ESSAYS ON DECISION THEORY

by

Yusufcan Demirkan

A thesis submitted to the

Graduate School of Social Sciences and Humanities

in partial fulfillment for the

degree of

Master of Arts

in

Economics

July 2017

Department of Economics

Koç University

Acknowledgements

I am greatly indebted to my thesis advisor Prof. Mert Kimya for his endless support and continuous guidance. I would like to show my greatest respect for Prof. İnsan Tunalı who supported me since the beginning of my undergraduate studies, and I am especially grateful for his invaluable contribution to the all aspects of my life.

I would like to thank Prof. Kemal Yıldız for being a member of my thesis committee and his valuable discussion.

I want to show my gratitude especially to my friend Umur Cem Akbaş, I have also benefited from various discussions with my friends Ayberk Arseven , İbrahim Ateşpare and Batuhan Solmaz.

It is for me a great pleasure to thank for the irreplaceable existence and unconditional support of Yeşim Odabaş.

Lastly, I want to thank to all my family members for everything they have done and they will do throughout my life.

This work is dedicated to my mom, Güler Demirkan, for her endless desire to raise me as a future scientist in every instance including her non-presence.

Abstract

This thesis consists of three different papers on the subject of decision theory. Economic theory tends to attract attention on binary relations existing between the objects of choice. Any rationalization is based upon a so called "Rationale" that solely depends on the binary relation over objects; however the unobservable concepts that are related to the choice set themselves may affect the choice process. In the first paper, we develop a model, where the decision maker has a concept set associated with each element in the choice set. Accordingly, she has a hierarchy of the concepts in her mind and orders the elements of the choice set with respect to each concept. Then she chooses the top element with respect to the most important concept. We will derive the empirical characterization and explain the scope of the procedure.

The second paper suggests a novel way to characterize a certain class of choice procedures which are defined by various conditions put on the consideration parameter. By showing the equality between the consideration parameter and the well-known hazard rate h, we will show that the machinery provided through the hazard rate h will characterize some of the famous stochastic choice procedures based on the consideration parameter directly. In addition, a new stochastic choice procedure based on a similarity relation is introduced and characterized.

The last paper connects two empirically observed facts through a new theory. The randomness of the choices people make and their lack of considering the full set of available alternatives are both well-known and empirically supported facts. We present a "theory of selves" approach to build a firm connection between these two observations by assuming that each decision maker (DM) has multiple deterministic consideration filters where she maximizes her well-defined preferences over each resulting consideration set. Each consideration filter corresponds to a different self that arises with some probability and thus the choice becomes random. We characterize this "Random Filtering" procedure using some of the influential consideration filter forms from the literature. Several well-known choice anomalies and context effects can be captured by the model.

Keywords: Consideration filter, random choice, bounded rationality, theory of selves, random utility models, revealed preference, concepts, consideration parameter, hazard rate.

Özet

Bu tez karar teorisi üzerine yazılmış üç adet makaleden oluşuyor. Ekonomik teori genellikle seçim objeleri arasında varolan ikili bağıntılara dikkat çeker. Herhangi bir rasyonalizasyon bu tip bir ikili bağıntıya dayanan “Rasyonal”e dayanır, fakat seçim kümesiyle bağlantılı olan gözlemlenemeyen konseptler de bu süreçte etkili olabilir. İlk makalede, her bir elemanla bir konsept kümesini bağdaştıran bir karar alıcıyı modelliyoruz. Karar alıcının aklında bir konseptler hiyerarşisi ve her bir elemanı bu konseptlere göre sıraladığı bir düzeni var. Kişi, en önemli konseptinin en üst sıraya koyduğu elemanı seçiyor. Bu modeli empirik karakterizasyonunu bulup gözlemlenebilir kapsamını açıklıyoruz.

İkinci makale dikkat parametreleri üzerine çeşitli kısıtlamalar konularak oluşturulan seçim prosedürlerini karakterize edebilmek için özgün bir yöntem öneriyor. Dikkat parametresi ve tehlike hızı arasındaki eşitlik gösterildikten sonra tehlike hızına dayanan bu yöntemin dikkat parametresi üzerinden tanımlanan önde gelen rassal seçim modellerinden bazılarını açıkladığını göstereceğiz. Ayrıca, benzerlik ilişkisine dayalı yeni bir rassal seçim prosedürü tanıtılıp karakterize edilecek.

Son makale empirik olarak gözlemlenmiş iki unsuru yeni bir teoriyle birbirine bağlıyor. İnsanların seçimlerinin rassal olduğu ve mevcut alternatiflerin hepsini dikkate almadıkları iyi bilinen ve ampirik olarak da gözlenmiş iki unsurdur. Bu iki gözlem arasında sıkı bir bağ kurmak adına bir “kişilikler teorisi” yaklaşımı sunuyoruz: Karar alıcı, iyi tanımlanmış tercihlerini, birden fazla olan deterministik ‘dikkat filtreleri’ sonucu oluşan ‘dikkat kümeleri’ üzerinde maksimuma çıkarıyor. Her bir ‘dikkat filtresi’, belirli bir olasılıkla ortaya çıkan bir kişiliğe tekabül ediyor ve birden fazla kişiliğin farklı olasılıklarla ortaya çıkması gözlenen seçimleri rassal hale getiriyor. Bu “Rassal Filtreleme” prosedürünü, literatürde etkin olan bazı ‘dikkat filtresi’ formları üzerinden tanımlayıp, empirik olarak karakterize ediyoruz. Model iyi bilinen birtakım seçim anormallikleri ve bağlam etkilerini açıklıyor.

Anahtar Sözcükler: Dikkat filtresi, rassal seçim, sınırlı rasyonalite, kişilikler teorisi, rassal fayda modelleri, açıklanan tercih, konseptler, dikkat parametresi, tehlike hızı.

Contents

1.Conceptual Choice **1**
1.1 Introduction 1
1.2 Model 3
1.3 Empirical Characterization 4
1.3.1 Characterization 4
1.3.2 Conceptual Dominance 5
1.4 Comparison to Other Models 6
1.5 Related Literature 7
1.6 A Quick Look at Two Extensions 9
1.6.1 Memory 9
1.6.2 Limited Concept Set 9
1.7 Conclusion 9
Appendix: Proofs 10
References 11

2.Characterization via Hazard Rate 14
2.1 Introduction 14
2.2 Model 15
2.3 Similarity- Based Stochastic Choice 19
2.4 Stochastic Consideration Set-MM 21
2.5 Monotonicity 21
2.6 A Luce Model 22
2.7 Duplicates 24
References 26
3. Random Filtering **27**
3.1 Introduction 27
3.2 Random Filtering 29
3.2.1 Results 31

3.3 Attention Filtering……33

3.4 Competition Filtering……….34

3.4.1 Rationalization/ Categorization……………………………………………………………………………………..35

3.5 Context Effects…….36

3.6 Literature Review………38

Appendix……..39

References………..42

Conceptual Choice

Yusufcan Demirkan *

Koc University, Rumelifeneri Yolu, 34450 Sariyer Istanbul Turkey.

August 14, 2017

Abstract

Economic theory tends to attract attention on binary relations existing between the objects of choice. Any rationalization is based upon a so called "Rationale" that solely depends on the binary relation over objects; however the unobservable concepts that are related to the choice set themselves may affect the choice process. In our model, the decision maker has a concept set associated with each element in the choice set. Accordingly, she has a hierarchy of the concepts in her mind and orders the elements of the choice set with respect to each concept. Then she chooses the top element with respect to the most important concept. We will derive the empirical characterization and explain the scope of the procedure. Several extensions and further improvements will be mentioned.

1 Introduction

Standard economic theory focuses on the relations, in particular on "preferences", that exists between the objects of the choice set for explaining decision problems. We, on the other hand, emphasize the importance of concepts that are related to the choice set. Working backwards, we define two primitive relations based upon the concept set itself: Every concept has a rank at the hierarchy, and each concept orders elements of the choice set. The decision maker chooses the element that is at the top according to the most important concept. Consider the following example taken from Cherepanov et al. (2013).

Assume an individual is facing an ethically sensible decision, whereby she should choose between watching movie 1 alone, movie 2 alone or watching movie 1 with a handicapped person. Assume also that she chooses movie 1 to movie 2 given only these two. If movie 1 alone and movie 1 with a handicapped person are given, then she chooses the latter one to the former. Finally, given all options she prefers to watch movie 2. Although such a profile is usually observed, according to the WARP this choice profile is contradictory. Observe that while the first choice is about choosing a movie alone, the second and third choices include an additional part whereby you should decide whether you will watch movie alone or not, because of the availability of the last alternative.

Intuitively, it is natural to associate the first choice with concepts such as movie and quality; whereas, the third option may evoke concepts such as social impact. Basically,

* *E-mail Address: ydemirkan@ku.edu.tr*

one can treat the first part as if the individual decides according to her own preferences over movies; while in the second part inclusion of the third option triggers social preferences. In such a case, a person who cares about the judgement of society upon her choices may consider social preferences as much more important than her own preferences, so that the concept of social impact lies above all other concepts. Such a person is motivated to choose watching movie 1 with a handicapped person rather than watching movie 1 alone; however, given the option of watching movie 2 alone that is socially acceptable, she changes her choice. Consequently, this contradictory choice we observe may be due to the importance of a concept associated with an option.

Another classic example from the literature is the "Attraction Effect".[1] This is the widely cited phenomenon in the psychology and marketing literature that is also called the "Asymmetric Dominance Effect".[2] A typical data corresponding to the attraction effect is provided below. Although y is chosen when x and y are the only options; x is chosen whenever d is present, since the alternative d is dominated only by x.

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

Our choice procedure reveals that a concept related to d is significant for this choice problem, since the removal of it leads to a choice reversal. In the simplest interpretation, it can be claimed that d is associated with the concept of dominance. Hence, the newly added object brings an important concept to the mind of the decision maker and causes a choice reversal, even though d itself is never chosen.

In general, we build a model where the decision maker has a concept set associated with each element in the choice set. The elements in the concept space are ranked according to a complete, asymmetric and transitive relation, $\rhd$. Besides, each concept in the concept space induces its own order on the choice set, which is also transitive, asymmetric and complete. The decision maker chooses the top element according to the most important concept in her mind.

To empirically test the model, we derive a characterization result that primarily depends on choice reversals. Through choice reversals, we can derive that a concept associated with a particular object x dominates all concepts associated with another object y. When this is the case, we say that x conceptually dominates y. If x conceptually dominates y, then removal of y in any set that contains x cannot lead to a choice reversal. We will call this condition *Dominated Reasoning* and it will completely characterize our model.

This is interestingly connected with Dietrich and List (2010), where a connection between formal rational choice theory and reasoning is built through usage of propositions. We believe that a "motivating reason" in their terminology can be constructed by the concepts that affect the choice problem. In the case of a single maximum concept as we assume here, it is the case that a single proposition that contains this concept is formed as a reason underlying choice. If we extend our model to the case where the foremost concepts is formed through logical connectives between multiple concepts, propositions can be extended respectively by logical connectives so that a more complex reasoning is constructed. Also, if we consider the last condition with respect to Dietrich and List

[1] For an interesting approach that captures this effect look at Clippel and Eliaz (2012), where choice is modeled as a cooperative solution to the bargaining problem between different selves of the decision maker.

[2] For notational ease, we suppressed set delimiters, i.e. $c(\{x, y\}) = c(xy)$ and $c(S \setminus \{x\}) = c(S \setminus x)$. In some circumstances where it can create confusion, standard notation is used.

(2010), it corresponds to not considering a dominated reason in any choice problem (on any set S) if the dominating reason is also present.

In a world where majority of decisions are made considerably fast because of the cost of time and complexity of the calculations, such a map that decisions are based on is meaningful. We know from psychology literature that our mind works associatively, and a conceptual map signaled by the choice set is compatible with this feature. Beyond these; conceptual approach provides a way to form complex justifications in propositional form, based upon the hierarchy of concepts. In this manner, extending the model into a stochastic version and considering the concept sets as the priors individuals form in their minds is useful. Building an extended model capturing both the formation and accumulation of concepts (in the process of constructing priors) through signals received from the objects of choice set is our future goal.

In the upcoming section, we will explain the model and the choice process. Then in Section 3 we will derive the empirical characterization of the model by *Dominated Reasoning*. The notion of *conceptual dominance* is introduced in Section 3.2, where we show our second result. In Section 4, we show the independence of our condition and Weak WARP. Besides, the connection between (LA)WARP and our condition is also demonstrated via additional examples. We will explore related literature in Section 5. Before the conclusion in Section 7, several possible extensions will be illustrated without going into further details in Section 6. The proofs of the results are in the Appendix.

2 Model

Let X be a finite set of elements. The set of all nonempty subsets of X is denoted by $\mathbb{X}$. As usual, the choice function is symbolized by $c : \mathbb{X} \rightarrow X$ where $c(S) \in S$ for all $S \in \mathbb{X}$. Denote by c_x the concepts associated with each element $x \in X$. The set of all concepts associated by the decision maker to all $S \subseteq X$ is denoted as $c^S = \cup_{x \in S} c_x$. There is a transitive, asymmetric and complete relation $\triangleright$ imposed upon c^X . $\forall c \in c^X$, $\exists >_c$ which is also asymmetric, transitive and complete such that it orders objects of X according to each concept.

Note that we only observe the choice data without any further information on the concept set and the hierarchy.[3] Next, we will define our choice procedure:

Definition 1. *(Conceptual Choice "CC") A choice function $c : \mathbb{X} \rightarrow X$ is (CC) if and only if there exists a set of concepts c_x $\forall x \in X$, $\triangleright$ on c^X and $>_c$ for all $c \in c^X$ such that $\forall S \subseteq X$:*

$$c(S) = \arg\max_S >_{\{ arg\, max_{c^S} \triangleright \}}$$

A special example where a choice anomaly observed is "Small-Large Donation" case outlined in Cherepanov et al. (2013). Consider the following choice data:

$$c(sn) = s, c(snl) = n$$

The decision maker chooses the option of small donation whenever the alternative of no donation is also given. On the other hand, if we add the option of large donation to the choice set, she chooses to donate nothing. In this case, the option of large donation behaves like a decoy as in the above example we gave. Let $c_l = \{cost\}$,

[3]Choice data includes the set of alternatives X and the choice function.

$c_s = \{social \quad impact\}$ and $c_n = \{parsimony\}$. It will reveal in the following sections that c_l has the most important concept. Since it is a singleton in this case, cost is considered to be the foremost concept for the decision maker. For now, assume that c_s is the second important concept. In short,

$$c_l \triangleright c_s \triangleright c_n$$

$$l >_s s >_s n$$

$$n >_l s >_l l$$

justifies this choice data according to our model.

3 Empirical Characterization

3.1 Characterization

In this part, we will try to observe the restrictions imposed by our model on the data. Then, we will show that our model is characterized empirically by C_1. In order to do this, we have to find an apparatus that works whenever a choice data suits *Conceptual Choice*. For this purpose, we will use choice reversals.

Assume that a choice reversal occurs when we remove an element from S that is not chosen from it, i.e. $z \neq c(S)$ and $c(S) \neq c(S \setminus z)$. Observe that this is only possible if z has the most important concept of the set c^S and according to this concept $c(S)$ is considered to be the best. Let y be any element in S such that $y \neq z$. Since by the above observation we know that z is associated with the maximum concept in c^S, removal of no element except $c(S)$ and z will not change the choice. Note that this is true for any set T containing such y and z, since the maximum concept associated with z ranks always better than the maximum concept associated with y. From this observation, we will state our necessary condition, which we will call simply C_1:

Condition 1. *(Dominated Reasoning "C_1")* $c(S) \neq c(S \setminus x)$ *where* $x \neq c(S) \implies$ $c(T)=c(T \setminus z)$ *for all* $z \in S \setminus \{x, c(T)\}$.

In addition to its necessity, this condition is also sufficient for the characterization of our choice procedure. With this condition at hand, we gain a powerful apparatus to analyze choice data. Lastly, we state the main result of this paper:

Theorem 1. *A choice function c satisfies Conceptual Choice if and only if C_1 holds.*

Recall the "Attraction Effect" example we gave in the introduction. In that example, there is only one choice reversal in the choice data given. This reversal occurs when we remove d from $\{x, y, d\}$. According to C_1, it must be the case that the removal of y does not change the choice. Indeed, it is so. On the other hand, as we showed in the introduction, it can be explained by *Conceptual Choice*. So, "Attraction Effect" is one of the wide range of examples we can explain with our model.

Let us also see an example where the choice data does not satisfy C_1. Taken from Masatlioglu et al. (2012), consider the following choice data:

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

This choice profile we observe is called as "Choosing Pairwisely Unchosen". Observe that a choice reversal occurs whenever we remove an element not equal to the chosen one. This implies that the maximum concept (call it c^*) in the concept set of $\{x, y, z\}$ is an element of $c_x \cap c_y$. From $x = c(xy) = c(xz)$, we derive that both x and y are ranked higher than z with respect to c^*. This is not in accordance with $z = c(xyz)$ which implies that z is ranked higher than others with respect to c^*. Therefore, *Conceptual Choice* is not able to explain such a choice anomaly. It clearly violates our C_1. To see this, note that removal of x from $\{x, y, z\}$ causes a choice reversal. According to C_1, it must be the case that the removal of y does not change the choice, but it is not the case. Hence, C_1 is violated.

3.2 Conceptual Dominance

Our goal here is to derive what we can learn from the choice data. In some examples, it may be the case that several different representations can lead to the same choice data we observe. To handle this problem, we define the following:

Definition 2. *Suppose c is a Conceptual Choice. We say that x conceptually dominates y if for all representations the maximum concept associated with x ranks higher than the maximum concept associated with y.*

The problem is to observe this dominance empirically. Following the discussion of the previous sections, we know that the tool of choice reversal gives us what we want. For example, if we remove an element that is not chosen from the respective set and the choice changes, then we conclude that this element must be associated with the maximum concept. From this observation, we can derive a relation that holds whenever a choice reversal occurs. Define the following:

Definition 3. *xDy if there exists a set S such that $c(S) \neq c(S \setminus x)$ where $x \neq c(S)$ and $y \in S$, where $y \neq x$.*

Observe that xDy implies that x *conceptually dominates* y. In addition, note that if xDy and yDz, xDz, since underlying $\rhd$ is transitive. Therefore, we can take transitive closure of D, which is denoted by $\overline{D}$. We can conclude that if $x\overline{D}y$, then x *conceptually dominates* y. Surprisingly, $\overline{D}$ completely characterizes *conceptual dominance.*

Theorem 2. *Suppose c is Conceptual Choice. Then x conceptually dominates y if and only if $x\overline{D}y$.*

Proof of the if part is shown above. For the proof of the only if part, please see the appendix.

Observe that our decoy element d in the "Attraction Effect" example demonstrates these notions directly. Since there is a choice reversal in the choice data, $d\overline{D}y$ and $d\overline{D}x$. Also, there may be two representations where $c_d^* \rhd_1 c_x^* \rhd_1 c_y^*$ and $c_d^* \rhd_2 c_y^* \rhd_1 c_x^*$.[4] In both cases, d conceptually dominates others by definition. Therefore, d in this example is a clear demonstration of what is shown in Theorem 2.

[4]Superscript * denotes the maximum element of the concept sets with respect to the corresponding hierarchy $\rhd_i$, $i = 1, 2$

4 Comparison to Other Models

In general, the literature of bounded rationality focuses on modeling heuristics that are shown to be widely used by individuals. We know that Weak WARP empirically characterizes wide range of models in the literature. These include Rationalization (Cherepanov et al. (2013)), CTC (Manzini and Mariotti (2012)) and Rational Shortlist (Manzini and Mariotti (2007)). Beyond going into the details of these models, let us state Weak WARP and show the independence between Weak WARP and C_1.

Definition 4. *(Weak WARP):* $\{x, y\} \subseteq M \subseteq S$, $x \neq y$, $c(S) = c(xy) = x \implies c(M) \neq y$.

By giving counterexamples, we will demonstrate the independence of C_1 and Weak WARP. First consider:

$$x = c(xywz) = c(xyw) = c(xwz) = c(xy) = c(xw) = c(xz)$$

$$y = c(xyz) = c(ywz) = c(yw) = c(yz)$$

$$z = c(wz)$$

This choice data definitely does not satisfy Weak WARP, since $\{x, y\} \subseteq \{x, y, z\} \subseteq \{x, y, w, z\}$, but $y = c(xyz)$. On the other hand, it does not violate our condition. Indeed; $c_w \rhd c_z \rhd c_x \rhd c_y$, $x >_w y >_w z >_w w$, $y >_z x >_z z$ and other suitable orders induce above choice data. Thus, Condition 1 is not subsumed by Weak WARP empirically. For the converse, consider following choice data:

$$x = c(xyz) = c(xz)$$

$$y = c(xy) = c(yz)$$

$$t = c(yzt) = c(yt)$$

$$z = c(yz) = c(zt)$$

Observe that our condition is violated. To see this, first note that z causes choice reversal in the first case where y is also contained. In such case, our condition says that removal of y from the second case where z also included should not lead to any choice reversal. Hence, this choice profile violates our condition. On the other hand, it does not violate Weak WARP. Thus, Weak WARP is not subsumed empirically by our condition. Therefore, we can conclude that these two conditions are independent empirically.

However, literature is not limited with the condition of Weak WARP. Another version of Weak WARP is Limited Attention Weak WARP -(LA)WARP- which characterizes Choice with Limited Attention (CLA) by Masatlioglu et al. (2012). This model differs from others by providing a very elegant procedure of eliciting preferences from choices using the notion of "attention", which is a fundamental cognitive constraint. Recall that we use choice reversals for empirical analysis of the choice data, so one can expect a connection between these models. However, this is not the case. First we define:

Definition 5. *WARP(LA): For any nonempty S, there exists $x^* \in S$ such that, for any T including x^*:*

$$if\ c(T) \in S\ and\ c(T) \neq c(T \setminus x^*),\ then\ c(T) = x^*$$

Remark 1. *C_1 implies $WARP(LA)$.*

From Masatlioglu et al. (2012), we know by $Lemma1$ that c satisfies $WARP(LA)$ if and only if P is acyclic. We will use this lemma to establish that our condition implies $WARP(LA)$. For the proof, please see the appendix.

Now consider the example given in Section 3.1 that is called "Choosing Pairwisely Unchosen". As demonstrated there, it cannot be explained by our model, however Masatlioglu et al. (2012) shows that their model can accommodate such a choice anomaly. Therefore, we can conclude that our model is empirically subsumed by their model.

Another example can be given from the dual-self theory (a special case of Kalai, Rubinstein and Spiegler (2002)). Consider

$$x = c(wxyz) = c(xy), y = c(xyz)$$

From our model, we can derive that $c_w \triangleright c_z$ and these two are the first and second concepts that maximize $\triangleright$ on the concept set of $\{w, x, y, z\}$. Also, we know that x maximizes $>_{c_w}$ on the whole set, while y maximizes $>_{c_z}$ on $\{x, y, z\}$. This is a further example that cannot be accomodated by Weak WARP (see Cherepanov et al.s (2013)), but it is suitable to our model.

As a last example, it is also important to mention that standard theory can be incorporated into our model. To see this, assume that the concept set is a singleton, i.e. $c^X = \{c\}$. In this case, there is only one order induced by the single concept c, so $>_c = \succ$.

5 Related Literature

There are two streams in economic theory, one is due to Samuelson and based on perfect rationality. The other is due to Herbert Simon and based on bounded rationality. Bounded rationality is a restricted version of perfect rationality, where the agent is constrained by a cognitive limitation. One way to model this is using consideration sets. Some of the well-known papers in "two-stage choice" literature are Categorize Then Choose (CTC) (Manzini and Mariotti (2012)), Rationalization (Cherepanov et al. (2013)), Choice through Attribute Filters (Kimya (2015)), Rational Shortlist (Manzini and Mariotti (2007)), and Choosing by Checklists (Mandler et al. (2012)) are examples of such heuristics. Categories considered in Manzini and Mariotti (2012) are closely related to the conceptual approach in the sense that formation of categories may be accomplished with reference to the concept space. If we determine a line that separates the concept space into two according to the ranking in the hierarchy, then one of these parts shade other as a category. Also, since it is noted that categories change in accordance to the context, defining power of concepts may determine the contextual derivation of categories. Rationalization due to Cherepanov et al.s (2013) provides a model where the decision maker chooses the alternative that is best among the ones that she can rationalize. As you may noted previously, several important examples in this paper are taken from their work.

Deriving its power from choice reversals, Masatlioglu et al. (2012) developed Revealed Attention which investigates the implications of a model based on scarce attention. Differing from the classical bounded rationality literature, it does not depend on a heuristic, but develops an elegant tool for analyzing the choice theoretic implications of a very fundamental cognitive constraint. Even though our model and Masatlioglu et al. (2012)

differ in essence, by the general approach they provide on decision problems, they are similar. Also, our empirical scope is subsumed by theirs, as shown in Section 4.

A widely used characterization of the heuristic based models is Weak WARP. Rationalization (by Cherepanov et al. (2013)), CTC (Manzini and Mariotti (2012)) and Rational Shortlist models (Manzini and Mariotti (2007)) are characterized empirically by Weak WARP, which does not characterize our model as shown in the previous part. A closely connected model which is developed by the same authors (Lleras et al. (2010)) is Choice with Limited Consideration which is characterized by Weak WARP (since Weak WARP is implied by Limited Consideration WARP as noted in Manzini and Mariotti (2012)), so it has the same empirical scope with others such as Rationalization.

Although we did not extend the model to the case where $\triangleright$ is stochastic and a search mechanism is imposed along time, the notion of concepts are very suitable for this approach. Caplin and Dean (2011), Caplin et al. (2011), Masatlioglu and Nakajima (2013) and provides examples of these models where the dimension of time included. An extension where an individual is more likely to realize the elements that are linked to the 'good' concepts is meaningful for analyzing search processes.

There is also a well-established literature that tries to model "reference-dependent" behavior. In these models, there is either an exogenously or endogenously given reference point, and this point affects the choice process and the choice. Some prominent papers trying to model this phenomenon are , Apesteguia and Ballester (2009), Ok et al. (2014), Masatlioglu and Ok (2013) and Dean et. al. (2016). The last two investigates a special case of reference, namely "status quo" bias. In essence, these patterns may originate from the mind of the individual decision maker. In the scope of our approach, the maximum concept in the mind of the decision maker may be a reference point. Given any choice set, such an individual picks the most preferred element with respect to the maximum concept. Another vein of research is modeling the choice from lists, i.e. what would happen if the DM is endowed with a list of alternatives. Rubinstein and Salant (2006) and Yldz (2016) are two papers about choosing from lists.

Salant and Rubinstein (2008) models the choice using frames where a frame conceptually captures any information that affects the choice behavior beyond the available alternatives (our motivation is similar in essence). In our model, concepts are evoked from the information processed by the individual. Furthermore, these concepts themselves affect how to process information along time. Therefore, our model and choice with frames are similar intuitively.

Finally, the construction of concepts is not separately mentioned throughout the paper. One obvious candidate for the formation of a concept set is the set of properties that are associated with the objects in X, but concepts may be more general and abstract than these. There is a connection to Dietrich and List (2010) in the sense that their propositions are derived from concepts, and a relation is imposed upon the set of propositions. In several areas such as computer science, cognitive science, philosophy and psychology the notions of concept and context are used. There is a sub-discipline in mathematics that is about the analysis of concepts and context, called Formal Concept Analysis. Gardenfors (2004) uses conceptual spaces to represent knowledge. Kamenica (2008) derives the informational content included in the markets from the context where his usage of context is similar to us. Tversky and Simonson (1993) develops an approach where they form context-dependent preferences.

6 A Quick Look at Two Extensions

6.1 Memory (inspired from Salant (2003))

Let us assume that an individual has a single memory cell for each object and a concept, so in our case this cell is of the form (x, c_x) where $x \in X$ and $c_x \in C^X$. Define:

$$f : X \times C^X \to X \times C^X$$

where $(x, c_x) \mapsto (y, c_y)$ if:

$$(i) c_y \triangleright c_x$$

and

$$(ii) y >_{c_y} x$$

For such a process to be optimal, it needs to be the case that the individual encounters all concepts sets and at each stage considers all the alternatives. Therefore, in this case we may not have the optimal element chosen according to this procedure. However, note that as the memory grows, the probability that the optimal one chosen increases. In an interpretation where intelligence is associated with memory, an individual who has a wider concept set in her memory has more chance to decide optimally, as expected. On the other hand, an individual that interprets the incoming data from choice objects less efficiently (i.e. a more intelligent person derives more information from the signals spread by the choice set) has less chance to decide optimally.

6.2 Limited Concept Set

Assume that we model memory directly with concept sets. An individual with limited memory will not be able to keep all concepts signaled to her. For example, consider a limited memory that after some point in time cannot preserve any further concepts. We can formulate this by claiming that

$$c^X = C_T$$

where $C_T := \bigcup_{t=1}^{T} c^{X_t}$, so after $t = T$ no further concepts are deposited in the memory of the individual. Of course, the case of an individual with full memory corresponds to the model presented above. Another possible representation is taking a T-memory cell instead of one as we did above. In that case, we approach the problem without considering the time dimension, but still we have the limited memory we want to model. It is obvious to see that an individual with limited memory decides suboptimal given it is possible for another individual to have a wider memory.

7 Conclusion

Our model is based on the intuition that a relational structure exists between the objects of the world. This is similarly true for the choice set itself. We model this through forming concept sets. The things a choice set is related to are themselves important for the choice process, so we defined the "Conceptual Choice" and have shown the relevance of the procedure through several examples. Empirical characterization and relation to the existing literature are investigated. Two extensions are briefly mentioned after we characterized the model with a condition taking its power from choice reversals. The next goal is to further improve the model to the point where we fundamentally describe the major parts of decision behavior.

8 Appendix: Proofs

We will prove Theorem 1 in 3 steps. In the first two steps, the proofs of the acyclicity of $\triangleright$ and $>_c$ are given. By the first step, we are able to extend $\triangleright$ to a linear order. Then we use this fact to prove acylicity of $>_c$ at step 2. In the last step the sufficiency proof is provided. Finally, note that we assume WLOG that each element $x \in X$ is associated with a concept set c_x that has a single element. The construction of the hierarchy of the concept sets are done through these sets.

Definition 6. *$c_z \triangleright c_x$ if there exists a set S such that $c(S) \neq c(S \setminus z)$ where $z \neq c(S)$ and $x \in S$.*

Proof of Theorem 1. Neccessity is already shown. Assume that c satisfies C_1.

Step 1: $\triangleright$ is acyclic

Assume to the contrary that $\triangleright$ is cyclic. Then there exist $x_1, x_2, ..., x_n$ such that $x_i \neq x_j$ for all $i \neq j$, and $c_1, c_2, ..., c_n$ are the respectively associated concepts such that $c_1 \triangleright c_2 \triangleright ... \triangleright c_n \triangleright c_1$. By definition and cyclicity, there exists S_i such that $c(S_i \setminus x_i) \neq c(S_i)$ for all i, where $x_{i+1} \in S_i$ such that $x_i \neq x_{i+1}$ and $c(S_i) \neq x_i$ for all i. Let $S^* = \bigcup_{i=1}^{n} S_i$. Note that by C_1, we can eliminate all $S_j \setminus \{x_j\}$, since $c(S_j) \neq c(S_j \setminus x_j)$ for all j. Assuming $c(S^*) \in S_i$, eliminate all $S_j \setminus \{x_j\}$ such that $j \neq i$ except elements in $S_j \cap S_i$. Also, note that x_j eliminates x_{j+1}, since $c(S_j) \neq c(S_j \setminus x_j)$ for all j and $x_{j+1} \in S_j$ such that $x_j \neq x_{j+1}$. Hence, except S_i, we can eliminate remaining elements and reduce S^* to S_i, since $c(S^*) \in S_i$. Thus, $c(S^*) = c(S_i)$. But $c(S_i) \neq x_i$ for all i by assumption, so $c(S^*) \neq x_i$ for all i. Since $c(S^*) \neq x_i$ and C_1 holds, we can reduce S^* to $S_i \setminus \{x_i\}$[5]. Thus, $c(S_i \setminus x_i) = c(S^*) = c(S_i)$, a contradiction to our assumption. Therefore, $\triangleright$ is acylic.

Remark 2. *Note that since we proved the acyclicity of $\triangleright$, we can extend $\triangleright$ into a linear order and denote it by $\bar{\triangleright}$.*

Definition 7. *$x >_{c_z} y$ if there exists an S such that $c(S) = x$ and $c_z \bar{\triangleright} c_x$ for all $S \setminus z$.*

Step 2: $>_c$ is acylic for all $c \in c^X$.

Assume to the contrary $>_c$ is cylic, i.e. $x_1 >_{c^*} x_2 >_{c^*} ... >_{c^*} x_n >_{c^*} x_1$ for some $c^* \in c^X$ and distinct x_i's from 1 to n in X. Let this c^* be a related concept to $x^* \in X$.

By cyclicity and the definition above, we know that there exist S_i's such that $c(S_i) = x_i$ for all i, where $x_{i+1} \in S_i$ such that $x_i \neq x_{i+1}$. Also from the definition, it must be the case that $c^* \bar{\triangleright} c_x$ for all $x \in S_i \setminus \{x^*\}$. Let $S^* = \bigcup_{i=1}^{n} S_i$. By the cycle above, this must hold for all

[5]In this case, eliminate all $S_j \setminus \{x_j\}$ such that $j \neq i$ except elements in $S_j \cap S_i$ as shown above. In the second part, hold x_{i+1} fixed and eliminate all other x_k's except the ones in $S_i \setminus \{x_i\}$

S_i defined as above. This implies that $c^* \bar{\triangleright} c_x$ for all $x \in S^* \setminus \{x^*\}$. Therefore, we can say that $c(S^*) >_{c^*} y$ for all $y \in S^* \setminus \{c(S^*)\}$. By C_1, S can be reduced to the partition S_i that contains $c(S^*)$. Note that $c(S^*) \in S_i$ for at least one i. Note that $c(S^*) = c(S_i) = x_i$, because by C_1 we can eliminate all $S_j \setminus \{x^*\}$ except $j = i$ and without eliminating anything inside S_i. Similarly, we can reduce S^* to S_{i-1} and thus $x_i = c(S^*) = c(S_{i-1}) = x_{i-1}$, a contradiction to the assumption of distinct x_i's from 1 to n in X. Thus, $>_{c^*}$ is acylic.

Step 3: Sufficiency

Assume C_1 holds. Then, we know that by definition $c_z \bar{\triangleright} c_x$ for all $x \in S \setminus \{z\}$ if $c(S) \neq c(S \setminus z)$ where $z \neq c(S)$, i.e. c_z has the maximum concept over c^S, which means that if we remove y fom any set that includes z, the choice does not change. Also, $c(S) >_{c_z} y$ for all $y \in S \setminus \{c(S)\}$. Thus, $c(S)$ is the maximum element over the order induced by the maximum concept, and by definition it is (CC). If $c(S) = c(S \setminus z)$ for all $z \in S \setminus \{c(S)\}$, then since $\bar{\triangleright}$ is a linear order we know that there is a maximal element, call it c^*. So, $c(S) >_{c^*} y$ for all $y \in S \setminus \{c(S)\}$ and we are done.

Proof of Theorem 2. For the proof of the only if part, assume to the contrary $x\overline{D}y$ does not hold. By the proof of the Theorem 1, there exists a hierarchy $\triangleright$ such that it puts y over x. By definition of, x does not *conceptually dominate* y, a contradiction.

Lemma 1. *C_1 implies that D is acyclic.*

Proof of Lemma 1. C_1 implies that $\triangleright$ is acyclic by the Step 1 of the proof of Theorem 1. Note that by Definition 3 and 6, $\triangleright$ is acyclic if and only if D is acyclic. So, D is acyclic.

Proof of Remark 1. Assume to the contrary $WARP(LA)$ does not hold. By Lemma 1 of Masatlioglu et al. (2012), this implies that P is cyclic. As in the proof of Lemma 1, suppose P has a cycle. Then for all $i = 1, ..., k$, there exists T_i such that $x_i = c(T_i) \neq c(T_i \setminus x_{i+1})$ and $x_k = c(T_k) \neq c(T_k \setminus x_1)$. By Definition 3, this implies that $x_{i+1}Dx_i$ for all i and x_1Dx_k, so we get $x_1Dx_kDx_{k-1}D...Dx_2Dx_1$, a cycle. This is a contradiction to Lemma 1.

References

Apesteguia, J. and M. Ballester. 2012. Choice by sequential procedures. Games and Economic Behavior 77 (1): 90-99

Apesteguia, J. and M. Ballester. 2009. A Theory of Reference-Dependent Behavior. Economic Theory 40 (3): 427-455

Caplin, Andrew, Mark Dean, and Daniel Martin. 2011. Search and Satisficing. American Economic Review 101 (7): 28992922.

Cherepanov, Vadim, Timothy Feddersen, and Alvaro Sandroni. 2010. Rationalization. Theoretical Economics 8 (3): 775-800

Dean, M., Kibris, O., Masatlioglu, Y. 2016. Limited Attention and Status Quo Bias. Journal of Economic Theory, forthcoming

De Clippel, G., Eliaz, K. 2012. Reason Based choice: A Bargaining Rationale for the Attraction and Compromise Effects. Theoretical Economics 7 (1): 125-162.

Dietrich, Franz and Christian List. 2010. A reason-based theory of rational choice. Unpublished paper

Ganter, Bernard and Rudolf Wille. 1999. Formal Concept Analysis: Mathematical Foundations. Springer, 1999.

Gardenfors, Peter. 2004. Conceptual spaces as a framework for knowledge representation. Mind and Matter 2(2): 927

Kalai, Gil, Ariel Rubinstein, and Ran Spiegler. 2002. Rationalizing Choice Functions by Multiple Rationales. Econometrica 70 (6): 2481-2488

Kamenica, E. 2008. Contextual Inference in Markets: On the Informational Content of Product Lines, American Economic Review, 98, 2127-2149

Kimya, Mert. 2015. Choice, Consideration Sets and Attribute Filters unpublished

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

Mandler, M., Manzini, P., Mariotti, M. 2012. A Million Answers to Twenty Questions: Choosing by Checklist. Journal of Economic Theory, 147(1), 71-92.

Manzini, Paola, and Marco Mariotti. 2007. Sequentially Rationalizable Choice. American Economic Review 97 (5): 182439.

Manzini, Paola, and Marco Mariotti. 2012. Categorize Then Choose: Boundedly Rational Choice and Welfare. Journal of the European Economic Association 10: 1141-1165.

Masatlioglu, Y., D. Nakajima, and E. Ozbay. 2011, Revealed attention American Economic Review 102 (5): 2183-2205

Masatlioglu, Y., Nakajima, D. 2013. Choice by Iterative Search. Theoretical Economics 8 (3): 701-728.

Ok, E. A., Ortoleva, P., G. Riella. 2014. Revealed (p) reference theory. The American Economic Review 105(1): 299-321.

Rubinstein, A. and Y. Salant. 2006 A model of choice from lists. Theoretical Economics 22 (1): 317

Rubinstein, A. and Y. Salant. 2008. (A, f): Choice with Frames. Review of Economic Studies 75 (4): 128796

Samuelson, P. A. 1938. A Note on the Pure Theory of Consumers Behaviour. Economica 5 (17): 6171.

Simon, Herbert. 1955. A Behavioral Model of Rational Choice Quarterly Journal of Economics, 69(1): 99-118.

Tversky, Amos, and Itamar Simonson. 1993. Context-Dependent Preferences. Management Science 39 (10): 117989.

Yldz, Kemal 2016 : “List-rationalizable choice.” Theoretical Economics 11 (2): 587-599

Characterization via Hazard Rate

Yusufcan Demirkan
and
Mert Kimya

Koc University, Rumelifeneri Yolu, 34450 Sariyer Istanbul Turkey.

August 14, 2017

Abstract

Following paper suggests a novel way to characterize a certain class of choice procedures which are defined by various intuitive conditions put on the consideration parameter δ. By showing the equality between δ and the well-known hazard rate h, we will show that the machinery provided through the hazard rate h will characterize some of the famous stochastic choice procedures based on δ directly. In addition, a new stochastic choice procedure based on a similarity relation is introduced and characterized.

1 Introduction

The simple observation that decision makers do not commit to a single choice led to the vast literature on stochastic choice. Although the number of papers on the subject is large and diverse in nature, we can concentrate on two currents that are highly affective in the literature: Luce value extensions and consideration set approach. The former one is fairly old and due to Luce(1959), and simply deals with the problem by the logic of comparing relative values of the objects in the choice set. The other one is new and due to Manzini and Mariotti (2014) as a stochastic model of the notion of consideration sets. These two approaches can be seen as rival approaches symbolizing the concepts of "perfect rationality" and "bounded rationality", respectively. Obvious from the abstract, we will mainly deal with the latter approach.

This paper is technical in its essence. Our goal is to introduce a machinery that is sufficient to characterize a certain class of choice procedures only from the observables, i.e. the hazard rate derived from the observed choice probabilities. Even though we remain silent about the relative plausibility of these stochastic choice procedures which we characterize (because of our technical focus), we introduce two new models based on familiar notions, namely "similarity" and "duplicates".

Let us call all choice procedures that are based on the notion of the consideration parameter δ, $\mathbb{C}$. This class is defined by the given pair $(\delta, \succ)$. These defining characteristics cannot be identified observationally, so in this sense a definition does not provide empirical characterization, as expected. On the other hand, a certain subset of $\mathbb{C}$ that satisfies what we call Condition 1 can reveal preferences from the data and thus the second component of the defining characteristics can be identified. This ensures that the hazard rate h that depends on preferences is well-defined. After showing the equality between δ and h, the unobservable defining characteristic of $\mathbb{C}$ turns into an observable characteristic, so the definition turns into a characterization that can be obtained empirically. Also note that all models that are not in $\mathbb{C}$ but can be translated into such models under certain conditions can be characterized through the machinery provided. This translation will be used as we characterize a simple Luce model.

In summary, the machinery is as follows:

1) Translate the choice procedure in such a way that is in $\mathbb{C}$ defined by $(\delta, \succ)$.

2) Identify preferences and define the hazard rate h.

3) $\mathbb{C}$ is now characterized by h.

2 Model

The machinery we will provide works for a certain class of choice procedures. These procedures can be distinguished by their dependence on the consideration parameter, δ. Let us first recall the basic definitions of a stochastic choice rule and a certain subset of it, namely menu dependent stochastic consideration set rule introduced by Manzini and Mariotti (2014) (From this point on we will use the abbreviation MM).

Definition 1. *A stochastic choice rule is a map* $p : \overline{X} \times M \to [0,1]$ *such that* $\sum_{x \in \overline{A}} p(x, A) = 1$*,* $p(a, A) = 0$ *for all* $a \notin \overline{A}$*,* $p(a, A) \in (0,1)$ *for all* $a \in \overline{A}$ *and* $A \in M \setminus \emptyset$ *where* M *is the set of all menus in* X *and* $\overline{A} = A \cup \{x^*\}$ *where* x^* *denotes the outside option.*

Note that this is the standard definition of a stochastic choice rule. One important thing to recall the existence of an outside option for the sake of the demonstration. Also note that $p(x^*, \emptyset) = 1$ trivially.

The following definition is the same definition as in MM.

Definition 2. *A menu-dependent stochastic consideration rule is a stochastic choice rule* $p_{\succ,\delta}$ *for which there exists a pair* $(\succ, \delta)$ *where* $\succ$ *is a strict total order and* $\delta : X \times M \setminus \emptyset \to (0,1)$ *such that*

$$p_{\succ,\delta}(x, A) = \delta(x, A) \prod_{y \in A : y \succ x} (1 - \delta(y, A))$$

for all $A \in M$ *and for all* $x \in A$.

It is clear from the definition that this specific stochastic choice rule is dependent upon two primitives, $(\succ, \delta)$. In general, one can put any set of conditions on the consideration parameter, δ. Thus, we can define any specific menu-dependent stochastic consideration rule first by identifying the preference relation and then putting certain intuitive conditions on δ. Before giving the full definition of a restricted stochastic consideration set rule, we need to define the following.

Definition 3. *A restriction function* f *is a truth function* $f : Prop(\delta) \to \{0,1\}$ *where* $Prop(\delta)$ *is the set of propositions defined over the consideration parameter* δ*.*

Definition 4. *A restricted stochastic consideration rule is a menu-dependent stochastic consideration rule* $p_{\succ,w}$ *that is defined by the set of conditions on* δ *for which* $f(w) = 1$ *where* $w \in Prop(\delta)$*.*

The method is as follows: First, with a genuine condition on δ, we will identify preferences from the observed choices. This condition, which we will call Condition 1 is a weak condition that is met by the examples we will characterize. After showing that δ and hazard rate are actually equal to each other, we will replace the set of propositions on δ with the set of propositions on the hazard rate.

We will also characterize some other Luce type models after a translation of their models to the language of our models, namely to the language of δ.

Without going further about the method, we need to define what a hazard rate is.

Definition 5. *(Hazard Rate)*

$$h(x, A) = \frac{p(x, A)}{1 - \sum_{y \in A : y \succ x} p(x, A)}$$

for all $A \in M$ and $x \in A$.

Remark 1. $h(x^*, \emptyset) = 1$

To see this, note that there is no element in empty set so that there is no strictly preferred element to x^*. This implies that the denominator is equal to 1 and therefore $h(x^*, \emptyset) = p(x^*, \emptyset)$. By assumption, $p(x^*, \emptyset) = 1$.

By the definition of a restricted menu dependent stochastic consideration rule, we know that the defining primitives of the rule are $(\delta, \succ)$. To characterize such a rule, we need to identify both parameters from the observed choice probabilities. The strategy is demonstrated above: First we will reveal preferences from a condition put on δ. This will help us to define hazard rate h as it is dependent upon the preference relation. After defining h, we will show the equality of h and δ, so the other defining parameter of the model is now equal to a parameter which is observed from the data. Thus, one can directly convert the conditions imposed upon δ to its equivalent condition imposed upon h after identifying preferences, and the model is immediately characterized.

The condition with which we will identify preferences is the following:

Condition 1.

$$\frac{\delta(x, A)}{\delta(x, A \setminus y)} \geq \frac{\delta(y, A)}{\delta(y, A \setminus x)} \times (1 - \delta(x, A))$$

for all $x, y \in X$ and $A \in M \setminus \emptyset$.

Note that this is a condition on δ and obviously an element of $Prop(\delta)$. This condition will help us to identify preferences from the observed choice probabilities. The first lemma will provide us identification of the preferences.

After showing the equality between δ and h, we will be able to replace Condition 1 with the equivalent stated in terms of h. Again, we will be able to identify preferences through Lemma 1 using this equivalent condition.

Theorem 1. *A restricted menu dependent stochastic consideration rule can be characterized by w^*:*

$$w^* := \{w \in Prop(h) : g(w) = 1\}$$

and if preferences cannot be identified from the model, then $\exists w_\succ \in Prop(h)$ s.t. $w_\succ$ implies either $x \succ y$ or $y \succ x$ (but not both) where $\succ$ is assumed to be acyclic and $w_\succ \in w^*$ for $g : Prop(h) \to \{0, 1\}$.

Proof of Theorem 1. *In the proof, we will first identify preferences from observed choice probabilities. This will be achieved through Condition 1. Note that this does not say we can identify preferences from h. To do this, we will show in Lemma 2 the equality between δ and h, and then it is obvious to see that preference identification can be done through Condition 1 stated in terms of h. Lastly, note that this converted condition is our needed $w_{\succ}$ stated in Theorem 1.*

Lemma 1.

$$\frac{p(x,xy)}{p(x,x)} \geq \frac{p(y,xy)}{p(y,y)} \rightarrow x \succ y$$

Proof of Lemma 1. *Assume to the contrary $y \succ x$. In this case, note that $p(x,xy) = \delta(x,xy)(1-\delta(y,xy))$, $p(x,x) = \delta(x,x)$, $p(y,y) = \delta(y,y)$ and $p(y,xy) = \delta(y,xy)$. So,*

$$\frac{p(x,xy)}{p(x,x)} \geq \frac{p(y,xy)}{p(y,y)}$$

if and only if

$$\frac{\delta(x,xy)}{\delta(x,x)} \times (1-\delta(y,xy)) \geq \frac{\delta(y,xy)}{\delta(y,y)}$$

.
However, this contradicts Condition 1, so $x \succ y$.

Lemma 2.

$$\delta(x,A) = h(x,A)$$

for all $x \in A$ and $A \in M \setminus \emptyset$.

Proof of Lemma 2.
The proof of the lemma is in two steps.

Step 1:
By the definition of a random choice rule, we know that $\sum_{x\in\overline{A}} p(x,A) = 1$. Using the definition of menu- dependent stochastic choice rule, we can write this in terms of δ. Let us assume that we enumerated the elements in such a way that $X = \{x_1, x_2, ..., x_n\}$ where $x_i \succ x_{i+1}$ for all $i \in \{1,2,...,n\}$.

$$1 = \sum_{x\in\overline{A}} \delta(x,A) \prod_{y\in A:y\succ x} (1-\delta(y,A))$$

$$= \delta(x_1,A) + \delta(x_2,A)(1-\delta(x_1,A)) + ... + \prod_{x_i\in A} (1-\delta(x_i,A))$$

where the last term is equal to $p(x^*,A)$. Note that $(1-\delta(x_1,A))$ is common to all terms except the first one. So:

$$= \delta(x_1,A) + (1-\delta(x_1,A))(\delta(x_2,A) + ... + \prod_{x_i\in A, i\neq 1} (1-\delta(x_i,A))$$

. *We can take $\delta(x_1,A)$ to the left and then we can cancel $(1-\delta(x_1,A))$ from both sides,* *since $\delta(x,A) \neq 1$ for all $x \in A$ where $A \in M \setminus \emptyset$ by definition. Thus:*

$$1 = \delta(x_2, A) + \delta(x_3, A)(1 - \delta(x_2, A)) + ... + \prod_{x_i \in A, i \neq 1} (1 - \delta(x_i, A))$$

. Note that a similar procedure can be applied also for x_2. Repeating this process for each $k \in \{1, ..., n\}$:

$$1 = \delta(x_k, A) + (1 - \delta(x_k, A))(\delta(x_{k+1}, A) + ... + \prod_{x_i \in \{x_{k+1}, ..., x_n\}} (1 - \delta(x_i, A))$$

for all $k \in \{1, ..., n\}$.

Step 2:

$$1 - \sum_{y \in A: y \succ x} p(x, A) = \sum_{z \in A: x \succeq z} p(x, A)$$

$$= \sum_{z \in A: x \succeq z} \delta(z, A) \prod_{y \in A: y \succ z} (1 - \delta(y, A))$$

Similar to the above notation, let us assume that $x = x_k$ for some $k \in \{1, ..., n\}$.Note that the most preferred element is x_k, so all elements that are strictly preferred to x_k are also strictly preferred to $\{x_{k+1}, ..., x_n\}$. This implies that we can write the above equation in the following way:

$$= \prod_{x_i \in A: x_i \succ x_k} (1 - \delta(x_i, A))[\delta(x_k, A) + \delta(x_{k+1}, A)(1 - \delta(x_k, A)) + ... + \prod_{x_j \in A: x_k \succeq x_j} (1 - \delta(x_j, A))]$$

$$= \prod_{x_i \in A: x_i \succ x_k} (1 - \delta(x_i, A))[\delta(x_k, A) + (1 - \delta(x_k, A))[\delta(x_{k+1}, A) + ... + \prod_{x_j \in A: x_k \succ x_j} (1 - \delta(x_j, A))]$$

Note that by Step 1, this last expression reduces to:

$$= \prod_{x_i \in A: x_i \succ x_k} (1 - \delta(x_i, A))$$

.

Thus, we proved that:

$$1 - \sum_{y \in A: y \succ x} p(x, A) = \prod_{x_i \in A: x_i \succ x_k} (1 - \delta(x_i, A))$$

.

It follows from the definition of menu-dependent stochastic choice rule that:

$$\delta(x, A) = \frac{p(x, A)}{\prod_{y \in A: y \succ x}(1 - \delta(y, A))} = \frac{p(x, A)}{1 - \sum_{y \in A: y \succ x} p(x, A)} = h(x, A)$$

.

By Lemma 2, we can prove Lemma 1 using the equivalent of the Condition 1 stated in terms of the hazard rate h.

Lastly; note that the equality between δ and h implies that $Prop(\delta) = Prop(h)$. Although δ is not observed from the data, we can observe h. The model is defined by conditions on δ, i.e. by w for which $f(w) = 1$ where $w \in Prop(\delta)$. Note that $f(w) = 1$ for some $w(\delta)$ if and only if $g(w^*) = 1$ for $w^* = w$, since $\delta = h$. Therefore, we can change conditions on δ with the same conditions on h. By the fact that h is observable, the model is directly characterized by h. Thus, the proof is complete. □

3 Similarity-Based Stochastic Choice

The first model we will characterize via hazard rate is a new model we developed. In the abstract we noted that the machinery provided will work for all models that depend on the consideration parameter δ. Models may differ from each other by the conditions they impose on this parameter, and in general these conditions rely on either intuition or experimental evidence.

The notion of similarity is not new to economics, although the models that are based upon it are small in number. The foremost work is from the psychology literature, due to Tversky. Tversky (1977) tries to develop a formal model where the similarity between objects is measured. In economics, Rubinstein (1988) develops a model based on two primitives, which is the pair $(\succ, \sim)$ where the latter one denotes the similarity relation. Actually, the two primitives of our model are the same with Rubinstein, and in this sense our intuition is closest to his. Also, a recent work by Payro and Ulku (2014) builds a formal model of mistakes in deterministic choice that are attributable to the similarity between the most preferred element and the others. Without going into the details, first we will define what the model is and then we will see the characterization of it through the machinery we provided.

Definition 6. *A similarity relation $\sim$ is a subset of $X \times X$ that satisfies reflexivity and symmetry.*

As Tversky (1977) notes, a similarity relation need not to be transitive in general. The following stochastic choice procedure we will define is intended to be a very general one in the sense that it both captures the well-known violations of rationality and reflects the intuition behind the systematic of our choice process. The only restriction is on the consideration parameter δ and it is only a mild one. Let us first define the procedure, and then argue what it captures.

Definition 7. *A similarity-based stochastic consideration rule is a menu-dependent stochastic consideration with additionaly given $\sim$ such that it satisfies the following:*
(i)

$$\delta(x, A) \neq \delta(x, A \cup y) \rightarrow [x \sim y]$$

given $x \neq y$ and
(ii)

$$\frac{\delta(x, A)}{\delta(x, A \setminus y)} = \frac{\delta(y, A)}{\delta(y, A \setminus x)}$$

for all $x, y \in A$ and $A \in M \setminus \emptyset$ such that $x \sim y$.

The first condition is an intuitive one, it says that if the consideration of an element x changed with the addition of a new element y, then y must be related to x somehow. Thus, they are similar to each other on some basis that is unknown to us.

The second condition is about the relative effect of two similar elements x and y to each other. It says that two similar elements affect each other relatively in the same proportion given the same set, so this condition does not say anything about beyond "context" effects of these.

As noted in MM, characterizing menu-dependent models is very hard and we need to put extra strong assumptions upon δ. Note that we put very simple and intuitive restrictions and it is still hard to get a simple characterization, since there are two sides of the change: one from the own consideration parameter of the object of choice (let us call it x), and the other from the uncontrolled changes coming from the strictly preferred elements to x. We characterized this model, however we do not propose here our characterization. This is because the machinery we provide propose a very simple characterizing without extracting any additional effort. The only condition we have to check whether *Condition*1 that gives the identification of preferences is satisfied or not.

Lemma 3. *Similarity-based stochastic consideration rule satisfies Condition 1.*

Proof of Lemma 3. *To see this, note that by definition $(1 - \delta(x, A))$ is always smaller than 1. So, multiplying the right hand side of Condition (ii) in Definition 7 can only make the right hand side smaller, and thus Condition 1 is satisfied. The similar logic applies to the other case*

Since we observed that Condition 1 is satisfied, we can identify preferences through Lemma 1 in the proof of the Theorem 1. After the preferences are identified, we can define the hazard rate h. Then by Theorem 1, the equivalent of Condition (ii) in terms of h is sufficient to characterize the model (We do not need the first one since it only reveals the similarity relation). Thus, the characterization is as follows:

Theorem 2. *A stochastic choice rule is a similarity-based stochastic consideration rule if and only if $\succ$ is acylic and it satisfies the following:*

$$\frac{h(x, A)}{h(x, A \setminus y)} = \frac{h(y, A)}{h(y, A \setminus x)}$$

for all $x, y \in A$ and $A \in M \setminus \emptyset$.

Proof of Theorem 2. *The characterization immediately follows from Theorem 1 by replacing Condition (ii) of Definition 7 with the one that stated in terms of h. We did not state the characterizing condition for all $x, y \in A$ and $A \in M \setminus \emptyset$ such that $x \sim y$, since the condition is also satisfied by x's that are not similar to y trivially by the contrapositive of the Condition (i) of the Definition 7 (Note that $\delta(x, A) = \delta(x, A \cup y)$ for all such x and y).*

4 Stochastic Consideration Set-MM

MM is a formal model that is founded upon the notion of "consideration sets". The most important contribution of the paper is to open a new direction in the research of stochastic choice rules that depends on the consideration of each alternative in the choice set. The model MM develops is the most basic version where the consideration of each alternative is constant through all menus, i.e. the consideration parameter is menu independent. The choice procedure is already defined in the Model section, but it was the menu-dependent version of their actual model. Let us define the stochastic consideration set rule:

Definition 8. *A stochastic consideration set rule is a stochastic choice rule $p_{\succ,\delta}$ such that given a pair $(\succ, \delta)$ where $\succ$ is a strict total order and $\delta : X \times M \setminus \emptyset \to (0,1)$ such that*

$$p_{\succ,\delta}(x, A) = \delta(x) \prod_{y \in A: y \succ x} (1 - \delta(y))$$

for all $A \in M$ and for all $x \in A$

As it is clearly seen from the Definition 8, the only difference from the menu-dependent stochastic consideration rule is the fact that the consideration of each alternative x is menu-independent, i.e. $\delta(x) = \delta(x, A)$ for all $x \in A$ and $A \in M \setminus \emptyset$.

To characterize MM, we need to first check that it satisfies Condition 1 with which we identify preferences.

Lemma 4. *Stochastic consideration set rule satisfies Condition 1.*

Proof of Lemma 4. *We know that the consideration parameter δ is menu-independent, so Condition 1 reduces to the following:*

$$1 \geq (1 - \delta(x))$$

which is trivially satisfied by the definition of δ. Hence, we can define the hazard rate h.

Theorem 3. *A stochastic choice rule is a stochastic consideration set rule if and only if $\succ$ is acyclic and $h(x, A) = h(x, B)$ for all $x \in A \cap B$ and $A, B \in M \setminus \emptyset$.*

Proof of Theorem 3. *Follows directly from the machinery.*

5 Monotonicity

Assuming monotonicity of the consideration parameter as it is defined in MM is a natural extension of the model. Note that such an extension corresponds to random version of what Lleras et al.(2017) calls a "Competition Filter". Recall that such a filter obeys the following rule: If an element x is considered in a superset of S, then it must be considered in S. In its random counterpart, this dictates that $\delta(x, S)$ is decreasing as we move to a superset.

Definition 9. *A monotonic stochastic consideration rule is a menu-dependent stochastic consideration with the following restriction: $f(w) = 1$ if and only if $\delta(x, S) \geq \delta(x, T)$ $\forall S \subseteq T$.*

For characterizing monotonic stochastic consideration rule, we need the following translation of Theorem 1. Recall that via Theorem 1 we reduced the definition of a menu dependent stochastic consideration set rule which consists of two unobserved parameters into a model with only one unobserved parameter. Note that the following theorem is only in terms of observables, so the model can be turned into a model only in terms of observables.

Theorem 4. *A stochastic choice data is consistent with menu-dependent consideration rule under restriction f if and only if for each $x \in X$ there exists a set $Y_x \subseteq X$ that does not include x and that satisfies:*

- $f(\frac{p(x,S)}{1-p(Y_x \cap S,S)}) = 1$
- *For any $x, y \in X$ with $x \neq y$, either $Y_x \subset Y_y$ or $Y_y \subset Y_x$.*

Proof of Theorem 4. *Necessity is obvious. For sufficiency, let xPy if $Y_x \subseteq Y_y$. Note that P is acyclic by the second condition. Extending P to a strict linear order we can find a preference relation $\succ$. This allows us to define the hazard rate h and by the equality of h and δ the proof is complete.*

As a corollary, we can see the characterization of monotonic menu dependent consideration set rule.

Corollary 1. *A stochastic choice data is consistent with monotonic menu-dependent consideration set rule if and only if for each $x \in X$ there exists a set $Y_x \subseteq X$ that does not include x and that satisfies:*

- $\frac{p(x,S)}{1-p(Y_x \cap S,S)}$ *is decreasing for each $x \in X$ under set inclusion.*
- *For any $x, y \in X$ with $x \neq y$, either $Y_x \subset Y_y$ or $Y_y \subset Y_x$.*

The characterization implies that there is at least one alternative whose probability always falls (weakly) as the set gets bigger, more generally, there is a set of alternatives Y_x for each alternative x (related by set inclusion) such that the hazard rate falls if Y_x is considered to be the upper contour set of x.

6 A Luce Model

Although models founded upon the Luce value type formulas are different from the ones demonstrated previously, we can turn them into each other by defining δ properly. Beyond its demonstrative purposes, it comes out that a consideration parameter that is defined in terms of a Luce value is also meaningful. To see this, note that in the way we define the consideration parameter in terms of the Luce value v, an element is likely to be considered more if it has a higher Luce value. On the other hand, as the number of inferior options rises, the sum of Luce values in the denominator rises and this effect will balance the higher consideration effect arising from the higher Luce value of the strictly preferred element.

We will show that it is possible to define δ in terms of Luce value v such that a Luce-type choice procedure turns into a menu-dependent stochastic consideration rule. First, we will define how such a choice procedure looks like. The below definition also incorporates the existence of an outside option, since the machinery we provide depends upon the existence of it. The definition is taken from Echenique, Saito, and Tserenjigmid (2013)(only some small changes are made).

Definition 10. *A stochastic choice function p satisfies Luce model if and only if there exists a utility function $v : X \cup M \to \mathbb{R}_{\geq 0}$ such that it is of the following form:*

$$p(x, A) = \frac{v(x)}{\sum_{x \in A} v(x) + v(A)}$$

Note that the term $v(A)$ enters with the existence of an outside option and it intuitively captures the opportunity cost of not choosing anything from the menu.

To apply our machinery, we need to convert the Luce model into a model that depends on δ. This can be done with the following theorem.

Theorem 5. *A menu-dependent stochastic consideration rule is a Luce model if and only if it satisfies the following:*

$$\delta(x, A) = \frac{v(x)}{\sum_{y \in A : x \succeq y} v(y) + v(A)}$$

for all $x \in A$ and $A \in M \setminus \emptyset$.

Proof of Theorem 5.

$$p_{\succ,\delta}(x, A) = \delta(x, A) \prod_{y \in A : y \succ x} (1 - \delta(y, A))$$

$$= \frac{v(x)}{\sum_{y \in A : x \succeq y} v(y) + v(A)} \prod_{y \in A : y \succ x} (1 - \frac{v(y)}{\sum_{z \in A : y \succeq z} v(z) + v(A)})$$

$$= \frac{v(x)}{\sum_{y \in A : x \succeq y} v(y) + v(A)} \prod_{y \in A : y \succ x} (\frac{\sum_{z \in A : y \succ z} v(z) + v(A)}{\sum_{z \in A : y \succeq z} v(z) + v(A)})$$

Let us enumerate the elements that are strictly preferred to x such that:

$$\{y \in A : y \succ x\} = \{x_1, x_2, ..., x_n\}$$

where $x_i \succ x_{i+1}$ for all i.

$$= \frac{v(x)}{\sum_{y \in A : x \succeq y} v(y) + v(A)} \times \frac{\sum_{z \in A \setminus \{x_1\}} v(z) + v(A)}{\sum_{z \in A} v(z) + v(A)} \times ... \times \frac{\sum_{z \in A \setminus \{x_1, ..., x_n\}} v(z) + v(A)}{\sum_{z \in A \setminus \{x_1, ..., x_{n-1}\}} v(z) + v(A)}$$

Note that above telescoping product the above multiplication reduces to the following:

$$= \frac{v(x)}{\sum_{y\in A:x\succeq y} v(y)+v(A)} \times \frac{\sum_{z\in A\setminus\{x_1,\ldots,x_n\}} v(z)+v(A)}{\sum_{z\in A} v(z)+v(A)}$$

But now note that:

$$\sum_{z\in A\setminus\{x_1,\ldots,x_n\}} v(z)+v(A) = \sum_{y\in A:x\succeq y} v(y)+v(A)$$

since $A \setminus \{x_1, ..., x_n\} =: \{y \in A : x \succeq y\}$ *because of the fact that* x_n *is the next best alternative that is strictly preferred to* x*, i.e.* $x_n \succ x$ *and there exists no* $t \in A$ *such that* $x_n \succ t \succ x$*.*

Thus:

$$= \frac{v(x)}{\sum_{y\in A} v(y)+v(A)}$$

So, finally we have:

$$p_{\succ,\delta}(x, A) = \delta(x, A) \prod_{y\in A:y\succ x} (1 - \delta(y, A)) = \frac{v(x)}{\sum_{y\in A} v(y)+v(A)}$$

.

Hence, $p_{\succ,\delta}$ *satisfies Luce model by Definition 9.* □

Why would one want to translate a model into a model of stochastic consideration sets? Besides the interesting relationships uncovered through such an exercise, we think that such an exercise is also useful as it allows for an immediate comparison of models that are not written as consideration models with consideration models.

For instance, the theorem above immediately implies the following corollary.

Corollary 2.

- *Luce model is incompatible with Example* 1 *(baseline MM model),* 2 *and* 4*.*
- *Luce model is a special case of Example* 3 *if and only if it can be rationalized by* v *that satisfies* $v(T) \geq v(S)$ *whenever* $S \subseteq T$*, i.e. the value of the outside option increases as the set gets bigger.*

7 Duplicates

Duplicates are element that are similar in attributes. One may argue that two elements are duplicate of each other if there is no distinguishing feature between them from the point of view of the decision maker. This can be further advanced to the level that given the conceptual representation system of two decision makers, their conceptual representations are isomorphic to each other.

The importance of duplicates in choice problems originates back to famous example of Debreu(1960). A very well-known work that is developed to solve this example within a Luce-like framework is Gul et al.(2014). We did not investigate yet the translatability of their framework into a consideration set framework, we only develop a simple model of duplicates where the condition put on the δ is related to it. First we will define what a duplicate is and then consideration parameter that is adjusted for the effect of duplicates:

Definition 11. *A duplicate relation d is a binary relation that is a subset of $X \times X$ and satisfies reflexivity and symmetry.*

Note that there is nothing different in the definitions of being duplicate and similar. In its meaning, being a duplicate is stronger than being similar. This effects the restriction put on the consideration parameter, in similarity-based stochastic choice the only effect of similarity relation was the relative constant effect of two similar elements given a context. Here, as you will see in the below definition, a duplicate necessarily affects the consideration of an element in a set. Also, note that a menu independent distribution of the consideration parameter over the set of elements is given; however, menu-dependence enter through the effect of the duplicates.

Definition 12. *A consideration parameter δ is a duplicate adjusted consideration parameter δ_d if and only if:*

$$\delta_d(x, A) = \frac{\delta(x)}{\eta(x, A)}$$

where $\eta(x, A) = \sum \mathbb{1}_{\{y \in A : ydx\}}$ for all $A \in M \setminus \emptyset$ and for all $x \in A$ and in the case $\eta(x, A) = 0$,

$$\delta_d(x, S) = \delta(x)$$

for all $S \subseteq A$.

The definition of the stochastic choice procedure is totally the same with the menu-dependent stochastic consideration rule except the consideration parameter is adjusted for the effect of the duplicates, i.e. the menu-dependent consideration parameter has a particular form dependent upon the number of duplicates given a menu..

Definition 13. *A duplicate adjusted consideration rule is a stochastic choice rule $p_{\succ,\delta_d}$ such that given a pair $(\succ, \delta_d)$ where $\succ$ is a strict total order and δ_d is duplicate adjusted consideration parameter such that:*

$$p_{\succ,\delta_d}(x, A) = \delta_d(x, A) \prod_{y \in A : y \succ x} (1 - \delta_d(y, A))$$

for all $A \in M$ and for all $x \in A$.

Lemma 5.

$$\frac{\delta_d(x, xy)}{\delta_d(x, x)} = 1$$

for all $x, y \in X$.

Proof of Lemma 5. *Note that $\delta_d(x, x) = \delta(x)$. By the definition of δ_d, if y is not a duplicate of x, it is equal to 1. If ydx, then $\eta(x, xy) = 1$ and again the fraction is equal to 1 and we have the lemma.*□

Since we have Lemma 4, the two element set version of the Condition 1 that we use for proving Lemma 1 reduces to the following condition:

$$1 \geq (1 - \delta_d(x, xy))$$

which is trivially satisfied by the definition of a consideration parameter. Therefore, by Lemma 1 we have the preference identification we need for defining the hazard rate. Since hazard rate is well-defined in this case, the machinery is directly applicable and thus we have the following:

Theorem 6. *A stochastic choice rule p is a duplicate adjusted consideration rule if and only if $\succ$ is acyclic and the following holds:*

$$h(x, A) = \frac{h(x)}{\eta(x, A)}$$

for all $A \in M$ and for all $x \in A$.

Note that if $\eta(x, X) = 0$, then by definition $\delta_d(x, S) = \delta(x)$ for all $S \subseteq X$, hence we have MM. Thus, according to the definition of duplicate adjusted consideration parameter, if there are no duplicates to a particular element $x \in X$, then x is chosen according to stochastic consideration set rule of MM.

8 References

References

Debreu, G. 1960. "Review of Individual choice behaviour by R. D. Luce." American Economics Review 50: 186-188.

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

Lleras, Juan S., Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay. 2017. "When More is Less: Choice by Limited Consideration." Journal of Economic Theory 170, 70-85.

Luce, R.D. 1959. "Individual choice behavior: a theoretical analysis." , Wiley, New York

Manzini, P. M. Mariotti 2014. "Stochastic choice and consideration sets." Econometrica, 82(3), 11531176.

Masatlioglu, Y., D. Nakajima, and E. Ozbay. 2011, "Revealed Attention" American Economic Review 102 (5): 2183-2205

Payr, F. L. lk 2015. "Similarity-Based Mistakes in Choice" Journal of Mathematical Economics, 61: 152-156.

Rubinstein, A. 1988. "Similarity and decision-making under risk (is there a utility theory resolution to the Allais paradox?)." Journal of Economic Theory, 46(1): 145-153.

Tversky, Amos (1977). "Features of similarity." Psychological Review, 84(659): 327352.

Random Filtering

Yusufcan Demirkan
and
Mert Kimya

Koc University, Rumelifeneri Yolu, 34450 Sariyer Istanbul Turkey.

August 14, 2017

Abstract

The randomness of the choices people make and their lack of considering the full set of available alternatives are both well-known and empirically supported facts. We present a "theory of selves" approach to build a firm connection between these two observations by assuming that each decision maker (DM) has multiple deterministic consideration filters where she maximizes her well-defined preferences over each resulting consideration set. Each consideration filter corresponds to a different self that arises with some probability and thus the choice becomes random. We characterize this "*Random Filtering*" procedure using some of the influential consideration filter forms from the literature. Several well-known choice anomalies and context effects can be captured by the model.

1 Introduction

Think of a decision maker (DM) who does not consider all available alternatives, either rationally or due to a cognitive lack. Imagine her in the marketplace. She may have different consideration sets applied to the whole market; in a rainy day she may be inclined to use her melancholic consideration set and ignore most of the alternatives due to her unwillingness to spend more time. In a sunny day she may be full of life and full of motivation, her consideration set might be greater than her melancholic one. Although unobserved for an outsider, these different selves may be reflected in her consideration sets, thus the choice of the decision maker becomes random.

The randomness of choices and lack of full consideration are both well-known and empirically supported facts.[1]Our approach brings these two insights together by modeling the DM as composed of multiple selves. Each self has a probability to arise and a corresponding deterministic consideration filter. The agent has well-defined stable preferences that she maximizes over each of her resulting consideration sets. Recall the example of the DM in marketplace, she has two selves: 'melancholic' and 'happy'. These two selves arise with some probability, so their corresponding consideration filters are used with the same probability.

[1]See Wright and Barbour (1977), Hauser and Wernerfelt (1990), Iyengar and Lepper (2000) Reutskaja et al. (2011) etc.

The model directly connects the influential literature on deterministic consideration sets[2] with random choice by presenting a "theory of selves" approach. The random choice is due to each self's probability of occurrence, i.e. probability of using a different consideration filter. Hence, the model allows us to characterize the random counterparts of influential deterministic models of consideration formation.

Conceptually our model is in the same spirit with the famous Random Utility Models (RUM) [3]. The underlying motive of these models is the variation in the preferences of the agents, and this is reflected as a probability distribution over the set of utility functions. We view preferences as a more fundamental and stable part of human decision-making, and assume that the variation in choice is due to the different consideration filters an agent uses depending on her psychological state. That is, to the contrary of "variation in preferences" approach adopted by RUM, our approach takes the view of "stable preferences, variable deterministic consideration". An advantage of assuming variation in consideration filters is that the model is able to accommodate certain well-known context effects unlike RUM which assumes full consideration. We show that the random counterpart of a deterministic consideration set model will be able to accommodate violations of regularity, as long as the corresponding deterministic model is able to accommodate violations of WARP. Hence, the empirical scopes of deterministic models of consideration are preserved under the assumption of multiple consideration filters.

This brings up the question of to what extent our approach preserves the empirical features of these deterministic models. We will show that the connection between the deterministic model and its random counterpart is rather strong: We show that certain features of the underlying model are carried over to the random model. For instance, if the underlying model has no empirical content then the corresponding deterministic model would also be free of empirical content (see Theorem 1). Similarly, if the underlying model is restricted so that what is considered can be directly identified, then the corresponding random model we construct is empirically equivalent to the underlying model (see Theorem 2).

We will completely characterize the random counterparts of the following well-known consideration sets used as the underlying deterministic models:

1)Perfect rationality: A decision maker who is perfectly rational has full consideration, i.e. she considers all of the available alternatives. This is the standard approach in choice theory.

2)Attention Filter: A form of consideration set is an "Attention Filter" due to Masatlioglu et al. (2012). Consider an agent who does not pay attention to an alternatives in a set S (think of the little store in your neighborhood). If I remove that alternative from S (or from that store), the set of alternative she pays attention to in this subset of S must be the same with the set of alternatives that takes her attention in S. So, nothing changes in what the agent considers if we move an alternative she does not consider previously.

3)Competition Filter: Another well-known consideration set is the so-called "Competition Filter" of Lleras et al.(2017). A competition filter reflects the intuition in its name, if an alternative x attracts attention in a bigger set, it must attract attention in a smaller set given it is available, since in a bigger set there is much more competition

[2]Masatlioglu et al. (2012), Lleras et al. (2017), Cherepanov et al. (2013), and Manzini and Mariotti (2012).

[3]see Luce (1959), Block and Marschak (1960), McFadden (1978), Falmagne (1978), and Gul, Natenzon and Pesendorfer (2010)

between alternatives to attract attention.

4)Rationalization: In this model, the DM needs a "Rationale" to consider an alternative, thus the set of such alternatives become "psychologically permissible". Hence, a consideration set is formed by underlying unobserved rationales of the decision maker.

5)Categorization: In this paper due to Manzini and Mariotti (2012), an alternative is considered if it belongs to an undominated category where each category is formed through an unobserved process.

A perfectly rational agent considers all the available alternatives. As we mentioned above, in this case multiple filters are equivalent to one single filter where the agent considers the whole given set. We will show that preference maximization over rational filters is equivalent to deterministic preference maximization and thus it must satisfy WARP.

The characterization of attention filters uses the 'revealed preference' implied by the model. Given this model, we show that if the probability of choosing alternative x decreases when we remove some alternative z, then it must be the case that x is preferred to z. In other words, the revealed preference relation is directly obtained from the violations of regularity. We show that this model is completely characterized by the acyclicity of this revealed preference. Here, we see that the identification procedure comes as a natural counterpart of its deterministic version, i.e. the revealed preference analysis is carried to the random case.

The characterization of an agent using competition filters shows that this is not the case in general, the probabilistic nature of the process brings further complications to account for. Correspondingly, the identification of preferences in this part is more complicated than the one in the case of attention filters. The algorithm of identification is in steps; in the first step we collect alternatives that never decrease in probability, and we call this set X^0. Then, we collect all elements that are decreasing in probability in a certain subset while elements in a subset of X^0 increase in probability in such a way to compensate this decrease. Such elements are collected in X^1. The algorithm continues in this way until all the alternatives are exhausted. The intuition is as follows: If the probability of an element x chosen in S decreases in $S \setminus z$, then there must be at least one better element than x the DM did not consider previously in S but started to consider in $S \setminus z$ for some of her competition filters. Thus, the choice probability of these better elements must increase. Furthermore, this increase in probability should compensate the decrease in choice probability of worse alternatives. Hence, random counterpart of the deterministic model imposes a new 'cardinal' condition on the choice data we observe.

The models of Categorization and Rationalization can be expressed in terms of competition filters and vice versa[4]. This allows us to characterize these models in the same way we characterize competition filters (see Theorem 7).

The organization of the paper is as follows: In Section 2 we will give a general definition of "*Random Filtering*" and build a connection between the probabilistic choice data and the model we use. In addition, a DM that uses perfectly rational filters is characterized in this part. Then, in Section 3 we will characterize a decision maker who uses multiple attention filters and present a revealed preference analysis. Section 4 gives the characterization of a DM using multiple competition filters, rationalization and categorization. Section 5 provides an analysis of 'context effects' using our approach. We will conclude with Section 6 that presents a complete literature review and Appendix which contains the proofs of the results.

[4]see Lleras et al. (2017)

2 "*Random Filtering*"

We model a decision maker who uses multiple consideration filters. Let X be the set of alternatives and $\mathbb{X}$ be the set of nonempty subsets of X. A menu A is a subset of alternatives, i.e. $A \in \mathbb{X}$. We will observe probabilistic data, meaning the data comes in the form of a probability function $p : X \times \mathbb{X} \to [0,1]$ such that it obeys the axioms of probability. The following is a definition of a random choice rule.

Definition 1. *A random choice rule is a map* $p : X \times \mathbb{X} \to [0,1]$ *such that* $\sum_{x \in A} p(x, A) = 1$*,* $p(a, A) = 0$ *for all* $a \notin A$*, and* $p(a, A) \in [0,1]$ *for all* $a \in A$ *and* $A \in \mathbb{X}$*.*

To define our procedure, we need to define what a consideration filter is.

Definition 2. *(Consideration Filter) A consideration filter is a mapping* $\Gamma : \mathbb{X} \to \mathbb{X}$ *where* $\Gamma(A) \subseteq A \ \forall A \in \mathbb{X}$*.*

Let $\mathbf{\Gamma}$ be the set of all consideration filters. The following will give a definition of a restriction put on this collection.

Definition 3. *(Restriction) A restriction is a function* $f : \mathbf{\Gamma} \to \{0,1\}$*,we say* $\Gamma \in \mathbb{\Gamma}$ *is restricted by* f *if* $f(\Gamma) = 1$*.*

The above definition will become clear with the following formulations of consideration filters coming from the well-known papers in the literature.

Definition 4. *(Rational Filter) A consideration set is a rational filter if and only if* $\Gamma(S) = S \quad \forall S \in \mathbb{X}$*. Hence, the restriction function corresponding to a rational filter is the following:* $f(\Gamma) = 1$ *if and only if* $\Gamma(S) = S \quad \forall S \in \mathbb{X}$*.*

Definition 5. *(Attention Filter, Masatlioglu et al. (2012))*
An attention filter is a consideration set such that it satisfies the following:

$$x \notin \Gamma(S) \implies \Gamma(S) = \Gamma(S \setminus x)$$

.

Similarly, Γ is restricted by attention filter whenever $f(\Gamma) = 1$ we have $x \in \Gamma(T) \implies x \in \Gamma(S) \quad \forall S \subseteq T$

Definition 6. *(Competition Filter, Lleras et al. (2017)) A competition filter is a consideration set such that it satisfies the following:*

$$x \in \Gamma(T) \implies x \in \Gamma(S) \quad \forall S \subseteq T$$

Definition 7. *(Rationalization Filter, Cherepanov et al. (2013))*

$$\Gamma^R(S) = \{x \in S : \exists R_i \in R \quad xR_iy \quad \forall y \in S \setminus x\}$$

where R *is the set of all rationales (no condition imposed, only a binary relation).*

Definition 8. *(Categorization Filter, Manzini and Mariotti (2012))*

$$\Gamma^{\gg}(S) = \{x \in S : \nexists C \in \mathbb{S} \quad s.t. \quad C \gg C^* \quad and \quad x \in C^*\}$$

where $\gg$ *is an asymmetric shading relation on the categories* $C \in \mathbb{X}$ *and* $\mathbb{S}$ *is the set of all nonempty subsets of* S*.*

Next, we define the general deterministic choice procedure for a single consideration filter. *Single Filtered Choice* models a DM who maximizes her well-defined preferences $\succ$ over what she considers, namely the consideration filter restricted by a condition f.

Definition 9. *"Single Filtered Choice"*
A choice function $c : \mathbb{X} \to X$ is called "Single Filtered Choice" (SFC) under restriction f if and only if there exists an asymmetric, transitive and complete preference relation $\succ$ on X and a single consideration filter $\Gamma : \mathbb{X} \to \mathbb{X}$ where Γ satisfies the restriction condition f such that:

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

for all $S \in \mathbb{X}$ for all $S \in \mathbb{X}$

Note that *Single Filtered Choice* reduces down to well-known deterministic models under restriction condition. For instance, *Single Filtered Choice* under Definition 5 (Attention Filter) reduces down to Choice with Limited Attention (CLA) by Masatlioglu et al.(2012).

Now think of a DM endowed with a collection of consideration sets where each one of them is restricted by the same condition f. Continuing with the same example, assume that she has a collection of attention filters, and each one of these filters (associated with a different self of the DM) has a probability to arise. DM maximizes $\succ$ over each filter, so we get the definition for the following procedure, which we call *Random Filtering.*

Definition 10. *"Random Filtering"*
A random choice rule is called "Random Filtering" (RF) under restriction f if and only if there exists an asymmetric, transitive and complete preference relation $\succ$ on X, a collection of filters $\mathbb{\Gamma}$ restricted by f where $|\mathbb{\Gamma}| \in \mathbb{Z}_+$ and a probability distribution γ over $\mathbb{\Gamma}$ such that:

$$p(a, A) = \sum_{\Gamma \in \mathbb{\Gamma}} \gamma(\Gamma) \mathbb{1}_{\{a \in A : a = \arg\max_{\Gamma(A)} \succ\}}$$

where $\mathbb{1}$ is the indicator function and $A \in \mathbb{X}$ and $a \in A$.

The intuition behind the choice procedure is the following: We have a DM that has filters that correspond to different selves of her. Each such self has a probability to arise in different environments, these are reflected by the probability distribution γ over the collection of the filters, $\mathbb{\Gamma}$. Given her well-defined preferences, the DM maximizes $\succ$ on each filter she has. Thus, the choice probability of $a \in A$ is equal to the total probability weight of the filters a maximizes $\succ$.

2.1 Results

Our first result builds a bridge between single (deterministic) and random filtering. A model is empirically testable if it cannot accommodate some observations. If it is valid for any observation, then we cannot empirically test it since it allows everything. The following result says that as long as the underlying deterministic filter gives testable implications, the random version is not empty, and vice versa.

Theorem 1. *Random Filtering restricted by f is empirically testable if and only if the underlying Single Filtered Choice with f is empirically testable.*

To understand the intuition behind the proof, note that if *Single Filtered Choice* under restriction f is empirically testable, then we have a choice function c that cannot be accommodated by this model. We can find a probabilistic model that is a direct counterpart of this choice function by defining $p(x,S) = 1$ given $c(S) = x$. Then this cannot be accommodated by the *Random Filtering* model under the same restriction. For the only if part, assume that *Single Filtered Choice* is not empirically testable. In this case, since the underlying model is not empirically testable, it is easy to show that the *Random Filtering* model also is not empirically testable.

The following definition is needed for the next result.

Definition 11. *A restriction f is a constant restriction if the following is true:* $\forall \Gamma \neq \Gamma'$ *such that* $f(\Gamma) = f(\Gamma') = 1$ *we have* $\Gamma(S) = \Gamma'(S)$ $\forall S \in \mathbb{X}$.

Note that if DM uses a *Random Filtering* procedure with a constant restriction f, then given a set S she always considers the same alternatives under different filters. Hence, random consideration due to different selves turns into a deterministic consideration under constant restriction. This in turn implies that the *Random Filtering* is no more random and equivalent to the underlying *Single Filtered Choice* under that restriction. If a DM considers all available alternatives as in the case of a perfectly rational agent, then this DM is restricted by a constant rule.

Theorem 2. *A probabilistic choice data is consistent with Random Filtering under a constant restriction f if and only if* $\forall S \in \mathbb{X}$, $p(x,S) = 1$ *for some* $x \in S$ *and the resulting choice function is consistent with Single Filtered Choice under the constant restriction f.*

Theorem 2 allows us to characterize *Random Filtering* restricted by a Rational Filter directly, since such a restriction is an example of a constant restriction in which the DM considers all available alternatives. *Random Filtering* restricted by a rational filter is called "*Rational Filtering*".

Corollary 1. *"Rational Filtering"*
A probabilistic choice data is consistent with "Rational Filtering" if and only if $\exists x \in S$ *for all* $S \in \mathbb{X}$ *such that* $p(x,S) = 1$ *and the corresponding choice function satisfies WARP.*[5]

For a probabilistic choice data to be consistent with *Random Filtering*, we need to have a collection of consideration sets that results in the probabilistic data we have. Note that maximizing $\succ$ on a consideration set corresponds to a choice function. So, given sufficient number of choice functions, we can represent $p(x,S)$ in terms of the frequency of choice functions that chooses x in S over the whole set of choice functions. Next result summarizes this connection:

Theorem 3. *A probabilistic choice data is consistent with Random Filtering under f if and only if there exists a preference relation $\succ$ and a set of choice functions $\mathbb{C}$ with* $|\mathbb{C}| \in \mathbb{R}$, *each consistent with SFC under f, such that the following holds:*

$$p(a,A) \approx \frac{|c \in \mathbb{C} : c(A) = a|}{|\mathbb{C}|}$$

for all $A \in \mathbb{X}$ *and* $a \in A$.

[5] For the resulting choice function.

Proofs of all the results in this section and similarly for all other sections are in the appendix. Note that beyond establishing a firm connection between single filtering and random filtering (i.e. the deterministic and random models of multiple filters), we characterized the first type of consideration sets, namely the consideration sets of a perfectly rational agent. Including the case of such a cognitively superior agent and other fully identified cases, we showed that using multiple filters does not create any randomness in the choice, so in that sense as long as we know what someone considers, it is insignificant how many filters someone uses, since the DM uses only one single filter in essence. Another important intuition comes from Theorem 3 and its Corollary, since it shows us that by creating at least as much choice functions as there are filters, we can consistently represent any probabilistic data. Beyond "theory of selves", it gives a further intuition into modelling a population of consumers consisting of different types. Assuming each choice function corresponds to a different type (not a single self of a unique individual), probabilistic choice represents the choices of the population. Constructing choice functions that correspond to the filters the DM uses is a crucial part of each proof, and as it is made clear on the appendix, one can construct these from the observed probabilistic choice data.

3 *Attention Filtering*

Consider a DM who is not perfectly rational, she has lack of attention, however her attention may not be deterministic. The usual approach to model the randomness of the attention (as in Masatlioglu and Cattanea 2017) is assuming that each subset of a set has a probability of attracting attention in that particular set. In contrast, we model random attention as if it arises due to randomness of the psychological state the DM is in. Thus, our model gives a direct and solid interpretation of what a random choice is. One can reach the attention probability of a set $S \in \mathbb{X}$ by summing over the probabilities of the filters that result in that particular subset, i.e. given $q(A,S)$ is the probability that $A \subseteq S$ attracts attention in S, we have:

$$q(A,S) = \sum_{\{\Gamma \in \mathbb{\Gamma} : \Gamma(S)=A\}} \gamma(\Gamma)$$

.

This natural interpretation provides an important insight: The randomness of attention paid to a certain set of alternatives does not originate from the alternatives themselves, they come from the cognitive capabilities of the agent, and these depend on the psychological state the DM is in.

In this section, we will assume that structure of the consideration set is an *Attention Filter* (see Example 2)). Such a collection of filters will be denoted by $\mathbb{\Gamma}_{\mathbb{AF}}$ where the subscript AF denotes the restriction of being a attention filter. The intuition of an attention filter is straightforward, it says that given an element does not attract attention in a set S, then removing that alternative should not change what the decision maker pays attention to. Note that since there are multiple filters for DM, the DM may pay attention to different subsets of a set of alternatives in each filter she uses. The following definition will be the basis of our characterization:

Definition 12.

$$aRb \quad if \quad p(a, A \setminus b) < p(a, A)$$

for some $A \subseteq X$.

The intuition behind the above definition is the following: Note that if we remove b from A, Γs in which $b \notin \Gamma(A)$ remains the same. Thus, given a maximizes some of these filters, the probability change does not come from these unchanged attention filters. This implies that the change in the choice probability of a comes from an attention filter Γ where $b \in \Gamma(A)$ and a is the most preferred element in $\Gamma(A)$. Observe what happened in such attention filters: As we removed b, these attention filters did not remain the same and changed so that we observed a decrease in the choice probability of a. Hence, another element which is not previously in $\Gamma(A)$ entered $\Gamma(A \setminus b)$ and strictly preferred to the best element of $\Gamma(A)$, which is a, and therefore we observed the decrease in the choice probability of a. Since a is the best element of $\Gamma(A)$ and we know that $b \in \Gamma(A)$, a is better than b, i.e. aRb. This implies that the acyclicity of R is a necessary condition for the data to be consistent with *Attention Filtering*. The following theorem shows that it is also sufficient.

Theorem 4. *"Attention Filtering" (AF)*
Random Filtering is an Attention Filtering if and only if R is acyclic.

To see the intuition behind the proof, note that we can construct sufficiently many consideration sets. Then, we can distribute each element to these filters so that the choice probabilities are compatible with this distribution. We need to ensure that each of these consideration set is an attention filter. To achieve this, we use the acyclicity of R and show that we can distribute these elements in such a way that they do not violate the property of being an attention filter. In this step, acyclicity ensures that there are enough cells to be filled by each element and these cells do not overlap with each other.

Note that one can also do a revealed preference analysis in line with the deterministic counterpart. The following is the definition of revealed preference:

Definition 13. *(Revealed Preference) Suppose a probabilistic choice data is consistent with Attention Filtering. Let $\{\mathbb{\Gamma}^{\mathbb{k}}, \succ_k\}$ be all possible representations consistent with the data. Then a is revealed preferred to b if and only if $a \succ_k b \ \forall k$.*

First, let us take the transitive closure of R, denote it by $\overline{R}$.The argument given before the characterization theorem shows that if $a\overline{R}b$ then a is revealed to be preferred to b. It turns out that this is the revealed preference in the model.

Theorem 5. *(Revealed Preference Theorem) a is revealed preferred to b if and only if $a\overline{R}b$.*

4 "*Competition Filtering*"

This section assumes another structure for the consideration sets which is the competition filter. Recall that a competition filter satisfies the following condition:

$$x \in \Gamma(T) \implies x \in \Gamma(S) \qquad \forall S \subseteq T$$

Such a collection of filters will be denoted by $\mathbb{\Gamma}_{\mathbb{CF}}$ where the subscript CF denotes the restriction of being a competition filter. The idea of a competition filter is to capture the competition between alternatives in a given choice set. If an alternative x attracts attention in a bigger set, it means that it is sufficiently competitive to attract attention in a smaller set. To characterize this model, we need to construct a preference relation over the alternatives. To do this, we need the following construction of the "level sets":

Definition 14. *(Level sets)*

$$X^0 := \{x \in X : p(x,S) \geq p(x,T)\ \forall S \subseteq T\}$$

$$X^1 := \{x \in X^{>0} : \exists S \subseteq T \in \mathbb{X}\ p(x,S) < p(x,T) \textit{ and } \sum_{y \in M} p(y,S) - p(y,T) \geq p(x,T) - p(x,S)\}$$

where $X^{>0} = X \setminus X^0$ *and* $M \subseteq X^0$.

$$X^{k+1} := \{x \in X^{>k} : \exists S \subseteq T \in \mathbb{X}\ p(x,S) < p(x,T) \textit{ and } \sum_{y \in M} p(y,S) - p(y,T) \geq p(x,T) - p(x,S)\}$$

where $X^{>k} = S^* = X \setminus \bigcup_{i=1}^{k} X^i$ *and* $M \subseteq \bigcup_{i=1}^{k} X^i$.

Define these level sets until n *such that* $X^n = \{\emptyset\}$. *Let* $X^* = \cup_{i=1}^{n-1} X^i$.

Although the above defined "level sets" may look complicated, the intuition behind the construction is clean-cut. The construction of the level sets depends on the observation that a decrease in the probability of y must be balanced at least by one element that is better. To see the logic, note that we observe a decrease in the probability of an element in a subset, because a better element entered the competition filter in that subset and preferred to the previous best element, y. Thus, such better elements than y must balance the decrease in y. Since the alternatives in X^0 never decrease in probability, these elements are candidate for this compensation and certainly the best element of X belongs to this set (Note that the choice probability of the best element can never decrease, since by the property of a competition filter it must be considered in every subset of it, and by the virtue of being the best element). If the data is consistent with *Competition Filtering* then we can repeat this construction until we exhaust the whole set of alternatives. It also turns out that if the above-defined level sets exhausts the set of all alternatives then the data is consistent with *Competition Filtering.*

Theorem 6. *Random Filtering is a Competition Filtering if and only if* $X^* = X$.

To prove the only if part, we need to construct sufficiently many competition filters where the observed choice probabilities are compatible with the implied construction. Intuitively, such construction is possible by the definition of level sets. Note that the first level set takes the best elements, i.e. the elements that never decrease in probability. The next level set is the set of elements that are balanced by the best elements, so in this sense they are second-best, because of the logic we explained in the previous paragraph before Definition 15: At least one better element should always balance out the decrease in a worse one by increasing in probability at least that in the same amount. The proof is in the appendix.

4.1 *Rationalization/Categorization*

As pointed out in Lleras et al.(2017), even though the model of rationalization is not a two-stage model, it can be expressed in terms of such a model. Recall that we can model the first stage with the following "Rationalization Filter": $\Gamma^R(S) = \{x \in S : \exists R_i \in R \quad xR_iy \quad \forall y \in S \setminus x\}$. An alternative x is considered in the first stage if there is a reason, i.e. a "Rationale" maximized by that alternative. In the second stage, the DM chooses according to an order (as we mentioned, we use the theory of order rationalization and the definition of an order coincides with our usual definition of a preference relation). Conversely, given $x \in \Gamma(S)$ and $y \notin \Gamma(S)$, we can define a "rationale" R_i such that xR_iy. Thus, these two models can be expressed in terms of each other. Hence, this model can be characterized by Theorem 6, since it is a general characterization for any finite number of competition filter.

Similarly, one can express Categorization(Manzini and Mariotti, 2012) in terms of a competition filter, and vice verse. Recall the competition filter needed to express, "Categorization Filter":

$$\Gamma^{\gg}(S) = \{x \in S : \nexists C \in \mathbb{S} \quad s.t. \quad C \gg C^* \quad and \quad x \in C^*\}$$

An alternative x is considered if there is no category that dominates the category x belongs to. A DM that uses *Categorization Filter"ing"* is has the following type collection of filters: $\Gamma(S) = \Gamma^{\gg}(S)$ $\forall \Gamma \in \mathbb{\Gamma}$ and $\forall S \in \mathbb{X}$. We will denote this collection by $\mathbb{\Gamma}_{\gg}$. For the converse, define a shading relation $\gg$ in the following way: $T \gg x$ if $x \notin \Gamma(T)$. Also similarly, Theorem 6 is sufficient to characterize this model.

Theorem 7. *The following are equivalent:*
1) A RF is a "Competition Filtering".
2) A RF is a "Rationalization Filtering".
3) A RF is a "Categorization Filtering".

5 *Context Effects*

One of the important merits of these various models is the fact that they do not satisfy the well-known regularity condition: $p(x, T) \geq p(x, S)$ $\forall T \subseteq S$. A classical violation of regularity is provided by Debreu (1960). Consider three alternatives: red bus (R), blue bus (B) and a train (T). Imagine that in every two element subset of this set, we observe that each element has the choice probability of 1/2, but in the whole set the choice probability of a train is higher than 1/2. Note that this example violates regularity, since train has a higher probability of choice in the whole set. The main point of this example that there are two very similar alternatives, R and B, which only differ by their color. As we move to the whole set, these very similar alternatives seem as a one and train apparently differentiates from these (assuming the color of the bus does not matter for the DM). Because of this feature, this example is called as the "duplicates problem" in the literature. Next, we will present two different approaches to explain this phenomena with our *Random Filtering* model that uses first attention filters and then competition filters, i.e. respectively *Attention Filtering* and *Competition Filtering.* In the following examples, assume that $\gamma(\Gamma_i) = 1/6$.

Attention Filtering						
	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Γ_6
RBT	R	RT	RBT	BT	BT	B
RB	R	R	R	B	B	B
RT	R	RT	RT	RT	R	R
TB	B	B	TB	TB	TB	B

Assume that each element has the probability of being chosen equal to 1/2. In addition to that, let

$$p(T, RBT) = 2/3, \qquad p(R, RBT) = p(B, RBT) = 1/6$$

The above table is an example of a DM who uses *Attention Filtering* and has 6 of them. Each column satisfies the condition of being an attention filter, but as you see what she pays attention to differ from filter to filter. The following preference profile of the DM would give the "duplicates problem" as a result: $T \succ R$ and $T \succ B$. Note that given this ordering, T maximizes the attention filters from 2 to 5 in the whole set. This gives us $p(T, RBT) = 2/3$. The rest follows similarly. The following table explains the same example from the perspective of a DM that uses *Competition Filtering*.

Competition Filtering						
	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Γ_6
RBT	R	B	T	T	T	T
RB	R	B	B	B	R	R
RT	R	R	RT	T	T	T
TB	B	B	TB	T	T	T

This above table summarizes a DM who uses 6 different competition filters. The logic is similar to the previous example about attention filters, however we need a different preference profile. Let it be the following: $B \succ T$ and $R \succ T$. To give an example from the table, look at the line of the subset R, B. From 3 to 6, the DM considers the options mixed in a such a way that train is always considered but buses equally likely. In the first one the DM considers only the red one and in the second only the blue one. Since the DM prefers the option of bus to the train no matter what the color is, we have the desired probabilities of choice. One can check that all 6 filters satisfy the condition of being a competition filter.

The final result shows us that it is not surprising to have a violation of regularity in these models, since the random model is directly connected to the underlying model, which allows for choice reversals.

Theorem 8. *RF violates "regularity" if and only if underlying SFC violates WARP.*

Theorem 8 puts a loose restriction for a *RF* to violate regularity. If the underlying model violates WARP, then we must observe a choice reversal as we move from a set to its subset in a way that the DM does not choose what she chooses in a smaller set. This amounts to saying that she will choose a particular element x in a bigger set more frequently and violates regularity. Conversely, if regularity is violated, then there must exist at least one choice function that selects x in a superset but not on its subset, recall Theorem 3 that gives such a connection. Such a choice function violates WARP.

6 Literature Review

As we particularly emphasized, our paper forms a connection between the consideration set and random choice literature by presenting a "theory of selves" approach. The recent literature of deterministic consideration models started to become highly developed, there are various papers on this subject. Although we presented some of these in the introduction, we will give a brief summary of these without falling into too much repetition.

It is well-known that people do not consider all available alternatives, so in a sense they form consideration sets. The papers in the literature put intuitive conditions on these consideration sets. Masatlioglu et al. (2012) models Attention Filter, where the removal of an alternative DM is unaware of does not change the set of considered alternatives. On the other hand, Lleras et al. (2017) models Competition Filter. In this model, an alternative that attracts attention in a bigger set must attract attention in any subset where this alternative is an element. Although these papers model a similar phenomena, their empirical implications differ substantively. Observe that these models directly put conditions on the consideration sets they form. The following models do not explicitly model a consideration set, however their choice procedure implicitly contains a one. We outlined such a relation in the previous parts. Categorize Then Choose is an example of such a model due to Manzini and Mariotti (2012). The DM in this model categorizes the set of available alternatives according to an unobserved process and there is a sort of dominance relation between categories which is asymmetric: a category can shade another. One can deduce a consideration set by collecting all alternatives that belong to an undominated (unshaded) category. Similarly, Rationalization due to Cherepanov et al.(2013), tells a completely different story but is connected to consideration set literature. An alternative is psychologically permissible in their model if there is a rationale to justify it. The DM has a set of rationales which are defined to be only binary relations, the consideration set can be deduced by collecting all alternatives that at least maximizes one relation. Besides these models, there are some other well-known models we did not explicitly model in this paper: Masatlioglu and Ok(2014) builds a reference-dependent model where the reference is taken to be the status-quo alternative, a consideration set is formed given the status-quo. Manzini and Mariotti (2007) first eliminates alternatives according to some rationale, and then maximizes another rationale among the remaining alternatives. The set of remaining alternatives are the ones who are considered. Caplin and Dean(2011), and Caplin, Dean, and Martin(2011) a DM forms implicitly a consideration set that depends on time. Eliaz, Richter, and Rubinstein (2011) models a DM who forms a consideration set of at most 2 alternatives. Eliaz and Spiegler (2009) uses the approach of consideration sets in a competitive marketing setting. Kimya (2015) uses attributes of the alternatives to form consideration sets.

The literature on random choice is also highly developed. One of the foremost works in this literature is Manzini and Mariotti (2014), where the DM has a fixed probability of considering each alternative. Brady and Rehbeck (2016) is in the same vein with Manzini and Mariotti (2014), where menu dependency of choice incorporated with some restrictions. The other vein of works is famous Random Utility Models (RUM), where the DM has a probability distribution over the set of utility functions. This in turn is equivalent to saying that there are different set of preference profiles, i.e. the preferences of the DM are not stable. As we put out, the approach of "stable preferences, changing consideration" approach we developed by using multiple deterministic filters is more intuitive. Moreover, these random utility models do not allow violation of regularity. The literature on (RUM) is very old and well-established, some works are: Luce (1959), Block and Marschak (1960), McFadden (1978), Falmagne (1978), and Gul, Natenzon and Pesendorfer (2010). A recent unpublished work in progress belongs to Masatlioglu and Cattaneo, where they model Random Attention by assuming that each menu has a certain probability of attracting attention. This model has the same empirical implication with *Attention Filtering*. Our main difference is the following: We model a DM that uses each deterministic attention filter with some probability in a set of unobserved collection of filters; whereas they directly assume the randomness, i.e. each subset has a certain probability to attract attention. Thus, our model gives an explanation for the probabilistic attention; it gives also a content to the random choice we observe.[6]

7 APPENDIX

Proof of Theorem 1. *If SFC is empirically testable, then it cannot accommodate every choice function. Assume that c is one of them. For any set S, define* $p(x, S) = 1$ *if* $c(S) = x$*. Note that RF cannot accommodate this, since the underlying model does not allow c.*

Conversely, assume that SFC is not empirically testable, i.e. it can accommodate any choice function. By definition, RF is a collection of finite number of filters that are restricted by the same condition. Therefore, RF can accommodate any random data since underlying choice function can accommodate anything.

Proof of Theorem 2. *A constant restriction f says that* $\Gamma(S) = \Gamma'(S) = T$ *where T is a certain subset of S for all* Γ, Γ' *for which* $f(\Gamma) = f(\Gamma') = 1$*. Thus, the collection of all such filters is equivalent to one single filter* Γ^* *and RF reduces to a SFC under f. The converse direction is true by definition, we can write any SFC in terms of a multiple filters under constant restriction.*

Proof of Corollary 1. *Directly follows from Theorem 2, since the restriction is a constant restriction.*

Proof of Theorem 3. *For the if part, note that for each choice function c under SFC restricted by f, there is a corresponding consideration set* Γ *under f. Take* $\mathbf{\Gamma}$ *to be the collection of these filters. Define* $\gamma(\Gamma) \approx 1/|\mathbb{C}|$*. Note that* $|c \in \mathbb{C} : c(A) = a| = |\{\Gamma \in \mathbb{\Gamma} : a = \arg\max_{\Gamma(A)} \succ\}|$ *Since* $1/|\mathbb{C}|$ *is a constant, we can take it out the summation (hence* $\gamma(\Gamma)$*) and we are left with* $\sum_{\Gamma \in \mathbb{\Gamma}} \mathbb{1}_{\{a \in A : a = \arg\max_{\Gamma(A)} \succ\}}$ *which is equal to* $|\{\Gamma \in \mathbb{\Gamma} : a = \arg\max_{\Gamma(A)} \succ\}|$ *and thus* $|c \in \mathbb{C} : c(A) = a|$*, so the proof of the if part is complete since* $\sum_{\Gamma \in \mathbb{\Gamma}} \gamma(\Gamma) \mathbb{1}_{\{a \in A : a = \arg\max_{\Gamma(A)} \succ\}} \approx \dfrac{|c \in \mathbb{C} : c(A) = a|}{|\mathbb{C}|}$.

For the only if part, multiply all $\gamma(\Gamma)$ *(denote it by m) and let* $\mathbb{C}$ *be the set of choice functions that correspond to each* Γ *where we produce* $\gamma(\Gamma)/m$ *of each such choice function. If this is not a rational number, then one can approximate the irrational* $\gamma(\Gamma)/m$

[6]Please refer to these articles to learn more about the marketing and psychology literature evidence on consideration sets and random choice.

with a sequence of rational numbers which converge to that irrational number by the denseness of rationals. Hence, we expressed given probabilities that are consistent with RF in terms of $\frac{|c \in \mathbb{C} : c(A) = a|}{|\mathbb{C}|}$.

Proof of Theorem 4. *("Attention Filtering")*

Necessity already shown, R's cyclicity implies that the revealed preference has a cycle by the revealed preference theorem and this is a contradiction to the definition of Attention Filtering. For the only if part, first extend $\overline{R}$ *which is to a strict linear order over* X*, denote this extension by* $\succ$*. Enumerate all elements in* X *by* x_i*, where* $x_i \succ x_{i+1}$ $\forall i$*. Define* $\underline{x_i}(S) = \{y \in S : x_i \succ y\} \cup \{x_i\}$[7] *for any set* S*. We will prove that we can construct a collection of attention filters* $\{\mathbb{\Gamma}_{AF}\}$ *recursively. First, construct a collection of attention filters such that* $|\mathbb{\Gamma}_{AF}| \leq |X| \times |\mathbb{X}|$*. Define* k_i *as the ith minimum probability that is observed in the data and let* $k_0 = 0$*. Let* $\gamma(\Gamma_i) = k_i - k_{i-1}$*. We will start the construction first by defining it for all 2-element subsets.* $\forall x_i, x_{i+k} \in X$, $\Gamma_j(x_i x_{i+k}) = x_i x_{i+k}$ $\forall j$ *from* 1 *to* k *where the sum of respective* $\gamma(\Gamma_j)$*'s are equal to the* $p(x_i, x_i x_{i+k})$*. The rest is filled with* x_{i+k}*. Suppose that for all sets* $S \in \mathbb{X}$ *with* $|S| < k$*, the attention filters are formed. Take* T *with cardinality* k*. Let* x_1^T *and* x_n^T *be the best and worst elements, respectively. Consideration sets are formed in the following way: First, put* x_n^T *to the last* Γ_i*'s such that the sum of respective* $\gamma(\Gamma_i)$*'s are equal to the* $p(x_n^T, T)$*. Then, for any* $x \in T \setminus x_1^T x_n^T$*, put* $\underline{x}$ *to the sufficient amount of* Γ_i*'s similar to the above procedure, but starting from the minimum index* k *where* $\Gamma_k(\underline{x}) = \underline{x}$ *and only to this type filters, in order. We need to guarantee that there is enough cells to fill in each step. To see this, note that we only eliminate better elements than* x_i *as we move from* $\underline{x_i}$ *to any superset* T*. So; it must be the case that* $p(x_i, \underline{x_i}) \geq p(x_i, T)$*, since otherwise we would have* $p(x_i, \underline{x_i}) < p(x_i, T)$ *and this implies that* $x_i \succ x_k$ *for* $k < i$*, a contradiction. Since* $p(x_i, \underline{x_i}) \geq p(x_i, T)$*, the cells that are needed by* $x \in T \setminus x_1^T x_n^T$ *decreases in number. This means that we have enough cells for all such* x*. Lastly, distribute* $\underline{x_1^T}$ *to the remaining cells. We need to show that in this procedure all* $x_i \neq x_j$ *where* $j > i$ *must stay on different cells. This follows by the construction, since we already put* x_j *into the cells in* $\underline{x_i}$ *following the order coming from the cells of the type* $\Gamma_k(\underline{x_j}) = \underline{x_j}$*. So, the order following from the* $\Gamma_k(\underline{x_i}) = \underline{x_i}$ *to distribute* x_i *in any set* T *uses completely different cells than* x_j*'s cells coming from* $\Gamma_k(\underline{x_j}) = \underline{x_j}$*.*

Now, we need to show that the constructed sets are attention filters. Assume to the contrary $\Gamma_i(S) \neq \Gamma_i(S \setminus x)$ *for* $x \notin \Gamma_i(S)$*.* $\Gamma_i(S) = \underline{y}$ *for some* $y \in S \setminus x$*.* $x \notin \Gamma_i(S)$ *implies that* $x \succ y$*. So, it need to be the case that* $p(y, S \setminus x) \geq p(y, S)$*, since otherwise* $y \succ x$ *which is a contradiction. Note that* $\underline{y} \subseteq S \setminus x$*. Since we put* $\underline{y}$ *in* S *and* $S \setminus x$ *according to the order coming from* $\Gamma_k(\underline{y}) = \underline{y}$*.* $\Gamma_i(S \setminus x) \neq \Gamma_i(S) = \underline{y}$*, so* $\underline{y}$ *fills more cells in* S *than* $S \setminus x$*, since they start from the same index but* y *does not fill the ith index in* $S \setminus x$*. This implies that* $p(y, S \setminus x) < p(y, S)$ *which implies* $y \succ x$*, a contradiction. Hence, this construction satisfies the condition of being an attention filter for each* i*.*

Proof of Theorem 5. *We already showed that if* $a\overline{R}b$ *then* a *is revealed to be preferred to* b*. For the other direction, observe that if it is not the case that* $a\overline{R}b$ *then we can find a preference profile* $\succ$ *that includes* $\overline{R}$ *(recall that it is the transitive closure of* R*) where* $b \succ a$*. Then, the proof of Theorem X shows that the data can be rationalized with this preference.*

[7] For ease of notation, we will use simply $\underline{x}$ and which set we take on the lower contour set will be clear from the context.

Proof of Theorem 6. *We need to prove the sufficiency part. Define $\triangleright$ in the following way: $x \triangleright y$ if $\in X^k$ and $y \in X^l$ where $l > k$. Take a preference profile that includes $\bar{\triangleright}$, such a preference profile exists since we can extend $\bar{\triangleright}$ to a strict linear order because of the acyclicity. Enumerate the elements in the same way as in the previous proof. For the whole set X, define $\Gamma_j(X) = \{x_i\}$ $\forall j$ as in the Proof of Theorem 4. We will construct recursively. Suppose that for all sets $S \in \mathbb{X}$ with $|S| > k$, the competition filters are formed. For $|S| = k$, define the following:*

$$\Gamma^0(S) = \bigcup_{S \subset T} \Gamma^0(T)$$

. The algorithm for constructing the final competition filters Γ^* is as follows: If $p(x_i, T) > |\{\Gamma^0 \in \mathbb{\Gamma}_{CF} : x_i = \arg\max_{\Gamma^0(T)} \succ\}|$, put x_i into $\Gamma \in \{\Gamma^0 \in \mathbb{\Gamma}_{CF} : x_j = \arg\max_{\Gamma^0_i(T)} \succ\}$ where $i > j$ and $p(x_j, T) < |\{\Gamma^0 \in \mathbb{\Gamma}_{CF} : x_j = \arg\max_{\Gamma^0(T)} \succ\}|$, i.e. we put the better element into a competition filter maximized by a worse element so that to distribute the remaining probability. Note that this construction satisfies to be a competition filter, since by definition all elements in the filters of supersets are contained in S's filter and the algorithm only adds new elements to the above defined filters. We only need to show that this algorithm results in a consistency between the probabilistic data and the construction. To the contrary, assume that the algorithm does not work, i.e. we cannot distribute the remaining probability from x_k in the described way after we successfully distributed it until x_k (starting from the most preferred until kth one). So, it must be the case that:

$\sum_{l \geq k} p(x_k, S) > \gamma(\{\Gamma_i \in \mathbb{\Gamma}_{CF} : \Gamma^0_i(S) = \{\emptyset\}\} \cup \bigcup_{i \geq k} \{\Gamma^0_i \in \mathbb{\Gamma}_{CF} : \Gamma^0_{i,x_i}\})$ [8]. (1)

Observe the following two equalities:

1$=\gamma(\{\Gamma_i \in \mathbb{\Gamma}_{CF} : \Gamma_{i^0,\{\emptyset\}}(S) \cup \bigcup_{i \geq k} \{\Gamma_i \in \mathbb{\Gamma}_{CF} : \Gamma^0_{i,x_i} \cup \bigcup_{i<k} \{\Gamma_i \in \mathbb{\Gamma}_{CF} : \Gamma^0_{i,x_i}\})$ (2)

1$=\sum_i p(x_i, S) = \sum_{i \geq k} p(x_i, S) + \sum_{i<k} p(x_i, S)$ (3)

But then, then by (1) and (3):

[8]For notational ease, we use the following: $\Gamma^0_{i,x_i} = \{x_i = \arg\max_{\Gamma^0_i(T)} \succ\}$. Also,

$$\gamma(\{\Gamma_i \in \mathbb{\Gamma}_{CF} : \Gamma^0_i(S) = \{\emptyset\}\} \cup \bigcup_{i \geq k} \{\Gamma^0_i \in \mathbb{\Gamma}_{CF} : \Gamma^0_{i,x_i}\})$$

denotes the sum of the probabilities of using respective Γ_i's

$$1 > \sum_{i<k} p(x_i, S) + \gamma(\{\Gamma_i \in \mathbb{\Gamma}_{CF} : \Gamma_i^0(S) = \{\emptyset\}\} \cup \bigcup_{i \geq k} \{\Gamma_i^0 \in \mathbb{\Gamma}_{CF} : \Gamma_{i,x_i}^0\})$$

Using (2) we are left with:

$$\gamma(\bigcup_{i<k} \{\Gamma_i \in \mathbb{\Gamma}_{CF} : \Gamma_{i,x_i}^0(S)) > \sum_{i<k} p(x_i, S)$$

Note the following:

$$p(x_i, S \cup x_j) \geq \gamma(\{\Gamma_i \in \mathbb{\Gamma}_{CF} : \Gamma_{i,x_i}^0(S)\})$$

Combining this last expression with the previous one, we have the following:

$$\sum_{i<k} p(x_i, S \cup x_j) > \gamma(\bigcup_{i<k} \{\Gamma_i \in \mathbb{\Gamma}_{CF} : \Gamma_{i,x_i}^0(S)\}) > \sum_{i<k} p(x_i, S)$$

Finally, the last expression implies:

$$\sum_{i<k} p(x_i, S \cup x_j) - p(x_i, S) > 0$$

Recall our necessary condition: There must be a better element that balances the decrease in probability, however note that here the sum of top k elements decrease. Thus, we cannot find any better element to balance this decrease, since there exists no better element than the ones in $\{1, ..., k\}$. *This is a contradiction to our necessary condition, hence the proof is complete.*

Proof of Theorem 7. *The equivalence between these models are shown in the text. Theorem 6 can be directly applied to "Categorization Filtering" and "Rationalization Filtering".*

Proof of Theorem 8. *Assume that it allows for regularity but there is no underlying SFC that violates WARP. Then, given* x *is chosen from a set* S, *it must be the case that* x *will be chosen in any subset* T *of* S *given* $x \in T$. *So, for any such subset, the probability that* x *is chosen can never decrease, and thus regularity cannot be violated.*

For the converse, assume WARP is violated for the underlying SFC, but regularity is not violated. Since WARP is violated, there can exist $x \in S \cap T$ *where* $T \subseteq S$ *such that* $c(S) = x \neq c(T)$ *for some* $\Gamma \in \mathbb{\Gamma}$. *By Theorem 3, this increases the probability that* x *is chosen in* S, *leading to a contradiction since regularity is violated.*

8 References

References

Block and Marschak 1960. "Random orderings and stochastic theories of responses." In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, and H. Mann (Eds.), Contributions to probability and statistics. Stanford, Calif.: Stanford University Press, 1960: 97-132

Brady, R. L. and John Rehbeck 2016. "Menu-dependent stochastic feasibility." Econometrica, 84(3), 1203-1223.

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.

Cherepanov, Vadim, Timothy Feddersen, and Alvaro Sandroni. 2010. “Rationalization.” Theoretical Economics 8 (3): 775-800

Debreu, G. 1960. “Review of Individual choice behaviour by R. D. Luce.” American Economics Review 50: 186-188.

Eliaz, Kfir, Michael Richter, and Ariel Rubinstein. 2011 . “Choosing the Two Finalists.” Economic The- ory 46 (2): 211-19

Eliaz, Kfir, and Ran Spiegler. 201 1. “Consideration Sets and Competitive Marketing.” Review of Eco- nomic Studies 78 (1): 235-62

Falmagne, J. C. 1978. “A representation theorem for finite random scale systems.” Journal of Mathematical Psychology 18(1): 5272

Gul, F., Natenzon, P., and Pesendorfer, W. 2014. “Random choice as behavioral optimization.” Econometrica 82(5): 1873-1912.

Hauser, John R., and Birger Wernerfelt. 1990. “An Evaluation Cost Model of Evoked Sets.” Journal of Consumer Research 16: 383-408

Iyengar, Sheena S., and Mark R. Lepper. 2000. “When Choice is Demotivating: Can One Desire Too Much of a Good Thing?” Journal of Personality and Social Psychology, 79(6): 995-1006.

Kimya, Mert. 2015. “Choice, Consideration Sets and Attribute Filters”, mimeo

Luce, R.D. 1959. “Individual choice behavior: a theoretical analysis.” , Wiley, New York

Masatlioglu, Y., and Matias D. Cattaneo, “A Random Attention Model: Identification, Estimation and Testing”, mimeo

Masatlioglu, Y., and Ok, E.A., 2014. “A canonical model of choice with initial endowments.” Rev. Econ. Stud.81 (2), 851-88

Lleras, Juan S., Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay. 2010. “When More is Less: Choice by Limited Consideration.” Journal of Economic Theory 170, 70-85.

Manzini, Paola, and Marco Mariotti. 2007. “Sequentially Rationalizable Choice.” American Economic Review 97 (5): 1824-39.

Manzini, Paola, and Marco Mariotti. 2012. “Categorize Then Choose: Boundedly Rational Choice and Welfare.” Journal of the European Economic Association 10: 1141-1165.

Manzini, P. Mariotti, M. (2014). Stochastic choice and consideration sets. Econometrica, 82(3), 11531176.

Masatlioglu, Y., D. Nakajima, and E. Ozbay. 2011, "Revealed attention" American Economic Review 102 (5): 2183-2205

McFadden, D. 1978. "Modelling the choice of residential location." In: S. Karlqvist, L. Lundquist, F. Snickers and J. Weibull, Eds., Spatial Interaction Theory and Planning Models, North-Holland, Amsterdam (1978).

Reutskaja, E., Nagel, R., Camerer, C.F., and Antonio Rangel. 2011. "Search dynamics in consumer choice under time pressure: an eye-tracking study." Am. Econ. Review 101 (2): 900-926

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.