xyzt) = y, \quad c(xyz) = x \text{ and } c(xy) = y.$$

To see how LC-WARP rules out the example above, take S equal to $\{x, y\}$. Either x or y should obey the condition in the axiom for S. Suppose that $x^* = y$, then for $T' = xyzt$, and $T = xyz$. Since $y = c(xyzt)$, $c(xyz) \in S$ if c satisfies LC-WARP then $y = c(xyz)$, which is not the case. If $x^* = x$, then consider $T' = xyz$ and $T = xy$. Since $c(xyz) = x$, then LC-WARP would require $x = c(xy)$, which is not true either. So there is no element in S that satisfies the condition for x^*. Now if $x = x^*$, $c(xyz) = x$ and $c(xy) = y$ imply that x does not satisfy the condition. Therefore, above example violates LC-WARP for $S = \{x, y\}$. Note that the axiom applies to any set of alternatives so it rules out more than the above example.

In other words, to put the same example in the context of our model of choice with limited consideration with a consideration filter, note that $c(xyzt) = y$ reveals y attracts attention at $\{x, y, z, t\}$. This implies that the decision maker aware of y in $\{x, y, z\}$. $c(xyz) = x$ requires that she prefers x to y ($x \succ y$) since she is aware of y. This observation also implies that x

is revealed to be preferred to y. By the similar argument, $c(xyz) = x$ and $c(xy) = y$ imply y is revealed to be preferred to x. This is a contradiction, since it would imply a cycle of 2.

Since the choice reversals in our model directly imply some revealed preference we can notice the following. Whenever her choices from a small set and a larger set is inconsistent, the former reflects her true preference under this framework. Formally, for any distinct x and y, define the following binary relation:

$$xPy \text{ if there exist } S \text{ and } T \text{ with } \{x, y\} \subseteq S \subset T \text{ such that } x = c(S) \text{ and } y = c(T). \quad (1)$$

This binary relation P determines observed choice reversals. We can see from the definition of CLC, and the welfare improving quality of choice reversals in our model, that if we observe y being chosen from a large menu, and x being chosen from a smaller menu containing y, then x must be preferred to y by the DM. In addition, we also should be able to conclude that she prefers x to z if xPy and yPz for some y, even when xPz does not hold (i.e. we don't see a choice reversal from z to x). Thus, we let P_T be the transitive closure of P.

The following lemma shows the link between LC-WARP and P_T, which furthermore we'll see is the only behavioral postulate needed to characterize our *Choice with Limited Consideration* model.

Lemma 1. *c satisfies Limited Consideration WARP if and only if P_T is acyclical.*

Proof. First, we show that P is acyclical, so is P_T. Assume that $x_n P x_{n-1} P \cdots P x_1 P x_n$ occurs. Then there exists no element in $\{x_1, \ldots, x_n\}$ serving the role of x^* in the axiom. For example, x_k cannot be x^* since $x_{k+1} P x_k$, i.e. there exist $S_k \subset T_k$ with $\{x_k, x_{k+1}\} \subset S_k \subset T_k$ s.t. $x_{k+1} = c(S_k)$ and $x_k = c(T_k)$. To show the other way, take $S \in \mathcal{X}$. Since P is acyclical, there exist P-undominated elements in S. Then it is routine to check that any of them serves the role of x^* in the axiom. □

The following theorem shows that a CLC with a consideration filter is captured by a single behavioral postulate: Limited Consideration WARP. This makes it possible to test our model non-perimetrically by using the standard revealed-preference technique a la Samuelson and to derive endogenously decision maker's preferences and consideration filter from observed choice data.

Theorem 1. *A choice function c satisfies LC-WARP if and only if c is a CLC with a consideration filter.*

Proof. The if-part is a direct implication of Lemma 1. If c violates Limited Consideration WARP, its revealed preference has a cycle. Let us prove the only-if part. By Lemma 1 and the existence of a linear order that is an extension of a partial order on a nonempty X,there is a preference that includes P_T. Take such a preference arbitrarily and define

$$\Gamma^m(S) = \{x \in S | \exists S \subset T \text{ s.t. } x = c(T)\}$$

We have already shown that this Γ^m is indeed a consireation filter and $c(S)$ is the $\succ$-best element in $\Gamma^m(S)$. Therefore, $(\Gamma^m, \succ)$ represents c. □

3.1 Revealed Preference and Revealed (In)attention

We now illustrate how to infer (1) a decision maker's actual preference and (2) to what she pays (and does not pay) attention from her choice behavior, given that is a CLC with a consideration filter. We are interested in this question because the revealed preference can be used for welfare analysis and the revealed attention/inattention can determine which marketing strategy is effective.

Imagine a decision maker chooses x when y is one of the available alternatives. The standard revealed preference argument immediately concludes that x is preferred to y. To justify such an inference, one must implicitly assume that she has paid attention to y. Without this hidden assumption, we cannot make any inference because she may prefer y but overlook it. Therefore, eliciting her ranking between x and y (preference) might no longer follow the standard revealed preference analysis because her choice can be attributed to her preference or her inattention.

This observation of the possible tension between preference and attention suggests that multiple pairs of a preference and a consideration filter can generate the same choice behavior. To illustrate this, consider the choice function with three elements exhibiting a cycle:

$$c(xyz) = y, \quad c(xy) = x, \quad c(yz) = y, \quad c(xz) = z.$$

One possibility is that her preference is $x \succ y \succ z$ and she overlooks x both at $\{x, y, z\}$ and $\{x, z\}$. Another possibility is that her preference is $z \succ x \succ y$ and she does not pay attention to x only at $\{x, y, z\}$. Consequently, we cannot determine which of them is her true preference from her choice data. Likewise, we cannot determine whether she pays an attention to x at $\{x, z\}$. On the other hand, both of the preferences rank x above y. Therefore, if these two pairs are only possibilities, we can unambiguously conclude that she prefers x to y, which we shall now define as the revealed preference.

Let c is a CLC with a consideration filter, suppose there are k different pairs of strict preferences and consideration filters representing c:

$$(\Gamma_1, \succ_1), (\Gamma_2, \succ_2), \ldots, (\Gamma_k, \succ_k).$$

So for each i, $c(S) = \max_{\succ_i} \Gamma_i(S)$.

If there are two $\succ_i$ and $\succ_j$ disagreing on the ranking of x and y, we cannot identify her true preference between x and y. On the contrary, if every $\succ_i$ ranks x above y, we can infer that she prefers x to y, which leads to the following definition:

Definition. *Suppose c is a CLC, represented by k different pairs:* $(\Gamma_1, \succ_1), \ldots, (\Gamma_k, \succ_k)$.

1. *x is **revealed preferred** to y, denoted by $x \succ_R y$, if $x \succ_i y$ for all i.*
2. *x is **revealed to attract attention** at S if $x \in \Gamma_i(S)$ for all i.*
3. *x is **revealed to NOT attract attention** at S if $x \notin \Gamma_i(S)$ for all i.*

If one wants to know whether x is revealed to be preferred to y, it seems to be necessary to check for every $(\Gamma, \succ)$ whether it represents her choice or not, which is not practical especially when there are many alternatives. We shall now provide characterization of her revealed preference, attention and inattention completely.

Let us go back to the previous example where $c(xyz) = y$ and $c(xy) = x$. We can immediately conclude that she pays attention to y at $\{x, y, z\}$ so does she at $\{x, y\}$ (revealed attention). Since she picks x from $\{x, y\}$, we can conclude that she prefers x over y (revealed preference). Then, we also learn that she does not pay attention to x at $\{x, y, z\}$ because she picked an inferior alterative y (revealed inattention).

We can see from the example, and the the definition of P_T that it is sufficient to have xP_Ty to conclude that x is revealed preferred to y, now the question we can ask ourselves is if there is some revealed preference not captured by P_T. The next proposition states that the answer is no: P_T is the revealed preference.

Proposition 1. *Suppose c is a CLC with a consideration filter. x is revealed to be preferred to y if and only if xP_Ty.*

Proof. We have already proven the if-part of Proposition 1. This immediately implies that P is acyclical since c is a CLC with a consideration filter. To see the only-if part, take any pair of x and y without xP_Ty. Then there exists a preference $\succ$ including P_T and $y \succ x$

since P_T is transitive. Define the consideration filter Γ^m as follows:[6]

$$\Gamma^m(S) = \{x \in S | \exists S \subset T \text{ s.t. } x = c(T)\} \tag{2}$$

By construction, Γ^m is a consideration filter. It is easy to see that $c(S)$ is $\succ$-best element in $\Gamma^m(S)$ because $c(S) \in \Gamma^m(S)$ and $x(\neq c(S)) \in \Gamma^m(S)$ only if $c(S)Px$ (so $c(S) \succ x$). □

Now we investigate when we can unambiguously conclude our decision maker pays (or does not pay) attention to an alternative. Note that if she chooses x from S, she must be paying attention to x at S. Therefore, we can determine that she pays attention to x at any smaller decision problem including x. On the other hand, suppose she has revealed to prefer x over y and chooses y from a set T including x. Then we can immediately conclude that she does not pay attention to x at T. Furthermore, this also implies that she does not pay attention to x at any decision problem larger than T.

Proposition 2 summarizes the above discussions and also provides full characterizations for revealed attention and inattention.

Proposition 2. *Suppose c is a Choice with Limited Consideration then:*

(i) x is revealed to attract attention at S if and only if x is chosen from some super set of S (possibly from S), i.e. $x = c(T)$ for $T \supseteq S$.

(ii) x is revealed not to attract attention at S if and only if x is revealed to be preferred to $c(T)$ for some T such that $x \in T \subset S$.

The if-parts of both statements have been already shown above. The proofs of the only-if-parts are given in Appendix.

A corollary of proposition 2 is that we can restate the condition of limited attention in terms of consideration filters, since we defined Γ^m to be the minimal consideration filter that is consistent with the choices and the structure we imposed on any Γ.

Corollary 1. *Let c be a CLC with a consideration filter. Then x is revealed to attract attention at S if and only if $x \in \Gamma^m(S)$.*

[6]We are going to use this construction of theconsideration filter Γ^m throughout the paper. Γ^m is the smallest possible consideration filter that we can infer from the choices, which is going to be important for the characterization for the characterization of revealed preference and attention.

4 CLC: Strong Consideration Filters

In this section we provide a characterization of CLC with strong consideration filters, which are consideration filters, that also satisfy the attention filter condition from Masatlioglu et al. (2009). Here we show that we gain predictive power, but not surprisingly we cannot incorporate some observed behavior such as "choosing pairwise unchosen" to the model. The additional condition that we add in this section to the consideration filter is that the removal of an alternative that is overlooked does not change the consideration set. Essentially the condition we add is that the DM maker is unaware that she is unaware.

Despite the similarities between our choice with limited consideration with a consideration filter, and MNO's CLC with an attention filter, the two models are independent. One particular feature of our model is that when we observe choice reversals, we conclude that a smaller menu (less options) are revealed to improve welfare, whereas this is not necessary the case in MNO.

Example 2. To show the independence between our model and Masatlioglu et al. (2009) we show here examples of consideration set formation where: (i) Γ is a consideration filter but not an attention filter, (ii) Γ' is an attention filter but not a consideration filter.

menu	xyz	xy	xz	yz
Γ :	x	xy	xz	y
Γ' :	xyz	xy	yz	z

Despite being two different models of consideration, most real world examples mentioned above and in Masatlioglu et al. (2009), satisfy both properties. Thus we study these properties in this section, where we define a Strong Filter, which satisfy both properties: consideration filter and attention filter. And derive all the revealed preference, and revealed attention results for those structures.

For instance an example of a consideration filter that is not an attention filter consider a DM looking for cars, she only considers those brands with at most 3 models. It is easy to see that this is a consideration filter, since if a car is considered for any subset of the available options, there will be at most 3 models of that brand, thus x will be considered. On the other hand, suppose x is not considered given the available options, thus there are 4 or more cars in the brand of x. If there are exactly 4 cars, removing x will lead the DM to consider the remaining 3 cars in the brand of x, thus this will not be an attention filter.

In addition, MNO give a formation of a consideration set that is an attention filter but not a consideration filter, which they call "Searching more when the decision is tough". Suppose the DM is looking for airline tickets. The DM considers alternatives that are easy to find (first page of search results in some travel website) and if there is an item that dominates all others in all dimensions (price, departure time, arrival, number of stops) then the DM picks that. Otherwise she spends time and does an extensive search to consider all available options (look at other sites, or airlines that are not affiliated with the travel website), and picks the preferred one. In this case to see why this might not be a consideration filter, suppose x is easy to find, but does not dominate any of the easy to find airline tickets, i.e. y is easy to find and cheaper, then DM will consider all the possible options. If we remove y, then x dominates all the easy options, thus the DM will not make the extensive search, and all the hard to find elements are considered in the first menu, and not considered in the submenu where we remove y.

The fact that many of the examples considered in this paper and in MNO satisfy both the attention and consideration filter properties, suggests that we should study the case where a consideration filter satisfies the following extra property that "an item that does not attract DM's attention does not affect her attention span at all"(Masatlioglu et al. 2009), this is called an attention filter by MNO.[7] This prompts the following definition of a strong filter:

Definition. *An consideration function* Γ *is called strong Consideration Filter if*

(i) For any S *and* T, $x \in \Gamma(T)$ *implies* $x \in \Gamma(S)$ *whenever* $x \in S \subset T$.

(ii) For any S, $x \notin \Gamma(S)$ *implies* $\Gamma(A) = \Gamma(A \setminus x)$.

Indeed, all the heuristic ways of generating a consideration set discussed in section 2 are strong filters, except the consideration filter where the DM considers all elements of a brand if there are most n goods of that brand. Thus is natural to characterize these type of choice functions.

Similarly we can define a choice function c as Choice with Limited Consideration with a strong consideration filter as

Definition. *A choice function* c *is a CLC with strong consideration filter if there exists a strict order and a consideration set mapping* Γ *such that*

$$c(S) = \max_{\succ} \Gamma(S)$$

and Γ *is a strong Consideration Filter.*

[7]The companion paper, Masatlioglu et al. (2009), extensively investigates consideration sets that satisfies only this property, called attention filter, and which is formally stated as $x \notin \Gamma(S) \Longrightarrow \Gamma(S) = \Gamma(S \setminus x)$.

First let us characterize the reveled preference when Γ is known to be a strong consideration filter. To do this, we revisit the cyclical choice behavior in the previous subsection[8] where we know that while x is revealed to preferred to y, there is no other revealed preference (see Proposition 1). Interestingly, we can uniquely pin down the preference for the cyclical choice example when Γ is a strong consideration filter.

To see this, first note that $c(xyz) = x$ implies that the DM pays attention to x at $\{x, y, z\}$ so does she at $\{x, z\}$ (revealed attention). Since she picks z from $\{x, z\}$, we can conclude that she prefers z over x (revealed preference). Since any strong consideration filter is an attention filter, we must have x is revealed to preferred to y. Therefore, her preference is uniquely pinned down: $z \succ x \succ y$.

Now we generalize this observation. Suppose $c(T) \neq c(T \setminus y)$. Then we conclude that y must be paid attention to at T. Since Γ is a strong filter, y must attract attention not only at T but also at any decision problem S smaller than T including y. Therefore, if $c(S) \neq y$, $c(S)$ is revealed to be preferred to y. Formally, for any distinct pair of x and y define:

$$xP'y \text{ if there exist } S \text{ and } T \text{ such that} \quad \begin{array}{l} (i)\ \ \{x, y\} \subset S \subset T \text{ and } x = c(S) \\ (ii)\ c(T) \neq c(T \setminus y) \end{array}$$

Notice that $c(T) \neq c(T \setminus c(T))$. This implies that $c(T)$ must have been considered not only at T but also at any decision problem S smaller than T including $c(T)$ since Γ is a strong filter. Therefore, whenever $\{c(T)\} \subseteq S \subset T$ and $c(T) \neq c(S)$, we have $c(S) \succ c(T)$. Indeed this is the way we infer z is better than x in the previous paragraph.

As before, if $xP'y$ and $yP'z$ for some y, we also conclude that she prefers x to z even when $xP'z$ does not hold. The following proposition states that the transitive closure of P', denoted by P'_T is the revealed preference.

Strong consideration filter captures the idea that an alternative that is not paid attention in a smaller set cannot attract attention when there are more alternatives. Hence, situations where presence of some alternatives reminds the DM the existence of some other alternatives are compatible with the attention filter but not with the strong consideration filter.

The example of "choosing pairwisely unchosen", studied in MNO, perfectly highlights this distinction between attention filter and strong consideration filter structures. Recall that we uniquely identify the attention filter for $\{x, y, z\}$, $\{x, z\}$, and $\{y, z\}$;

$$\Gamma(xyz) = xyz, \quad \Gamma(yz) = y, \text{ and } \Gamma(xz) = x.$$

[8] Recall the cyclical choice problem: $c(xyz) = y, \quad c(xy) = x, \quad c(yz) = y, \quad c(xz) = z.$

Here, z attracts attention only when both x and y are present. In other words, while z draws the attention from a big selection, it is not considered from a restrictive selection. Hence this is not a strong filter. A strong filter requires $z \in \Gamma(xyz)$ but $z \notin \Gamma(yz)$, hence we can immediately conclude that this choice behavior cannot be explained by a strong filter. We can also reach the same conclusion by using revealed preference: the DM's choice exhibits two choice reversals: (1) between $\{x, y, z\}$ and $\{x, z\}$ and (2) between $\{x, y, z\}$ and $\{y, z\}$. Based on Proposition 3, the first one implies that her preference must be $x \succ z \succ y$ and the second reveals $y \succ z \succ x$, which are contradicting.

Next, we provide a characterization for CLC with a strong consideration filter. The axiom we propose is a stronger version of LC-WARP. Remember that LC-WARP requires that every set S has the "best" alternative x^* and it must be chosen from any other decision problem T where the choice is part of S and is chosen in some problem that is more complex, i.e. $T' \supset T$. Remember that, with a consideration filter, an alternative, say x^*, attracts attention at a choice set, T if it is chosen from a superset $T' \supset T$.

Now that we assume the attention filter is strong, we can also conclude it when we know x^* is paid attention to at some decision problem $T' \supset T$, if removing it causes a choice reversal, by observing $c(T') \neq c(T' \setminus x^*)$. Therefore, we need to modify the requirement in LC-WARP: if the removal of x^* changes the choice in some super set of T, then it attracts attention at T (even if it is not chosen at T' or T).

Note that throughout this paper we use the term consideration as a concept related to the formation of consideration sets when too many options might overwhelm the DM, and she might not pay attention to some. In the companion paper, MNO, use the term *attention* as a concept related to formation of consideration sets in the presence of unawareness (unawareness of being unaware). Thus the similarity of the names of the behavioral postulates, LC-WARP and LA-WARP, as choice patterns that capture behavior under the different consideration structures. Therefore, it is only natural to define behavioral postulate that captures behavior under consideration sets that satisfy both properties (strong filters), as Limited Attention and Consideration WARP.

(A2) LCA-WARP. If for any nonempty S, there exists $x^* \in S$ such that for any $T \ni x^*$,

$$c(T) = x^* \text{ whenever } \quad \begin{array}{l} (i)\ c(T) \in S,\ \textit{and} \\ (ii)\ c(T') \neq c(T' \setminus x^*) \text{ for some } T' \supset T \end{array}$$

It turns out that LCA-WARP is the necessary and sufficient condition for CLC with a strong consideration filter. Indeed, it is equivalent to the acyclicity of the revealed preference, P'_T.

Theorem 2. *(Characterization) A choice function satisfies LCA-WARP if and only if it is a CLC with a strong consideration filter.*

Theorem 2 characterizes a special of class of choice behavior studied we studied earlier. Similar to Theorem 1, the characterization involves a single behavioral postulate which is stronger that WARP with Limited Consideration. We show that while this model has higher predictive power, which comes with diminishing explanatory power: "choosing pairwisely unchosen" is no longer within the model.

We finalize this section by revisiting the Attraction Effect. Consider the following observed choice behavior:

$$c(xyz) = y, \quad c(xy) = x, \quad c(yz) = y, \quad c(xz) = x.$$

It is routine to verify that this choice behavior satisfies LCA-WARP.[9] Hence Theorem 2 implies that it is consistent with a CLC with a strong consideration filter. The choice reversal between $\{x, y, z\}$ and $\{x, y\}$ yields that her preference must be $x \succ y \succ z$.

In addition, one can derive the unique consideration set mapping. To see this, consider the set $\{x, y, z\}$. First of all, the choice, which is y, must be in the consideration set. Since removing z changes the choice, therefore z is also in it (attention filter). Finally, we know x is better than the choice from above discussion, x does not belong the consideration set of $\{x, y, z\}$. Hence $\Gamma(xyz) = yz$. In addition, the strong consideration filter assumption requires that y and z attract attention whenever they are available, which pins down the consideration set mapping uniquely for this example.

$$\Gamma(xyz) = yz, \quad \Gamma(xy) = xy, \quad \Gamma(yz) = yz, \text{ and } \Gamma(xz) = xz.$$

We argue that the benefit of a CLC with a strong consideration filter, is that we gain in predictive power, in terms of what we can pin down as revealed attention and revealed preference. We can see that in both, CLC with a consideration filter and CLC with a strong consideration filter, we can conclude revealed attention from choice reversals, however with the consideration filter we can only conclude attention for the chosen options that constitute the choice reversal, whereas in the CLC with a strong attention filter, the fact that removing an alternative that is not chosen causes a choice reversal also reveals attention. Moreover,

[9] One can show that x serves the role of x^* for $\{x, y, z\}$. For the rest, $c(S)$ does the job.

if we can pin down more of the revealed attention, if the choice function is a CLC, we can also pin down more of the preference as shown in the Attraction Effect example above.

The cost of having better identification in terms of what we can conclude that the DM observed, is that we lose some types of behavior considered under the two models, since as mentioned above, some behavior is captured by a consideration filter and not an attention filter and vice versa. However most examples from this paper, and MNO, satisfy both properties.

In the CLC with a consideration filter, we can only pin down part of the attention and preference, and in the case where choices satisfy WARP we cannot conclude anything. We provide a full characterization of the extent of characterizing attention and consideration, and the tension between attention and preference in these models, especially when the choice rules satisfy WARP in appendix A.2.

Similarly to the CLC with a consideration filter from section 3, when we have a strong consideration filter, instead of a consideration filter, we can also characterize revealed preference from LCA-WARP.

Proposition 3. *Suppose c is a CLC with a strong consideration filter. Then, x is revealed to be preferred to y if and only if xP'_Ty.*

Proof. The if-part has been already demonstrated. The only-if part can be shown paralleled with Theorem 2, where we shall show that any $\succ$ including P'_T represents c by choosing Γ properly. □

5 Additional Discussion

In this section two independent special cases of our model. The first one assumes that the decision maker has no limited attention problem in a binary set, hence she pays attention to both alternatives. In the second one, the consideration filter is generated by a transitive order, which might conflict with the preferences.

5.1 Full Consideration in Binary Comparisons

Although choice overload is usually attributed to the number of options presented to decision makers, another source of choice overload is the number of attributes. Therefore, even with small number of alternatives, one may not compare all available alternatives. Our framework allows such an extreme case.

However, if the source of attention is just abundance of alternatives, it is more likely that she considers all of alternatives in smaller decision problems. As a benchmark case, we consider a decision maker who has an attention filter but pays attention to both alternatives in every binary decision problem. That is, $\Gamma(S) = S$ whenever $|S| = 2$. We now provide a characterization for this class of choice function.

With the full consideration assumption, our decision maker's choices in binary sets must be consistent with preference maximization for some preference. Hence, it is needed to assume that

(A3) Pairwise Consistency. If $c(xy) = x$ and $c(yz) = y$ imply $c(xz) = x$.

As we discussed earlier, an alternative, $c(T)$, is revealed to attract attention at a set S whenever T is a super set of S. If $c(T)$ is not the chosen one, then it is strictly worse than $c(S)$. This information should not conflict with binary data, that is, $c(S)$ must be the choice from $\{c(S), c(T)\}$.

(A4) Weak Contraction. If $c(S) \in T$ and $S \subset T$ then $c(S) = c(\{c(S), c(T)\})$.

This axiom is trivially satisfied in the standard theory, where $c(S) = c(T)$. Here, we require a weaker version of it, because we need to know whether the alternative attracts attention to reach the same conclusion.

Theorem 3. *A choice function satisfies A3 and A4 if and only if it is a CLC with a consideration filter where there is full consideration at binary sets.*

There are two important implications of this theorem: (i) more predictive power and (ii) a unique preference. As previously mentioned, on the appendix A.2 we show that there are limitations to the predictive power of the model, since many consideration filters, and preferences can represent a CLC with a consideration filter. For instance, our earlier model, CLC with consideration filter, can accommodate any type of behavior with three elements, so it is not falsifiable. Because of Pairwise Consistency, cyclical choice behavior in binaries is ruled out in this model. Hence the model can be tested even with three options.

As in the standard theory, where the DM is assumed to pay attention to all the alternatives, it is possible to infer her preference by asking the choice from the sets of two alternatives:

$$xP^*y \text{ if } c(xy) = x \tag{3}$$

In section 3, we illustrate that different preferences might generate the same behavior. In the extreme case, where WARP is satisfies, there is no revealed preference at all. This is because a choice can be attributed either to preference or to inattention. To illustrate this, consider the choice function with three elements satisfying WARP:

$$c(xyz) = x, \quad c(xy) = x, \quad c(yz) = y, \quad c(xz) = x.$$

One possibility is that her preference is $x \succ y \succ z$ and she considers everything. Another possibility is that her preference is $z \succ y \succ x$ and she considers $c(S)$ at S. By full consideration assumption at binaries, one can pin down the true preference even WARP is satisfied (see Appendix A.2 to see the limitations of the standard CLC with consideration filter model when c satisfies WARP).

5.2 Attention Filters Generated by a Transitive Order

Here, we also consider a natural special case whereby the decision maker overlooks or disregards an alternative because it is dominated by another item in some aspect. Imagine Maryland's economics department is hiring one tenure-track theorist. Since there are too many candidates in the market, the department asks other departments to recommend their best theory student. Therefore, a candidate from Michigan is ignored if and only if there is another Michigan candidate who is rated better by Michigan. In this case, Maryland's filter is represented by a irreflexive and transitive order as long as each department's ranking over its students is rational. However, the order does not compare any two candidates from different schools so it is not complete.[10] Notice that this order may not be consistent with the preference of Maryland. It is possible that Michigan evaluates its job candidates differently than Maryland, in which case Maryland may eliminate its preferred candidate. Therefore, the order and the preference may be inconsistent.

Formally, let $\rhd$ be an irreflexive and transitive order over X and $\Gamma_{\rhd}$ be an consideration filter generated by $\rhd$, that is:

$$\Gamma_{\rhd}(S) = \{x \in S | \nexists y \in S \ \text{ s.t. } \ y \rhd x\},$$

for all $S \in \mathcal{X}$. Here, the decision maker does not consider x at decision problem S if and only if there is another alternative $y \in S$ that dominates x according to the transitive order. It is easy to see that $\Gamma_{\rhd}$ is indeed a special class of strong consideration filters.

[10]The special case in which the rationale always yields a unique maximal element corresponds to the standard model of rationality.

Here we illustrate that $\Gamma_{\rhd}$ is a strong consideration filter for any irreflexible transitive order $\rhd$. First $\Gamma_{\rhd}(T)$ is a subset of $\Gamma_{\rhd}(S)$ for all $S \subset T$. To see this, assume $z \in \Gamma_{\rhd}(T)$. Then there exists no alternative in T $\rhd$-dominates z, which implies that z is $\rhd$-undominated in any subset of T, so $z \in \Gamma_{\rhd}(S)$ for all $S \subset T$. Particularly, we have $\Gamma_{\rhd}(S) \subset \Gamma_{\rhd}(S \setminus x)$ for any $x \in S$. Now we need to show that $\Gamma_{\rhd}(S \setminus x) \subset \Gamma_{\rhd}(S)$ when $x \notin \Gamma_{\rhd}(S)$. Suppose $x, y \in S \setminus \Gamma_{\rhd}(S)$. Then, there must exist $z \in S \setminus x$ such that $z \rhd x$. If $x \rhd y$, then by the transitivity, $z \rhd y$ as well so $y \notin \Gamma_{\rhd}(S \setminus x)$. If it is not $x \rhd y$, then what eliminates y at S is also included in $S \setminus x$ so $y \notin \Gamma_{\rhd}(S \setminus x)$. Therefore, $\Gamma_{\rhd}$ is a strong consideration filter.

Since $\Gamma_{\rhd}$ is a strong consideration filter, LCA-WARP is a necessary condition for CLC with an attention filter generated by a transitive order. In addition to that there is another necessary condition:

(A5) Expansion. If $x = c(S) = c(T)$, then $x = c(S \cup T)$.

Manzini and Mariotti (2007) dub this property Expansion, and it directly rules out Attraction Effect type of anomalies. It says that an alternative chosen from each of two sets is also chosen from their union. To see that it is necessary, assume $(\Gamma_{\rhd}, \succ)$ represents c and $x = c(S) = c(T)$. The latter implies that x is the $\succ$-best element in both $\Gamma_{\rhd}(S)$ and $\Gamma_{\rhd}(T)$. Hence x is $\rhd$-undominated in both S and T, so x is in $\Gamma_{\rhd}(S \cup T)$. Since $\Gamma_{\rhd}(S \cup T) \subset \Gamma_{\rhd}(S) \cup \Gamma_{\rhd}(T)$, x is also the $\succ$-best $\Gamma_{\rhd}(S \cup T)$. Hence $x = c(S \cup T)$.

Therefore, if a consideration filter of our decision maker is generated by a transitive order, then her choice must satisfy Expansion, as well as A2. Indeed its converse is true so these two axioms characterize such choice functions.

Theorem 4. *A choice function satisfies LCA-WARP and Expansion if and only if it is a CLC with a strong consideration filter, which is generated by a transitive order.*

Now, we discuss the revealed preference and the revealed order. Notice that this is a special case of CLC with a strong consideration filter, P'_R, which is the revealed preference for the strong consideration filter must be a part of the revealed preference for this model and it turns out that there is no extra inference of DM's preference.

However, we can now obtain the revealed order: if all of attention filters generated by some transitive order that can represent the choice agree on $x \rhd y$, we call it the revealed order. The revealed order can be obtained in a simple way. If she picks x from $\{x, y\}$ but reveals that she prefers y over x, it must be the case that y is disregarded at $\{x, y\}$. Therefore, we can conclude $x \rhd y$.

Proposition 4. *Suppose c is a CLC with a consideration filter that is generated by a transitive order.*

- *x is revealed to be preferred to y if and only if $xP'_R y$.*
- *$x \rhd y$ is revealed if and only if yP'_R but $x = c(xy)$.*

Finally, we argue that the class of choice behavior characterized in Theorem 4 is a specific subclass of Manzini and Mariotti (2007)'s rational shortlist method. Similar to our model, the shortlist method operates through two binary relations P_1 (acyclic) and P_2 (asymmetric): the decision maker filters out P_1-dominated alternatives and selects P_2-best among them. Unlike our model, P_1 may be intransitive and P_2 may have a cycle. Hence, our model is a rational shortlist method where P_1 is transitive and P_2 is a preference. Indeed, it is a strict subset of shortlist method. Before we illustrate this, we remind the second axiom, Weak WARP, used in the characterization of the rational shortlist method. The axiom says that if an alternative x is chosen both when only y is also available and when y and other set of alternatives, T, are available, then y is not chosen from any subset of T whenever x is available. Formally,

Weak WARP. Suppose $\{x, y\} \subset S \subset T$. If $x = c(xy) = c(T)$, then $y \neq c(S)$.

We end this section, with a list of examples that show the independence of our axioms and the ones presented by

Here is the example in which Weak WARP and Expansion are satisfied but not LCA-WARP, hence our model is a strict subclass of their model.

Example 3. There are five alternatives: a, b, c, d, x. The decision maker has two rationales: one is acyclic $P_1 = \{(c, a), (d, b), (a, x)\}$ and the second one is asymmetric including $P_2 = \{(a, b), (b, c), (c, d), (d, a), (x, d)\}$. Note that P_2 is cyclical. The decision maker sequentially applies P_1 and P_2 to make a choice as in the shortlisting method.[11]

Now we show that this choice behavior violates A2 at $S = \{a, b, c, d\}$. In other words, there is no alternative in S which serves the role of x^* in the axiom. For example, the alternative a changes the choice ($c(dx) \neq c(adx)$) but it is not chosen ($c(ad) \neq a$ and $\{a, d\} \subset \{a, d, x\}$).

[11] One can define P_2 completely so that there is a unique survivor of this two-stage elimination, hence this is a well-defined choice function that is a rational shortlist method. Since our aim is to show that it violates LCA-WARP, we define necessary part of P_2.

Likewise, the following example, show behavior that satisfies Weak WARP and Expansion but violates LC-WARP.

Example 4. There are six alternatives: a, b, c, d, x, y. The decision maker has two rationales: P_1: aP_1c, bP_1d, xP_1a, and yP_1b. P_2: aP_2b, bP_2c, cP_2d, dP_2a, xP_2y and all of a, b, c, d P_2-dominates x and y.

Notice that P_1 is acyclic and P_2 is asymmetric (but cyclic). The decision maker sequentially applies P_1 and P_2 to make a choice. We argue that exactly one element survives this process so her choice is always uniquely determined. Since P_1 is acyclic, at least one element survives in the first round of elimination. By inspecting P_2, we can see no alternative survives in the second round if and only if all of a, b, c, d survive in the first round. Similarly, we can see that more than one alternative survive in the second round only if the survivors of the first round include either "a and c" or "b and d." However neither of them is possible because of aP_1c and bP_1d. Therefore, there is a unique survivor of this two-stage elimination, so this is a well defined choice function that is a rational shortlist method. Hence, it satisfies both Weak WARP and Expansion. However, this violates LC-WARP at $S = \{a, b, c, d\}$ because "$a = c(ab)$ but $b = c(abx)$," "$b = c(bc)$ but $c = c(bcy)$," "$c = c(cd)$ but $d = c(acd)$," and "$d = c(ad)$ but $a = c(abd)$."

Finally, this last example satisfies LC-WARP and Expansion but not LCA-WARP.

Example 5. The decision maker's preference $\succ$ is $a \succ b \succ c \succ d$. Her attention filter is generated by the following $\rhd$: $d \rhd c \rhd b \rhd a$ but it is not transitive. Then, $\succ$ and $\rhd$ represent a choice with limited attention whose attention filter is generated by $\rhd$ so it satisfies LC-WARP and Expansion. However, it violates LCA-WARP at $S = \{b, d\}$ because "$d = c(abcd)$ but $a = c(acd)$" and "$b = c(bd)$ but $d = c(bcd)$.

6 Related Literature

Samuelson (1938) first introduced the Weak Axiom of Revealed Preference (WARP) as a novel approach to derive demand functions. Based on this idea of revealed preference analysis, choice theory developed with the pioneering work of Arrow (1959), Richter (1966), Hansson (1968) and Sen (1971). There are several other papers that use the classical choice framework to analyze choice behavior that does not fit the rational model, but it is nonetheless motivated by experimental and psychological evidence.

Kalai et al. (2002) study choice behavior in terms of number of preference orders necessary to rationalize choice data. Cherepanov et al. (2008) present a model of rationalization, in which preferences over outcomes are well defined, but there are some rationalization (psychological) constraints. Hence a good element might not be chosen because it cannot be rationalized represented as "being able to give an explanation for". de Clippel and Eliaz (2009) offer a model of personal bargaining across selves that generate choices that might fail to be rational in the classical sense. Masatlioglu and Ok (2005) consider choice with statu-quo bias, and how the presence of an element considered as the statu-quo, can influence choice behavior. This idea of reference dependence is choices is studied also by Ok et al. (2008).

Manzini and Mariotti (2007), Mandler et al. (2008), and Manzini and Mariotti (2009) provide models where the DM sequentially eliminates objects, according to some process, and the makes a final choice from the remaining ones. In this models, unlike in our model, the DM considers all the elements, and uses some process (binary relation) to eliminate some before having to make the final choice.

The lack of consideration of some options, plays an relevant role in several papers. Eliaz and Ok (2006) present a choice framework where it is possible to distinguish indifference and indecisiveness from choice data. Masatlioglu and Nakajima (2010) consider a model of choice by elimination, where the consideration sets evolve as some alternatives are sequentially eliminated. Eliaz et al. (2009) consider a model where the consideration sets of DMs are directly observed, rather than partially identified like in our model.

Finally, the effect on final choices of a large set of options is featured in some other papers like Dean and Caplin (2008), and Dean (2008), where the consideration of elements is a procedure based on search costs. Also, Tyson (2008), offers a model of "Satisfycing" (see Simon (1955)) where the preferences are generated by systems of binary relations, which are indexed by choice sets. He also requires some consistency between preferences for going from large choice sets to its subsets, similar to our behavioral requirement on Consideration Filters. Finally, as previously mentioned, the companion paper Masatlioglu et al. (2009) studies a similar choice environment, with different requirements on the consistency of consideration filter to analyze some other types of observed behavior.

7 Conclusion

Consumers do not consider all the available alternatives. They intentionally or unintentionally ignore some of the alternatives and focus on a limited number of alternatives. In this paper, we relax the full consideration of the standard choice theory to allow for the choice with limited consideration.

Marketing and finance literatures argue that the abundance of alternatives is the basic motive for limiting the consideration set. While limiting the consideration, it is well documented that different types of filters have been used by the consumers. As motivated by the real life examples, we provide characterization for these different types filters: (i) If a consumer considers an alternative among a large set of option, he will stil continue to considering the same alternative when some alternatives become unavailable. (ii) If an alternative that the consumer does not consider becomes unavailable, his consideration set will not be affected. (iii) Consumer has some categories, and he only considers the top options according to his categories. (iv) Consumer is able to fully consider only limited number of alternatives, e.g. when he is confronted with two alternatives, he can consider both of the alternatives, but when there are more than two alternatives, he may not be able to consider some of them. Although the consideration sets are not observable, our axiomatic approach enables the identify which filters are used by the decision makers simply by observing their choices. Identification of the filters will help companies to develop new marketing strategies such that their products will attract attention by the consumers. Additionally, we show that choice with limited consideration is capable of explaining behavioral anomalies that look puzzling under standard choice theory.

A Appendix

A.1 CLC: Choice Correspondences

In the following section we show that the approach can be extended to choice correspondences. Here we generalize the main characterization and revealed preference/attention results from the paper, the characterization of CLC, by allowing choices to be multivalued, and choice sets be arbitrary. The intuition behind the representations is the same, thus we present them first for functions to gain better intuition without the technical complications of the correspondences and arbitrary choice sets.

All the proofs of this section are included in this section.

In this section we consider the general case where X is the (possible infinite) choice set. Again $\mathcal{X}$ is the set of all proper subsets of X. Similarly, a choice correspondence will be given by $c : \mathcal{X} \to X$, such that $c(S) \in S$ for every $S \in \mathcal{X}$. Now we let $\succsim$ be a weak order linear order on X. Also we denote the best element in S with respect to $\succsim$ by $\max_{\succsim} S$. Here we show that all the results hold for the more general case with only a few changes and additions since we have revealed indifference as well. Nonetheless the characterization of CLC for correspondences is not as clean an straightforward as the one for functions and finite X, thus here we just give the results and all the motivation and discussion in on section 3.

The difference between this setup and what we presented in section 3 is that we are allowing for choices to be multi-valued (choice correspondences), and the choice set to be arbitrary. When choices are multi-valued, instead of the linear order presented before we need to consider weak orders to allow for the possible indifference relation.

We showed that the only behavioral postulate that characterized CLC with a consideration filter was (LC)WARP. Thus we need to modify the LC-WARP presented before. The first natural modification allows for multivalued choices, and as we will see, extends the concept of revealed indifference. This extension, given by the following axiom, requires choices to be consistent across choice sets. Making a parallel with LC-WARP, if there is more than one element, x^*, y^* satisfying the property for LC-WARP for a set S, then those two elements must be satisfy (or not) LC-WARP together for any set that contains both of them. In other words, it does not allow for choice reversals for one element but not the other, once both elements have been chosen at some point.

> **(A6) - Weak Revealed Indifferences (WRI).** If for $\{x, y\} \subseteq T \subset S$, $x, y \in c(S)$, then $x \in c(T)$ implies $y \in c(T)$.

One consequence of WRI, is that whenever there is a choice reversal then the intersection of the two choice sets must be empty, and thus we are be able to distinguish between (strict) revealed preference and revealed indifference from choice data, as we will see.

Lemma 2. *Let c satisfy WRI. Then $x \in c(T)$ and $x \notin C(S)$ for some $S \subset T$ implies $c(S) \cap c(T) = \emptyset$.*

Proof. Suppose there exists $y \in C(T) \cap c(S)$, then by WRI we have $x \in c(S)$, since $x, y \in c(T)$ and $y \in c(S)$ which would be a contradiction. Therefore $c(T) \cap c(S) = \emptyset$. □

The second axiom that we want to introduce to characterize CLC for correspondences is called No Cyclic Choice Reversals. This is a stronger condition than LC-WARP, since it guarantees that not only LC-WARP is satisfies, but also, once there is a choice reversal we cannot find a chain of pairwise comparisons by either indifference or more choice reversals (which in this model characterize preference), that would imply a reversal going the other way. So if we reverse the choices from $c(T')$ to something in $c(T)$, so $x \in c(T')$ and $x \notin c(T)$ for some $T' \supset T$, then we won't be able to indirectly reverse the choice the other way, from something in $c(T)$ to x or anything in $c(T')$. It is straightforward to see that NCCR implies LC-WARP for choice functions.

(A7)- No Cyclic Choice Reversal(NCCR). If for any set of menu-submenu pairs $\{S_i, T_i\}$ (so that $T_i \subseteq S_i$), where $c(T_{i+1}) \cap c(S_i) \neq \emptyset$, for $i = 1, ..., n-1$. Then if $c(T_i) \cap c(S_i) = \emptyset$ for one i we have $c(T_1) \cap c(S_n) = \emptyset$

Again, we define two binary relations in a similar fashion to the choice function case: P (an asymmetric relation) which is the same P as the function case given in 1, and I, a symmetric relation that we will see is going to capture the idea of revealed indifference.

Definition. *Given a choice correspondence c define two binary relations, P and I the following way.*

1. *xPy if there exists $\{x, y\} \subseteq S \subset T$ such that $x \in c(S)$, $y \notin c(S)$, and $y \in c(T)$.*
2. *xIy if there exists $\{x, y\} \subseteq S$ such that $x, y \in c(S)$.*

First of all, we can see that if a choice correspondence satisfies our two axioms, we have that P and I are disjoint.

Proposition 5. *If c satisfies NCCR and WRI, then $P \cap I = \emptyset$.*

Proof. Let xPy, then there exists $\{x, y\} \subseteq S \subset T$ such that $x \in c(S)$, $y \notin c(S)$, and $y \in c(T)$. Suppose there exists $T' \in \mathcal{X}$ such that $x, y \in c(T')$. By lemma 2, $c(T) \cap c(S) = \emptyset$, and $\{x, y\} \subseteq T'$, then $c(xy) = xy$ by WRI. And by NCCR, since $c(xy) \cap c(S) \neq \emptyset$, we have $c(xy) \cap c(T) = \emptyset$, which is a contradiction. So there does not exists such a T', and therefore $\neg(xIy)$.

Let xIy, then there exists $S \supseteq \{x, y\}$ such that $x, y \in C(S)$. By WRI we must also have $c(xy) = xy$. Suppose there exists $\{x, y\} \subseteq S \subset T$ such that $x \in c(S)$, $y \notin c(S)$, and $y \in c(T)$. Then by lemma 2 we must have $c(S) \cap c(T) = \emptyset$. And $c(S) \cap c(xy) = x$ implies $c(T) \cap c(xy) = \emptyset$, which is a contradiction since $y \in c(T)$ and $c(xy) = xy$. Therefore no such $S \subset T$ exists, thus $\neg(xPy)$. □

We can define a new binary relation R by taking the union of P and I. Intuitively, we are taking a symmetric relation that captures some notion of indifference when there is limited observation and and asymmetric relation that captures some notion of strict preference (in terms of choice reversals as discussed in section 3).

Lemma 3. *Let $R = P \cup I$, then xRy if and only if there exists $\{x, y\} \subseteq S \subseteq T$ such that $x \in c(S)$ and $y \in c(T)$.*

Proof. This follows from the definitions of P and I. □

Hence we can see that the two axioms are equivalent to not being able to find two conflicting choice reversals. In parallel to the function case, this will imply that once we take the transitive closure of R, we will not have cycles.

Proposition 6. *c satisfies NCCR, and WRI if and only if* $x_nRx_{n-1}R...Rx_2Rx_1$ *implies* $\neg(x_1Px_n)$.

Proof. ($\Rightarrow$). Consider $x_i \in X$ such that $x_nRx_{n-1}R...Rx_2Rx_1$. Then by the definition of R, there must exists for each $i = 2, ..., n$ sets and subsets $S_i \supseteq T_i$ with $x_i \in c(S_i)$ and $x_{i-1} \in (T_i)$ (see lemma 3).

Suppose x_1Px_n, then there exists $\{x, y\} \subseteq S' \subset T'$ such that $x_1 \in c(S')$, $x_n \notin c(S')$, and $x_n \in c(T')$. By WRI, $c(S') \cap c(T') = \emptyset$. Let $S_1 = S'$ and $T_1 = T'$, then by NCCR $c(S_n) \cap c(T_1) = \emptyset$, but $x_n \in c(S_n)$ by definition, and $x_n \in c(T_1)$ by x_1Px_n. Contradiction. Therefore $\neg(x_1Px_n)$.

($\Leftarrow$). Let $x, y \in C(T)$ and $x \in c(S)$ for some $S \supseteq \{x, y\}$. Then we have yIx, which by definition of R implies yRx. If $y \notin c(S)$ then by definition of P we have xPy, but this is a contradiction since yRx implies $\neg(xPy)$ by the condition. So c satisfies WRI.

Let $S_i \subseteq T_i$ such that $x_i \in c(S_i)$ and $x_{i-1} \in c(T_i)$ for $i = 2, ..., n$ and $x_1 \in c(S_1)$. So we have $c(S_i) \cap c(T_{i+1}) \neq \emptyset$ for all $i = 2, ..., n$. This implies $x_nRx_{n-1}R...Rx_2Rx_1$. Now we prove the contrapositive, let c fail NCCR. WLOG let $x_n \in c(T_1)$ and $c(T_1) \cap (S_1) = \emptyset$, so $c(S_n) \cap c(A_1) \neq \emptyset$, and for one of the S_i, T_i, $c(S_i) \cap c(T_i) = \emptyset$. Then we have $x_nRx_{n-1}R...Rx_2Rx_1$, and since $x_n \in c(T_1)$ and $x_n \notin c(S_1)$, and $x_1 \in c(S_1)$, by definition of P x_1Px_n. This fails the condition that $x_nRx_{n-1}R...Rx_2Rx_1$ implies $\neg(x_1Px_n)$. □

Just like in the function case, let R_T be the transitive closure of R. Just like in the functions case, we show that if R satisfies the condition of proposition 6, then we can extend R_T so that c is represented by some $(\Gamma, \succsim)$ where Γ is a consideration filter and $\succsim$ is a weak order that contains R_T.

First, we show that for any $\succsim$ that can possibly represent a CLC correspondence c, then $\succsim$ needs to include R_T.

Proposition 7. *Suppose c is a CLC represented by* $(\Gamma, \succsim)$. *Then* $R_T \subseteq \succsim$.

Proof. Let xR_Ty, then $z_1, ..., z_n$ such that $x = z_1Rz_2R...Rz_n = y$ (possibly $z_2 = y$, in which case xRy). For any z_iRz_{i+1}, there exists $\{z_i, z_{i+1}\} \subseteq S_i \subseteq T_i$ such that $z_i \in c(S_i)$ and $z_{i+1} \in c(T_i)$. Since Γ_i is an attention filter, $z_i, z_{i+1} \in \Gamma(S_i)$; and $z_i \in c(S_i)$ implies that z_i is $\succsim$ $-$maximal in $\Gamma(S')$, i.e. $z_i \succsim z_{i+1}$ because c is a CLC correspondence. Therefore $x = z_1 \succsim z_2 ... \succsim z_n = y$, and by transitivity of $\succsim$, $x \succsim y$ follows. Therefore $R_T \subseteq \succsim$. □

The following theorem shows that CLC behavior when allowing for choice correspondences is completely characterized by the two axioms NCCR and WRI.

Theorem 5. *A choice correspondence c is a CLC if and only if c satisfies NCCR and WRI.*

Proof. ($\Rightarrow$). First we show necessity of the two axioms. Let c be a CLC represented by $(\succsim, \Gamma)$, where Γ is a consideration filter.

To prove NCCR, let $T_i \subseteq S_i$ be a set of menus such that $c(S_i) \cap c(T_{i+1}) \neq \emptyset$. Without loss let $c(S_1) \cap c(T_1) = \emptyset$. Let $x_i \in c(S_i)$ and $y_i \in c(T_i)$ be elements of the respective choice sets.

Since c is a CLC, the information $c(S_1) \cap c(T_1) = \emptyset$, $c(S_i) \cap c(T_{i+1}) \neq \emptyset$, and $c(S_i) \cap c(T_{i+1}) \neq \emptyset$ tells us $x_i, y_i \in \Gamma(T_i)$ for all i and therefore we can conclude

$$y_1 \succ x_1 y_i \succsim x_i \ \forall i x_{i+1} \sim y_i \ \forall i$$

Therefore we have $y_n \succsim x_n \sim y_{n-1} \succsim ... \succsim y_2 \succsim x_2 \sim y_1 \succ x_1$. Since $\succsim$ is a weak order, we must have $y_n \succ x_1$. For any $z \in c(S_n)$, $z \sim y_n$ since c is a CLC. So $z \succ x_1$ and $z \succ w$ for any $w \in c(T_1)$. This implies that for all $w \in c(T_1)$, $w \notin c(S_n)$. Similarly for any $z \in c(T_1)$, $z \sim x_1$ and since c is a CLC represented by $(\succsim, \Gamma)$, $w \succ z$ for all $w \in c(S_n)$, and we get $z \notin c(T_1)$ since c is a CLC. Therefore $c(S_n) \cap c(T_1) = \emptyset$.

Now we prove the necessity of WRI. Let c be a CLC represented by $(\succsim, \Gamma)$. Suppose $x, y \in c(S)$ for some $\{x, y\} \subseteq T \subset S$. Then $x, y \in \Gamma(S)$ and since Γ is a consideration filter, $x, y \in \Gamma(T)$. Given that $x, y \in \Gamma(S) \cap (S)$, and there is a weak order $\succsim$ such that $c(S) = \max_{\succsim} \Gamma(S)$ for all S, we must have $x \sim y$. Let $x \in c(T)$, then for all $z \in \Gamma(T)$, $x \succsim z$. By transitivity $y \succsim z$ for all $z \in \Gamma(T)$, since $x \sim y$ and $y \in \Gamma(T)$. Therefore $y \in c(T)$. Which means that c satisfies **WRI**.

($\Leftarrow$). We can construct a weak order $\succsim$ if c satisfies NCCR and WRI, such that $(\Gamma^m, \succsim)$ represent c. This is proved in propositions 8 and 9. □

Similarly to the definition of revealed preference on section 3. We can define revealed (strict) preference and revealed indifference.

Definition. *Let c be a CLC correspondence and that there are k different attention filter, weak orders representing c*

$$(\Gamma_i, \succsim_1), (\Gamma_2, \succsim_2), ..., (\Gamma_k, \succsim_k)$$

For such a c we define the use the same definition for revealed preference, attention and inattention as for the choice functions but also distinguish between the revealed strict preference and revealed indifference.

1. *x is **revealed preferred** to y, $x \succsim_R y$, if $x \succsim_i y$ for all i.*
   - *x is (strictly) **revealed preferred** to y, $x \succ_R y$, if $x \succ_i y$ for all i.*
   - *x is **revealed indifferent** to y, $x \sim_R y$, if $x \sim_i y$ for all i.*
2. *x is **revealed to attract attention** at S if $x \in \Gamma_i(S)$ for all i.*
3. *x is **revealed NOT to attract attention** at S if $x \notin \Gamma_i(S)$ for all i.*

In a similar fashion to the characterization for CLC functions and revealed attention and preferences in sections 3.1 and 3, we show that our axioms guarantee that the binary relation R, does not have a cycle. So if there is a chain of weakly related elements, then there cannot be a strict relation that creates a cycle.[12] The following result is the analogue for correspondence as the result proved in proposition 6.

Now we know that from observe choice data, R_T is identifiable. The next propositions show the analogue Revealed Preference and Revealed Attention results for correspondence that tell us that R_T actually all you can distinguish form choice data.

Proposition 8. *Let $(\Gamma, \succsim)$ represent c. Let R_T be the transitive closure of R. Let c be a CLC correspondence. Then x is revealed preferred to y ($x \succsim_R y$), if and only if xR_Ty.*

Proof. ($\Rightarrow$). We prove the contrapositive. Suppose $\neg(xR_Ty)$. Note R_T is not necessarily complete we can have i) yR_Tx or ii) $\neg(yR_Tx)$. First define

$$\Gamma^m(S) = \{x \in S | x \in c(T) \text{ for some } T \supseteq S\}$$

Since $\neg(xR_Ty)$, there is no $S \subseteq T$ such that $x \in c(S)$ and $y \in c(T)$, thus there is no S such that $x, y \in \Gamma^m(S)$ by the construction of Γ^m.

Now it is possible to construct a weak order, $\succsim$ such that $y \succ x$ and $(\Gamma^m, \succsim)$ represents c. First add (y, x) to R_T,[13] Thus we have $R' = R_T \cup (y, x)$, and denote R'_T to be its transitive closure. We claim that $(x, y) \notin R'_T$. To see this, $\neg(xR_Ty)$ means that $(x, y) \notin R_T$. So suppose $(x, y) \in R'_T$, since $(x, y) \notin R_T \cup (y, x)$, then there exists $z_1 = x, z_2, ..., z_n = y$ such that $x = z_1R'z_2R'...R'z_n = y$. $(x, y) \notin R_T$ implies that we must have (y, x) somewhere in this chain. So $z_i = y, z_{i+1} = x$ for some $i < n$. But then we have $x = z_{i+1}R_Tz_{i+2}...R_Tz_n = y$, which implies that $(x, y) \in R_T$, contradiction. So adding (y, x) ro R_T doesn't add (x, y) to the transitive closure of $R_T \cup (y, x)$, R'_T.

[12]Recall that R is not necessarily complete.
[13](x, y) might be in R_T already.

Then by the existence of a complete extension of R'_T as a corollary of Szpilrajn (1930) (see Ok (2004)), we can construct a weak order $\succsim$, where $y \succ x$. Therefore we have $\neg(xR_Ty)$, since $y \succ x$, if $(\Gamma^m, \succsim)$ in fact represents c. Now we show that this is a CLC representation of c.

First let $x \in c(S)$, we want to show that x is $\succsim$-maximal in $\Gamma^m(S)$. First, $x \in \Gamma^m(S)$ by construction. Let $y \in \Gamma^m(S)$, then $y \in (T)$ for some $T \supseteq S$, therefore xRy by definition of R. Since $\succsim$ is a complete extension of the transitive closure of R, $x \succsim y$ follows by construction.

Now let $x \in S$ such that $x \succsim y$ for all $y \in \Gamma^m(S)$, we want to show that $\nexists z \in c(S)$ such that $z \succ x$. This is almost immediate by the way we constructed Γ^m. Note that $x \in \Gamma^m(S)$ implies $x \in c(T)$ for some $T \supseteq S$. Likewise if $\exists\ z \in c(S)$ such that $z \succ x$ we have $x, z \in \Gamma^m(S)$. However, $x \succsim z$ for all $z \in \Gamma^m(S)$, therefore no such z exists, and we can conclude that $(\Gamma^m, \succsim)$ represents c.

($\Leftarrow$). Let xR_Ty then by proposition 7, for any $(\Gamma_i, \succsim_i)$ representing c, $x \succsim_i y$ which equivalent to $x \succsim_R y$, x is revealed preferred to y. □

Corollary 2. *Let I_T and P_T be the symmetric and asymmetric components of R_T respectively.*

(i) x is revealed indifferent to y if and only if xI_Ty

(ii) x is (strictly) revealed preferred to y if and only if xP_Ty.

Proposition 9. *Suppose c is a CLC correspondence with an attention filter.*

(i) x is revealed to attract attention at S if and only if x is chosen from some super set of S (possibly from S).

(ii) x is revealed not to attract attention at S if and only if x is revealed to be preferred to $y \in c(T)$ for some T such that $x \in T \subset S$, and $x \notin c(T)$.

Proof. (i) ($\Rightarrow$). Let x be revealed to attract attention at S. Thus for all $(\Gamma_i, \succsim_i)$, $x \in \Gamma_i(S)$. We prove the contrapositive. Suppose for all $T \subseteq S$, $x \notin c(T)$. Then $x \notin \Gamma^m(S)$, as previously defined ($\Gamma^m(S) = \{x \in S | x \in c(T) \text{ for some } T \supseteq S\}$), and we know from proposition 9 that there exists a weak order $\succsim$ such that $(\Gamma^m, \succsim)$ represents c. Therefore x is not revealed to attract attention at S.

($\Leftarrow$). $x \in c(T)$ for some $T \supseteq S$, since c is a CLC, x is $\succsim_i$-maximal in $\Gamma_i(T)$ for all i. Therefore for all $S \subseteq T$ with $x \in S$, $x \in \Gamma_i(S)$ since G_i is an attention filter for all i, otherwise if $x \in \Gamma_i(S)$, then $x \in \Gamma_i(T)$. Therefore x is revealed to attract attention at S.

(ii) ($\Leftarrow$). Let $x \succsim_i y$ for all i, where $y \in c(T)$, and $x \in T \subseteq S$. Then for any Γ_i, y is $\succsim_i$-maximal in $\Gamma_i(T)$. $x \notin c(T)$ implies that if $x \in \Gamma_i(T)$, then $y \succ_i x$, which is impossible since $x \succsim_R y$. Therefore $x \notin \Gamma_i(T)$ for all i. Since Γ_i is an attention filter $x \notin \Gamma_i(S)$ for all i since $T \subseteq S$. Therefore x is revealed not to attract attention at S.

($\Rightarrow$). We prove the contrapositive. Let $x \in S$. Suppose for all $T \subseteq S$ with $x \notin c(T)$, $\neg(xR_Ty)$ for any $y \in c(T)$. We want to show that there exists a pair $(\Gamma, \succsim)$ with $x \in \Gamma(S)$ representing c. Therefore x is not revealed *not* to attract attention at S.

We use $\Gamma^m(S) = \{x \in S | x \in c(T) \text{ for some } T \supseteq S\}$ as well. Define Γ' and $\succsim'$ as follows:

$$\Gamma'(T') = \begin{cases} \Gamma^m(T') \cup \{x\} & \text{if } x \in T' \subseteq S \\ \Gamma^m(T') & \text{otherwise} \end{cases}$$

Let $R' = R_T \cup (y,x)$ for all $y \in c(T)$ for $T \subseteq S$ (for all these y we have $\neg(xR_Ty)$ by the assumption). Following the proof of proposition 8 that this will imply $(x,y) \notin R'_T$, the transitive closure of R',there exists a complete extension of R'_T, call it $\succsim$, where $(x,y) \notin \succsim$, and moreover $(\Gamma^m, \succsim)$ represents c.

Now we show $(\Gamma', \succsim)$ also represents c and $x \in \Gamma(S)$. For any $T \notin \{T \in \mathcal{X} : T \subseteq S\}$, $\Gamma'(T) = \Gamma^m(T)$, and we know from proposition 8 that $c(T) = \max_{\succsim} \Gamma'(T)$. Now let $T \in \{T \in \mathcal{X} : T \subseteq S\}$, so $x \in T$. Suppose $y \in c(T)$, then $y \in \Gamma^m(T)$, and $y \succsim z$ for all $z \in \Gamma^m(T)$, and by construction of $\succsim$, $y \succ x$. Therefore $y \succsim z$ for all $z \in \Gamma'(T)$. Similarly, if y is $\succsim$-maximal in $\Gamma^m(T)$ (we know $y \in c(T)$), y is also $\succsim$-maximal in $\Gamma'(T)$ since $y \succ x$ and $\Gamma'(T) = \Gamma^m(T) \cup \{x\}$. Therefore $(\Gamma', \succsim)$ represent c and $x \in \Gamma'(S)$, therefore x is not revealed to not attract attention at S. $\square$

So the revealed preference, indifference, and revealed attention inattention results can be summarized by the fact that R_T gives use the revealed preference and indifference, Γ^m, as defined in 2, gives us the revealed attention, and R_T along with Γ^m give us the revealed inattention (when elements are revealed not to attract attention in a menu).

Thus we can see that this exercise of considering choice correspondences has the same intuitive results than the function setting, but lacks some clarity and tractability.

A.2 Uniqueness and relation to WARP

The following proposition explains the relation between the revealed preference, revealed attention, and revealed inattention, and moreover the relationship between all the parts of our model and the classical choice model and WARP.

Proposition 10. *Let c be a CLC with a consideration filter. Then the following are equivalent.*

(1) c satisfies WARP.

(2) For all $S \in \mathcal{X}$, Revealed Inattention is empty (i.e. for all S, for all $x \in S$ there exists a $(\Gamma, \succsim)$ that represent c and $x \in \Gamma(S)$.

(3) $P_T = \emptyset$.

(4) For all $S \in \mathcal{X}$, Revealed Attention corresponds to $c(S)$.

Proof. We prove (1) ⇒ (2) ⇒ (3) ⇒ (4) ⇒ (1).

(1) ⇒ (2). Let c satisfy WARP, thus for all S there exists x^* such that $c(T) \in S$ implies $c(T) = x^*$ for all $T \ni x$, in particular $x^* = c(S)$. Suppose for some $S \in \mathcal{X}$ there exists $y \in S$ that is revealed not to attract attention at S, then by Proposition 2, we must have $yP_Tc(T)$, for some $T \subset S$, and $y \notin c(S)$. Thus $c(T) = y \ni S$, which implies $x^* = c(T)$ by WARP, contradiction. Hence there does not exist any y that reveals to attract no attention at S, for all $S \in \mathcal{X}$.

(2) ⇒ (3). We prove the contrapositive, so if xPy for some $x, y \in X$ (we know $P \subseteq P_T$, and if xP_Ty there exist a chain $x_1Px_2P...Py$, so without loss we can consider the case where xPy), then Revealed Inattention is nonempty for some $S \in \mathcal{X}$. Let xPy, then there exist $T \subset S$ such that $x = c(T)$ and $y = c(S)$, so x is revealed preferred to $y = c(T)$ and $T \subset S$, thus by Proposition 2, we have that x is revealed not to attract attention at S. Hence Revealed Inattention is nonempty for S.

(3) ⇒ (4). Let $P_T = \emptyset$. Let $y \in S$ such that y is revealed to attract attention at S, we show that $y = c(S)$. Since y is revealed to attract attention at S, then $y = c(T)$ for some $T \supseteq S$ (the characterization of revealed attention is given by Prop. 2.) Since $P_T = \emptyset$, $P = \emptyset$ too, and if $y \neq c(S)$, then $c(S) \in T$ where $T \supseteq S$, thus $c(S)Py$, which is a contradiction since $P = \emptyset$, thus $y = c(S)$. Therefore revealed attention at S is only $c(S)$.

(4) ⇒ (1). We prove the contrapositive. Suppose WARP fails, then we show that there exists some S^* such that x is revealed to attract attention at S and $x \neq c(S^*)$. Let WARP fail, thus there exists $S \in \mathcal{X}$, such that for every $x \in S$, there exists $T_x \in x$ such that $c(T_x) \in S$ and $c(T_x) \neq x$. Since c is nonempty, let $x = c(S)$, and $c(T_x) = y$. Consider $\{xy\}$. Since c is nonempty we have two consider two cases:

- $x = c(xy)$, then since $x \in T_x$, we have $\{xy\} \subseteq T_x$. Given that $x = c(xy)$ and $y = c(T_x)$ where $T_x \supset \{xy\}$, we have that y reveals to attract attention at $\{xy\}$ and $y \neq c(xy)$ by the characterization of revealed attention (prop. 2).

- Similarly when $y = c(xy)$, given that $c(T_x) \in S$, we have $\{xy\} \subseteq S$. And $x = c(S)$ and $y = c(xy)$ implies that x is revealed to attract attention at $\{xy\}$.

In either case the $S^* = \{xy\}$, since x and y reveal to attract attention at $\{xy\}$ and one of the two is not chosen. □

This proposition gives a direct link between revealed attention and the minimal consideration structure.

Proposition 11. *Recall the definition for Γ^m, where*

$$\Gamma^m(A) = \{x \in S | \exists S \subset T \ s.t. \ x = c(T)\}$$

Let c be a CLC with a consideration filter, then for any $(\Gamma, \succ)$ that rationalize c, $\Gamma^m(A) \subseteq \Gamma(A)$, moreover, $(\Gamma^m, \succ)$ also rationalizes c.

Similarly, we can define a maximal consideration structure, to see how much of the preference order we can pin down.

Definition 1. Let c be a CLC. Define $\mathcal{I}(S) = \{x \in S : x$ is revealed not to attract attention at $S\}$, it is the set of elements at S not considered for any $(\Gamma, \succ)$ representing c.

This definition of the set of ignored elements prompts the following definition of the maximal possible filter.

Definition 2. Let c be a CLC. Then define the maximal filter as

$$\Gamma^*(S) = S \setminus \mathcal{I}(S)$$

Lemma 4. *$\Gamma^*(S)$ is a consideration filter.*

Proof. Let $x \in \Gamma^*(S)$, then x is not revealed to not attract attention at S, so for any $T \subset S$, either $c(T) P_T x$ or $\neg(c(T) P_T x)$ and $\neg(x P_T c(T))$, since x is not revealed preferred to any $c(T)$ for all subsets $T \subset S$. Given $S' \subset S$, where $x \in S'$ we know already that x is not revealed preferred to any $T \subset S$, therefore for any $T' \subset S' \subset S$, x is not revealed preferred to $c(T')$, thus x does not reveal to attract no attention at S', hence $x \in \Gamma^*(S')$ for $S' \subset S$. So Γ^* is a consideration filter. □

Even with maximal possible consideration we cannot pin down the preference completely, except for the case of WARP.

Proposition 12. *Let c be a CLC with a consideration filter. Then $\exists!, \succ^*$ such that $(\Gamma^*, \succ^*)$ rationalizes c if and only if c satisfies WARP.*

Proof. The only-if part is the standard result form classical choice theory. By proposition 10 the revealed inattention is empty for all $S \in \mathcal{X}$. Thus by the construction of Γ^*, $\Gamma^*(S) = S$ for all S. Thus by the std. result WARP is equivalent to the existence of a unique $\succ^*$ such that $c(A) = \max_{\succ^*} A$.

For the if-part. Let there be a unique $\succ^*$ such that $(\Gamma^*, \succ^*)$ rationalize c. It is first of all necessary that $\Gamma^*(S) = S$ for all S, else by the construction of Γ^*, $\Gamma^*(S) \neq S$ for some S implies that there is revealed inattention at S, and by proposition 10, c cannot satisfy WARP. But a unique preference order is equivalent to WARP under full consideration, so the result follows. □

A.3 Proofs

Proof of Proposition 2

Proof. "Only if part of i)" In the proof of Theorem 1, we show that

$$\Gamma(S) = \{y \in S | \exists T \supset S \text{ s.t. } y = c(T)\}$$

represents c along with some $\succ$. This means if x is never chosen from any superset of S, there is one attention filter representing c but it does not include x at S.

"Only if part of ii)" Suppose there exists no $T \subset S$ such that $xP_Rc(T)$ and $x \in T$. Define

$$\Gamma'(T') = \begin{cases} \Gamma(T') \cup \{x\} & \text{if } x \in T' \subset S \\ \Gamma(T') & \text{otherwise} \end{cases}$$

and $a \succ' b$ if (i) aP_Rb or (ii) $a \neq b = x$ and not xP_Ra. First, we show that $\succ'$ is acyclical by contradiction. Suppose $\succ'$ has a cycle, then it must involve x because P_R is acyclical, say $x \succ' a_1 \succ' \cdots \succ' a_n \succ' x$. By definition of $\succ'$ it implies $xP_Ra_1P_R \cdots P_Ra_n \succ' x$. Since P_R is transitive, it is reduced to $xP_Ra_n \succ' x$. Since P_R is asymmetric, $a_n \succ' x$ is because of the second condition in defining $\succ'$, which is not xP_Ra_n. This is a contradiction. Therefore, $\succ'$ is acyclical. Let $\succ$ be any completion of $\succ'$.

We argue that $(\Gamma, \succ)$ represents c case by case. Take $y \in \Gamma'(T) \setminus \{c(T)\}$.

Case 1: $y \neq x$. Then $y \in \Gamma(T)$. We know that any completion of P_R represents c along with Γ (see the proof of Theorem **??**). Because of that, y is P_R-dominated by $c(T)$. Since

$\succ$ is one of P_R's completion, y is also $\succ$-dominated by $c(T)$.

Case 2: $y = x$. If $x \in \Gamma(T)$, the logic used in Case 1 is applicable. Consider the case where $x \notin \Gamma(T)$. In this case, it must be $x \in T \subset S$ by the construction of Γ'. By the assumption, we have $not x P_R c(T)$. Therefore, $c(T) \succ x$ follows from the definition of $\succ'$ and $\succ$. Therefore, $(\Gamma', \succ)$ represents c. □

The Proof of Theorem 2

Define $xP''y$ if and only if there exist T and T' with $x, y \in T \subset T'$ such that

$$x = c(T) \text{ and } c(T') \neq c(T' \setminus y)$$

Lemma 5. *P'' is acyclic if and only if c satisfies Strong Consistency.*

The proof of Lemma 5 is completely analogous to earlier Lemma, so we skip it.

Let P''_R be the transitive closure of P'' and let $\succ$ be any arbitrary completion of P''_R. For every S, we call $B \subset S$ is a minimum block of S if and only if $c(S) \neq c(S \setminus B)$ but $c(S) = c(S \setminus B')$ for any $B' \subsetneq B$. Given this, define Γ recursively as follows:

1. $\Gamma(X)$ consists of the $\succ$-worst element of each of X's minimum block.
2. Suppose Γ has been already defined for all proper supersets of S. Then, define $\Gamma(S)$
   (a) First, put $x \in S$ into $\Gamma(S)$ if $x \in \Gamma(T)$ for some $T \supsetneq S$.
   (b) If there is a minimum block of S that does not have an element in $\Gamma(S)$ according to the above, pick the $\succ$-worst element into $\Gamma(S)$.

Lemma 6. *For any S,*

(i) $\{c(S)\}$ is a minimum block of S. There is no other minimum block that includes $c(S)$.

(ii) If B is a minimum block of S other than $\{c(S)\}$, then $c(S) \succ x$ for all $x \in B$.

(iii) If $c(T) \neq c(S)$ and $T \supsetneq S$, then T has a minimum block that is a subset of $T \setminus S$.

Proof. Part (i) and (iii) are obvious so only prove Part (ii). Let $B' = B \setminus x$ (it may be empty). Then we have

$$c(S) = c(S \setminus B') \neq c((S \setminus B') \setminus x)$$

Therefore, we have $c(S)P''x$ so it must be $c(S) \succ x$. □

Claim 1. Γ *is a strong consideration filter.*

Proof. Γ is an attention filter by construction so we shall prove that Γ is a strong consideration filter. Suppose $x, y \in S$, $x, y \notin \Gamma(S)$, but $y \in \Gamma(S \setminus x)$. Then there exists $T \supset S$ such that (i) $T \setminus x$ has a minimum block B and y is the worst element in B and (ii) none of elements in B is included in $\Gamma(T')$ for any $T' \supsetneq T \setminus x$.

Then, we must have $c(T) = c(T \setminus x)$. Otherwise $\{x\}$ is a minimum block of T' so we have $x \in \Gamma(T')$ that implies $x \in \Gamma(S)$. Therefore, we have

$$c(T) = c(T \setminus x) \neq c((T \setminus x) \setminus B) = c(T \setminus (\{x\} \cup B))$$

Therefore, by Lemma 6 (iii), T has a minimum block that is a subset of $x \cup B$ so at least one element in $x \cup B$ must be in $\Gamma(T)$, which is a contradiction. □

Now we want to show that $(\succ, \Gamma)$ represents c. Since Lemma 6 (i) implies that $c(S) \in \Gamma(S)$, all we need to show is that $c(S) \succ y$ for all $y \in \Gamma(S) \setminus c(S)$.

Claim 2. *If* $y \in \Gamma(S)$ *and* $y \neq c(S)$, *then* $c(S) \succ y$.

Proof. Since $y \in \Gamma(S)$, there exists $T \supset S$ such that $y \in \Gamma(T)$. Furthermore, T has a minimum block B where y is the worst element and none of elements in B is in $\Gamma(T')$ for any $T' \supsetneq T$. There are three easy cases: (i) if $c(S) = c(T)$ then by Lemma 6 (ii) we have $c(S) = c(T) \succ y$, (ii) if $y = c(T)$ then we have $c(S)P''y$ so it must be $c(S) \succ y$, and finally (iii) if $c(S) \in B$, then $c(S) \succ y$ by the construction. Therefore, we only need investigate the case when $y \neq c(T) \neq c(S)$ and $c(S) \notin B$. Note that $c(T) \succ y$ in this case by Lemma 6 (ii).

Now let $S' = S \setminus B$. Since $y \in B$, S' is a proper subset of S.

Case I: $c(S'') \neq c(S)$ for some S'' where $S' \subset S'' \subset S$.

By Lemma 6 (iii), S has a minimum block B' that is a subset of $S \setminus S'' \subset B$. Since $c(S) \notin B' (\subset B)$, every element in B' is worse than $c(S)$ by Lemma 6 (ii). Since y is the worst element in B that is a superset of B', we conclude $c(S) \succ y$.

Case II: $c(S'') = c(S)$ for all S'' where $S' \subset S'' \subset S$.

Since $y \neq c(T) = c(T \setminus (B \setminus y)) \neq c(T \setminus B)$, and $c(S \setminus (B \setminus y)) \in T \setminus (B \setminus y)$, we have $c(S \setminus (B \setminus y))\ P''\ y$. Therefore, $c(S) \succ y$ because of $c(S \setminus (B \setminus y)) = c(S)$. □

The Proof of Theorem 3

The if-part is demonstrated in the main body. For the only-if part, notice that A3 and A4 imply that P is acyclic so the proof of Theorem 1's only-if part is applicable by setting $\succ$ to be equal to P^*. Given this, it is easy to see that Γ constructed in the proof has the property $\Gamma(S) = S$ whenever $|S| = 2$. □

The Proof of Theorem 4

We have already shown the if-part of the statement in the main text so we shall show the only-if part. Take any completion of P'', denoted by $\succ$. (P'' is defined in the proof of Theorem 2. Such $\succ$ exists because Strong Consistency guarantees that P'' is acyclic.) Then define $x \rhd' y$ if and only if

$$y \succ x \text{ and } x = c(xy)$$

Since $x \rhd' y$ only if $y \succ x$ and $\succ$ is a preference, $\rhd'$ is acyclic. Let $\rhd$ be the transitive closure of $\rhd'$. It is acyclic as well. Therefore, all we need to show is that $(\Gamma_{\rhd}, \succ)$ represents c.

Lemma 7. *If $y \succ c(S)$, then $y \notin \Gamma_{\rhd}(S)$*

Proof. Let $x = c(S)$. The statement is trivial when $y \notin S$ so assume $y \in S$. Since $\rhd$ is the transitive closure of $\rhd'$, it is $\Gamma_{\rhd}(S) \subset \Gamma_{\rhd'}(S)$ so it is enough to show $y \notin C_{\rhd'}(S)$.

The statement is true when $|S| = 2$ by the definition of $\rhd'$. Suppose there exists S and $y \in S$ such that $y \succ x$ and $y \in \Gamma_{\rhd}(S)$. Without loss of generality, assume that S has the smallest cardinality among such sets. We shall lead a contradiction by showing several claims.

Claim 3. *For any $S' \subsetneq S$, $x \neq c(S')$ whenever $y \in S'$.*

Proof. Since $y \in \Gamma_{\rhd'}(S)$, it must be $y \in \Gamma_{\rhd'}(S')$ as well. If $x = c(S')$ then this violates the assumption that S has the smallest cardinality at which the statement of Lemma 7 is violated. □

Now consider all budget sets that can be obtained by removing one element from S. Notice that there are $|S|$ such decision problems and only elements in S may be chosen from those sets.

Claim 4. *For any $z \in S$, there exists $z' \in S \setminus \{z\}$ such that $z = c(S \setminus z')$.*

Proof. Suppose not. By the pigeonhole principle, there must exist $\alpha \in S$ such that

$$\alpha = c(S \setminus \beta) = c(S \setminus \gamma)$$

for some distinct $\beta, \gamma \in S$. By Expansion, $c(S) = \alpha$ so it must be $\alpha = x$. Since y must be included either in $S \setminus \alpha$ or $S \setminus \beta$, Claim 3 implies that $x \neq c(S \setminus \alpha)$ or $x \neq c(S \setminus \beta)$. This is a contradiction. $\square$

Claim 5. *$x = c(S \setminus y)$ and $y = c(S \setminus x)$.*

Proof. The combination of Claim 3 and 4 immediately implies $x = c(S \setminus y)$. By Claim 4, y must be chosen from $S \setminus z$ for some $z \in S$. If $z \neq x$, then then we have $yP'x$ so it must be $y \succ x$. Therefore, it cannot be $y \rhd' x$. Hence, $y = c(S \setminus x)$. $\square$

Now take any $z \in S \setminus \{x, y\}$. (Notice that $|S| \geq 3$). If $|S| = 3$, Claim 4 requires $z = c(S \setminus x)$ or $z = c(S \setminus y)$ but both possibilities are excluded by Claim 5. Suppose $|S| \geq 4$. Let $\alpha = c(S \setminus z)$. By Claim 4 and Claim 5, $\alpha \in S \setminus \{x, y, z\}$. Hence, we have $\alpha P'x$ so it must be $\alpha \succ x$. Now consider $c(S \setminus \alpha)$, which must be something other than x. Hence, we have $xP'\alpha$ so is $x \succ \alpha$. This is a contradiction. Therefore, there is no S such that $x = c(S)$ but $y \in S$ so Lemma 7 is proven. $\square$

Lemma 8. *$c(S) \in \Gamma_{\rhd}(S)$.*

Proof. Let $x = c(S)$ but there exists $y \in S$ such that $y \rhd x$. If $y \rhd' x$ (i.e. before taking the transitive closure), then it must be $x \succ y$ and $c(xy) = y$. $c(xy) = y$ and $c(S) = x$ imply $yP'x$ so we cannot have $x \succ y$, which is a contradiction. Therefore, it cannot be $y \rhd' x$. so there must exist $z_1, \ldots, z_k$ such that

$$y \rhd' z_1 \rhd' z_2 \rhd' \cdots \rhd' z_k \rhd' x.$$

By definition of $\rhd'$ it must be

$$x \succ z_k \succ \cdots \succ z_2 \succ z_1 \succ y$$

and

$$c(yz_1) = y, \ c(z_1 z_2) = z_1, \ldots, c(z_{k-1} z_k) = z_{k-1}, \ c(z_k x) = z_k$$

Since $x \succ y$, we cannot have $y = c(xy)$ (if so, it would be $y \rhd' x$ and we have shown that it would lead a contradiction.) so it must be $x = c(xy)$.

Now consider $c(xyz_1 \ldots z_k)$. It cannot be x because, if so, it must be $z_k P'x$ so is $z_k \succ x$. It cannot be z_i (because if so $z_{i-1} \succ z_i$ or $y \succ z_1$). Therefore, it must be $y = c(xyz_1, \ldots, z_k)$.

Since $x = c(xy)$, there must exist i such that:

$$y = c(xyz_i z_{i+1} \dots z_k) \neq c(xyz_{i+1} \dots z_k)$$

which implies $yP'z_i$ so $y \succ z_i$. This is a contradiction. Therefore, we conclude that there is no $y \in S$ such that $y \rhd x$. □

These two lemmas prove that c is represented by $(\Gamma_{\rhd}, \succ)$. □

The Proof of Proposition 4

The if-parts of both the revealed preference and the revealed order are shown in the main body. To prove the only-if parts, the proof of the only-if part of Theorem 4 is applicable. If not $xP'_R y$, it has been shown that a preference with $y \succ x$ can represent c. If $yP'_R x$ but $x = c(xy)$, then we indeed define $x \rhd y$ to represent c. □

References

K. Arrow. Rational choice functions and orderings. *Economica*, 26(102):121–127, May 1959.

R. J. Aumann. Musings on information and knowledge. *Econ Journal Watch, Atlas Economic Research Foundation*, 2(1):88–96, 2005.

G. Basar and C. Bhat. A parameterized consideration set model for airport choice: an application to the san francisco bay area. *Transportation Research Part B: Methodological*, 38(10):889–904, 2004.

A. Caplin, M. Dean, and D. Martin. Search and satisficing. 2009. Working Paper.

V. Cherepanov, T. Feddersen, and A. Sandroni. Rationalization. mimeo,Nothwestern University, 2008.

A. Chernev. When More Is Less and Less Is More: The Role of Ideal Point Availability and Assortment in Consumer Choice. *Journal of Consumer Research*, 30(2):170–183, September 2003.

J. Chiang, S. Chib, and C. Narasimhan. Markov chain monte carlo and models of consideration set and parameter heterogeneity. *Journal of Econometrics*, 89(1-2):223–248, November 1998.

G. de Clippel and K. Eliaz. Reason-based choice: A bargaining rationale for the attraction and compromise effects. Working papers, Brown University, 2009.

M. Dean. Status quo bias in large and small choice sets. mimeo, 2008.

M. Dean and A. Caplin. Search, choice, and revealed preference. Mimeo, 2008.

U. Dulleck, F. Hackl, B. Weiss, and R. Winter-Ebmer. Buying online: Sequential decision making by shopbot visitors. Economics working papers 2008-10, Department of Economics, Johannes Kepler University Linz, Austria, Aug. 2008.

K. Eliaz and E. A. Ok. Indifference or indecisiveness? choice-theoretic foundations of incomplete preferences. *Games and Economic Behavior*, 56(1):61–86, July 2006.

K. Eliaz and R. Spiegler. Consideration sets and competitive marketing. Mimeo, 2007.

K. Eliaz, M. Richter, and A. Rubinstein. An etude in choice theory: Choosing the two finalists. Mimeo, 2009.

H. Ergin. Costly contemplation. Mimeo, 2003.

H. Ergin and T. Sarver. A unique costly contemplation representation. Working paper, 2007.

J. T. Gourville and D. Soman. Overchoice and assortment type: When and why variety backfires. *Marketing Science*, 24(3):382–395, 2005. ISSN 1526-548X.

B. Hansson. Choice structures and preference relations. *Synthese*, 18(4):443–458, October 1968.

J. R. Hauser and B. Wernerfelt. An evaluation cost model of evoked sets. *Journal of Consumer Research*, 16:383–408, March 1990.

G. Huberman and T. Regev. Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *Journal of Finance*, 56(1):387–396, 02 2001.

S. Iyengar, G. Huberman, and W. Jiang. How much choice is too much? contributions to 401(k) retirement plans. In *Pension Design and Structure*, chapter 5, pages 83 –97. Oxford University Press, 2004.

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

L. Jarvis and J. Wilcox. A model for predictive measurements of advertising effectiveness. *Industrial Marketing Management*, 6:9–14, 1977.

G. Kalai, A. Rubinstein, and R. Spiegler. Rationalizing choice functions by multiple rationales. *Econometrica*, 70(6):2481–2488, November 2002.

K. V. Katsikopoulos and L. Martignon. Nave heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology*, 50(5):488 – 494, 2006.

E. Lapersonne, G. Laurent, and J.-J. L. Goff. Consideration sets of size one: An empirical investigation of automobile purchases. *International Journal of Research in Marketing*, 12 (1):55 – 66, 1995.

M. Laroche, J. Rosenblatt, and I. Sinclair. Brand categorization strategies in an extensive problem solving situation: A study of university choice. *Advances in Consumer Research*, 11:175–179, 1984.

R. J. Lavidge and G. A. Steiner. A model for predictive measurements of advertising effectiveness. *Journal of Marketing*, 25:59–62, 1961.

M. Mandler, P. Manzini, and M. Mariotti. A million answers to twenty questions: Choosing by checklist. Working Papers 622, Queen Mary, University of London, Department of Economics, 2008.

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

P. Manzini and M. Mariotti. Choice over time. In *Oxford Handbook of Rational and Social Choice*, chapter 10. Oxford University Press, 2009.

L. Martignon and U. Hoffrage. Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, 52(1):29–71, February 2002.

Y. Masatlioglu and D. Nakajima. Choice by iterative search. Mimeo, Department of Economics, University of Michigan, 2010.

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

Y. Masatlioglu, D. Nakajima, and E. Ozbay. Revealed attention. Mimeo, 2009.

G. A. Miller. The magic number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63:81–97, March 1956.

E. A. Ok. *Real Analysis with Economic Applications*. SUNY-Oswego, Department of Economics, January 2004.

E. A. Ok, P. Ortoleva, and G. Riella. Revealed (p)reference theory. mimeo, 2008.

E. A. Pessemier. Stochastic properties of changing preferences. *American Economic Review*, 68(2):380–85, May 1978.

M. D. Richards, J. E. Sheridan, and J. W. Slocum. Comparative analysis of expectancy and heuristic models of decision behavior. *Journal of Applied Psychology*, 60(3):361–368, 1975.

M. Richter. Revealed preference theory. *Econometrica*, 34(3):635–645, July 1966.

J. H. Roberts and J. Lattin. Consideration: Review of research and prospects for future insights. *Journal of Marketing Research*, 34:406–410, August 1997.

D. E. Rosen, J. M. Curran, and T. B. Greenlee. College choice in a brand elimination framework: The administrator's perspective. *Journal of Marketing for Higher Education*, 8(4), 1998.

P. Samuelson. A note on the pure theory of consumers behavior. *Economica*, 5:61–71, 1938.

B. Scheibehenne. *The Effect of Having Too Much Choice*. PhD thesis, Humboldt Universität zu Berlin, 2008.

B. Scheibehenne, R. Greifeneder, and P. M. Todd. What moderates the too-much-choice effect? *Psychology and Marketing*, 26(3):229–253, 2009. ISSN 1520-6793.

B. Schwartz. *The Paradox of Choice : Why More Is Less*. Harper Perennial, January 2005.

A. K. Sen. Choice functions and revealed preference. *Review of Economic Studies*, 38(115): 307–17, July 1971.

S. M. Shugan. Cost of thinking. *Journal of Consumer Research*, 7:99–111, 1983.

H. A. Simon. A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1): 99–118, 1955.

G. J. Stigler. The economics of information. *Journal of Political Economy*, 69(3):213–225, 1961.

E. Szpilrajn. Sur l'extension de l'ordre partiel. *Fund. Math.*, 16:386–389, 1930.

A. Tversky. Elimination by aspects. *Psychological Review*, 79:281–299, 1972.

A. Tversky and D. Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, 1974.

C. J. Tyson. Cognitive constraints, contraction consistency, and the satisficing criterion. *Journal of Economic Theory*, 127(1):51–70, January 2008.

D. Weinberger. *Everything is Miscellaneous*. Times Books, Henry Holt and Co. LLC, 2007.

J. Wu and A. Rangaswamy. A fuzzy set model of search and consideration with an application to an online market. *Marketing Science*, 22(3):411–434, 2003.