American Economic Review 2012, 102(5): 2183–2205
http://dx.doi.org/10.1257/aer.102.5.2183

Revealed Attention†

By Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay*

The standard revealed preference argument relies on an implicit assumption that a decision maker considers all feasible alternatives. The marketing and psychology literatures, however, provide well-established evidence that consumers do not consider all brands in a given market before making a purchase (*Limited Attention*)*. In this paper*, *we illustrate how one can deduce both the decision maker's preference and the alternatives to which she pays attention and inattention from the observed behavior. We illustrate how seemingly compelling welfare judgments without specifying the underlying choice procedure are misleading. Further*, *we provide a choice theoretical foundation for maximizing a single preference relation under limited attention.* (*JEL* D11, D81)

Revealed preference is one of the most influential ideas in economics and has been applied to a number of areas of economics, including consumer theory.[1] According to standard revealed preference theory, x is revealed to be preferred to y if and only if x is chosen when y is also available (Samuelson 1938). Any choice reversal, therefore, observed both empirically and experimentally, is attributed to irrationality since it cannot be expressed as a preference maximization.

The revealed preference argument relies on the implicit assumption that a decision maker (DM) considers all feasible alternatives. Without the full consideration assumption, the standard revealed preference method can be misleading. It is possible that the DM prefers x to y but she chooses y when x is present simply because she does not realize that x is also available (Hausman 2008). For example, while using a search engine, a DM might only pay attention to alternatives appearing on the first page of the results since it takes too much time to consider all the search results. She then picks the best alternative of those on the first page, say y. It is possible that her most preferred item, x, does not appear on the first page. Therefore we, as outside observers, cannot conclude that y is better than x even though y is chosen when x is

* Masatlioglu: University of Michigan, Department of Economics, 611 Tappan Street, Ann Arbor, MI 48109 (e-mail: yusufcan@umich.edu); Nakajima: University of Michigan, Department of Economics, 611 Tappan Street, Ann Arbor, MI 48109 (e-mail: nakajima@umich.edu); Ozbay: University of Maryland, Department of Economics, 3105 Tydings Hall, College Park, MD 20742 (e-mail: ozbay@umd.edu). We would like to thank Tilman Börgers, Kfir Eliaz, Emel Filiz-Ozbay, Chris House, Stephan Lauermann, Paola Manzini, Marco Mariotti, Collin Raymond, Lones Smith, Rani Spiegler, and Neslihan Uler for their helpful comments. Valuable comments were also received from numerous seminar and conference participants. Masatlioglu and Nakajima thank the National Science Foundation for financial support under Grant SES-1024544.

† To view additional materials, visit the article page at http://dx.doi.org/10.1257/aer.102.5.2183.

[1] Varian (2006) provides a nice survey of revealed preference analysis.

available. Nevertheless, as in the above example, the DM may have a well-defined preference and is maximizing her preference within her bounded understanding of what is available.[2]

This example immediately raises a question: how can we elicit her (stable) preference without the full attention assumption? We consider a DM who picks her most preferred item from the alternatives she pays attention to, not from the entire feasible set. Then we shall illustrate when and how one can deduce both the DM's preferences and the alternatives to which she does or does not pay attention from her observed choices. Furthermore, we illustrate the problem of the welfare judgment without specifying the underlying choice procedure by showing an example where our method and the conservative criterion of Bernheim and Rangel (2007, 2009) result in the completely opposite implication.

The marketing literature calls the set of alternatives to which a DM pays attention in her choice process the *consideration set* (Wright and Barbour 1977). The formation of the consideration set has been studied extensively in the marketing and finance literatures (e.g., Hauser and Wernerfelt 1990; Roberts and Lattin 1991). It has been argued that due to cognitive limitations, DMs cannot pay attention to all the available alternatives. As Simon (1957) pointed out, being able to consider all possible alternatives is as hard as comparing them for decision makers. Therefore, a DM with limited cognitive capacity (possibly stemming from unawareness, as demonstrated in Goeree 2008[3]), restricts her attention to only a small fraction of the objects present in the associated market (Stigler 1961; Pessemier 1978; Chiang, Chib, and Narasimhan 1999).[4] In sum, a DM intentionally or unintentionally filters out some alternatives to prevent her cognitive capacity from being overloaded (Broadbent 1958).

The common property in the formation of consideration sets is that it is unaffected when an alternative she does not pay attention to becomes unavailable. This basic property of the attention filter, which is also documented in the psychology literature (Broadbent 1958), can be interpreted as the minimal condition. This property is trivially satisfied in classical choice theory where it is assumed that the DM is able to pay attention to all the available alternatives. Additionally, it is normatively appealing especially when a DM pays attention to all of the items she is aware of and is unaware that she is unaware of other items. For example, if a personal computer (PC) buyer is not only unaware of a particular PC, but she is also unaware that she overlooks that PC, then, even when that PC becomes unavailable, she will not recognize such a change. Therefore, her consideration set will stay the same.

Interestingly, this property is also satisfied when the DM actually chooses the consideration sets by taking the cost of investigation and the expected benefit into account. Suppose the DM excludes x from her consideration. If x becomes unavailable, she has no reason to add or remove any alternative to her consideration set because she could have done so when x was available. Therefore, her

[2] As argued in Aumann (2005), this behavior is still considered rational (at least boundedly rational) since she is choosing the best alternative under her limited information about what is available.

[3] Lavidge and Steiner (1961) presented awareness of an item as a necessary condition to be in the consideration set. How unawareness alters the behavior of the DM has been studied in various contexts such as game theory (Heifetz, Meier, and Schipper 2010; Ozbay 2008), and contract theory (Filiz-Ozbay 2010).

[4] In addition, in financial economics it is shown that investors reach a decision within their limited attention (Huberman and Regev 2001). Similar examples can be found in job search (Richards, Sheridan, and Slocum 1975), university choice (Dawes and Brown 2005), and airport choice (Basar and Bhat 2004).

consideration set is not affected when x becomes unavailable. Furthermore, this property is also satisfied when the formation is based on many decision heuristics, such as paying attention only to the N-most advertised alternatives or the products appearing in the first page of search results. As a result, our property is appealing from both normative and descriptive point of views.

In this paper, we refer to the consideration sets satisfying this property as *attention filters*. Under this structure, it is possible to elicit the DM's preference whenever a choice reversal is observed.[5] For example, assume that she chooses x, but removing y changes her choice. This can happen only when her consideration set has changed. This would be impossible if she did not pay attention to y. Hence, y must have been considered (Revealed Attention). Given the fact that x is chosen while y draws her attention, we conclude that she prefers x over y (Revealed Preference). In sum, whenever her choice changes as a consequence of removing an unchosen alternative, the initially chosen alternative is preferred to the removed one.

Given that our identification strategy relies on the particular choice procedure, where she maximizes her preference within her attention filter, it is natural to ask the falsifiability of our model. We show that our model is fully characterized by weakening the Weak Axiom of the Revealed Preference (WARP). This result renders our model behaviorally testable.

Our method to distinguish between a preference and attention/inattention generates several policy implications. For instance, if a product of a firm is unpopular in the marketplace, there could be two different explanations: (i) the product has a low evaluation by consumers; or (ii) it does not attract attention of consumers. Identifying the right reason will lead to different strategies for the firm to improve sales.

Our paper also contributes to the recent discussion about welfare analysis under nonstandard individual behavior.[6] We elicit the DM's preference in a positive approach, which is based on a particular choice procedure. Bernheim and Rangel (2009) criticize such an approach by arguing that it is not necessary to explain the behavior to make a welfare analysis. Instead, they make welfare arguments directly from the choice data without assuming any choice procedure (model-free). Particularly, they claim that y is strictly welfare-improving over x if y is sometimes chosen when x is available but x is never chosen when y is present. This intuitive criterion of welfare analysis is meaningful, however, only if the DM considers all the presented alternatives.[7] In Section II, we discuss this issue in detail to illustrate the problem of the naive use of the model-free approach. Indeed, we provide an example where their welfare implication contradicts our revealed preference (hence the actual preference); that is, y is revealed to be preferred to x even when x is strictly welfare-improving over y in the sense of Bernheim and Rangel (2009).

So far we have discussed how one can elicit DM's preference and consideration sets in our model. In doing so, we impose a relatively weak condition on the

[5] Without any structure on the formation of the consideration sets, any choice behavior can be rationalized by any preference (Hausman 2008).

[6] See Ambrus and Rozen (2010); Apesteguia and Ballester (2010); Cherepanov, Feddersen, and Sandroni (2010); Chambers and Hayashi (2008); Green and Hojman (2007); Manzini and Mariotti (forthcoming); Masatlioglu and Nakajima (2009); Noor (2011); Rubinstein and Salant (2012).

[7] Indeed, Bernheim and Rangel (2007) mention that if we know the DM believes that she is choosing from a set that is other than the objective feasible set, we should take it into account for the welfare analysis (Section IIIB).

formation of consideration sets so that our approach is applicable to a wide range of choice data. As a result, although our model is refutable, it provides an alternative explanation for several frequently observed behaviors that cannot be captured by the standard choice theory: Attraction Effect, Cyclical Choice, and Choosing Pairwisely Unchosen (see Section IV). Our explanations for these choice patterns depend solely on limited attention, hence seemingly irrational behaviors can be explained without introducing changing preference. Nevertheless, depending on the intended application, it is possible to analyze this framework under different restrictions on consideration sets.

There are several related models where the final choice is made after eliminating several items, which can be interpreted as a choice with limited consideration such as applying a rationale to eliminate alternatives (Manzini and Mariotti 2007; Apesteguia and Ballester 2009; Houy 2007; Houy and Tadenuma 2009), considering only the *N*-most eye-catching alternatives (Salant and Rubinstein 2008), focusing only on alternatives a decision maker can rationalize to choose by some other criterion (Cherepanov, Feddersen, and Sandroni 2010), and considering only alternatives belonging to undominated categories (Manzini and Mariotti forthcoming). Our model is both descriptively and behaviorally distinct from these models. In addition, unlike our model, these models implicitly assume that a DM considers all feasible alternatives at the first stage and *intentionally* eliminates several alternatives. Therefore, their stories are not applicable to cases where the source of limited consideration is unawareness of some alternatives.

Finally, we would like to compare our model to several other models involving consideration sets in decision theory. Lleras et al. (2010) study a different model of choice under limited consideration where a product attracting attention on a crowded supermarket shelf will be noticed when there are fewer products.[8] Masatlioglu and Nakajima (2009) propose a model of an iterative search where a decision maker cannot consider all alternatives, which can be because of unawareness like our model. The difference is that they emphasize that a consideration set depends on the initial starting point and evolves dynamically during the course of search. In the models of Caplin and Dean (2011), and Caplin, Dean, and Martin (2011), a DM goes through alternatives sequentially and, at any given time, chooses the best one among those she has searched. Unlike our model, their "choice process" data includes not only the DM's choice without time limit, but also what she would choose if she were suddenly forced to quit the search at any given time.

Eliaz and Spiegler (2011) analyze a market where firms would like to manipulate consumers' consideration sets by using costly marketing devices. Eliaz, Richter, and Rubinstein (2011) study a very concrete and reasonable way to construct a consideration set. Indeed, some of the consideration sets we shall present as examples are within their models. Contrary to our model, however, in their paper, the DM's consideration set (called *finalists*) is observed and is directly investigated. In our model the consideration set is an object that must be inferred from the DM's final choice.

The outline of this paper is as follows: Section I introduces the basic notations and definitions. In Section II, we provide two characterizations for the revealed

[8] While this paper is complementary to our paper, their implications are completely different. We discuss it in the Section VI.

preference and the revealed (in)attention from observed choice data. Section III provides a simple behavioral test for our model and discusses the related literature. Then, in Section IV, we illustrate that our limited attention model is capable of accommodating several frequently observed behaviors. Finally, Sections V and VI conclude the paper.

I. The Model

Throughout this paper, let X be a finite set of alternatives that may be available for a DM to choose; $\mathcal{X}$ denotes the set of all nonempty subsets of X, which is interpreted as the collection of all the (objective) feasible sets observed by a third party.

A. *Attention Filters*

In our model, a DM picks the best element from those she pays attention to (her consideration set). Our goal is to elicit her preference along with her attention and inattention from her actual choice data. This is impossible, however, without any knowledge about her attention and inattention. One can always claim that she picks an alternative because she ignores everything else, so one cannot infer her preference at all.

We now propose a property of how consideration sets change as feasible sets change, instead of explicitly modeling how the feasible set determines the consideration set. This approach makes it possible to apply our method to elicit the preference without relying on a particular formation of the consideration set. We shall explain that this property is normatively compelling in several situations and is indeed true in many heuristics people actually use in real life.

Let S be a feasible set the DM is facing. She does not pay attention to all alternatives in S. Let $\Gamma(S)$ be the (nonempty) set of elements to which she pays attention. Formally, Γ is a mapping from $\mathcal{X}$ to $\mathcal{X}$ with $\emptyset \neq \Gamma(S) \subset S$. We call Γ a consideration set mapping. Of all consideration set mappings, we focus on those having the following property:

DEFINITION 1: *A consideration set mapping* Γ *is an* attention filter *if for any* S, $\Gamma(S) = \Gamma(S \backslash x)$ *whenever* $x \notin \Gamma(S)$.[9]

This definition says that if an alternative does not attract the attention of the decision maker, her consideration set does not change when such an item becomes unavailable.

To illustrate that this is a normatively appealing property, we shall provide two examples where the DM's consideration set mapping should be an attention filter. The first example is based on unawareness. Imagine a DM (wrongly) believes $\Gamma(S)$ is her feasible set (S is the actual one). That is, she is not only unaware of alternatives in $S \backslash \Gamma(S)$ but unaware that she is unaware of these alternatives. If so, she will not recognize the change of the feasible set when such an item becomes unavailable, so her consideration set should not change. This is exactly what the property dictates.

[9] Throughout the paper, unless it leads to confusion, we abuse the notation by suppressing set delimiters, e.g., writing $c(xy)$ instead of $c(\{x, y\})$ or $\Gamma(xy)$ instead of $\Gamma(\{x, y\})$ or $S \backslash x$ instead of $S \backslash \{x\}$.

The second one is choosing rationally what to consider (or not to consider). Because of scarcity of time and/or complexity of decision problems, a DM focuses selectively on a smaller set of alternatives and ignores the rest. Suppose she knows S is her entire feasible set. Then, she picks her consideration set $\Gamma(S)$ based optimally on her prior beliefs about the value of alternatives and the cost of inspecting them. Then, her consideration set mapping must satisfy our property. To see this, imagine that she considers only a and b when her feasible set is $\{a,b,c,d\}(\Gamma(abcd) = \{a,b\})$. Assume that d becomes unavailable now. She has no reason to add c to her consideration set because she could have done so when d was available. For the same reason, it is not rational to remove b (or a) from her consideration set. Therefore, it must be $\Gamma(abc) = \{a,b\}$. That is, her consideration set mapping is an attention filter. Notice that this must be true whatever beliefs and cost function she has.[10]

Furthermore, in addition to being normatively appealing, our condition is also descriptively appealing. Many heuristics that are actually used to narrow down the set of choosable options generate attention filters. We list some of them.

Top *N*: A DM considers only top *N* alternatives according to some criterion that is different from her preference. For instance:

- Consider only the three cheapest suppliers in the market (Dulleck et al. 2008).
- Consider the *N*-most advertised products in the market.
- Consider the products that appear on the first page of the web search and/or sponsored links (Hotchkiss et al. 2004).
- Consider the first *N* available alternatives according to an exogenously given order (Salant and Rubinstein 2008).[11]

Top on each criterion: A DM has several criteria and considers only the best alternative(s) on each criterion (modeled as a complete and transitive binary relation). For instance:

- Consider only a job candidate if she is the best in a program. Or consider the top-two job candidates from all first-tier schools and the top candidate from second-tier schools.
- Consider only the cheapest car, the safest car, and the most fuel-efficient car on the market.[12]

[10] The only exception is that the feasible set itself conveys some information that affects her belief or cost function.

[11] Salant and Rubinstein (2008) characterizes this class of choice functions by assuming *N* is observable.

[12] This heuristic is very close to the "Rationalization" of Cherepanov, Feddersen, and Sandroni (2010). Indeed, it is a special version of Rationalization. In their model, unlike "the top on each criterion," depending on the feasible set, different sets of criteria might be utilized to eliminate alternatives in the first stage. See Section III for further discussion.

Most popular category: A DM considers alternatives that belong to the most popular "category" in the market. For instance:

- There are several bike shops in the DM's town. The DM first checks online to find the store offering the largest variety of bikes and goes to that store. Therefore, the DM only considers bikes sold in the selected store.[13] Zyman (1999) provides real-world evidence for such behavior. The sale of Sprite is increased dramatically when it is simply repositioned from the category of lemon-limes (less popular category) to soda (more popular category).

B. *Choice with Limited Attention*

In the previous subsection, we defined the concept of the attention filter and discussed features that make it both normatively and descriptively appealing. Now we define the choice behavior of a DM who picks the best element from her consideration set according to the complete and transitive preference. Formally, a choice function assigns a unique element to each feasible set. That is, $c : \mathcal{X} \rightarrow X$ with $c(S) \in S$ for all $S \in \mathcal{X}$.

DEFINITION 2: *A choice function c is a* choice with limited attention (CLA) *if there exists a complete and transitive preference* $\succ$ *over X and an attention filter* Γ *such that* $c(S)$ *is the* $\succ$*-best element in* $\Gamma(S)$.[14]

In the following sections, we answer the following questions under the assumption that a DM follows a choice with limited attention but her preference and attention filter is not observable: (i) How can we identify her preference and attention filter through her choice data? (ii) Which choice functions are compatible with the model of a choice with limited attention?

II. Revealed Preference and (In)Attention

In this section, we illustrate how to infer (i) the DM's preference and (ii) what the DM pays (and does not pay) attention to from her observed choice that is a CLA. The standard theory concludes that x is preferred to y when x is chosen while y is available. To justify such an inference, one must assume implicitly that she has paid attention to y. Without this hidden assumption, we cannot make any inference because she may prefer y but overlooks it. Therefore, eliciting the DM's preference is no longer trivial because her choice can be attributed to her preference or to her inattention.[15]

[13] For instance, suppose store A deals with Makers 1 and 2's bikes while store B sells bikes from Makers 2 and 3. Then, the DM compares the number of Makers 1 and 2's bikes with that of Makers 2 and 3's to choose which store to visit.

[14] That is, $c(S) \in \Gamma(S)$ and $c(S) \succ x$ for all $x \in \Gamma(S) \setminus c(S)$.

[15] In the extreme case where the choice data satisfy the weak axiom of revealed preference, we have no way of knowing whether the DM is aware of all alternatives and maximizing a particular preference, or whether she only pays attention to the one she chooses. In the latter, her preference has no significant importance. In Section V, we discuss the situations where one can pin down the preference even in this extreme case.

Table 1—Two Possible Representations for the Cyclical Choice

Preference		Attention filter			
		$\{x,y,z\}$	$\{x,y\}$	$\{y,z\}$	$\{x,z\}$
$z \succ_1 x \succ_1 y$	Γ_1	xy	xy	y	xz
$x \succ_2 y \succ_2 z$	Γ_2	xyz	xy	yz	z

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

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

One possibility is that the DM's preference is $z \succ_1 x \succ_1 y$ and she overlooks z both at $\{x,y,z\}$ and $\{y,z\}$. Another possibility is that her preference is $x \succ_2 y \succ_2 z$ and she does not pay attention to x only at $\{x,z\}$ (see Table 1 for the corresponding attention filters).

We cannot identify which of them is her true preference. Nevertheless, if only these two pairs represent c, we can unambiguously conclude that she prefers x to y because both of them rank x above y. For the same reason, we can infer that she pays attention to both x and y at $\{x,y,z\}$ (Table 1). This example makes it clear that we need to define revealed preference when multiple representations are possible.

DEFINITION 3: *Assume c is a choice by limited attention and there are k different pairs of preference and attention filter representing c, $(\Gamma_1, \succ_1), (\Gamma_2, \succ_2), \ldots, (\Gamma_k, \succ_k)$. In this case,*

- *x is revealed to be preferred* to *y if $x \succ_i y$ for all i,*
- *x is revealed to attract attention at S if $\Gamma_i(S)$ includes x for all i,*
- *x is revealed* **not** *to attract attention at S if $\Gamma_i(S)$ excludes x for all i.*

This definition is very conservative: we say x is revealed to be preferred to y only when all possible representations agree on it. We do not want to make any false claims or claims that we are not sure are true. This conservative approach makes it possible that a social planner is always safe to follow our welfare recommendations.

If one wants to know whether x is revealed to be preferred to y, it would appear necessary to check for every $(\Gamma_i, \succ_i)$ whether it represents her choice or not. This is not practical, however, especially when there are many alternatives. Instead we shall now provide a handy method to obtain the revealed preference, attention, and inattention completely.

In the example above, when Γ is an attention filter, it is possible to determine the relative ranking between x and y. To see this, note that if the DM pays attention to x and z at both $\{x,z\}$ and $\{x,y,z\}$, then we should not observe choice reversal. If there is a choice reversal, then this means that her attention set changes when y is removed from $\{x,y,z\}$. This is possible only when she pays attention to y at $\{x,y,z\}$ (Revealed

Attention). Given the fact that x is chosen from $\{x, y, z\}$ we conclude that the DM prefers x over y (Revealed Preference). This observation can be easily generalized. Whenever the choices change as a consequence of removing an alternative, the initially chosen alternative is preferred to the removed one. Formally, for any distinct x and y, define:

$$(1) \qquad xPy \text{ if there exists } T \text{ such that } c(T) = x \neq c(T \setminus y).$$

By the argument analogous to the one above, if xPy then x is revealed to be preferred to y. In addition, since the underlying preferences are transitive, we also conclude that she prefers x to z if xPy and yPz for some y, even when xPz is not revealed directly from the choice. Therefore, the transitive closure of P, denoted by P_R, must also be part of her revealed preference. One may wonder whether some revealed preference is overlooked by P_R. The next theorem states that the answer is no: P_R is the revealed preference in our model.

THEOREM 1 (Revealed Preference): *Suppose c is a CLA. Then*, *x is revealed to be preferred to y if and only if* $xP_R y$.

Theorem 1 illustrates that welfare analysis is possible even with nonstandard choices. In addition, it provides a guideline for a policymaker.

The revealed preference characterized by Theorem 1 is independent of *how* her consideration set is formed, as long as her consideration set mapping is an attention filter. Therefore, it is applicable to many situations. Depending on how her consideration set is formed, however, it may appear to be inappropriate to base the welfare analysis solely on our revealed preference. For instance, one can interpret her attention/inattention as some reflection of her preference and argue that it should be incorporated to the welfare analysis. We do not disagree with such attempts, but to do so the policymaker must have more concrete views about the DM's actual consideration set formation. In those cases, our revealed preference is what the policymaker can say without knowledge of the DM's underlying consideration-set formation process.

Notice that our analysis is a model-based approach as the welfare criterion is obtained assuming a particular underlying choice procedure: a choice with limited attention. On the other hand, Bernheim and Rangel (2009) propose that one should make a welfare judgment only when the choices are unambiguous. Their intuition is that if x is never chosen while y is present and y is chosen at least once when x is available, then y should be strictly welfare-improving over x. Since this intuitive criterion is independent of the underlying model, their approach is called model-free. Using Theorem 1, we are able to illustrate in a reasonable example that the above intuition might deceive us. In the next example, while x is never chosen when y is present, y is chosen at least once over x. Nevertheless, Theorem 1 dictates that x is revealed to be preferred to y.

EXAMPLE 1: *There are four products* x, y, z, *and t. Each of them is packed in a box. Consider a supermarket that displays these products in its two aisles according to the following rules:* (*i*) *Each aisle can carry at most two products*; (*ii*) *x and y cannot be placed into the same aisle because they are packed in big boxes*; (*iii*) *The supermarket fills the first aisle first and uses the second aisle only if it is necessary*; (*iv*) *y and z are* *put into the first aisle whenever they are available*; (*v*) *t is placed in the first aisle only after all other available items are put in an aisle and the first aisle still has a space. Consider a costumer with preference* $t \succ x \succ z \succ y$ (*not observable*) *and she visits only the first aisle and picks her most preferred item displayed in that aisle.*

It is easy to see that her consideration set mapping is an attention filter as the supermarket does not changes its lineups in its first aisle when something in the second aisle become unavailable. Hence, Theorem 1 is applicable.

Since x never appears in the first aisle when y is available, she never chooses x whenever y is feasible (and y is chosen when only x and y are available). Thus, the criterion by Bernheim and Rangel (2009), although it is very conservative to make a welfare statement, conclude that y is welfare-improving over x, which is opposite to her true preference.

In contrast to Bernheim and Rangel (2009), our model correctly identifies her true preference between y and x by Theorem 1. To see this, suppose all of four products are available. Then, y and z are placed in the first aisle, so z is chosen. When y becomes unavailable, then x is moved to the first aisle and is chosen. Furthermore, when z is also sold out, then x and t are placed in the first aisle, so she picks t. In sum, her choices will be $c(xyzT) = z$, $c(xzT) = x$ and $c(xT) = t$. Then, when only choice is observable, our model concludes that the DM prefers z over y and x over z. Therefore, we can identify her preference between x and y correctly.

This example highlights the importance of knowledge about the underlying choice procedure when we conduct welfare analysis.[16] In other words, welfare analysis is more delicate a task than it looks.

Next, we investigate when we can unambiguously conclude that the DM pays (or does not pay) attention to an alternative. Consider the choice reversal above, from which we have concluded that she prefers x to y. Therefore, whenever y is chosen, she must not have paid attention to x (*Revealed Inattention*).

As we illustrate, we infer that x is revealed to attract attention at S whenever x is chosen from S or removing x from S causes a choice reversal. Furthermore, it is possible to reach the same conclusion even when removing x from S does not cause a choice reversal. Imagine that the DM chooses the same item, say $\alpha \neq x$, from S and T and removing x from T causes a choice reversal, so we know $x \in \Gamma(T)$ for sure. Now collect all items that belong to either S or T but not to both. Suppose all of those items are revealed to be preferred to α. Then, those items cannot be in $\Gamma(S)$ or $\Gamma(T)$. Therefore, removing those items from S or T cannot change her consideration set. Hence, we have

$$\Gamma(S) = \Gamma(S \cap T) = \Gamma(T)$$

and can conclude that x is considered at S.

The following theorem summarizes this observation and also provides the full characterization of revealed attention and inattention.

[16] For a detailed discussion of this subject, see Manzini and Mariotti (2009b).

THEOREM 2 (Revealed (In)Attention): *Suppose c is a CLA. Then*,

(i) *x is revealed not to attract attention at S if and only if* $xP_R c(S)$,

(ii) *x is revealed to attract attention at S if and only if there exists T* (*possibly equal to S*) *such that:*

$$(a) \qquad c(T) \neq c(T \backslash x),$$

$$(b) \qquad yP_R c(S) \text{ for all } y \in S \backslash T, \quad zP_R c(T) \text{ for all } z \in T \backslash S.$$

Theorem 2 identifies both revealed attention and inattention. This information is as important as the revealed preference. For example, if a product is not popular in a market, it is very important for a firm to know the reason, which can be either that it is not liked by consumers or that it does not attract the attention of consumers.

III. Characterization

The two preceding theorems characterize revealed preference and revealed (in) attention. They are not applicable, however, unless the observed choice behavior is a CLA. Therefore, a question to ponder is: how can we test whether a choice data is consistent with CLA? Surprisingly, it turns out that CLA can be characterized simply by only one behavioral postulate of choice.

Before we state the postulate, recall the sufficient and necessary condition for observed behavior to be consistent with the preference maximization under the full attention assumption: the Weak Axiom of Revealed Preference (WARP). WARP is equivalent to stating that every set S has the "best" alternative x^* in the sense that it must be chosen from any 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^*,

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

Because of the full attention assumption, being feasible is equal to attracting attention. This is no longer true when we allow for the possibility of limited attention, however. To conclude that x^* is chosen from T, we not only need to make sure that the chosen element from T is in S and x^* is available but also that x^* attracts attention. As we have discussed, we can infer this when removing x^* from T changes the DM's choice, which is the additional requirement for x^* to be chosen from T. This discussion suggests the following postulate, which is a weakening of WARP:

WARP with Limited Attention (WARP(LA)): For any nonempty S, there exists $x^* \in S$ such that, for any T including x^*,

$$\text{if } c(T) \in S \text{ and } c(T) \neq c(T \backslash x^*), \text{ then } c(T) = x^*.$$

WARP(LA) indeed guarantees that the binary relation P defined in equation (1) is acyclic and fully characterizes the class of choice functions generated by an attention filter. The next lemma makes it clear that WARP(LA) is equivalent to the fact that P has no cycle.

LEMMA 1: *P is acyclic if and only if c satisfies WARP with Limited Attention.*

PROOF:

(The if-part) Suppose P has a cycle: $x_1 P x_2 P \cdots P x_k P x_1$. Then for each $i = 1, \ldots, k-1$ there exists T_i such that $x_i = c(T_i) \neq c(T_i \backslash x_{i+1})$ and $x_k = c(T_k) \neq c(T_k \backslash x_1)$. Consider the set $\{x_1, \ldots, x_k\} \equiv S$. Then, for every $x \in S$, there exists T such that $c(T) \in S$ and $c(T \backslash x) \neq c(T)$ but $x \neq c(T)$, so WARP(LA) is violated.

(The only-if part) Suppose P is acyclic. Then every S has at least one element x such that there is no $y \in S$ with yPx, which means that there is no $y \in S$ with $y = c(T) \neq c(T \backslash x)$. Equivalently, whenever $c(T) \in S$ and $c(T) \neq c(T \backslash x)$, it must be $x = c(T)$, which is WARP(LA).

THEOREM 3 (Characterization): *c satisfies WARP(LA) if and only if c is a CLA.*

Theorem 3 shows that a CLA is captured by a single behavioral postulate. This makes it possible to test our model nonparametrically by using the standard revealed-preference technique à la Samuelson and to derive the DM's preferences and attention filter based on Theorem 1 and 2 from the observed choice data.

As we mentioned in the introduction, there are several related decision theoretic models where the final choice is made after eliminating several items, which are similar to a CLA such as Manzini and Mariotti (2007, forthcoming), Cherepanov, Feddersen, and Sandroni (2010), and Lleras et al. (2010). We shall illustrate that our model is different from these models both in a descriptive sense and in a behavior sense.

To show the difference more starkly, we compare our model with the "Rationalization" concept in Cherepanov, Feddersen, and Sandroni (2010). At first glance, Rationalization would appear to be a special case of our model. In fact,this is not the case. In the Rationalization model, the DM chooses the best alternative among those she can rationalize. The set of rationalizable alternatives is defined by her set of rationales. Each rationale is a transitive binary relation that may or may not be complete. The set of rationalizable alternatives in S consists of all the alternatives that dominate all other alternatives according to at least one of her rationales. Formally,

$$\Gamma_{CFS}(S) = \{y \in S \,|\, \exists R_i \text{ such that } y R_i x \text{ for all } x \in S\},$$

where each R_i is a rationale (a transitive binary relation).

In general, Γ_{CFS} is not an attention filter. To see this, consider three alternatives x, y, z and two rationales: $x R_1 y R_1 z$ and $y R_2 x$. First, observe that when all options are present, then x is rationalizable but z is not. On the other hand, y is rationalizable only when z is removed because R_2 does not compare y and z. That is, $z \notin \Gamma_{CFS}(xyz)$ but

$\Gamma_{CFS}(xyz) \neq \Gamma_{CFS}(xy)$—whereas our framework requires $\Gamma_{CFS}(xyz) = \Gamma_{CFS}(xy)$. This example shows that there are rationales that do not satisfy the conditions of our model. At the same time, it is easy to show that for any rationalization,

$$x \in \Gamma_{CFS}(S) \Rightarrow x \in \Gamma_{CFS}(T) \text{ for all } x \in T \subset S.$$

This property does not necessarily hold in our framework (e.g., Most Popular Category). Hence, there are attention filters that do not satisfy the conditions of their model. In short, neither model is a special case of the other.

One can modify Rationalization to make it a proper special case of our model. The necessary modification requires that the admissible rationales are not only transitive but are also complete.[17] If Rationalization were restricted in this way, each rationalizable alternative is an attention filter (though the converse is still not true).

We now demonstrate how these models differ from the CLA model behaviorally by means of examples. First, we shall present an example of a CLA that cannot be explained by any of these models. Although these models have different characterizations, all of them satisfy the axiom called Weak WARP (Manzini and Mariotti 2007) so we only need to show that it violates that axiom. The Weak WARP states that if x is chosen over y both from the pair and from a larger set, y cannot be chosen from anywhere between. Formally,

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

Consider the following example of a CLA:

EXAMPLE 2: *There are four alternatives* x, y, a, b. *The alternatives* a *and* b *are never chosen* (*unless there is no other alternative*) *but they alter the attention of the DM. Her preference is* $y \succ x \succ a \succ b$ *and she picks the best alternative from those she considers. She considers* y *only when either* a *or* b *is feasible, but not both, and always considers all other alternatives. It is easy to see that her consideration set mapping is an attention filter so her choice function satisfies WARP*(*LA*). *It does not satisfy Weak-WARP, however, because* $c(xy) = c(xyab) = x$ (*y is not considered*) *but* $c(xya) = y$.

Conversely, none of the above alternative models is a special case of the CLA model. In Example 3, we present a model of the Rational Shortlist Method of Manzini and Mariotti (2007) that cannot be a CLA. One can easily verify that exactly the same choice function can be generated by other models mentioned above. The rational shortlist model consists of two rationales, P_1 and P_2, where P_1 has no cycle (not necessarily transitive) and P_2 is a complete and transitive order.[18] The decision

[17] "The top on each criterion" introduced in Section IA coincides with the rationalization model when all rationales are complete.

[18] Actually, Manzini and Mariotti (2007) do not require the second rationale (P_2) to be complete and transitive (it only requires P_2 to be asymmetric). We put the stronger requirement on P_2 in order to highlight that the difference between these models is generated by the first stage, not by the incompleteness or intransitivity of the second rationale, which corresponds to the DM's preference in our model.

is made applying these rationales sequentially to eliminate alternatives. Consider the following example of the rational shortlist model:

EXAMPLE 3: *The first rationale* (*not transitive*[19]) *and the second rationale* (*transitive*) *are:*

$$tP_1y, yP_1x, zP_1x, zP_1s,$$

$$sP_2xP_2yP_2zP_2t.$$

For instance, if the feasible set is $\{s, y, z\}$, *s is eliminated in the first stage by z and she picks y in the second stage by comparing y and z according to* P_2. *This choice function, however, would generate contradictory revealed preferences if it were a CLA:*

- zPt *since* $z = c(yzt)$ *and* $y = c(yz)$,
- tPy *since* $t = c(xyt)$ *and* $x = c(xt)$,
- yPz *since* $y = c(syz)$ *and* $s = c(sy)$.

Thus, it cannot be explained by a CLA by Lemma 1. Hence, this choice cannot be a part of our model.

IV. Anomalies

Our limited attention model is capable of accommodating several frequently observed behaviors: Attraction Effect, Cyclical Choice, and Choosing Pairwisely Unchosen. Our explanations for these choice patterns depend solely on limited attention; hence, seemingly irrational behaviors can be explained without introducing changing preference. We will overview them and illustrate how our model accommodates them. In addition, we elicit the DM's preference, attention, and inattention from such choice data.

Attraction Effect.—The attraction effect refers to a phenomenon where adding an irrelevant alternative to a choice set affects the choice.[20] A typical attraction effect choice patterns is

$$c(xyd) = y, \quad c(xy) = x, \quad c(yd) = y, \quad c(xd) = x.$$

Here d is the irrelevant alternative that shifts the choice from x to y.[21] Thus, d is the decoy of y. Lehmann and Pan (1994) show experimentally that introducing new

[19] One can show that if P_1 is transitive, the first-stage elimination generates an attention filter so the resulting choice will be a CLA as long as P_2 is complete and transitive.

[20] This phenomenon is well-documented and robust in behavioral research on marketing (Huber, Payne, and Puto 1982; Tversky and Simonson 1993), including choices among monetary gambles, political candidates, job candidates, environmental issues, and medical decision making. Advertising irrelevant alternatives is commonly used as a marketing strategy to invoke the attraction effect on the customers.

[21] The standard continuity is inconsistent with the attraction effect: $x = c(x, d_n, y)$ for all n but y is chosen at the limit $(y = c(x, y))$ where $\{d_n\}$ is a sequence of x's decoys converging to x. Nevertheless, the model can still enjoy a

products causes an attraction effect particularly by affecting the composition of consideration sets. How the CLA model accommodates the attraction effect is in line with their findings. One possible representation is that the DM's preference is $y \succ x \succ d$ and she considers y only when d is present (otherwise, she considers everything). It is clear that her consideration set mapping is an attention filter.

Now we elicit the preference of a DM whose choice behavior follows the same pattern above without knowing her preference and consideration sets. By Theorem 1, $y = c(xyd) \neq c(xy)$ implies that y is revealed to be preferred to d. That is, our model judges that she prefers y over its own decoy.

Although most of the research on attraction effect is centered around one decoy option, a natural extension of the attraction effect is to include additional decoys. In particular, what happens if a decoy of x is introduced in addition to the aforementioned example? Teppan and Felfernig (2009) demonstrated that displaying both a decoy of x and a decoy of y along with x and y will lead the DM to choose as if there were no decoys.[22]

Formally, suppose that there are two decoys, d_x and d_y of x and y, respectively. That is,

$$c(xyd_xd_y) = x, \quad c(xyd_y) = y, \quad c(xy) = x.$$

Most of the theoretical literature, including that which can accomodate the attraction effect with one decoy option, cannot accomodate this choice behavior.[23] Nevertheless, the CLA model can accommodate this behavior: she considers y only when d_y is present but d_x is not. She ignores x when d_y is available but not d_x. Then she will exhibit the above choice as long as she prefers x over d_x and y over d_y.

Again, assume we have no prior information about the DM's preference and consideration sets. The first two choices reveal that she pays attention to d_x at $\{x, y, d_x, d_y\}$ so prefers x over d_x. Similarly, the second and third tell us she prefers y over d_y. Therefore, our approach again elicits her preference between an alternative and its decoy.

Here we rely on the paper by Lehmann and Pan (1994), which suggests experimentally that attraction effect is due to the composition of consideration sets. There are other explanations, however, for attraction effect (Huber, Payne, and Puto 1982). For example, one explanation concerns the DM being able to "give a reason" for the choice of x over y or vice versa. An asymmetrically dominated alternative gives such a reason. It seems that each explanation could be more appropriate than the others depending on the environment.

weaker continuity along with the attraction effect. For example, assume $y_n \to y$ and $y, y_n \notin S$, then

$$\text{If } y_n \notin c(S \cup y_n) \text{ then } \{y\} \neq c(S \cup y_n).$$

Indeed, one can show that the CLA is continuous in this sense if $\succ$ is continuous and the attention filter satisfies: (a) $y_n \notin \Gamma(S \cup y_n)$ implies $y \notin \Gamma(S \cup y)$; and (b) $z \in \Gamma(S \cup y_n)$ implies $z \in \Gamma(S \cup y)$ when $y_n \to y$.

[22] Eliaz and Spiegler (2011) studied a game theoretical model where firms would like to influence consumers' consideration sets by introducing costly decoys.

[23] This generalized attraction effect is another example that lies outside of recent models provided in Cherepanov, Feddersen, and Sandroni (2010), Manzini and Mariotti (forthcoming), and Lleras et al. (2010) since it does not satisfy Weak WARP. There are two exceptions: Ok, Ortoleva, and Riella (2010), and de Clippel and Eliaz (2012). These two models, however, can accommodate neither Cyclical nor Choosing Pairwisely choice patterns.

Table 2—Choosing Pairwisely Unchosen

Revealed preference	zP_Rx and zP_Ry			
	$\{x, y, z\}$	$\{x, y\}$	$\{y, z\}$	$\{x, z\}$
Revealed attention	xyz	x	y	x
Revealed inattention	—	—	z	z

Cyclical Choice.—May (1954) provides the first experiment where cyclical choice patterns are observed and these results have been replicated in many different choice environments (e.g., Tversky 1969; Loomes, Starmer, and Sugden 1991; Manzini and Mariotti 2009a; Mandler, Manzini, and Mariotti 2010). Consider a cyclical choice pattern:

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

We have already illustrated that this choice pattern can be captured by our model at the beginning of Section II. Now let us elicit the preference. Since the DM exhibits a choice reversal when y is removed from $\{x, y, z\}$, we can identify that y attracts her attention when these three elements are present. So, we can conclude that she prefers x over y. As illustrated before, however, we cannot determine the ranking of z.

Choosing Pairwisely Unchosen.—In this choice pattern, the DM chooses an alternative that is never chosen from pairwise comparisons:

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

Since removing x or y from $\{x, y, z\}$ changes her choice, it is revealed that z is better than x and y but we cannot determine her preference between x and y. Since her revealed preference has no cycle, her behavior is captured by our model through Lemma 1 and Theorem 3.

Note that the best element, z, is not chosen in any binary choice, so we can conclude that she pays attention to z only when x and y are present. Applying Theorem 2, we can pin down her consideration set uniquely except when her feasible set is $\{x, y\}$. (See Table 2.)

One possible story that generates such an attention pattern is "searching more when the decision is tough." Several items are hard to find even if they are feasible. The DM first considers alternatives that are feasible and easy to find and if there is an item that dominates all others, she chooses it immediately. Otherwise, she makes an extensive search to find all feasible items. In the former case, the consideration set consists only of easily found (and feasible) alternatives and in the latter case it coincides with the feasible set. Given this story, suppose her true preference is $z \succ x \succ y$ where the decision between x and y is very tough and z is hard to find. She makes an extensive search to find z only if she sees both x and y. If either x or y is missing, she does not bother to search, and therefore overlooks z.

V. Further Comments on Revealed Preference

In this section, we discuss the boundaries of our revealed preference approach. First of all, our revealed preference could be very incomplete; in other words, it only provides coarse welfare judgments. In the extreme case where the choice data satisfies WARP, Theorem 1 and Theorem 2 do not provide any identification of the preference and attention/inattention. This is because the DM's behavior can be attributed fully to her preference or to her inattention (never considering anything other than her actual choice). Thus, we cannot make any statement without imposing any additional assumption. This extreme example illustrates the limitation of choice data, which alone is not enough to identify her preferences. Notice that the classical revealed preference is not an exception since it assumes implicitly full attention.

Nevertheless, a policymaker may be *forced* to make a welfare judgment even when our revealed preference is silent. There are three directions to deal with incompleteness of our revealed preference: (i) looking for additional data other than choice data; (ii) imposing additional structures on attention filter; and/or (iii) utilizing other methods as long as the resulting revealed preference includes ours. We will discuss each of them in detail.

Additional Data.—The idea of our (direct) revealed preference is that we can conclude that x is preferred to y if x is chosen while y receives attention, which is inferred because removing y changes her choice. If we know y is considered for some other reason, however, we will naturally make the same conclusion even without observing such a choice change. One can obtain such information from many sources, such as eye-tracking, functional magnetic resonance imaging, and the tracking system in internet commerce.[24] If the policymaker believes that these sources are trustworthy, he can utilize them to obtain additional information about preferences.

Furthermore, additional information about preferences can also have a cascading effect. For instance, the choice data may not reveal the ranking between x and y but some laboratory experiment or survey study may have already found that x is better than y. In such a case, the policymaker can add $x \succ y$ to the revealed preference generated by our method (Theorem 1), say $P' = P_R \cup \{(x,y)\}$. By using the transitive closure of P', denoted by P'_R, the policymaker can obtain more attention/inattention information as in Theorem 2. Indeed, Theorem 1 and 2 are exactly applicable by replacing P_R with the transitive closure of P'_R.

Similarly, a policymaker may know that the consumer pays attention to x under certain decision problems. This information immediately generates more information about her preference (the chosen element here is better than x), which tells more about her attention/inattention, as in the previous case.

Further Restrictions on Consideration Sets.—The other direction is to impose additional restrictions on Γ. For example, if the source of limited attention is simply the abundance of alternatives, one reasonable restriction is that the DM considers

[24] In this regard, our theory highlights the importance of other tools (besides observed choice) that can shed light on the choice process rather than outcome.

at least two alternatives for each decision problem. That is, $|\Gamma(S)| \geqslant 2$. Under this restriction, the choice data reveals the consumer's preference completely. This result is trivial but still it is important in order to identify whether an unchosen alternative attracts attention. Our approach will provide an answer for the revealed attention. The revealed attention and inattention will be characterized by Theorem 2 by replacing P_R with the completely identified preference.

Notice that the classical revealed preference can be seen as one of such an attempt with the strongest assumption on the consideration set $\Gamma(S) = S$. Our model highlights that we need to assume *how* choices are made in order to make a meaningful revealed preference exercise. The assumption about *what* is chosen (like WARP) is not enough.

Other Methods.—One can combine our methodology with others which try to make the welfare analysis without relying on a particular choice procedure, such as Apesteguia and Ballester (2010), and Bernheim and Rangel (2009). What is common between our model and theirs is that all try to respect the consumer's choice for the welfare judgment as much as possible. The difference is that our model does so only when the consumer actually considers other unchosen alternatives.

Now imagine that a policymaker knows/believes a consumer behaves according to our model. Then, he should first elicit her preference based on our method. Admittedly, it only provides an incomplete ranking (and empty if the choice data satisfies WARP). If the policymaker is *forced* to make a complete welfare judgment with a risk of making mistakes, he can apply the other methods with the constraint of respecting the revealed preference generated by our model. In other words, these methods should be used to break the incompleteness of our revealed preference.

For instance, consider Apesteguia and Ballester's approach. They first axiomatically construct an index to measure the consistency between choice data and a certain preference, and of all complete and transitive preferences pick the one that minimizes the inconsistency for the welfare analysis. If the policymaker knows the DM follows a choice with limited attention, however, he should first elicit her preference based on our method and then pick the inconsistency-minimizing preference *only from those that are consistent with our revealed preference*. The resulting welfare judgment can be wrong (can be different from her actual preference). Nevertheless, this sequential process eliminates certain mistakes the policymaker would make if he simply applied the other model-free methods. For instance, applying Apesteguia and Ballester's approach directly to Example 1 will lead to the wrong conclusion: y is welfare-improving over x but this sequential advocacy certainly kills such a mistake.

VI. Conclusion

Limited attention has been widely studied in economics: neglecting the nontransparent taxes (Chetty, Looney, and Kroft 2009); inattention to released information (DellaVigna and Pollet 2007); costly information acquisition (Gabaix et al. 2006); and rational inattention in macroeconomics (Sims 2003). For example, Goeree (2008) shows that relaxing the full attention assumption by allowing customers to be unaware of some computers in the market is enough to explain the high markups in the PC industry.

In this paper, we study the implications of limited attention on revealed preference. We illustrate when and how one can deduce both the preference and consideration sets of a DM who follows a CLA. The distinction between a preference and an (in)attention is crucial. For instance, if a product is not popular in a market, it is very important for a firm to know the reason, which can either be that it is not liked by consumers or that it does not attract the attention of consumers. Our model provides a theoretical framework to distinguish these two possibilities. Similarly, a social planner can find a proper strategy to make sure that people choose the right option in 401(K) plans and health insurance. Hence, in a welfare analysis it is important to understand the underlying model of the DM.

Since revealed preference and (in)attention are the main focus of the paper, we impose a rather weak restriction on consideration sets. Such a weak condition allows us to apply our revealed preference and (in)attention theorems to seemingly irrational choice patterns (i.e., Attraction Effect, Cyclical Choice, and Choosing Pairwisely Unchosen). Nevertheless, depending on the intended application, our framework can be used to analyze choices under different restrictions on consideration sets.

In many real-world markets, products compete with each other for the space in the consideration set of the DM, who has cognitive limitations. In these situations, if an alternative attracts attention when there exist many others, then it is easier to be considered when some of other alternatives become unavailable. If a product is able to attract attention on a crowded supermarket shelf, the same product will be noticed when there are fewer alternatives; i.e., $x \in \Gamma(T)$ implies $x \in \Gamma(S)$ whenever $x \in S \subset T$. Lleras et al. (2010) extensively study consideration sets that satisfy this property. They also consider the cases where both conditions are satisfied.

Lleras et al. (2010) also consider another special case whereby the DM overlooks or disregards an alternative because it is dominated by another item in some aspect. Imagine that Maryland's economics department is hiring one tenure-track theorist. Since there are too many candidates in the job market to consider all of them, the department asks other departments to recommend their best theory students. 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 an irreflexive and transitive order as long as each department's ranking over its students is rational. Formally, given an irreflexive and transitive order $\rhd$[25], the attention filter consists of alternatives that are undominated with respect to this order, $\Gamma_{\rhd}(S) = \{x \in S \mid \nexists y \in S \text{ such that } y \rhd x\}$.

Appendix: Proofs

Notice that the if parts of Theorem 1 and Theorem 2 have already been shown in the main text. The following proofs use these results.

[25] This order is not necessarily complete, as in this example; Michigan does not compare its students with candidates from other schools.

PROOF OF THEOREM 3:

Suppose c is a CLA represented by $(\succ, \Gamma)$. Then Theorem 1(if part) implies that $\succ$ must include P so P must be acyclic. Therefore, by Lemma 1, c must satisfy WARP(LA).

Now suppose that c satisfies WARP(LA). By Lemma 1, P is acyclic so there is a preference $\succ$ that includes P. Pick any such preference arbitrarily and define

$$\Gamma(S) = \{x \in S : c(S) \succ x\} \cup \{c(S)\}. \tag{2}$$

Then, it is clear that $c(S)$ is the unique $\succ$-best element in $\Gamma(S)$, so all we need to show is that Γ is an attention filter. Suppose $x \in S$ but $x \notin \Gamma(S)$ $(\text{so } x \neq c(S))$. By construction, $x \succ c(S)$ so it cannot be $c(S)Px$. Hence, it must be $c(S) = c(S \backslash x)$ so we have $\Gamma(S) = \Gamma(S \backslash x)$.

PROOF OF THEOREM 1 (The Only-If Part):

Suppose $xP_R y$ does not hold. Then there exists a preference that includes P_R and ranks y better than x. The proof of Theorem 3 shows that c can be represented by such a preference so x is not revealed to be preferred to y.

PROOF OF THEOREM 2 (The Only-If Parts):

(Revealed Inattention) Suppose x is not revealed to be preferred to $c(S)$. Then pick a preference that includes P_R and puts $c(S)$ above x. The proof of Theorem 3 shows that c can be represented by such a preference and an attention filter Γ with $x \in \Gamma(S)$.

(Revealed Attention) Suppose there exists no T that satisfies the condition. We shall prove that if c is a CLA then it can be represented by some attention filter Γ with $x \notin \Gamma(S)$. If $c(S)\, P_R x$ does not hold, we have already shown that c can be represented with $x \succ c(S)$ and $x \notin \Gamma(S)$, so x is not revealed to attract attention at S, so we focus on the case when $c(S)\, P_R x$.

Now construct a binary relation, $\tilde{P}$, where $a \tilde{P} b$ if and only if "$a P_R b$" or "$a = c(S)$ and not $b P_R c(S)$." That is, $\tilde{P}$ puts $c(S)$ as high as possible as long as it does not contradict P_R. Since P_R is acyclic and c is represented by an attention filter, one can show that $\tilde{P}$ is also acyclic. Given this, take any preference relation $\succ$ that includes $\tilde{P}$, which includes P_R as well. We have already shown that $\tilde{\Gamma}(S) \equiv \{z \in S : c(S) \succ z\} \cup \{c(S)\}$ is an attention filter and $(\tilde{\Gamma}, \succ)$ represents c. Now define Γ as follows:

$$\Gamma(S') = \begin{cases} \tilde{\Gamma}(S') & \text{for } S' \notin \mathcal{D} \\ \tilde{\Gamma}(S') \backslash x & \text{for } S' \in \mathcal{D} \end{cases},$$

where $\mathcal{D}$ is a collections of sets such that

$$\mathcal{D} = \left\{ S' \subset X : \begin{array}{cc} c(S') = c(S) & \text{and} \\ z P_R c(S) \text{ for all } z \in (S \backslash S') \cup (S' \backslash S) & \end{array} \right\}.$$

That is, Γ is obtained from $\tilde{\Gamma}$ by removing from x any budget set S' where $c(S) = c(S')$ and any item that belongs to S or S' but not to both is revealed to be better than $c(S)$. Notice that x cannot be $c(S)$ because if this true, the condition of the statement is satisfied for $T = S$. Hence, $\Gamma(S') \subset \tilde{\Gamma}(S')$ always includes $c(S')$. Furthermore, the proof of Theorem 3 shows that $(\tilde{\Gamma},\succ)$ represents c. Therefore, $(\Gamma,\succ)$ also represents c, so we only need to show that Γ is an attention filter.

To do that, it is useful to notice that $\tilde{\Gamma}$ is an attention filter and $c(T') = c(T'')$ whenever $\tilde{\Gamma}(T') = \tilde{\Gamma}(T'')$ because $(\tilde{\Gamma},\succ)$ represents c.

Suppose $y \notin \Gamma(T)$. We shall prove $\Gamma(T) = \Gamma(T \setminus y)$.

Case I: $y = x$

If $T \notin \mathcal{D}$, then we have $\Gamma(T) = \tilde{\Gamma}(T) = \tilde{\Gamma}(T \setminus x) = \Gamma(T \setminus x)$. If $T \in \mathcal{D}$, then it must be $c(T) = c(T \setminus x)$ (otherwise, the condition of the statement is satisfied) so by construction of $\tilde{\Gamma}$ and Γ, we have $\Gamma(T) = \tilde{\Gamma}(T) \setminus x = \tilde{\Gamma}(T \setminus x) = \Gamma(T \setminus x)$.

Case II: $T \in \mathcal{D}$ and $y \neq x$

Since $y \notin \Gamma(T)$ is equivalent to $y \notin \tilde{\Gamma}(T)$, we have $\tilde{\Gamma}(T) = \tilde{\Gamma}(T \setminus y)$. Therefore, $c(T \setminus y) = c(T) = c(S)$. By construction of Γ and $\tilde{\Gamma}$, it must be $y \succ c(S)$, which implies $y P_R c(S)$ by construction of $\succ$. Therefore, $T \setminus y \in \mathcal{D}$. Therefore, $\Gamma(T) = \tilde{\Gamma}(T) \setminus x = \tilde{\Gamma}(T \setminus y) \setminus x = \Gamma(T \setminus y)$.

Case III: $T \notin \mathcal{D}$ and $y \neq x$

If $T \setminus y \in \mathcal{D}$, analogous to the previous case, we have $c(T) = c(T \setminus y) = c(S)$ and $y P_R c(S)$, so it must be $T \in \mathcal{D}$, which is a contradiction. Hence, $T \setminus y \notin \mathcal{D}$, so we have $\Gamma(T) = \tilde{\Gamma}(T) = \tilde{\Gamma}(T \setminus y) = \Gamma(T \setminus y)$.

REFERENCES

Ambrus, Attila, and Kareen Rozen. 2010. “Rationalizing Choice with Multi-Self Models.” Unpublished.

Apesteguia, Jose, and Miguel A. Ballester. 2009. “Choice by Sequential Procedures.” Unpublished.

Apesteguia, Jose, and Miguel A. Ballester. 2010. “A Measure of Rationality and Welfare.” Unpublished.

Aumann, Robert J. 2005. “Musings on Information and Knowledge.” *Economics Journal Watch* 2 (1): 88–96.

Basar, Gozen, and Chandra Bhat. 2004. “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.

Bernheim, B. Douglas, and Antonio Rangel. 2007. “Toward Choice-Theoretic Foundations for Behavioral Welfare Economics.” *American Economic Review* 97 (2): 464–70.

Bernheim, B. Douglas, and Antonio Rangel. 2009. “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics.” *Quarterly Journal of Economics* 124 (1): 51–104.

Broadbent, Donald E. 1958. *Perception and Communication*. New York: Pergamon Press.

Caplin, Andrew, and Mark Dean. 2011. “Search, Choice, and Revealed Preference.” *Theoretical Economics* 6 (1): 19–48.

Caplin, Andrew, Mark Dean, and Daniel Martin. 2011. “Search and Satisficing.” *American Economic Review* 101 (7): 2899–2922.

Chambers, Christopher P., and Takashi Hayashi. 2008. “Choice and Individual Welfare.” California Institute of Technology Division of the Humanities and Social Sciences Working Paper 1286.

Cherepanov, Vadim, Timothy Feddersen, and Alvaro Sandroni. 2010. “Rationalization.” Unpublished.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. “Salience and Taxation: Theory and Evidence.” *American Economic Review* 99 (4): 1145–77.

Chiang, Jeongwen, Siddhartha Chib, and Chakravarthi Narasimhan. 1999. "Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity." *Journal of Econometrics* 89 (1–2): 223–48.

Dawes, Philip L., and Jennifer Brown. 2005. "The Composition of Consideration and Choice Sets in Undergraduate University Choice: An Exploratory Study." *Journal of Marketing for Higher Education* 14 (2): 37–59.

De Clippel, Geoffroy, and Kfir Eliaz. 2012. "Reason-Based Choice: A Bargaining Rationale for the Attraction and Compromise Effects." *Theoretical Economics* 7 (1): 125–62.

DellaVigna, Stefano, and Joshua M. Pollet. 2007. "Demographics and Industry Returns." *American Economic Review* 97 (5): 1667–1702.

Dulleck, Uwe, Franz Hackl, Bernhard Weiss, and Rudolf Winter-Ebmer. 2008. "Buying Online: Sequential Decision Making by Shopbot Visitors." Institut für Höhere Studien Wien Economics Series 225.

Eliaz, Kfir, Michael Richter, and Ariel Rubinstein. 2011. "Choosing the Two Finalists." *Economic Theory* 46 (2): 211–19.

Eliaz, Kfir, and Ran Spiegler. 2011. "Consideration Sets and Competitive Marketing." *Review of Economic Studies* 78 (1): 235–62.

Filiz-Ozbay, Emel. 2010. "Incorporating Awareness into Contract Theory." Unpublished.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg. 2006. "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model." *American Economic Review* 96 (4): 1043–68.

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

Green, Jerry, and Daniel Hojman. 2007. "Choice, Rationality and Welfare Measurement." Harvard University John F. Kennedy School of Government Working Paper 2144.

Hauser, John R., and Birger Wernerfelt. 1990. "An Evaluation Cost Model of Evoked Sets." *Journal of Consumer Research* 16: 383–408.

Hausman, Daniel. 2008. "Mindless or Mindful Economics: A Methodological Evaluation." In *The Foundations of Positive and Normative Economics: A Handbook*, edited by Andrew Caplin and Andrew Schotter, 125–55. New York: Oxford University Press.

Heifetz, Aviad, Martin Meier, and Burkhard C. Schipper. 2010. "Dynamic Unawareness and Rationalizable Behavior." Unpublished.

Hotchkiss, Gord, Steve Jensen, Manoj Jasra, and Doug Wilson. 2004. "The Role of Search in Business to Business Buying Decisions A Summary of Research Conducted." Unpublished.

Houy, Nicolas. 2007. "Rationality and Order-Dependent Sequential Rationality." *Theory and Decision* 62 (2): 119–34.

Houy, Nicolas, and Koichi Tadenuma. 2009. "Lexicographic Compositions of Multiple Criteria for Decision Making." *Journal of Economic Theory* 144 (4): 1770–82.

Huber, Joel, John W. Payne, and Christopher Puto. 1982. "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *Journal of Consumer Research* 9 (1): 90–98.

Huberman, Gur, and Tomer Regev. 2001. "Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar." *Journal of Finance* 56 (1): 387–96.

Lavidge, R. J., and G. A. Steiner. 1961. "A Model for Predictive Measurements of Advertising Effectiveness." *Journal of Marketing* 25 (6): 59–62.

Lehmann, Donald R., and Yigang Pan. 1994. "Context Effects, New Brand Entry, and Consideration Sets." *Journal of Marketing Research* 31 (3): 364–74.

Lleras, Juan S., Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay. 2010. "When More is Less: Choice by Limited Consideration." Unpublished.

Loomes, Graham, Chris Starmer, and Robert Sugden. 1991. "Observing Violations of Transitivity by Experimental Methods." *Econometrica* 59 (2): 425–39.

Mandler, Michael, Paola Manzini, and Marco Mariotti. 2010. "A Million Answers to Twenty Questions: Choosing by Checklist." Unpublished.

Manzini, Paola, and Marco Mariotti. 2007. "Sequentially Rationalizable Choice." *American Economic Review* 97 (5): 1824–39.

Manzini, Paola, and Marco Mariotti. 2009a. "Choice over Time." In *The Handbook of Rational and Social Choice*, edited by Paul Anand, Prastanta Pattanaik, and Clemens Puppe, 239–71. New York: Oxford University Press.

Manzini, Paola, and Marco Mariotti. 2009b. "Choice Based Welfare Economics for Boundedly Rational Agents." Unpublished.

Manzini, Paola, and Marco Mariotti. Forthcoming. "Categorize Then Choose: Boundedly Rational Choice and Welfare." *Journal of the European Economic Association.*

Masatlioglu, Yusufcan, and Daisuke Nakajima. 2009. "Choice by Iterative Search." Unpublished.

May, K. O. 1954. "Intransitivity, Utility, and the Aggregation of Preference Patterns." *Econometrica* 22 (1): 1–13.

Noor, Jawwad. 2011. "Temptation and Revealed Preference." *Econometrica* 79 (2): 601–44.

Ok, Efe A., Pietro Ortoleva, and Gil Riella, 2010. "Revealed (P)Reference Theory." Unpublished.

Ozbay, Erkut Y. 2008. "Unawareness and Strategic Announcements in Games with Uncertainty." Unpublished.

Pessemier, Edgar A. 1978. "Stochastic Properties of Changing Preferences." *American Economic Review* 68 (2): 380–85.

Richards, Max D., John E. Sheridan, and John W. Slocum. 1975. "Comparative Analysis of Expectancy and Heuristic Models of Decision Behavior." *Journal of Applied Psychology* 60 (3): 361–68.

Roberts, John H., and James M. Lattin. 1991. "Development and Testing of a Model of Consideration Set Composition." *Journal of Marketing Research* 28 (4): 429–40.

Rubinstein, Ariel, and Yuval Salant. 2012. "Eliciting Welfare Preferences from Behavioral Datasets." *Review of Economic Studies* 79 (1): 375–87.

Salant, Yuval, and Ariel Rubinstein. 2008. "(A, f): Choice with Frames." *Review of Economic Studies* 75 (4): 1287–96.

Samuelson, P. A. 1938. "A Note on the Pure Theory of Consumer's Behaviour." *Economica* 5 (17): 61–71.

Simon, Herbert A. 1957. *Models of Man: Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting.* New York: John Wiley and Sons, Inc.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–90.

Stigler, G. J. 1961. "The Economics of Information." *Journal of Political Economy* 69: 213–25.

Teppan, Erich Christian, and Alexander Felfernig. 2009. "Minimization of Product Utility Estimation Errors in Recommender Result Set Evaluations." In *WI-IAT 2009 Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, Vol. 1, 20–27. Milan, Italy: University of Milano-Bicocca.

Tversky, Amos. 1969. "Intransitivity of Preferences." *Psychological Review* 76: 31–48.

Tversky, Amos, and Itamar Simonson. 1993. "Context-Dependent Preferences." *Management Science* 39 (10): 1179–89.

Varian, Hal R. 2006." Revealed Preference." In *Samuelsonian Economics and the Twenty-First Century*, edited by M. Szenberg, L. Ramrattan, and A. A. Gottesman, 99–115. New York: Oxford University Press.

Wright, Peter, and Fredrick Barbour. 1977. "Phased Decision Strategies: Sequels to an Initial Screening." In *Multiple Criteria Decision Making,* edited by Martin K. Starr and Milan Zeleny, 91–109. Amsterdam: North-Holland Publishing Co.

Zyman, Sergio. 1999. *The End of Marketing as We Know It.* New York: Harper Collins.