Satisficing and Stochastic Choice

Victor H. Aguiar* María José Boccardi[†] and Mark Dean[‡]

July 12, 2016

Abstract

Satisficing is a hugely influential model of boundedly rational choice, yet it cannot be easily tested using standard choice data. We develop necessary and sufficient conditions for *stochastic* choice data to be consistent with satisficing, assuming that preferences are fixed, but search order may change randomly. The model predicts that stochastic choice can only occur amongst elements that are always chosen, while all other choices must be consistent with standard utility maximization. Adding the assumption that the probability distribution over search orders is the same for all choice sets makes the satisficing model a subset of the class of random utility models.

1 Introduction

People often do not pay attention to all the available alternatives before making a choice. This fact has lead to an extensive recent literature aimed at understanding the observable implications of models in which the decision maker (DM) has limited attention.[1] In an important recent paper, Manzini & Mariotti (2014) characterize the stochastic choice data generated by a decision maker (DM) who has standard preferences, but only notices each alternative in their choice set with some probability. The chosen item is therefore the best alternative in the 'consideration set' of noticed items, which may be a strict subset of the items which are actually available.

The idea that a DM may not search exhaustively through all available alternatives is not new. Simon (1955) introduced the concept of satisficing: an intuitively plausible choice procedure by which the DM searches through alternatives until they find one that is 'good

*Department of Economics, Brown University. Email: victor_aguiar@brown.edu

[†]Department of Economics, Brown University. Email: maria_jose_boccardi@brown.edu

[‡]Department of Economics, Columbia University. Email: mark.dean@columbia.edu

[1]Notable examples include Masatlioglu et al. (2012), Caplin et al. (2011), Eliaz & Spiegler (2011), and Salant & Rubinstein (2008).

enough', at which point they stop and choose that alternative.[2] This model has been hugely influential, both within economics, and in other fields such as psychology (Schwartz et al. (2002)) and ecology (Ward (1992)).

Despite the popularity of the satisficing model, testing its predictions can prove challenging. It has long been known that standard choice data, which records only the choices made from different choice sets, cannot be used to disentangle satisficing from utility maximization (see Caplin et al. (2011) for a discussion). Researchers have therefore typically resorted to richer data sets in order to test the satisficing model. For example, Caplin et al. (2011) make use of 'choice process' data, which records the evolution of choice with decision time, while Santos et al. (2012) use the order in which alternatives were searched as recorded from their internet browsing history.

In this paper, we characterize the observable implications of the satisficing choice procedure for *stochastic* choice data. Such data has been heavily studied in the economics literature.[3] We assume that the DM has a fixed utility function and satisficing level. In any given choice set, they search sequentially until they find an alternative which has utility above their satisficing level, at which point they stop and choose that alternative. If they search the entire choice set and do not find a satisficing alternative then they choose the best available option. We assume that search order varies randomly, leading to stochasticity in choice. On the one hand, our paper is related to the work of Manzini & Mariotti (2014) (henceforth MM). It specifies a procedure by which attention is allocated, while MM is agnostic in this regard. On the other, it provides an alternative test of the satisficing model to that of Caplin et al. (2011) and Santos et al. (2012), using a data set which is readily available in many settings.

Our main observation is that the satisficing model implies that choice is stochastic only in choice sets where there are multiple alternatives above the satisficing level. If this is the case, then the order of search will affect the chosen alternative. If not, then either the choice set will be fully searched and the best option deterministically chosen, or the single satisficing alternative will always be chosen. This allows us to behaviorally identify the alternatives that are satisficing for the decision maker.

Without further restriction, any stochastic choice data set can trivially be made

[2]Caplin et al. (2011) show that satisficing behavior can be optimal under some circumstances

[3]See for example Block & Marschak (1960); Luce & Suppes (1965); Falmagne (1978); Gul & Pesendorfer (2006); Gul et al. (2014); Manzini & Mariotti (2014)

commensurate with the satisficing choice procedure by assuming that all alternatives are above the satisficing level, and the resulting distribution of choices reflects the distribution of search orders in that choice set. In order to generate meaningful behavioral implications, we must place further restrictions on the satisficing model. For our main theorem we make the assumption that the distribution of search orders has a full support property (i.e., each item has a positive probability of being searched first), and also rule out the possibility of indifference. This allows us to identify the set of above-reservation alternatives and characterize satisficing with two simple intuitive conditions. The first states that choice can be stochastic only amongst elements that are always chosen (with some probability) when available. The second says that revealed preference, defined via the support of the random choice rule in each set, must satisfy the Strong Axiom of Revealed Preference (SARP). Under these conditions, the data will admit a satisficing representation and the satisficing set, utility function and distribution over search orders can be identified to a high degree of precision.

Our baseline specification puts no restrictions on the relationship between the distribution over search orders across different choice sets. We next consider a refinement of the satisficing model in which the distribution of search order in each choice set is a manifestation of the same underlying search distribution. In order to guarantee such a representation we need a third axiom: the Total Monotonicity condition of Block & Marschak (1960). This condition on its own is necessary and sufficient for the data to be commensurate with the random utility model (RUM). Thus, the fixed distribution satisficing model is the precise intersection between satisficing and random utility.

We next discuss extensions to our results in which we relax the assumptions of full support, no indifference and the observation of a complete data set. We show that a satisficing model without full support, but with fixed distribution is equivalent to the random utility model. Allowing for indifference (but maintaining the full support assumption) is equivalent to dropping the requirement that stochasticity only take place amongst always chosen alternatives. If data is incomplete, our necessary and sufficient conditions are unchanged, but our ability to identify above-satisficing elements is reduced.

Our final extension considers what happens if we allow for further sources of stochasticity - specifically random variations in the satisficing threshold and in the utility function. One reason to consider these cases is to determine whether they help in recovering the utilities of above-satisficing alternatives, which are not identified in our baseline model. Perhaps surprisingly, the answer is no. Adding a random satisficing level has no effect on the observable implications of the model, and so cannot be distinguished from the fixed satisficing level case. Adding stochasticity to utility does change the observable implications, but not in a way that allows us to improve our identification of preferences. A model in which random variations in utility or threshold are the *only* sources of stochasticity might allow such identification. We leave such an extension for future work.

Section 2 describes our set up. Section 3 characterizes the satisficing model. Section 4 considers the extensions described above, while section 5 discusses the related literature.

2 Set Up

2.1 Data

We consider a finite abstract choice set X, and let $\mathcal{D} \subseteq 2^X \setminus \emptyset$ be the set of menus in which behavior is observed. We assume that data comes in the form of a *random choice rule*, $p : X \times \mathcal{D} \mapsto [0,1]$, which specifies for each menu $A \in D$ the probability of choosing each element $a \in A$ (for example, if the DM has a one third probability of choosing x from $\{x, y, z\}$ then $p(x, \{x, y, z\}) = \frac{1}{3}$).

Definition 1 (Data set) *A data set consists of a set if menus $\mathcal{D} \subseteq 2^X \setminus \emptyset$ and a random choice rule $p : X \times \mathcal{D} \mapsto [0,1]$ such that $\sum_{a \in A} p(a, A) = 1 \; \forall A \in \mathcal{D}$. We say a data set is **complete** if $\mathcal{D} = 2^X \setminus \emptyset$.*

Random choice rules have been heavily studied in the theoretical, as well as the applied literature.[4] In practice, while a random choice rule is not directly observable, it can be estimated from observed choice frequencies, pooling either across repeated choices by the same individual, or by aggregating across the choices of different individuals.

2.2 The Satisficing Model

The satisficing choice procedure can be described as follows: when faced with a menu of options to choose from, the DM searches through the available alternatives one by one. If, at any point, they come across an alternative which is 'good enough', they stop searching and select that alternative from the menu. If they exhaustively search all alternatives without finding an element which satisfies their criteria, then they choose the best available alternative from the set. Note that the standard model of rational choice is a limiting case

[4] Examples of early theoretical work include Block & Marschak (1960) and Luce & Suppes (1965). More recent work includes Gul & Pesendorfer (2006); Manzini & Mariotti (2014); Gul et al. (2014).

of the satisficing model in which no alternative is 'good enough'.

As a concrete example, consider a DM searching for a book to buy in a bookshop prior to a flight. They examine the available books one by one, looking for one which satisfies their requirements (humorous, has good reviews, long enough to last the flight, not by Dan Brown). If they find such a book, they immediately go to the checkout and buy it. If they search the entire selection and don't find a book which matches this criteria then they go back and choose the best of the books that they did see.

The satisficing choice procedure therefore has three building blocks. The first is a fixed utility function $u : X \rightarrow \mathbb{R}$, which describes the preferences of the DM. Following Manzini & Mariotti (2014), for our main results we rule out indifference, and therefore assume that u is injective. We discuss the implications of allowing indifference in section 4.2.

The second model element is a utility threshold u^*, which we will refer to as the *reservation utility.* This defines the concept of 'good enough': an alternative $x \in X$ is good enough if $u(x) \geq u^*$. We define $U^* = \{a \in X | u(x) \geq u^*\}$ as the set of satisficing elements according to u and u^*. For convenience, we will assume that there is at least one satisficing element: i.e. $u^* \leq max_{x \in X} u(x)$. This assumption has no behavioral implication: a model in which only the best available alternative is above the reservation utility is indistinguishable from one in which there is no such alternative. However, it will streamline the statement of identification results in section 3.

The third element of the satisficing model is the order in which search occurs. A search order for a choice set A is defined by a linear order on that set.[5] We use R_A to denote the set of linear orders on A, with r_A a typical element in R_A. Our key assumption is that the order of search is determined stochastically: we use $\gamma_A : R_A \rightarrow [0, 1]$ to denote the probability distribution over the set R_A, which we call a 'stochastic search order'. Abusing notation slightly, we will use $\gamma_A(x \ r_A \ y)$ to denote the probability of all search orders in which x appears before y: i.e. $\gamma_A(r_A \in R_A | x \ r_A \ y)$.

We are agnostic about the source of this stochasticity. It could be that the DM randomly decides the order of search - in our example, sometimes the DM search through the books alphabetically, while sometimes they do so by genre. Alternatively, it could be that the random choice rule is generated by a DM who is faced by choice situations

[5]i.e. a complete, transitive and antisymmetric binary relation on A with the interpretation 'searched no later than'.

which are framed in different ways,[6] with the framing unobservable to the researcher. For example, sometimes the bookstore puts the thrillers at the front of the store, while sometimes they put the romantic comedies at the front. These 'frames' affect the order in which the DM searches (though not their preferences), but are not known to the researcher.

A data set can be represented by the satisficing model if there exists a utility function, satisficing level and family of stochastic search orders which would generate the observed choice probabilities:

Definition 2 (General Satisficing Model (GSM)) *A data set $(\mathcal{D}, p)$ has a Generalized Satisficing Model (GSM) representation if there exists an injective $u : X \to \mathbb{R}$, $u^* \in \mathbb{R}$ such that $u^* \leq max_{x \in X} u(x)$, and $\{\gamma_A\}_{A \in \mathcal{D}}$ such that, for any $A \in \mathcal{D}$ and $a \in A$*

$$p(a, A) = \begin{cases} \gamma_A \left(r_A | a \ r_A \ b \ \forall \ b \in A \backslash \{a\} \ s.t. \ u(b) \geq u^* \right) & \textit{if } u(a) \geq u^* \\ 1 & \textit{if } a = \arg\max_{x \in A} u(x) < u^* \\ 0 & \textit{Otherwise} \end{cases} \quad (1)$$

To illustrate how the model works consider the following example.

Example 1 *Let $A = \{a, b, c\}$ and γ_A be as displayed in table 1. Consider first the case where a, b are satisficing alternatives, while c is not, that is $u(a) > u(b) > u^* > u(c)$. Then, no matter the search order, c will never be chosen, and so $p(c, A) = 0$. However, as the DM will chose a if a is seen before b and b otherwise, their frequency depends on γ_A. In particular, $p(a, A) = \frac{3}{8}$ and $p(b, A) = \frac{5}{8}$. If instead all alternatives are below the satisficing level (i.e. $u^* > max_{x \in A} u(x)$) then choice will be independent of search order: all alternatives will always be searched, and the best subsequently chosen. In this case, this means that $p(a, A) = 1$.*

3 Characterizing the Satisficing Model.

3.1 A Negative Result

The aim of this paper is to describe properties of a stochastic choice data set which are necessary and sufficient to guarantee a satisficing representation. However, our first observation is negative: without further refinement, the GSM model provides no restriction.

[6] In the sense of Salant & Rubinstein (2008).

Order	(1)	(2)	(3)	(4)	(5)	(6)
1st	a	a	b	b	c	c
2nd	b	c	a	c	a	b
3rd	c	b	c	a	b	a
γ_A	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{1}{4}$

Table 1: An example of the satisficing model

Observation 1 *Any data set* $(\mathcal{D}, p)$ *has a GSM representation.*

In order to construct a GSM representation for any data set $(\mathcal{D}, p)$, set $U : X \to [0, 1]$ to be any arbitrary one to one real valued function and set $u^* = -2$; then $U^* = X$. For any menu $A \in \mathcal{D}$, let r_A^a be the set of linear orders on A such that $a \ r \ b$ for all $b \in A$. $\{r_A^a\}_{a \in A}$ therefore defines partition on R_A. Define

$$\gamma_A(r_A) = \frac{p(a, A)}{|r_A^a|} \text{ for each } r_A \in r_A^a$$

Such a representation will generate p as, for any a, $u(a) \geq u^*$

$$\gamma_A \left(r_A | a \ r_A^a \ b \ \forall \ b \text{ s.t. } u(b) \geq u^* \right) = \gamma_A \left(r_A \in r_A^a \right) = p(a, A)$$

The GSM is flexible enough to match any data set because it places no restriction on the distribution over search orders in each decision problem. Thus one can always construct a distribution of search orders that will match the data by assuming that all alternatives are above the satisficing level.

To derive testable restrictions for the satisficing model, we introduce a 'full support' condition on the distribution of search orders. This restrictions will allow us to identify satisficing alternatives as those which are chosen with positive probability in every choice set in which they appear. We can then utilize the underlying structure of the GSM model to derive behavioral restrictions. Intuitively, the stochastic nature of search generates stochastic behavior among satisficing alternatives. In contrast, we expect to observe deterministic utility maximizing behavior among choice sets which consist only of non-satisficing alternatives.

For the remainder of this section we concentrate on the simple case of full support, no indifference and complete data. We identify the behavioral conditions which characterize the resulting model. We also consider a special case of the model in which there is consistency in the distribution of search orders between choice sets. This additional restriction ensures that our model is behaviorally equivalent to a subset of the class of random utility models (RUMs). In section 4 we discuss extensions in which we drop the full support, no indifference, and complete data conditions.

3.2 Full Support Satisficing Models

Our model adds to the GSM the assumption that, in each choice set, any item will be searched first with some positive probability.

Assumption 1 (Full Support) *For any $a \in A$ and all $A \in \mathcal{D}$: $\gamma_A(r_A \in R_A : a\ r_A\ b \quad \forall b \in A \backslash \{a\}) > 0$.*

We describe a GSM which additionally satisfies this assumption as a *Full Support Satisficing Model (FSSM)*

Definition 3 (Full Support Satisficing Model (FSSM)) *A data set $(\mathcal{D}, p)$ has a Full Support Satisficing Model (FSSM) representation if it has a GSM representation in which the stochastic search order satisfies Full Support.*

The assumption of Full Support has an important implication: we can identify above-reservation alternatives as those which are always chosen with positive probability in any choice set in which they appear. This is because Full Support implies that, for each such alternative, a search order in which it is searched first occurs with positive probability in each choice set, ensuring that it will be chosen. Furthermore, any alternatives that are *not* above reservation utility will be chosen with zero probability in any choice set which contains an above reservation utility alternative.

We define the set of alternatives which are always chosen:

Definition 4 (Always Chosen Set) *For any data set $(\mathcal{D}, p)$, we define the* ***always chosen*** *set as $W^* = \{a \in X | p(a, A) > 0$ for all $A \in \mathcal{D}$ such that $a \in A\}$.*

For any complete data set generated by a FSSM, W^* must be equivalent to the set of above-reservation alternatives.

Lemma 1 *Assume a complete data set $(\mathcal{D}, p)$ admits an FSSM representation. Then, for any such representation $W^* = U^*$.*

All subsequent proofs are relegated to the appendix.

As we discuss in section 4.3, if the data set is not complete then W^* may be a strict superset of U^*: a below satisficing alternative may always be chosen because it is only observed in choice sets containing other below satisficing alternatives.[7]

Using the (observable) set W^*, we can define the first of two behavioral conditions which characterize the FSSM. It states that stochastic choice must only occur amongst elements of W^*. This follows from the fact that stochasticity in the satisficing model occurs only from stochasticity in search order.

Axiom 1 (Deterministic no satisficing choice) *If $a \in X \backslash W^*$ then for all $A \in \mathcal{D}$ either $p(a, A) = 0$ or $p(a, A) = 1$.*

The second condition ensures that the preference information revealed by a data set is well behaved. In order to state the condition, we introduce the following definitions.

Definition 5 (Stochastic Revealed Preference) *Define $C(A) = \{a \in A | p(a, A) > 0\}$. a is **stochastically revealed directly preferred** to b if, for some $A \in \mathcal{D}$ $a, b \in A$ and $a \in C(A)$. a is **stochastically revealed preferred** to b if $\{a, b\}$ is in the transitive closure of the stochastically revealed directly preferred relation. a is **stochastically strictly revealed preferred** to b if, for some $A \in \mathcal{D}$, $a \in C(A)$ and $b \notin C(A)$.*

Notice that, for data generated by a FSSM, these revealed preference concepts align with the underlying utility function except in the case of two alternatives above that satisficing level. Such objects will be revealed indifferent to each other, yet may in fact be ranked according to the utility function. It is a defining feature of this version of the satisficing model that utility differences above the threshold u^* are unimportant for behavior. Nevertheless, the FSSM implies that the stochastic revealed preference information must obey the Strong Axiom of Revealed Preference.

Axiom 2 (SARP) *$C(A)$ must obey SARP: if a is stochastically revealed preferred to b then b must not be stochastically strictly revealed preferred to a.*

Our first result is that axioms Axiom 1 and Axiom 2 are necessary and sufficient for a data set to have a FSSM representation

Theorem 1 *The following are equivalent:*

[7] Completeness can be replaced for a weaker condition on the richness of the data set which requires observing choices from all two and three element sets.

1. *A stochastic choice dataset $(\mathcal{D}, p)$ has an FSSM representation.*

2. *A stochastic choice dataset $(\mathcal{D}, p)$ satisfies Axiom 1 and Axiom 2.*

To understand the sufficiency of the two axioms - Axiom 1 and Axiom 2 -, note first that SARP allows us, through Afriat/Richter's theorem (1966), to construct a utility function which represents the stochastic revealed preference relation. Moreover, the elements of W^* will be maximal according to that utility representation, allowing for a u^* such that all elements of the always chosen set can be assigned a utility greater or equal than u^*; while all the elements that are not always chosen are assigned an utility level below u^*. Axiom 1 guarantees deterministic choice in sets which contain at most one above-reservation alternatives, and SARP again ensures that such choices are utility maximizing. For all other choice sets, Axiom 1 ensures that alternatives with utility below u^* (and so outside W^*) are not chosen, and a suitable stochastic search order can be constructed from the random choice rules to explain the pattern of choice amongst above satisficing alternatives.

Notice that, while Lemma 1 relies on the completeness of the data set, Theorem 1 does not. The behavioral content of the FSSM model is the same regardless of whether the data set is complete. However, the degree to which elements of the representation can be identified will be reduced in incomplete data sets, as we discuss in section 4.3.

The following examples illustrate the empirical content of the FSSM by presenting data sets which violate each of our axioms.

Example 2 (Violation of Axiom 1) *Let $X = \{a, b, c\}$, and let $p(a, \{a, b\}) = 1$, $p(b, \{b, c\}) = \frac{1}{2}$, $p(a, \{a, c\}) = 1$ and $p(a, \{a, b, c\}) = 1$. This does not satisfy Axiom 1 since $W^* = \{a\}$, but $p(b, \{b, c\}) = \frac{1}{2} \notin \{0, 1\}$. This behavior is incommensurate with the FSSM because the fact that b is chosen probabilistically from $\{b, c\}$ indicates that it must be above the satisficing level, yet this means that it should be chosen some of the time from $\{a, b, c\}$ due to the full support assumption.*

Example 3 (Violation of Axiom 2) *Let $X = \{a, b, c\}$, and let $p(a, \{a, b\}) = 1$, $p(b, \{b, c\}) = 1$, $p(a, \{a, c\}) = 0$ and $p(a, \{a, b, c\}) = 1$. This does not satisfy Axiom 2 since $p(a, \{a, b\}) = 1$ means that a is stochastically strictly revealed preferred to b, $p(b, \{b, c\}) = 1$ means that b is stochastically revealed preferred to c, while $p(c, \{a, c\}) = 1$ means that c is stochastically strictly revealed preferred to a. Such behavior is also incommensurate with the FSSM as, in each case, the uniquely chosen object must have a utility strictly higher than* *those which are not chosen, either because all are below the satisficing level, in which case the best option is chosen, or because only the chosen object is above the satisficing level.*

Theorem 1 shows the extent to which the FSSM can be tested and differentiated from other models. First, note that any data set in which $p(a, A) > 0$ for all $A \in \mathcal{D}$ will trivially satisfy both Axiom 1 and Axiom 2, and so admit an FSSM representation. This is because any data set in which the random choice rule has full support in every choice set can be rationalized by an FSSM in which every alternative is above the satisficing level, and the resulting pattern of choice is driven by the choice-set specific distribution over search orders. Second, note that the standard model of utility maximization (without indifference) is a limit case of the FSSM in which $|W^*| = 1$. Third, notice that an alternative interpretation of the FSSM is a model in which attention is complete and choices are governed by a preference relation which has indifference only amongst maximal elements. A model in which such indifference is resolved using a random tie breaking rule with full support amongst maximal elements is equivalent to the FSSM.

3.2.1 Recoverability in the FSSM

In the case of a complete data set which satisfies Axiom 1 and Axiom 2, many of the elements of the FSSM can be uniquely identified.

Theorem 2 *Let $(\mathcal{D}, p)$ be a complete data generated by an FSSM $(u, u^*, \{\gamma_A\}_{A\in\mathcal{D}})$. For any FSSM representation of the data $(\bar{u}, \bar{u}^*, \{\bar{\gamma}_A\}_{A\in\mathcal{D}})$*

1. $U^* = \bar{U}^*$
2. *For all* $a, b \notin U^*$, $u(a) > u(b) \Rightarrow \bar{u}(a) > \bar{u}(b)$
3. $\gamma_A(ar_Ab \quad \forall b \in \{A \cap U^*\} \setminus \{a\}) = \bar{\gamma}_A(ar_Ab \quad \forall b \in \{A \cap U^*\} \setminus \{a\})$ *for all* $A \in D$, $a \in A \cap U^*$

Theorem 2 tells us that, in a complete data set we can uniquely identify the above-satisficing elements, the preference ordering over non-satisficing elements, and the probability that one satisficing element will be seen before another in any choice set.

3.3 Fixed Distribution Satisficing Models

So far, we have allowed stochastic search order to vary arbitrarily between choice sets: an alternative that is likely to be searched first in choice set A may be very unlikely to be searched first in choice set B. However, in some cases such an assumption may be inappropriate. For example, consider the case in which the probability of search is governed by the 'salience' of different alternatives: a book with a bright pink cover may be more likely to be looked at before one with a dark brown cover regardless of the set of available alternatives.[8]

We now consider the implications of a satisficing model with full support in which the probability distribution over search orders is invariant to the set of available alternatives. We call this the 'fixed distribution' property.

Assumption 2 (Fixed Distribution) *There exists a $\Gamma_X : R_X \to [0,1]$ such that, for every $A \in \mathcal{D}$ and $r_A \in R_A$*

$$\gamma_A(r_A) = \Gamma_X(r_X | r_A \subset r_X)$$

For every choice set A, it is as if the DM draws a search order from a distribution Γ_X over linear orders on the grand set of alternatives X. They then follow that search order, ignoring any alternatives that are not in fact available in A.

Definition 6 (Fixed Distribution Satisficing Model (FDSM)) *A data set $(\mathcal{D}, p)$ has a Fixed Distribution Satisficing Model (FDSM) representation if it has a FSSM representation in which the family of stochastic search orders $\{\gamma_A\}_{A\in\mathcal{D}}$ satisfy Fixed Distribution.*

The conditions Axiom 1 and Axiom 2 are necessary for FDSM but not sufficient, implying that the FDSM is a strict subcase of FSSM. In order to obtain sufficiency, we make use of the Total Monotonicity condition of Block & Marschak (1960). Total Monotonicity by itself is a sufficient and necessary condition for Random Utility Maximization in our environment. This implies that the FDSM is the exact intersection of the FDSM model with the Random Utility model of Block-Marschak and Falmagne (1978).

In order to define the total monotonicity condition, we first need to define the following function for each $A \in \mathcal{D}$ and $a \in A$:

$$f(a, A) = \sum_{D \in \mathbf{B}(A)} (-1)^{|D \setminus A|} p(a, D)$$

[8] It is of course possible to think of cases in which such a property might not hold. For example, it could be that a brown book would be more salient than a pink book if all other books are pink, but not otherwise. In such cases the model presented in this section would not be appropriate.

where $\mathbf{B}(A)$ is the class of supersets of A (i.e., $\mathbf{B}(A) \equiv \{D \in \mathcal{D} | A \subseteq D\}$).

Block & Marschak (1960) and Falmagne (1978) proved that the following behavioral axiom called Total Monotonicity (or Block-Marschak Monotonicity) is necessary and sufficient for a RUM representation:

Axiom 3 (Total Monotonicity) *$f(a, A) \geq 0$ for all $a \in X$, for all $A \in \mathcal{D}$.*

Note that Total Monotonicity implies 'standard' monotonicity: the probability of choosing any given alternative falls as more alternatives are added to the choice set - that is $p(a, A) \geq p(a, B)$ when $A \subseteq B$.[9] However, it is also stronger than this condition as we can see in the following example: Set $X = \{a, b, c, d\}$, let $p(a, \{a, b\}) = 0.2$, and $p(a, \{a, b, c\}) = p(a, \{a, b, d\}) = 0.19$ and $p(a, \{a, b, c, d\}) = 0.17$. We check $f(a, \{a, b\}) = p(a, \{a, b\}) + p(a, \{a, b, c, d\}) - [p(a, \{a, b, c\}) + p(a, \{a, b, d\})]$ and observe that $f(a, \{a, b\}) = -0.01$ negative and violating Total Monotonicity. However, standard monotonicity holds in this example.

Clearly, a FDSM cannot lead to a failure of standard monotonicity. If it did that would mean that a given satisficing item is more likely to be found first in a bigger menu than in a smaller one, which is not consistent with the idea that the probability of any search order is fixed across menus. The higher order monotonicity conditions implied by total monotonicity can be interpreted as saying that the likelihood of a satisficing item being found first decreases with the size of the menu, but the marginal effect of adding a new option to the menu decreases with its size. In the example above $f(a, \{a, b\}) \geq 0$ means that the impact of adding one additional item c in the menu $\{a, b\}$ on the probability of choosing a – i.e. $p(a, \{a, b\}) - p(a, \{a, b, c\})$ – is bigger that the impact of adding the same item c in the bigger menu $\{a, b, d\}$ on the probability of choosing a – i.e. $p(a, \{a, b, d\}) - p(a, \{a, b, c, d\})$).

We are ready to state the main result of this section.

Theorem 3 *The following are equivalent:*

1. *A complete stochastic choice dataset $(\mathcal{D}, p)$ has an FDSM representation.*
2. *A complete stochastic choice dataset $(\mathcal{D}, p)$ satisfies Axiom 1, Axiom 2 and Total Monotonicity (Axiom 3).*

Note that, unlike Theorem 1, Theorem 3 requires a complete data set.

[9] To see this take the Mobius inverse representation of $p(a, A) = \sum_{D \in \mathbf{B}(A)} f(a, D)$ and $p(a, B) = \sum_{D' \in \mathbf{B}(B)} f(a, D')$ and note that if $D \in \mathbf{B}(B)$ then $D \in \mathbf{B}(A)$ and $f(a, D) \geq 0$ by total monotonicity, then we have that $p(a, A) \geq p(a, B)$.

3.3.1 Recoverability in the FDSM

In the case of a complete data set which satisfies Axiom 1, Axiom 2 and Total Monotonicity (Axiom 3) several of the elements of the FDSM can be identified. In particular, the identification of the search orderings is improved upon the FSSM recoverability.

Theorem 4 *Let $(\mathcal{D}, p)$ be a complete data generated by an FDSM (u, u^*, Γ_X) such that $X \setminus U^* \neq \emptyset$. For any FDSM representation of the data $(\bar{u}, \bar{u}^*, \overline{\Gamma}_X)$*

1. $U^* = \bar{U}^*$
2. *For all $a, b \notin U^*$, $u(a) > u(b) \Rightarrow \bar{u}(a) > \bar{u}(b)$*
3. $\Gamma_X(ar_X b) = \overline{\Gamma}_X(ar_X b)$ *all $A \in D$ and $a, b \in U^*$*

Theorem 4 tells us that, in a complete data set, we can once again identify the above-satisficing elements, and the preference ordering over the alternatives that are surely non-satisficing elements. We can also identify the probability that one revealed satisficing element will be seen before another in any choice set, for those elements that are satisficing.

3.4 Comparative Statics

In this section we study the comparative statics with respect to the primitives of the model.

We first study the behavioral implications of a change in the utility threshold. Intuitively, a lower utility threshold is associated with a decision maker who is satisfied 'more often'. In our model, this implies a larger 'always chosen' set. In order to make the comparison clear, we consider only DMs who otherwise exhibit the same revealed preference information and same search behavior.

Definition 7 (More satisficing than) *Let p and $\tilde{p}$ be two random choice rules on $\mathcal{D} \subseteq 2^X \setminus \emptyset$ We say that p is **more easily satisfied than** $\tilde{p}$ if*

1. $\tilde{W}^* \subseteq W^*$
2. $C_p(A) = C_{\tilde{p}}(A)$ *for all A such that $A \cap W^* = \emptyset$*
3. *If $a \in W^*$ and b is revealed preferred to a according to $\tilde{p}$, then $b \in W^*$*
4. *For any choice set A and $a \in \tilde{W}^* \cap W^* \cap A$, $p(a, A) \leq \tilde{p}(a, A)$*

*We say that DM p is **strictly more easily satisfied than** $\tilde{p}$ if (i) p is more easily satisfied than $\tilde{p}$ and (ii) $W^* \setminus \tilde{W}^* \neq \emptyset$*

We now show that if a random choice rule p is more easily satisfied than $\tilde{p}$ then there exists representations for them with the same utility functions and distribution over search orders but only differ on the utility threshold. Formally,

Claim 1 *Let p and $\tilde{p}$ be two random choice rules on $\mathcal{D} \subseteq 2^X \setminus \emptyset$ which admit an FSSM representation, and such that p is more easily satisfied than $\tilde{p}$. Then there exists a representation $(u, u^*, \{\gamma_A\}_{A\in\mathcal{D}})$ of p and $(\tilde{u}, \tilde{u}^*, \{\tilde{\gamma_A}\}_{A\in\mathcal{D}})$ of $\tilde{p}$ such that:*

1. $u = \tilde{u}$
2. $u^* \leq \tilde{u}^*$
3. $\{\gamma_A\}_{A\in\mathcal{D}} = \{\tilde{\gamma_A}\}_{A\in\mathcal{D}}$

If p is strictly more easily satisfied than $\tilde{p}$ then the inequality in part 2 is strict.

One would expect that a DM who is more easily satisfied does 'worse' in the sense that they would choose worse options. In order to formalize this intuition, we first define the probability distribution over utility realizations.

Definition 8 (Utility Distribution) *Let p be a random choice rule on $\mathcal{D} \subseteq 2^X \setminus \emptyset$ that admits a FSSM (FDSM) representation $(u, u^*, \{\gamma_A\}_{A\in\mathcal{D}})$. We define the utility distribution for menu A as the distribution over utility levels implied by the model*

$$p_u(\tilde{u}, A) = \sum_{a\in A} \mathbf{1}\{u(a) = \tilde{u}\} p(a, A)$$

The associated cumulative distribution function

$$F_{u|p}(\overline{u}, A) = \sum_{u\leq\overline{u}} p_u(u, A)$$

We now show that if a random choice rule p is more easily satisfied than $\tilde{p}$, and these admit a FSSM representation, then the random choice rule p would lead to worse decisions in terms of the implied utility distributions. To understand this, consider two FSSM representations of these random choice rules $\{u, u^*, \{\gamma_A\}_{A\in\mathcal{D}}\}$ and $\{u, \tilde{u}^*, \{\gamma_A\}_{A\in\mathcal{D}}\}$. From Claim 1, we have that $u^* \leq \tilde{u}^*$ which in turn implies that, for any $A \in \mathcal{D}$, $A \cap U^* \supseteq A \cap \tilde{U}^*$. Therefore the random choice rule with the lower utility threshold is randomizing over alternatives with lower utility levels. The following example illustrates how this mechanism works.

Example 4 *Let* $A = \{a, b, c\}$ *and for simplicity assume that* $u(a) = 1$*,* $u(b) = 2$ *and* $u(c) = 3$*. Let* $u^* = 1.5$ *and* $\tilde{u}^* = 2.5$*, and consider the search orders from Example 1. For the random choice rule* $\tilde{p}$*,* $\tilde{p}(c, A) = 1$ *and* $\tilde{p}(a, A) = \tilde{p}(b, A) = 0$*,*

$$F_{u|\tilde{p}}(u, A) = \begin{cases} 0 & \text{if } u < 3 \\ 1 & \text{if } u \geq 3 \end{cases}$$

For the random choice rule p*,* $p(a, A) = 0$*,* $p(b, A) = \frac{11}{24}$ *and* $p(b, A) = \frac{13}{24}$*, therefore*

$$F_{u|p}(u, A) = \begin{cases} 0 & \text{if } u < 2 \\ \frac{11}{24} & \text{if } 2 \leq u < 3 \\ 1 & \text{if } u \geq 3 \end{cases}$$

.

This result is formalized in Proposition 1.

Proposition 1 (Utility Threshold) *Let* p *and* $\tilde{p}$ *be two random choice rules on* $\mathcal{D} \subseteq 2^X \setminus \emptyset$ *which admit a FSSM representation,* $(u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$ *and* $(\tilde{u}, \tilde{u}^*, \{\tilde{\gamma}_A\}_{A \in \mathcal{D}})$*. Moreover, assume that* p *is more easily satisfied than* $\tilde{p}$ *and that we consider representations such that* $u = \tilde{u}$ *and* $\{\gamma_A\}_{A \in \mathcal{D}} = \{\tilde{\gamma}_A\}_{A \in \mathcal{D}}$*. Then, for any* $A \in \mathcal{D}$ *and* $\bar{u} \in \mathbb{R}$*,* $F_{u|p}(\overline{u}, A) \geq F_{u|\tilde{p}}(\overline{u}, A)$*. Moreover, if* p *is strictly more easily satisfied than* $\tilde{p}$ *then there exists an* $A \in \mathcal{D}$ *and* $\bar{u} \in \mathbb{R}$ *such that* $F_{u|p}(\overline{u}, A) > F_{u|\tilde{p}}(\overline{u}, A)$

Alternatively, we can study the effect of changes in the search orders. Intuitively, if the relative probability of search orders changes in such a way that search orders congruent with the preference ordering become more likely, then the choices of the DM improve. The following example illustrates how this mechanism works.

Example 5 *Consider again example 1 with* $u(a) > u(b) > u^* > u(c)$ *and search orders* γ_A *and* $\tilde{\gamma}_A$ *as shown in Table 2; where* $\tilde{\gamma}_A(\{a, b, c\}) = \gamma_A(\{a, b, c\}) + \frac{1}{12}$ *and* $\tilde{\gamma}_A(\{b, a, c\}) = \gamma_A(\{b, a, c\}) - \frac{1}{12}$*. That is,* $\tilde{\gamma}_A(\cdot)$ *induces the same probability distribution over search orders except of orders* $\{a, b, c\}$ *and* $\{b, a, c\}$*. The only effect of this change is to alter the relative probability of* a *and* b *being chosen from the set. Note that the random choice rule* p *induced by search orders* γ_A *is given by* $p(a, A) = \frac{3}{8}$*,* $p(b, A) = \frac{5}{8}$ *and* $p(c, A) = 0$*; while the random choice rule* $\tilde{p}$ *induced by search orders* $\tilde{\gamma}_A$ *is given by* $\tilde{p}(a, A) = \frac{11}{24}$*,* $\tilde{p}(b, A) = \frac{13}{24}$ *and* $\tilde{p}(c, A) = 0$*. For simplicity assume that* $u(a) = 3$ *and* $u(b) = 2$*. Correspondingly, the cumulative distribution*

Order	(1)	(2)	(3)	(4)	(5)	(6)
1st	a	a	b	b	c	c
2nd	b	c	a	c	a	b
3rd	c	b	c	a	b	a
γ_A	$\mathbf{\frac{1}{12}}$	$\frac{1}{6}$	$\mathbf{\frac{1}{3}}$	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{1}{4}$
$\tilde{\gamma}_A$	$\mathbf{\frac{1}{6}}$	$\frac{1}{6}$	$\mathbf{\frac{1}{4}}$	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{1}{4}$

Table 2: Changes in the probability of search orders and its effect on DM's well-being.

functions over utility values are given by

$$F_{u|p}(u, A) = \begin{cases} 0 & \text{if } u < 2 \\ \frac{5}{8} & \text{if } 2 \leq u < 3 \\ 1 & \text{if } u \geq 3 \end{cases}$$

$$F_{u|\tilde{p}}(u, A) = \begin{cases} 0 & \text{if } u < 2 \\ \frac{13}{24} & \text{if } 2 \leq u < 3 \\ 1 & \text{if } u \geq 3 \end{cases}$$

The change in distribution of search orders from γ_A to $\tilde{\gamma}_A$ makes it more likely that search occurs in the same order as the DMs preferences, and as a result, there is an improvement in the distribution over realized utility levels.

In Appendix A.8 we provide a formal definition of changes in the search order distribution that improve congruency, and show that they are indeed related to an improvement in the distribution of utility for the DM.

4 Extensions

In this section we extend our model by relaxing in turn the assumptions of full support, no indifference, and complete data

4.1 Fixed Distributions Without Full Support

As discussed in section 3.1, the GSM model is vacuous without the full support assumption. Here we consider the empirical implication of dropping full support but maintaining the fixed distribution assumption. In such a case the identification of satisficing elements as those that are always chosen breaks down. A satisficing element a may not be chosen in some sets if it is always searched after another satisficing element b. To illustrate this point consider the following example.

Example 6 *Let $X = \{a, b, c, d\}$, $U^* = \{a, b\}$ and Γ a distribution over search orders on X with full support and where each possible search order is equally likely; moreover assume that $u(c) > u(d)$. Then, for any menu such that $U^* \subseteq A$, $p(a, A) = p(b, A) = \frac{1}{2}$, and if $A \cap U^* \neq \emptyset$, then $p(U^*, A) = 1$ and $p(c, A) = p(d, A) = 0$. Finally, $p(c, \{c, d\}) = 1$ and $p(d, \{c, d\}) = 0$. Note that this is the standard case describe in section 3.3. Now, notice that since we do not assume that Γ needs to have full support on the set of search orders on X, the same data set can be generated by the following fixed distribution satisficing model without full support: $\bar{U}^* = X$, $\Gamma((a, b, c, d)) = \Gamma((b, a, c, d)) = \frac{1}{2}$ and $\Gamma(r_X) = 0$ for all r_X linear order on X, such that $r_X \notin \{(a, b, c, d), (b, a, c, d)\}$. Furthermore, this alternative representation is not unique.*

Because it is not possible to identify the satisficing alternatives the only implication of the satisficing model without full support, but with fixed distribution is Total Monotonicity - in other words it is behaviorally indistinguishable from the Random Utility model. This can be seen by noting that a RUM can be reinterpreted as a satisficing model with fixed distribution by assuming that all alternatives are above the reservation level, and treating the preference orderings from the random utility model as search orders in the satisficing model.

Theorem 5 *The following are equivalent:*

1. *A complete stochastic choice dataset $(\mathcal{D}, p)$ is generated by a FDSM without Full Support.*

2. *A complete stochastic choice dataset $(\mathcal{D}, p)$ satisfies Axiom 3.*

4.2 Allowing for Indifference

Here we relax the no indifference assumption while keeping the Full Support conditions. Allowing for indifference potentially introduces stochasticity among non-satisficing alternatives due to the DM's rule to break ties. We assume that tie breaking works as follows, if the DM is indifferent between two or more alternatives, and needs to choose one of them, she chooses at random from the set of indifferent alternatives with probabilities induced by the tie-breaking rule T.

Definition 9 (Tie-breaking rule) *Let $T : X \times \mathcal{D} \to \mathbb{R}_{++}$[10] be a function that assigns tie breaking weights to alternatives. In case of indifference between two or more alternatives in menu A, the DM applies the induced tie breaking rule as follows:*

$$T\left(a|A^{\sim_a}\right) = \frac{T(a, A)}{\sum_{b \in A^{\sim_a}} T(b, A)} \tag{2}$$

where $T\left(a|A^{\sim_a}\right) > 0$ is the always positive probability that a is chosen when a is indifferent to all the elements in the set $A^{\sim_a} \equiv \{b \in A : u(b) = u(a)\}$, and superior to all other elements in A (i.e. $u(a) \geq u(b)$ for all $b \in A$).

Note that if $|A^{\sim_a}| = 1$ then $T\left(a|A^{\sim_a}\right) = 1$, and that $\sum_{b \in A^{\sim_a}} T\left(a|A^{\sim_a}\right) = 1$ in general.

We now extend the Full Support Satisficing Model to allow for indifference.

Definition 10 (Full Support Satisficing Model with Indifferences (FSSMI)) *A data set $(\mathcal{D}, p)$ has a Full Support Satisficing Model with Indifferences (FSSMI) representation if there exists $u : X \to \mathbb{R}$, $u^* \in \mathbb{R}$ such that $u^* \leq \max_{a \in X} u(a)$, stochastic search orders $\{\gamma_A\}_{A \in \mathcal{D}}$ that satisfies Full Support and tie breaking rule $T : X \times \mathcal{D} \to \mathbb{R}_{++}$, such that, for any $a \in A$*

$$p(a, A) = \begin{cases} \gamma_A\left(a r_A b \ \forall \ b \ s.t. \ u(b) > u^*\right) & \text{if } u(a) \geq u^* \\ T\left(a|A^{\sim_a}\right) & \text{if } a \in \arg\max_{x \in A} u(x) < u^* \\ 0 & \text{Otherwise} \end{cases} \tag{3}$$

The following example illustrates the FSSMI.

Example 7 *Consider again example 1, and assume that DM's choices can be represented by a FSSMI. Let $U^* = \{a\}$, and let $u(b) = u(c)$. Then, $p(a, A) = 1$, $p(b, A) = p(c, A) = 0$ for all A such that $a \in A$. Let the tie breaking rule be generated by $T(b, \{\{b, c\}) = 1$ and $T(c, \{b, c\}) = 2$, then $p(b, \{b, c\}) = \frac{1}{3}$, and $p(c, \{b, c\}) = \frac{2}{3}$*

Axiom 1 is no longer necessary for the FSSMI model: stochasticity can occur amongst alternatives that are not always chosen due to indifference. In fact, it turns out that the

[10] Note that we rule out deterministic tie breaking since, the behavior of a DM that is indifferent between two alternatives a and b and always chooses a over b is behaviorally indistinguishable from a DM that prefers a over b.

behavioral implication of allowing for indifference is precisely the removal of this axiom from our set of necessary and sufficient conditions.

Theorem 6 *The following are equivalent:*

1. *A complete stochastic choice dataset* $(\mathcal{D}, p)$ *is generated by a FSSMI.*

2. *A complete stochastic choice dataset* $(\mathcal{D}, p)$ *satisfies Axiom 2.*

Theorem 6 highlights that the satisficing model *without* indifference can be reinterpreted as a standard optimizing model with random tie breaking, but allowing for indifference only amongst the best alternatives.

4.2.1 Recoverability in the FSSMI

Given $u^* \leq \max_{a\in X} u(a)$, the extension of the model to allow for ties does not obscure the identification result for the satisficing set as in definition 4, where the always chosen set coincides with the satisficing set, i.e. $W^* = U^*$. The following theorem describes the degree to which the other elements of the model can be identified.

Theorem 7 *Let* $(\mathcal{D}, p)$ *be a complete data generated by an FSSMI* $(u, u^*, \{\gamma_A\}_{A\in\mathcal{D}}, T)$*. For any FSSMI representation of the data* $(\bar{u}, \bar{u}^*, \{\overline{\gamma}_A\}_{A\in\mathcal{D}}, \bar{T})$

1. $U^* = \bar{U}^*$

2. *For all* $a, b \notin U^*$, $u(a) \geq u(b) \Rightarrow \bar{u}(a) \geq \bar{u}(b)$

3. $\gamma_A(ar_A b \quad \forall b \in \{A \cap U^*\} \setminus \{a\}) = \bar{\gamma}_A(ar_A b \quad \forall b \in \{A \cap U^*\} \setminus \{a\})$ *for all* $A \in D$, $a \in A \cap U^*$

4. $T(a|A^{\sim a}) = \bar{T}(a|A^{\sim a})$ *for all* $a \in A$, $A \in \mathcal{D}$.

Theorem 7 tells us that in a complete data set we can uniquely identify the above-satisficing elements, the preference ordering among non-satisficing elements, the tie breaking rule when used and the probability that one satisficing element is seen before another in any choice set. Its proof follows from Theorem 2 when the identification of the tie breaking rule is established. The identification of the tie breaking rule holds because whenever it is used, it is calibrated from the empirical choice probability among elements that are not revealed to be satisficing.

4.3 Incomplete Datasets

Here we relax the complete data set assumption while keeping the full support distribution condition and assuming no indifference. We do not work with the fixed distribution assumption since Total Monotonicity is not well defined for incomplete data sets, and the literature on Random Utility models has not dealt with this extension.

Notice that complete data is not a necessary assumption for Theorem 1. Thus, if we drop completeness, the implications of the model are not affected, but identification becomes weaker. To see this note W^* may be a strict superset of U^* since a below-satisficing alternative may be always chosen because it is only observed in choice sets containing below-satisficing alternatives. The accuracy with which we can identify the primitives of the model given observed data depends on the richness of the data set.

Theorem 8 *Let $(\mathcal{D}, p)$ be a data set (that needs not to be complete) that satisfies Axiom 1 and Axiom 2. Then for a FSSM $(u, u^*, \{\gamma_A\}_{A\in\mathcal{D}})$ represents the data if and only if*

1. $\widetilde{W} \subseteq U^* \subseteq W^*$
2. *u must represent the stochastic revealed preference relation on $X \setminus U^*$: that is if a is stochastically strictly revealed preferred to b then $u(a) > u(b)$, and if a is revealed preferred to b then $u(a) \geq u(b)$ for any $a, b \in X \setminus U^*$*
3. $\gamma_A(a\ r_A\ b \quad \forall b \in (U^* \cap A) \setminus \{a\}) = p(a, A)$ *for all $A \in D$, $a \in U^* \cap A$.*

where $\widetilde{W} \equiv \{a \in X | \exists A \in \mathcal{D},\ s.t\ p(a, A) \in (0, 1)\}$.

Theorem 8 tells us that, when dealing with incomplete data sets one can only identify with certainty satisficing choices, if these have been observed chosen when other satisficing alternatives were available as well; that is, if we see them being chosen stochastically. As we established before, the satisficing set is a subset of the set of always chosen alternatives, but with incomplete data it may be a strict subset. Furthermore, one can only partially recover the preference order for the revealed not satisficing alternatives. Finally, the search orders can only be identify up to the set of those that coincides with the one generated from the relative probabilities of the elements that surely are in U^*.

4.4 Random Utility and Random Threshold

Now we turn to the study of two variants of the satisficing model where we allow randomness in tastes and in the threshold rule. The first generalizes the constant threshold assumption and allows it to vary randomly. The second variant focuses on letting utility, so far taken as fixed, to vary randomly (i.e., random utility).

4.4.1 Random Threshold

Here we explore the satisficing model when we allow for a random threshold. Specifically, we consider a model in which, at each decision, the DM draws a search order and a threshold independently from two distributions. She then searches the menu until finding something with utility above the threshold, if such an object exists in the menu.

We establish the somewhat surprising result that assuming a random threshold adds no generality to the satisficing model. In fact, these variants are indistinguishable from the FSSM/FDSM with constant threshold.

Definition 11 (FSSM with Random Threshold (FSSM-RT)) *A data set $(\mathcal{D}, p)$ has a Full Support Satisficing Model with Random Threshold (FSSM-RT) if there exists an injective $u : X \to \mathbb{R}$, a continuous random variable $u^* \sim F_{u^*}(\cdot)$ such that $\tau(a) = Pr(u(a) > u^*) = 1 - F_{u^*}(u(a))$ with support$(u^*) \subseteq \mathbb{R}$ such that $U^{*,RT} = \{x \in X : \tau(x) > 0\} \geq 1$, and $\{\gamma_A\}_{A \in \mathcal{D}}$ with the Full Support property, such that, for any $A \in \mathcal{D}$ and $a \in A$:*

$$p(a, A) = \sum_{r_A \in R_A} \tau(a) \prod_{b \in A: b r_A a} (1 - \tau(b)) \gamma_A(r_A) + \prod_{c \in A} (1 - \tau(c)) \mathbf{1}(u(a) > u(b) \; \forall b \in A \backslash \{a\}). \quad (4)$$

Notice that the term $\tau(a) \prod_{b \in A: b r_A a} (1 - \tau(b))$ measures the probability of item $a \in A$ being satisficing given that all items searched before it in menu A under fixed search r_A where not satisficing, similarly $\prod_{c \in A} (1 - \tau(c))$ is the probability of not finding anything satisficing in A,[11] and $\mathbf{1}(u(a) > u(b) \; \forall b \in A) = 1$ when a is maximal under u in A and zero otherwise. For notational convenience, we assume also that the productory over the empty set is 1 - i.e. $\prod_{\emptyset} (1 - \tau(b)) = 1$.

We first study the implications of a random threshold with the Full Support assumption on the search process.

Lemma 2 *A complete stochastic choice dataset $(\mathcal{D}, p)$ that can be generated by a FSSM-RT satisfies Deterministic no satisficing choice (Axiom 1) and SARP (Axiom 2).*

[11] Notice that the summation of the first term of the expression over all items $P(A) = \sum_{c \in A} \sum_{r_A \in R_A} \tau(c) \prod_{b \in A: b r_A c} (1 - \tau(b)) \gamma_A(r_A) = 1 - \prod_{c \in A} (1 - \tau(c))$, thus the $1 - P(A) = \prod_{c \in A} (1 - \tau(c))$ is the residual probability.

The previous lemma allows us to establish the following equivalence result.

Theorem 9 *The following statements are equivalent:*

1. *A complete stochastic choice dataset $(\mathcal{D}, p)$ can be generated by a FSSM.*

2. *A complete stochastic choice dataset $(\mathcal{D}, p)$ can be generated by a FSSM-RT.*

The equivalence theorem just presented may be surprising, however we note that even when the FSSM-RT is more flexible than the FSSM, it imposes exactly the same restrictions in the primitive stochastic choice dataset. The reason is that the FSSM-RT and FSSM have at their heart the separation of stochastic choice for satisficing items and deterministic rational choice for non satisficing items, this remains true under an extended definition of the satisficing set $U^{RT*} = \{x \in X : \tau(x) > 0\}$.

Definition 12 (FDSM with Random Threshold (FDSM-RT)) *A data set $(\mathcal{D}, p)$ has a Fixed Distribution Satisficing Model with Random Threshold (FDSM-RT) representation if it has a FSSM-RT representation in which the family of stochastic search orders $\{\gamma_A\}_{A\in\mathcal{D}}$ satisfy the Fixed Distribution property.*

Lemma 3 *A complete stochastic choice dataset $(\mathcal{D}, p)$ that can be generated by a FDSM-RT satisfies Deterministic no satisficing choice (Axiom 1), SARP (Axiom 2) and Total Monotonicity (Axiom 3).*

The previous lemma allows us to establish the empirical equivalence between FDSM with and without RT.

Theorem 10 *The following statements are equivalent:*

1. *A complete stochastic choice dataset $(\mathcal{D}, p)$ can be generated by a FDSM.*

2. *A complete stochastic choice dataset $(\mathcal{D}, p)$ can be generated by a FDSM-RT.*

4.4.2 Random Utility FDSM

Here we focus on a variant of the FDSM where we allow for random utility.[12] We consider a model where there is a probability measure $\rho : \mathcal{U} \mapsto [0, 1]$ defined over a finite domain of injective utilities defined on X denoted by $\mathcal{U}$ (we assume that $\mathcal{U}$ is such that there is a bijection from it to the set of linear orders in X). For a given element in the support

[12] We omit the variant of FSSM with random utility as it is too non-restrictive, in fact, the reader can easily check using the results in this section that the necessary and sufficient condition that characterizes FSSM with random utility is the Degeneracy condition (A 4).

$u \in \mathcal{U}$, we assume that the DM is consistent with a FDSM representation or p_u^{FDSM} with $\{u, u^*, \{\gamma_A\}_{A \in \mathcal{D}}\}$ (i.e. u^*, and $\{\gamma_A\}_{A \in \mathcal{D}}$ the same for each $u \in \mathcal{U}$). We assume that the distribution of utility is independent of the distribution of linear search orders:[13]

Definition 13 (Random Utility FDSM representation (RU-FDSM)) *A data set $(\mathcal{D}, p)$ has a Random Utility FDSM representation (RU-FDSM) if it there exists a $\rho : \mathcal{U} \mapsto [0, 1]$ such that the data is consistent with the average of the FDSM models over all possible utilities:*

$$p(a, A) = \sum_{u \in \mathcal{U}} \rho(u) p_u^{FDSM}(a, A)$$

Note that this is equivalent to assuming that, each time the DM is faced with a choice, a utility function is drawn from $\mathcal{U}$, a search order is drawn from $\{\gamma_A\}_{A \in \mathcal{D}}$, and the DM applies the satisficing algorithm using this search order and utility function.

We note that the new model does not satisfy SARP (Axiom 2) nor Deterministic non satisficing choice (Axiom 1). It satisfies however:

Axiom 4 (Degeneracy) *For $x \in X \backslash W^*$ and $A \cap W^* \neq \emptyset$ then $p(x, A) = 0$.*

Notice also the following direct results for RU-FDSM:

Claim 2 *If the data set $(\mathcal{D}, p)$ has a RU-FDSM representation, then $W^* = \{a \in X : p(a, A) > 0,\ \forall A \in \mathcal{D}\} \equiv \{a \in X : \exists\ u \in \mathcal{U}, \rho(u) > 0, u(a) > u^*\}$ corresponds to the set of "sometimes" satisficing items. Also $X \backslash W^*$ corresponds to the set of never satisficing items.*

We are ready to characterize the RU-FDSM.

Theorem 11 *The following are equivalent:*

1. *A complete stochastic choice dataset $(\mathcal{D}, p)$ can be generated by a RU-FDSM.*
2. *A complete stochastic choice dataset $(\mathcal{D}, p)$ satisfies Axiom 3 and Degeneracy (Axiom 4).*

[13] This representation admits two interpretations: First, we can think of a particular DM that has changing or random tastes, that at each trial draws a utility function and a linear search order independently and then given a fixed threshold she searches through a menu until she finds something satisficing else she picks the best item according to her drawn utility. Another interpretation is at the level of the population we can think of infinite number of decision makers with mass equal to 1, each of them endowed with a utility function and linear search order drawn independently and identically from ρ and γ and a common or homogeneous threshold. The probability of choice can be thought as a market share or the mass of DMs that pick a given item in a menu.

4.4.3 Identification of the Utility of Satisficing Items with Random Utility and Random Threshold.

A less than satisfactory feature of the FSSM and FDSM models is that we cannot recover information about the utility of satisficing elements. It is an interesting question to explore whether this characteristic of the FSSM/FDSM is inherent to the general satisficing procedures or not. In this section, we use the exploration of random threshold/random utility variants of the satisficing model to try to answer this question. In general, we find that inferring information about the utility of satisficing items is still problematic.

For the case of a random threshold, Theorem 10 says that the FDSM and FDSM-RT are indistinguishable from one another. A direct corollary of this result is that the random threshold distribution $u^* \sim F_{u^*}$ is not identified. We can only separate the elements that have a positive probability of being over the threshold from those that have a zero probability of surpassing it (this follows from the fact that $W^* \equiv U^*$ in the case of FDSM and $W^* \equiv \{a \in X : \tau(a) > 0\}$ in the FDSM-RT). This partially identifies the support of F_{u^*} (up to monotone transformations of the random variable) but not the actual distribution. The following corollary follows trivially from Theorem 10.

Corollary 1 *If a complete stochastic choice dataset $(\mathcal{D}, p)$ can be generated by a FDSM-RT, then the utility of satisficing elements such that $U^{*RT} = \{a \in X : \tau(a) > 0\}$ is not identified: in particular, for every preference ordering over such elements there exists a FDSM-RT representation of the data with $u : X \to \mathbb{R}$ which is consistent with those preferences.*

From Theorem 10 it is also clear that without further restrictions in the search procedure captured by $\{\gamma_A\}_{A\in\mathcal{D}}$ we cannot hope to obtain more information about the random variable u^* and about the utility level of satisficing items.

The RU-FDSM contains the FDSM-RT/FDSM and is not contained by the latter by Theorem 11. In consequence, we expect no gains from allowing random utility in identify the utility of satisficing items. However, RU-FDSM is interesting because one could argue that the fact that some items are falling below the fixed threshold u^* randomly could help to identify some form of average utility, for example the quantity $Pr(u : u(a) > u^*) = \sum_{u\in\mathcal{U}} \rho(u)\mathbf{1}(u(a) \geq u^*)$. However, we can establish in the basis of Theorem 11 that $Pr(u : u(a) > u^*)$ is not identified.

To see this is true we need to considering the following property, where any item that is satisficing for some fixed utility is satisficing for all utilities in the domain of the random utility $(\rho,\mathcal{U})$. This property is closer to the FDSM as it restricts the taste variation for satisficing items.

Definition 14 (Always satisficing random utility distribution) *A random utility distribution is always satisficing if for any item $x \in X$ such that for some $u \in \mathcal{U}$ with $\rho(u) > 0$, and $u(x) > u^*$ it follows that for any other $\hat{u} \in \mathcal{U}$ with $\rho(\hat{u}) > 0$ it also holds that $\hat{u}(x) > u^*$.*

It is clear that this property allows for a sharper interpretation of W^*. In fact, consider the following immediate claim.

Claim 3 *If complete stochastic choice dataset $(\mathcal{D}, p)$ can be generated by a RU-FDSM with Always satisficing random utility distribution, then if $a \in W^*$ it follows that $Pr(u : u(a) > u^*) = 1$ and $Pr(u : u(b) > u^*) = 0$ for $b \in X \backslash W^*$. Thus $W^* = \{a \in X : u(a) > u^* \ \forall u \in \mathcal{U}, \rho(u) > 0\}$ is the set of always satisficing items.*

The proof of Theorem 11 shows that a complete stochastic choice dataset that can be represented as a RU-FDSM also admits a RU-FDSM representation with the Always satisficing random utility distribution assumption.

Corollary 2 *If a complete stochastic choice dataset $(\mathcal{D}, p)$ can be represented by a RU-FDSM it can also be represented RU-FDSM with Always satisficing property.*

Recall, that FDSM allows us to recover $\mathbf{1}(u(a) > u^*)$ only, which implies the lack of identification of utility (or utility intensity) for satisficing items. Thus the RU-FDSM fares no better than the FDSM in identifying satisficing items utility levels.

The above discussion shows that, as long as one allows relatively unrestricted stochasticity in the search order, adding randomness to the utility function or the satisficing threshold is of no use in identifying preferences for satisficing alternatives. One possible approach which might allow for such identification would be to restrict the allowable distributions over search orders (for example to degenerate, or uniform distributions). We leaving this possibility as an interesting avenue for future research.

5 Relation to Existing Literature

The paper closest in spirit to ours is Manzini & Mariotti (2014), which characterizes the random choice generated by a DM who makes choices by optimizing on a stochastically generated consideration set. As in our model, preferences are deterministic, with randomness in choice coming from stochastic changes is attention. However, the behavioral implications of the two models are quite different, with the satisficing model being the more general. In the set up of MM, all alternatives are always chosen with positive probability in each set. In such a data set, axioms Axiom 1 and Axiom 2 are always satisfied, and so the FSSM is trivially more general than the stochastic consideration set model. Moreover, the FDSM also nests the stochastic consideration set model. This follows from the fact that, when restricted to the class of data in which all alternatives are chosen with positive probability, the FDSM model is equivalent to the class of all RUMs, and the model of MM is a strict subset of this class. Moreover, the FSSM and FDSM can accommodate data sets in which not all alternatives are chosen with positive probability.

Our work also contributes to the literature aimed at testing the satisficing model. It is well known that standard deterministic choice data cannot be used to distinguish rational choice from satisficing behavior, implying that richer data is needed. Caplin et al. (2011); Caplin & Dean (2011) showed how to test the satisficing model using 'choice process' data, which records not just final choice made by a decision maker, but also how choices change with contemplation time. Santos et al. (2012) utilize data in which the sequence of search is recorded to test the satisficing model. Our paper describes the implication of the satisficing model for stochastic choice data, which is arguably easier to collect that either choice process, search or list data.

Another relevant paper is Rubinstein and Salant (2006) that studies the implications of choices from lists that include as a special case a variant of the satisficing model where the DM chooses the first element in a list (that could be put one to one with a linear search/ordering) that is above a threshold, if none she chooses the last one. We differ from this effort in that we assume we do not observe the lists or linear orderings, instead we infer the distribution of the linear search from the frequency of choice.

Ours is not the first paper to characterize the behavior of random choice rules. Much of the previous work has focused on random utility models (RUMs), in which the DM chooses in order to maximize a utility function, drawn from some distribution (see for example Block & Marschak (1960); Falmagne (1978); Gul et al. (2014)). As discussed above, the FSSM is behaviorally distinct from the class of RUMs. It is easy to construct examples of FSSMs which violate regularity, and so cannot be modeled as the resulting from random utility maximization. Moreover, RUMs are not guaranteed to satisfy either axioms Axiom 1 or Axiom 2. In contrast the FDSM is behaviorally a subset of the class of RUMs. Total monotonicity Axiom 3 is necessary and sufficient for a RUM representation (Falmagne, 1978).

References

Block, H. D. & Marschak, J. (1960). *Contributions to Probability and Statistics*, chapter Random Orderings and Stochastic Theories of Response. In Olkin (1960).

Caplin, A. & Dean, M. (2011). Search, choice, and revealed preference. *Theoretical Economics*, 6(1), 19–48.

Caplin, A., Dean, M., & Martin, D. (2011). Search and satisficing. *The American Economic Review*, (pp. 2899–2922).

Eliaz, K. & Spiegler, R. (2011). Consideration sets and competitive marketing. *The Review of Economic Studies*, 78(1), 235–262.

Falmagne, J. C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, 18(1), 52–72.

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

Gul, F. & Pesendorfer, W. (2006). Random Expected Utility. *Econometrica*, 74(1), 121–146.

Luce, R. & Suppes, P. (1965). *Handbook of Mathematical Psychology, Vol. III*, chapter Preference, Utility, and Subjective Probability. In R. D. Luce & Galanter (1965).

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

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

McClellon, M. (2015). Unique Random Utility Representations.

Molchanov, I. (2005). *Theory of Random Sets*. London: Springer.

Olkin, I., Ed. (1960). *Contributions to Probability and Statistics*. Stanford: Stanford University Press.

R. D. Luce, R. R. B. & Galanter, E., Eds. (1965). *Handbook of Mathematical Psychology, Vol. III*. New York: Wiley.

Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *The American Economic Review*, (pp. 900–926).

Richter, M. K. (1966). Revealed preference theory. *Econometrica: Journal of the Econometric Society*, (pp. 635–645).

Rubinstein, A. & Salant, Y. (2006). A model of choice from lists. *Theoretical Economics*, 1(1), 3–17.

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

Santos, B. D. L., Hortacsu, A., & Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6), 2955–80.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: happiness is a matter of choice. *Journal of personality and social psychology*, 83(5), 1178.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, (pp. 99–118).

Ward, D. (1992). The role of satisficing in foraging theory. *Oikos*, (pp. 312–317).

A Proofs

A.1 Proof of Lemma 1

Proof. ($W^* \subseteq U^*$) Let $a \in W^*$ then for any given $A \in \mathcal{D}$ we have $p(a, A) > 0$ then either: (i) $a \in U^*$ or (exclusive) (ii) $u(a) > u(b)$ for all $b \in A$ and $U^* \cap A = \emptyset$. Assume $a \notin U^*$ then (since $U^* \neq \emptyset$) there exists $b \in U^*$ and, by completeness of the data, there is a menu $A' \in \mathcal{D}$ such that $a, b \in A'$, and therefore $p(a, A') = 0$, so we have a contradiction.

($U^* \subseteq W^*$) Let $a \in U^* \Rightarrow u(a) > u^*$, by FSSM $p(a, A) = \gamma_A \left(r_A | a r_A b \;\; \forall b \in A \text{ s.t. } u(b) > u^*\right) > 0$ where the last inequality follows from Full Support assumption. ■

A.2 Proof of Theorem 1

Proof. First we prove that (1) implies (2). To prove that a data set $(\mathcal{D}, p)$ that admits a FSSM representation satisfies Axiom 1 first notice that $U^* \subseteq W^*$ even for incomplete data sets. To see this note that if $a \in U^*$ then $u(a) \geq u^*$ and since the data has a FSSM representation then $p(a, A) = \gamma_A(r_A | a\ r_A\ b\ \forall b \in A \cap U^* \backslash \{a\})$ for all $A \in \mathcal{D}$. Given the full support assumption, $\gamma_A(r_A | a\ r_A\ b\ \forall b \in A \cap U^* \backslash \{a\}) > 0$ for all $A \in \mathcal{D}$, therefore $p(a, A) > 0$ for all $A \in \mathcal{D}$ which in turn implies that $a \in W^*$.

Then if $a \notin W^*$ we have that $a \notin U^*$, which in turn implies that $u(a) < u^*$. Since u is injective, there exists a $a_A^* = argmax_{a \in A} u(a)$. Then either (i) $a = argmax_{b \in A} u(b)$ or (ii) $u(a) < \max_{b \in A} u(b)$. If (i) since the data has a FSSM representation $p(a, A) = 1$; while if (ii) $p(a, A) = 0$. In either case Axiom 1 follows.

To show that Axiom 2 holds, assume, by the way of contradiction, that (i) a is stochastically revealed preferred to b and (ii) b is stochastically strictly revealed preferred to a. From (ii), given that data admits a FSSM we must have (by the full support assumption) that $u(a) < u^*$ and $u(a) < u(b)$. If a is stochastically revealed preferred to b then there must exist a sequence of alternatives $c_1, \ldots, c_N$ and choice sets $A_1, ..., A_{N-1}$, such that $c_1 = a$, $c_N = b$, $c_n, c_{n+1} \in A_n$ and $c_n \in C(A_n)$. If $u(b) > u^*$ then it must be the case that $u(c_{N-1}) > u^*$ (otherwise c_{N-1} could not be chosen with positive probability when c_N was available). Iterating on this argument implies that $u(a) > u^*$. If $u(b) < u^*$ then this implies that $u(c_{N-1}) > u(b)$. If $u(c_n) < u^*$ for all n then iterating on this argument implies that $u(a) > u(b)$. Otherwise, the previous argument implies that $u(a) > u^*$. Either provides a contradiction.

Now we prove that (2) implies (1). For $A \subseteq X \setminus W^*$, $C(A) \equiv \{a \in A | p(a, A) = 1\}$ from Axiom 1. Given Axiom 2, we can generate an injective utility function using Afriat/Richter's theorem, such that $u : X \setminus W^* \to [0, \overline{u}]$ and $C(A) = argmax_{a \in A} u(a)$ for all $A \in \mathcal{D}$ such that $A \cap W^* = \emptyset$. Fix $u^* > \overline{u}$, enumerate the elements in W^* as $a_1, \ldots, a_N$ and let $u(a_n) = u^* + n$ for all $a_n \in W^*$. Thus $u : X \to \mathbb{R}$ is injective and $U^* \equiv W^*$. Furthermore, notice that Axiom 2 implies that W^* is non-empty, so that $u^* \leq max_{x \in X} u(x)$.

For every A such that $A \cap W^*$ is non-empty, let $A^* \equiv A \cap W^*$ and define R_{A^*} the set of all linear orders on A^*, and let $R_A(r_{A^*})$ be the set of all linear orders $r_A \in R_A$ that induce the linear order $r_{A*} \in R_{A^*}$.Then, set the probability of the set of linear orders that generate each r_{A^*} as

$$\gamma_A\left(r_A \in R_A(r_{A^*})| \quad a_i r_{A^*} a_j \ \forall a_j \in A^* \backslash \{a_i\}\right) = \frac{p(a_i, A)}{|\left\{r_{A^*} \in R_{A^*} : a_i r_{A^*} a_j \ \forall a_j \in A^* \backslash \{a_i\}\right\}|}$$

for all $a_i \in A^*$. Finally, distribute the probability mass above uniformly across the elements in $R_A(r_{A^*})$. For any $r_A \in R_A(r_{A^*})$, for a given r_{A^*}:

$$\gamma_A(r_A) = \gamma_A\left(r_A \in R_A | a_i r_{A^*} a_j \ \forall a_j \in A^* \backslash \{a_i\}\right) / |R_A(r_{A^*})|$$

Where $|\cdot|$ stands for the cardinality map. Note that, as $r_A \in R_A(r_{A^*})| \quad a_i r_{A^*} a_j \ \forall a_j \in A^* \backslash \{a_i\}$ for some $a_i \in A^*$, and as $p(a_i, A) > 0$ by construction of W^* this distribution will have full support on R_A.

If $A \cap W^* = \emptyset$ then for all $r_A \in R_A$ define

$$\gamma(r_A) = \frac{1}{|R_A|}$$

Thus, between them u, u^* and $\{\gamma_A\}_{A \in \mathcal{D}}$ satisfy the requirements of an FSSM. To verify that we can generate $(\mathcal{D}, p)$ notice that if we face a menu A we have the following cases:

(i) if $A \cap W^* = A$ then $u(a) > u^*$ for all $a \in A$ and, for each a, $p(a, A) = \gamma_A(r_A \in R_A | a \ r_A \ b \quad \forall b \in A \backslash \{a\})$, to see that this is true observe that $\gamma_A(r_A \in R_A | a \ r_A \ b \quad \forall b \in A \backslash \{a\}) = \sum_{r_A \in R_A} \mathbf{1}\left[r_A : r_{A^*} \text{ such that } a_i r_{A^*} a_j \ \forall a_j \in A^*\right] \gamma_A(r_A) = p(a, A)$.

(ii) If $A \cap W^* = \emptyset$ then $p(a, A) = 1$ if $u(a) > u(b)$ for all $b \in A \backslash \{a\}$ and zero otherwise. This follows directly from the fact that u was constructed to represent choice on such sets.

(iii) If $A \cap W^* \subset A$ and $A \cap W^* \neq \emptyset$ then we have $p(a, A) = \gamma_A(r_A | a \ r_A \ b \quad \forall b \in (A \cap W^*) \backslash \{a\})$ if $u(a) \geq u^*$ and $p(a, A) = 0$ if $u(a) < u^*$. To see that this is true observe that by definition of W^* and Axiom 1 $p(A \cap W^*, A) = 1$. To see that this is true, observe that if we assume that $p(A \cap W^*, A) < 1$ we must have that $p(a, A) > 0$ for some $a \notin W^*$ but that means by Axiom 1 that $p(a, A) = 1$ which is a contradiction of the fact that $p(A \cap W^*, A) > 0$.

Then $p(a, A) = 0$ if $u(a) < u^*$. For $a, b \in A \cap W^*$, the result follows as in (i). ■

A.3 Proof of Theorem 2

Proof. To prove (1) assume, by contradiction, that $U^* \neq \overline{U}^*$ and let $v \in \overline{U}^* \setminus U^*$. Then, it must be the case that $u(v) < u^*$ and $\overline{u}(v) \geq \overline{u}^*$. By completeness and the fact that $U^* \neq \emptyset$, $\exists a \in U^*$ and a $A \in \mathcal{D}$ such that $A = \{a, v\}$. Given that the data is represented by $(u, u^*, \{\gamma_A\}_{A\in\mathcal{D}})$, $p(a, A) = 1$. On the other hand, since $(\overline{u}, \overline{u}^*, \{\overline{\gamma}_A\}_{A\in\mathcal{D}})$ also represents the data it is the case that $p(a, A) \in (0, 1)$, which establishes a contradiction.

To prove (2) notice that from (1) $U^* = \overline{U}^*$. Since both, $(u, u^*, \{\gamma_A\}_{A\in\mathcal{D}})$ and $(\overline{u}, \overline{u}^*, \{\overline{\gamma}_A\}_{A\in\mathcal{D}})$ are generated by the same FSSM, then u and u^* represent the preferences given by definition 5 for all $a \notin U^*$. Therefore, it must be the case that u is a strictly increasing transformation of $\overline{u}$. on X/U^*

To prove (3) assume, by contradiction, that for some $A \in D$, $a \in A \cap U^*$

$$\gamma_A(ar_Ab \quad \forall b \in \{A \cap U^*\} \setminus \{a\}) \neq \bar{\gamma}_A(ar_Ab \quad \forall b \in \{A \cap U^*\} \setminus \{a\})$$

then $p(a, A|\gamma_A) \neq p(a, A|\overline{\gamma}_A)$ which in turn implies that both, $(u, u^*, \{\gamma_A\}_{A\in\mathcal{D}})$ and $(\overline{u}, \overline{u}^*, \{\overline{\gamma}_A\}_{A\in\mathcal{D}})$ cannot represent the same data. ■

A.4 Proof of Theorem 3

Proof. First we prove (1) implies (2). If the complete data is generated by a FDSM then there is a triple (u, u^*, Γ_X), take a realization of Γ_X with support on R_X and call it r_X, then define the linear ordering on $X \succ_X$:

(1) $a \succ_X b$ if ar_Xb and $a, b \in U^*$, (2) $a \succ_X b$ if $u(a) > u(b)$ and $a, b \notin U^*$ and (3) $a \succ_X b$ if $a \in U^*, b \notin U^*$. Now, assign this linear ordering $\succ_X$ the probability $\Gamma_X(r_X)$. It is direct to see that there is a Random Utility Maximization model without indifference with realizations $\succ_X$ with probability $\Gamma_X(r_X)$. By Block & Marschak (1960) it follows that the generated data set $(\mathcal{D}, p)$ satisfies Total Monotonicity (Axiom 3). The fact that Axiom 1 and Axiom2 hold follows from Theorem 1.

Second we prove (2) implies (1). Because FDSM is a subcase of FSSM and Axiom 1 and Axiom 2 hold we can build an utility $u : X \mapsto \mathbb{R}$ and a threshold $u^* \in \mathbb{R}$ such that $u(x) \geq u^*$ for all $x \in W^*$ and $u(x) < u^*$ for all $x \in X \backslash W^*$. Finally, if a complete stochastic choice dataset $(\mathcal{D}, p)$ satisfies Axiom 3 then the model is a Random Maximization Utility model so we can recover a distribution over linear orders on X $\Gamma^Q_X : \overline{R}_X \mapsto [0, 1]$ such that

$p(a, A) = \Gamma_X^Q(a\overline{r}_X b \ \forall\, b \in A)$ thanks to Falmagne (1978). Call its support $\overline{R}_X$ the set of quasi-search ordering, because they are the results of both deterministic (degenerate) utility maximization choice and fixed random searching and satisficing behavior. To construct the Γ_X we build each element of its support by taking an element of the support of the Quasi-search ordering $\overline{r}_X \in \overline{R}_X$ and we restrict it to W^*. Here we define an equivalence class on $\overline{R}_X$ if they have the same restriction $\overline{r}_X|W^*$, if we have any two elements $\overline{r}_X, \overline{r}'_X \in \overline{R}_X$ that have the same restriction to W^* (i.e., $x\overline{r}_X y \iff x\overline{r}'_X y$ for $x, y \in W^*$) we say $\overline{r}_X \equiv_{W^*} \overline{r}'_X$, the equivalence class set is denoted as $[\overline{r}_X]_{\equiv_{W^*}} = \{\overline{r}'_X \in \overline{R}_X | \overline{r}_X \equiv_{W^*} \overline{r}'_X\}$ then we assign to the representative of the equivalence class or the restriction $r_X|W^*$ the probability corresponding to the sum $\sum_{\overline{r}'_X \in [\overline{r}_X]_{\equiv_{W^*}}} \Gamma_X^Q(\overline{r}'_X)$. For any given restricted ordering $\overline{r}_X|W^*$ we build its transitive closure or the set of transitive extensions to X and call this set $R_X(\overline{r}_X) \subset X \times X$. We assign each of the elements of this set $\hat{r}_X \in R_X(\overline{r}_X)$ the probability $\sum_{\overline{r}_X \in [\overline{r}_X]_{\equiv_{W^*}}} \Gamma_X^Q(\overline{r}_X)/|R_X(\overline{r}_X)|$ where the numerator is the probability of the restricted to W^* quasi-search ordering $\overline{r}_X|W^*$ and the denominator is the cardinality of the previously defined set. Doing this for all elements of $\overline{R}_X$ we build a new support R_X with probabilities as indicated that provide us with Γ_X.

Note that Γ_X has full support due to how W^* is constructed. Because W^* is the always chosen set, we know that any element of the set of restrictions $\overline{r}_X|W^*$ such that for each $a \in W^*$, $a\overline{r}_X b$ for all $b \in W^* \backslash \{a\}$ has positive probability. It follows that by definition, for any $A \in \mathcal{D}$, and for any $x \in W^*$ we have $p(x, A) > 0$, this means that $\Gamma_X^Q(\overline{r}_X \in \overline{R}_X : x\overline{r}_X y \quad \forall y \in X\backslash\{x\}) > 0$. This implies that all representatives of the equivalence class of restricted orderings $r_X|W^*$ where x is searched first in W^* have positive probability. Then we have extended them to X with the uniform distribution for each set $R_X(\overline{r}_X)$ thus preserving the full support for the whole X. The reason is that the transitive closure to X of any $r_X|W^*$ contains linear search orders that have each $x \in W^*$ as the first searched element, because it contains the ordering that preserves the elements in W^* in the top and the rest at the bottom. But also it contains search orders with each element of $X\backslash W^*$ at the top for each of such elements and W^* at the bottom all with positive probability.

We have extended each restriction $\overline{r}_X|W^*$ such that we can let Γ_X be the fixed distribution of search orders. We have built a FDSM or a triple (u, u^*, Γ_X). To verify that this FDSM model generate $(\mathcal{D}, p)$ notice that if we face a menu A we have the following cases:

(i) if $A \cap W^* = A$ then $p(a, A) = \Gamma_X(r_X \in R_X|\ a\, r_X\, b \quad \forall b \in A\backslash\{a\})$, to see that this is true observe that $\Gamma_X(r_X \in R_X|\ a\, r_X\, b \quad \forall b \in A\backslash\{a\}) = \Gamma_X^Q(\overline{r}_X \in \overline{R}_X|a\, \overline{r}_X\, b \quad \forall b \in A\backslash\{a\})$.

In this case the first equality follows from the equivalence to random utility when restricted to W^*.

(ii) If $A \cap W^* = \emptyset$ then $p(a, A) = 1$ if $u(a) > u(b)$ for all $b \in A \setminus \{a\}$ and zero otherwise. This is direct from the fact that u represents C in such sets.

(iii) If $A \cap W^* \subset A$ then we have $p(a, A) = \Gamma_X(r_X | a \ r_X \ b \quad \forall b \in (A \cap U^*) \setminus \{a\})$ if $u(a) \geq u^*$ and $p(a, A) = 0$ if $u(a) < u^*$. To see that this is true observe that by definition of W^* and Axiom 1 $p(A \cap W^*, A) = 1$ then $p(a, A) = 0$ if $u(a) < u^*$. For $a, b \in A \cap W^*$ observe that by construction $\Gamma_X(r_X \in R_X | a r_X b \quad \forall b \in (A \cap U^*) \setminus \{a\}) = \Gamma_X^Q(\overline{r}_X \in \overline{R}_X | a \overline{r}_X b \quad \forall b \in (A \cap W^*) \setminus \{a\})$. Finally, observe that $\Gamma_X^Q(\overline{r}_X \in \overline{R}_X | a \overline{r}_X b \quad \forall b \in (A \cap W^*) \setminus \{a\}) = \Gamma_X^Q(\overline{r}_X \in \overline{R}_X | a \overline{r}_X b \quad \forall b \in A \setminus \{a\}) = p(a, A)$ because the facts that $p(c, A) = 0$ for any $c \notin W^*$ and that Γ_X^Q represents choice implies that $\Gamma_X^Q(\overline{r}_X \in \overline{R}_X | c \overline{r}_X a) = 0$. ■

A.5 Proof of Theorem 4

Proof. For (1) and (2) we use the results of Theorem 2.

(3) follows from the fact that $W^* \equiv U^*$ and the FDSM behaves as Random Utility in this for the elements in W^*. Assume by contradiction that $\Gamma_X(x r_X y) \neq \overline{\Gamma}_X(x r_X y)$ for some $x, y \in W^*$ but that means that in the menu $\{x, y\}$, $p(x, \{x, y\} | \Gamma_X) \neq p(x, \{x, y\} | \overline{\Gamma}_X)$ which is a contradiction. ■

A.6 Proof of Claim 1

Proof. First note that if p is more easily satisfied than $\tilde{p}$, and $\tilde{W}^* = W^*$, then the claim is trivially satisfied since both random choice rules are behaviorally indistinguishable, and therefore there exists representations $(u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$ and $(\tilde{u}, \tilde{u}^*, \{\tilde{\gamma_A}\}_{A \in \mathcal{D}})$ with (i) $u = \tilde{u}$, (ii) $u^* = \tilde{u}^*$ and, (iii) $\{\gamma_A\}_{A \in \mathcal{D}} = \{\tilde{\gamma_A}\}_{A \in \mathcal{D}}$.

Consider now the case where p is strictly more easily satisfied than $\tilde{p}$, i.e. $\tilde{W}^* \subset W^*$. First we prove that there exists representations for p and $\tilde{p}$ such that $u(a) = \tilde{u}(a)$ for any $a \in X \setminus W^*$. Since p and $\tilde{p}$, admit a FSSM (FDSM) representation, there exist injective utility functions u and $\tilde{u}$ that represents observed choices $C_p(A)$, $C_{\tilde{p}}(A)$ for all A such that $A \cap W^* = \emptyset$. Moreover, since $C_p(A) = C_{\tilde{p}}(A)$ for all A such that $A \cap W^* = \emptyset$ then there exist representations u and $\tilde{u}$ with $u(a) = \tilde{u}(a)$ for all $a \in X \setminus \tilde{W}^*$. Pick one such representation and set $u^* = \max_{a \in X \setminus W^*} u(a) + 1$.

Now we show that these representations can be also chosen such that $u(a) = \tilde{u}(a)$ for any $a \in X \setminus \tilde{W}^*$. For any $a \in W^* \setminus \tilde{W}^*$, p more easily satisfied than $\tilde{p}$ implies that if b is revealed preferred to a according to $\tilde{p}$, then $b \in W^*$, and therefore, any representation $\tilde{u}$ of $C_{\tilde{p}}(A)$ with $A \cap \tilde{W}^* = \emptyset$ also represents choices $C_{\tilde{p}}(A)$ with $A \cap \tilde{W}^* = \emptyset$. For such representations set $\tilde{u}^* = \max_{a \in X \setminus \tilde{W}^*} u(a) + 1$. Since $\tilde{W}^* \subset W^*$, then $\tilde{u}^* > u^*$. Finally, enumerate the elements in $\tilde{W}^*$ as $a_1, a_2, \ldots, A_N$ and let $u(a_n) = \tilde{u}(a_n) = \tilde{u}^* + n$.

Finally we need to prove that p is more easily satisfied than $\tilde{p}$ and $\tilde{W}^* \subset W^*$, there exists representations for p and $\tilde{p}$ that admit a FSSM representation with the above utility functions and utility thresholds and such that $\{\gamma_A\}_{A \in \mathcal{D}} = \{\tilde{\gamma_A}\}_{A \in \mathcal{D}}$. First notice that search orders can be trivially set such that $\gamma_A = \tilde{\gamma_A}$ for any A such that $A \cap W^* = \emptyset$. Second, if $A \cap (W^* \setminus \tilde{W}^*) = \emptyset$, p is more easily satisfied than $\tilde{p}$ implies that for any choice set A and $a \in \tilde{W}^* \cap W^* \cap A$, $p(a, A) \leq \tilde{p}(a, A)$, which in turn implies that $p(a, A) = \tilde{p}(a, A)$, and therefore we can find representations with $\gamma_A = \tilde{\gamma_A}$ for any A such that $A \cap (W^* \setminus \tilde{W}^*) = \emptyset$. Third, for any A such that $A \subset (X \setminus \tilde{W}^*)$, search orders are not identified for the random choice rule $\tilde{p}$, while search orders for random choice rule p are identified, up to equivalent lineal search orders among elements in $A \cap W^*$, and therefore we can set $\gamma_A = \tilde{\gamma}_A$ for all A such that $A \subset (X \setminus \tilde{W}^*)$.

Finally, consider menus A such that $A \cap \tilde{W}^* \neq \emptyset$ and $A \cap (W^* \setminus \tilde{W}^*) \neq \emptyset$. Define $\alpha \equiv A \cap (W^* \setminus \tilde{W}^*)$ and $\tilde{\alpha} \equiv A \cap \tilde{W}^*$, and define $p_\alpha = \sum_{a \in \alpha} p(a, A)$. Let α_i be the typical element in α, i.e. $\alpha \equiv \{\alpha_i\}_{i=1}^{|\alpha|}$; and $\tilde{\alpha}_i$ be the typical element of $\tilde{\alpha}$, i.e. $\tilde{\alpha} \equiv \{\alpha_i\}_{i=1}^{|\tilde{\alpha}|}$. Define $\tilde{R}_{\tilde{\alpha}_i}$ as the set of linear orders for the elements in $\tilde{\alpha}$ such that $\tilde{\alpha}_i \; \tilde{R}_{\tilde{\alpha}_i} \; b$ for all $b \in \tilde{\alpha} \setminus \{\tilde{\alpha}_i\}$ and $R^{\alpha}_{\alpha_i}$ be the set of linear orders for the elements in α such that $\alpha_i \; R^{\alpha}_{\alpha_i} \; b$ for all $b \in \alpha \setminus \{\alpha_i\}$. Then, define the linear orders $R_{\alpha_i, \tilde{\alpha}_i}$ as the linear orders for the alternatives in $\alpha \cup \tilde{\alpha}$ such that it is consistent with $\tilde{R}_{\tilde{\alpha}_i}$ and with $R^{\alpha}_{\alpha_i}$ and that any element in α is seen first than any element in $\tilde{\alpha}$; correspondingly define $R_{\tilde{\alpha}_i, \alpha_i}$ as the set of linear orders that are also consistent with $\tilde{R}_{\tilde{\alpha}_i}$ and with $R^{\alpha}_{\alpha_i}$ but such that any element in α is seen after any element in $\tilde{\alpha}$. Finally define

$$\gamma\left(r \in R_{\alpha_i, \tilde{\alpha}_i}\right) = \frac{p(\alpha_i, A) \frac{\tilde{p}(\tilde{\alpha}_i, A) - p(\tilde{\alpha}_i, A)}{p_\alpha}}{|R_{\alpha_i, \tilde{\alpha}_i}|}$$

and

$$\gamma\left(r \in R_{\tilde{\alpha}_i, \alpha_i}\right) = \frac{p(\tilde{\alpha}_i, A) \frac{p(\alpha_i, A)}{p_\alpha}}{|R_{\tilde{\alpha}_i, \alpha_i}|}$$

Now we extend these search orders to linear orders over all elements in A. Let $R_A\left(r_{\alpha_i, \tilde{\alpha}_i}\right)$

be the set of all linear orders $r_A \in R_A$ that induce the linear order $r_{\alpha_i,\tilde{\alpha}_i} \in R_{\alpha_i,\tilde{\alpha}_i}$. For any $r_A \in R_A(r_{\alpha_i,\tilde{\alpha}_i})$, for a given $r_{\alpha_i,\tilde{\alpha}_i}$:

$$\gamma_A(r_A) = \frac{\gamma_A\left(r_A \in R_A | a_i \; r_{\alpha_i,\tilde{\alpha}_i} \; a_j \; \forall a_j \in \alpha \cup \tilde{\alpha} \backslash \{a_i\}\right)}{\left|R_A\left(r_{\alpha_i,\tilde{\alpha}_i}\right)\right|}$$

Notice that the above collection of search orders generate both random choice rules since

$$\begin{aligned}
p(\alpha_i, A) &= \sum_{\tilde{\alpha}_i} \gamma\left(r \in R_{\alpha_i,\tilde{\alpha}_i}\right) \times |R_{\alpha_i,\tilde{\alpha}_i}| \quad \text{for any } \alpha_i \in \alpha \\
&= \sum_{\tilde{\alpha}_i} p(\alpha_i, A) \frac{\tilde{p}(\tilde{\alpha}_i, A) - p(\tilde{\alpha}_i, A)}{p_\alpha} \\
&= p(\alpha_i, A) \frac{1 - (1 - p_\alpha)}{p_\alpha} \\
&= p(\alpha_i, A)
\end{aligned}$$

$$\begin{aligned}
p(\tilde{\alpha}_i, A) &= \sum_{\alpha_i} \gamma\left(r \in R_{\tilde{\alpha}_i,\alpha_i}\right) \times |R_{\tilde{\alpha}_i,\alpha_i}| \quad \text{for any } \tilde{\alpha}_i \in \tilde{\alpha} \\
&= \sum_{\alpha_i} p(\tilde{\alpha}_i, A) \frac{p(\alpha_i, A)}{p_\alpha} \\
&= p(\tilde{\alpha}_i, A) \sum_{\alpha_i} \frac{p(\alpha_i, A)}{p_\alpha} \\
&= p(\tilde{\alpha}_i, A)
\end{aligned}$$

$$\begin{aligned}
\tilde{p}(\tilde{\alpha}_i, A) &= \sum_{\alpha_i} \left[\gamma\left(r \in R_{\alpha_i,\tilde{\alpha}_i}\right) |R_{\alpha_i,\tilde{\alpha}_i}| + \gamma\left(r \in R_{\tilde{\alpha}_i,\alpha_i}\right) |R_{\tilde{\alpha}_i,\alpha_i}|\right] \\
&= \sum_{\alpha_i} \left[p(\alpha_i, A) \frac{\tilde{p}(\tilde{\alpha}_i, A) - p(\tilde{\alpha}_i, A)}{p_\alpha} + p(\tilde{\alpha}_i, A) \frac{p(\alpha_i, A)}{p_\alpha}\right] \\
&= \left[p_\alpha \frac{\tilde{p}(\tilde{\alpha}_i, A) - p(\tilde{\alpha}_i, A)}{p_\alpha} + p(\tilde{\alpha}_i, A) \frac{p_\alpha}{p_\alpha}\right] \\
&= \left[\tilde{p}(\tilde{\alpha}_i, A) - p(\tilde{\alpha}_i, A) + p(\tilde{\alpha}_i, A)\right] \\
&= \tilde{p}(\tilde{\alpha}_i, A)
\end{aligned}$$

and therefore we can set $\{\tilde{\gamma}_A\}_{A \in \mathcal{D}} = \{\gamma_A\}_{A \in \mathcal{D}}$. ■

A.7 Proof of Proposition 1

Proof. First notice that, by Claim 1 we have that $u^* \leq \tilde{u}^*$. It is trivial to show the result if $u^* = \tilde{u}$, since both random choice rules are indistinguishable and it follows that $F_{u|p}(u, A) = F_{u|\tilde{p}}(u, A)$ for all $A \in \mathcal{D}$. Consider the case where, $u^* < \tilde{u}^*$. We have two cases (i) $W^* = \tilde{W}^*$ and $u^* < \tilde{u}^*$ and (ii) $\tilde{W}^* \subset W^*$ and $u^* < \tilde{u}^*$. If (i), then the random choice rules p and $\tilde{p}$ are behaviorally indistinguishable and we trivially get $F_{u|p}(u, A) = F_{u|\tilde{p}}(u, A)$. We focus now on (ii).

Let $\alpha \equiv W^* \setminus \tilde{W}^*$. For any menu $A \in \mathcal{D}$, differences in the utility threshold level only affects decisions if $A \cap \alpha \neq \emptyset$; otherwise $F_{u|p}(\cdot, A) = F_{u|\tilde{p}}(\cdot, A)$ and the result is trivially true. Now consider the case where $A \cap \alpha \neq \emptyset$. Consider two cases (i) $\max_{a\in A} u(a) < \tilde{u}^*$ and (ii) $\max_{a\in A} u(a) \geq \tilde{u}^*$. Since $A \cap \alpha \neq \emptyset$ these two cases suffice to cover all possible scenarios. If (i) then

$$F_{u|p}(\overline{u}, A) = \begin{cases} 0 & \text{for } \overline{u} < u^* \\ \sum_{a\in A: u(a)\leq \overline{u}} p(a, A) & \text{for } u^* \leq \overline{u} < \max_{a\in A} u(a) \\ 1 & \text{for } \max_{a\in A} u(a) \leq \overline{u} < \tilde{u}^* \\ 1 & \text{for } \overline{u} \geq \tilde{u}^* \end{cases}$$

On the other hand,

$$F_{u|\tilde{p}}(\overline{u}, A) = \begin{cases} 0 & \text{for } \overline{u} < u^* \\ 0 & \text{for } u^* \leq \overline{u} < \max_{a\in A} u(a) \\ 1 & \text{for } \max_{a\in A} u(a) \leq \overline{u} < \tilde{u}^* \\ 1 & \text{for } \overline{u} \geq \tilde{u}^* \end{cases}$$

From where it follows that

$$F_{u|p}(\overline{u}, A) - F_{u|\tilde{p}}(\overline{u}, A) = \begin{cases} 0 & \text{for } \overline{u} < u^* \\ \sum_{a\in A: u(a)\leq \overline{u}} p(a, A) & \text{for } u^* \leq \overline{u} < \max_{a\in A} u(a) \\ 0 & \text{for } \max_{a\in A} u(a) \leq \overline{u} < \tilde{u}^* \\ 0 & \text{for } \overline{u} \geq \tilde{u}^* \end{cases}$$

where $\sum_{a\in A: u(a)\leq \overline{u}} p(a, A) > 0$ for $u^* \leq \overline{u} < \max_{a\in A} u(a)$ by the Full Support assumption. Obtaining the desired result.

If (ii) then,

$$F_{u|p}(\overline{u}, A) = \begin{cases} 0 & \text{for } \overline{u} < u^* \\ \sum_{a \in A: u(a) \leq \overline{u}} p(a, A) & \text{for } u^* \leq \overline{u} < \tilde{u}^* \\ \sum_{a \in A: u(a) \leq \tilde{u}^*} p(a, A) + \sum_{a: \tilde{u}^* < u(a) \leq \overline{u}} p(a, A) & \text{for } \overline{u} \geq \tilde{u}^* \end{cases}$$

On the other hand,

$$F_{u|\tilde{p}}(\overline{u}, A) = \begin{cases} 0 & \text{for } \overline{u} < u^* \\ 0 & \text{for } u^* \leq \overline{u} < \tilde{u}^* \\ \sum_{a \in A: \tilde{u}^* < u(a) \leq \overline{u}} \tilde{p}(a, A) & \text{for } \overline{u} \geq \tilde{u}^* \end{cases}$$

From where it follows that

$$F_{u|p}(\overline{u}, A) - F_{u|\tilde{p}}(\overline{u}, A) = \begin{cases} 0 & \text{for } \overline{u} < u^* \\ \sum_{a \in A: u(a) \leq \overline{u}} p(a, A) & \text{for } u^* \leq \overline{u} < \tilde{u}^* \\ \sum_{a \in A: u(a) \leq \tilde{u}^*} p(a, A) + \sum_{a \in A: \tilde{u}^* < u(a) \leq \overline{u}} (p(a, A) - \tilde{p}(a, A)) & \text{for } \overline{u} \geq \tilde{u}^* \end{cases}$$

We need to prove that for any $\overline{u} \geq \tilde{u}^*$ it is the case that $\sum_{a \in A: u(a) \leq \tilde{u}^*} p(a, A) + \sum_{a: \tilde{u}^* < u(a) \leq \overline{u}} (p(a, A) - \tilde{p}(a, A)) \geq 0$ or equivalently

$$\sum_{a \in A: u(a) \leq \tilde{u}^*} p(a, A) \geq \sum_{a \in A: \tilde{u}^* < u(a) \leq \overline{u}} (\tilde{p}(a, A) - p(a, A))$$

Define $p_\alpha \equiv \sum_{a \in A: u(a) \leq \tilde{u}^*} p(a, A)$ then we need to prove that

$$p_\alpha \geq \sum_{a \in A: \tilde{u}^* < u(a) \leq \overline{u}} (\tilde{p}(a, A) - p(a, A))$$

Since p is more easily satisfied than $\tilde{p}$ we have that for any a such that $\tilde{u}^* < u(a)$, $p(a, A) \leq \tilde{p}(a, A)$, which in turn implies that the sum $\sum_{a \in A: \tilde{u}^* < u(a) \leq \overline{u}} (\tilde{p}(a, A) - p(a, A))$ is increasing in $\overline{u}$. Also note that

$$\begin{aligned}\lim_{\overline{u}\to\infty} \sum_{a\in A:\tilde{u}^*<u(a)\leq\overline{u}} (\tilde{p}(a,A)-p(a,A)) &= \\ \lim_{\overline{u}\to\infty} \sum_{a\in A:\tilde{u}^*<u(a)\leq\overline{u}} \tilde{p}(a,A) - \lim_{\overline{u}\to\infty} \sum_{a\in A:\tilde{u}^*<u(a)\leq\overline{u}} p(a,A) &= \\ 1-(1-p_\alpha) &= p_\alpha\end{aligned}$$

Thus, $p_\alpha \geq \sum_{a\in A:\tilde{u}^*<u(a)\leq\overline{u}} (\tilde{p}(a,A)-p(a,A))$ ■

A.8 Comparative Statics with Respect to Search Orders

First, we introduced the following class of equivalent search order for a given $A \in \mathcal{D}$.

Definition 15 ($\{a, b\}$-equivalent search orders) *Let $\gamma_A, \tilde{\gamma}_A$ be two probability distributions over search orders R_A for some $A \in \mathcal{D}$. We say that $\gamma_A, \tilde{\gamma}_A$ are $\{a,b\}$-equivalent search orders for menu A if $\gamma_A(r_A) = \tilde{\gamma}_A(r_A)$ for all r_A such that $r^*_{A\setminus\{a\}} = \tilde{r}_{A\setminus\{a\}}$ and $r^*_{A\setminus\{b\}} = \tilde{r}_{A\setminus\{b\}}$ for some $a, b \in A$.*

Proposition 2 (Salience of Satisficing Elements) *Let p and $\tilde{p}$ be two random choice rules that admit identical up to search orders FSSM (FDSM) representations $(u, u^*, \{\gamma_A\}_{A\in\mathcal{D}})$ and $(u, u^*, \{\tilde{\gamma}_A\}_{A\in\mathcal{D}})$. Let $F_{u|p}(\overline{u}, A)$ and $F_{u|\tilde{p}}(\overline{u}, A)$ be their respective probability distribution over utility levels. Let γ_A and $\tilde{\gamma}_A$ be $\{a,b\}-$ equivalent search orders as in Definition 15 with $\gamma_A(r^*_A) = \tilde{\gamma}_A(r^*_A) + \varepsilon$ and $\gamma_A(\tilde{r}_A) = \tilde{\gamma}_A(\tilde{r}_A) - \varepsilon$ for some $\varepsilon \in (\tilde{\gamma}(\tilde{r}_A), 1 - \tilde{\gamma}_A(r^*_A))$ and $ar^*_A b$ and $b\tilde{r}_A a$.*

If $u(a) \geq u(b)$ then $F_{u|p}(\cdot, A)$ FOSD $F_{u|\tilde{p}}(\cdot, A)$. Moreover, if $F_{u|p}(\overline{u}, A) \leq F_{u|\tilde{p}}(\overline{u}, A)$ with $F_{u|p}(\overline{u}, A) \neq F_{u|\tilde{p}}(\overline{u}, A)$ then $u(a) > u(b) \geq u^*$

Proof. First we prove that if $u(a) > u(b)$ then $F_{u|p}(\cdot, A)$ FOSD $F_{u|\tilde{p}}(\cdot, A)$. There are four possible cases: (i) $a, b \notin U^*$ then $F_{u|p}(\cdot, A) = F_{u|\tilde{p}}(\cdot, A)$. (ii) If $a \in U^*$ and $b \notin U^*$, under the conditions of the proposition then $F_{u|p}(\cdot, A) = F_{u|\tilde{p}}(\cdot, A)$. (iii) If $b \in U^*$ and $a \notin U^*$, under the conditions of the proposition then $F_{u|p}(\cdot, A) = F_{u|\tilde{p}}(\cdot, A)$. (iv) Finally, consider the case where $a, b \in U^*$. Under the conditions of the proposition $p(a, A) > \tilde{p}(a, A)$ and $p(b, A) < \tilde{p}(b, A)$ while $p(c, A) = \tilde{p}(c, A)$ for all $c \in A \setminus \{a, b\}$ and the results follows from $u(a) > u(b)$

Now we prove that if $F_{u|p}(\overline{u}, A) \leq F_{u|\tilde{p}}(\overline{u}, A)$ for all $\overline{u}$ with $F_{p|u}(\overline{u}, A) \neq F_{p|\tilde{u}}(\overline{u}, A)$ then $u(a) > u(b) \geq u^*$. Notice that the change in the probabilities of search orders only affects the relative probability of a, b being seen first, leaving unaltered all other relative probabilities. Therefore, the change from $\{\tilde{\gamma}_A\}_{A\in\mathcal{D}}$ to $\{\gamma_A\}_{A\in\mathcal{D}}$ only has an effect on the distribution over utility levels if $\{a, b\} \subseteq U^*$. Moreover, given the conditions of the proposition, $p(a, A) > \tilde{p}(a, A)$ and $p(b, A) < \tilde{p}(b, A)$. Then it must be that $u(a) > u(b)$. To show the latter, assume by contradiction that $u^* \leq u(a) < u(b)$, then $F_{u|\tilde{p}}(u(a), A) < F_I(u(a), A)$ which leads to a contradiction. ■

A.9 Proof of Theorem 5

Proof. First we prove that (1) implies (2). If a complete data has a FDSM without full support representation then there is a triple (u, u^*, Γ_X), take a realization of Γ_X with support on some subset or all R_X and call it r_X, then define the linear ordering on X $\succ_X$: (1) $a \succ_X b$ if $a r_X b$ and $a, b \in U^*$, (2) $a \succ_X b$ if $u(a) > u(b)$ and $a, b \notin U^*$ and (3) $a \succ_X b$ if $a \in U^*, b \notin U^*$. Now, assign this linear ordering $\succ_X$ the probability $\Gamma_X(r_X)$. It is direct to see that there is a Random Utility Maximization model without indifference with realizations $\succ_X$ with probability $\Gamma_X(r_X)$. By Block & Marschak (1960) it follows that the generated data set $(\mathcal{D}, p)$ satisfies Total Monotonicity (Axiom 3).

Now we prove that (2) implies (1). If a complete stochastic choice dataset $(\mathcal{D}, p)$ satisfies Axiom 3 then the model is a Random Maximization Utility model so we can recover a distribution over linear orders on X $\Gamma_X : \overline{R}_X \mapsto [0, 1]$ such that $p(a, A) = \Gamma_X^Q(a\overline{r}_X b \,\forall\, b \in A)$ thanks to Falmagne (1978). Interpret Γ_X as a fixed distribution search order, and select any injective $u : X \to \mathbb{R}$ and u^* such that $u(x) > u^*$ for all $x \in X$. By construction this triple (u, u^*, Γ_X) generates the observed data. ■

A.10 Proof of Theorem 6

Proof. First we prove that (1) implies (2). To show that Axiom 2 holds, assume, by the way of contradiction, that (i) a is stochastically revealed preferred to b and (ii) b is stochastically strictly revealed preferred to a. From (ii), given that data admits a FSSMI we must have (by the full support assumption) that $u(a) < u^*$ and $u(a) < u(b)$. If a is stochastically revealed preferred to b then there must exist a sequence of alternatives $c_1, \ldots, c_N$ and choice sets $A_1, ..., A_{N-1}$, such that $c_1 = a$, $c_N = b$, $c_n, c_{n+1} \in A_n$ and $c_n \in C(A_n)$. If $u(b) > u^*$ then it must be the case that $u(c_{N-1}) > u^*$ (otherwise c_{N-1} could not be chosen with positive probability when c_N was available). Iterating on this argument implies that $u(a) > u^*$. If $u(b) < u^*$ then this implies that $u(c_{N-1}) > u(b)$. If $u(c_n) < u^*$ for all n then iterating on this argument implies that $u(a) \geq u(b)$. Otherwise, the previous argument implies that

$u(a) > u^*$. Either provides a contradiction.

Now we prove that (2) implies (1). We say two items are revealed stochastically indifferent if a is revealed stochastically preferred to b and b is revealed stochastically preferred to a, in that case we denote aI^*b. We modify this relation $I^* \subseteq X \times X$ by removing its elements that have at least one item from always chosen set W^* that is non-empty by SARP. Formally we define the relation $I = \{(a,b) \in X \times X : (a,b) \in I^* \quad a \in X \backslash W^* \quad or$ $b \in X \backslash W^*\} \cup D(W^*)$ where $D(W^*)$ is the diagonal ordering in W^* (i.e., it contains only the elements $(a,a) \in W^* \times W^*$). I is an equivalence relation because it is reflexive, symmetric and transitive. Because Axiom 2 holds I^* is an equivalence relation, and I is still an equivalence relation because it only eliminates the indifference of the items in W^* except for reflexivity, namely the elements $(a,a) \in I^*$ for $a \in W^*$. By Axiom 2 and the definition of W^* no item in $X \backslash W^*$ is revealed indifferent to an item in W^*. The relation I induces an equivalence class that we denote as $[a]$. We concentrate on the quotient set $X^I = X/I$, we define the canonical projection $j : X \mapsto X/I$ and its inverse mapping $j^{-1} : X/I \mapsto X$. We let $\mathcal{D}^I \equiv \{j(A)\}_{A \in \mathcal{D}}$ be the indexed set by $\mathcal{D}$. In particular define $p_I : X^I \times \mathcal{D}^I \mapsto [0,1]$ as $p(a^I, A^I) = \sum_{a \in j^{-1}(a^I) \cap A} p(a, A)$ for $A \in \mathcal{D}$ such that $j(A) = A^I$ and $a^I \in A^I$, this mapping is well defined. If Axiom 2 holds it follows that the quotient dataset $\{p(a^I, A^I)\}_{a^I \in X^I, A^I \in \mathcal{D}^I}$ also satisfies SARP. Also observe, that in the quotient dataset the always chosen set $W^{I,*} = \{a^I \in X^I : p(a^I, A^I) > 0 \quad \forall A^I \in \mathcal{D}^I\}$ is such that $j^{-1}(a^I) \in W^*$for all $a^I \in W^{I,*}$, this follows from the construction of I because the equivalence classes in W^* are singletons. Observe also that Axiom 1 holds in the quotient dataset $\{p(a^I, A^I)\}_{a^I \in X^I, A^I \in \mathcal{D}^I}$, because if $b^I \in X^I \backslash W^{I,*}$ then either $p(b^I, A^I) = 0$ or (exclusively) $p(b^I, A^I) = 1$. In fact, the set X^I is a finite choice set with elements associated with degenerates probabilities of choice $p([a] \cap A, A) = \sum_{a \in [a] \cap A} p(a, A) \in \{0, 1\}$ if $[a] \subseteq X \backslash W^*$. To see this is true assume that $p([a] \cap A, A) \in (0,1)$ and $[a] \subseteq X \backslash W^*$, this means that there is a third element $c \in A$ such that $c \in A \backslash [a]$, that is stochastically revealed preferred to all $a \in [a]$ (i.e, $p(c, A) > 0$) and of course all elements $a \in [a]$ are stochastically revealed preferred to c, but that means that aIc for all $a \in [a]$ which means that $c \in [a]$, this is a contradiction. By theorem 1 we conclude that the quotient dataset $\{p(a^I, A^I)\}_{a^I \in X^I, A^I \in \mathcal{D}^I}$ can be generated by a FSSM without indifference, thus we build a triple $(\overline{u}, u^* \{\overline{\gamma}\}_{A^I \in \mathcal{D}^{\mathcal{I}}})$, that generates the quotient dataset.

With this in hand we build the FSSMI in the actual dataset $\{p(a, A)\}_{a \in x, A \in \mathcal{D}}$. (i) We build a utility function $u : X \mapsto \mathbb{R}$, by the composition $u = \overline{u} \circ j$ where j is the canonical projection defined above. By construction $u(a) > u^*$ for all $a \in W^*$ and $u(a) = u(b)$ if

$b \in [a]$. Moreover, $u(b) < u^*$ for all $b \in X \backslash W^*$.

(ii) The search probabilities are defined over the quotient set, we build search probabilities for the actual set X. $\overline{\gamma}$ defines a full support search distribution on each $A^I \in \mathcal{D}^I$. Now we define γ by the following algorithm: For each menu $A \in \mathcal{D}$, we obtain the menu $A^I \equiv j(A) \in \mathcal{D}^I$ in the quotient dataset, then take R_{A^I} the support of $\overline{\gamma}_{A^I}$, now for any element $r_{A^I} \in R_{A^I}$ define the restriction $r_{A^I}|W^{I,*}$. Now build the set of linear search orders on A $R_A(r_{A^I}|W^{I,*}) = \{r_A \in R_A : a, b \in W^*, ar_Ab \quad if \quad j(a)r_{A^I|W^{I,*}}j(b)\}$. We assign to each element $r_A \in R_A(r_{A^I}|W^{I,*})$, the probability $\gamma_A(r_A) = \sum_{r'_{A^I} \in R_{A^I}} \overline{\gamma}_{A^I}(r'_{A^I})\mathbf{1}(r'_{A^I}|W^{I,*} = r_{A^I}|W^{I,*})/|R_A(r_{A^I}|W^{I,*})|$. This construction provides as with $\{\gamma_A\}_{A \in \mathcal{D}}$ that defines a FS random linear ordering on each $\mathcal{D}$.

(iii) To build the menu dependent tie breaking rules we calibrate them as follows $T(a, A) = p(a, A)$ if $a \in X \backslash W^*$ and $a \in [a]$ such that $p([a] \cap A, A) = 1$. If $p([a] \cap A, A) = 0$ then we let $T(a|A^{\sim a}) = 1/|[a] \cap A|$. This guarantees a tie breaking rule that is always positive and that adds up to 1 as required.

We have generated a tuple $(u, u^*, \{\gamma_A\}_{A \in \mathcal{D}}, T)$ or a FSSMI representation that generates the complete dataset $\{p(a, A)\}_{a \in X, A \in \mathcal{D}}$. To verify this claim, notice that this follows immediately from applying Theorem 1 to generate the quotient dataset, and noticing that for non-satisficing elements we can generate the actual dataset using the calibrated tie breaking rule directly and observing that the elements in $W^{I,*}$ have a one to one correspondence to the elements in W^*. ■

A.11 Proof of Theorem 7

Proof. (1)-(3) follows from Theorem 4.

To prove (4) notice that, given $U^* = \bar{U}^*$, for all $a \notin U^*$, if $T(a|A^{\sim a}) \neq \bar{T}(a|A^{\sim a})$ then, from the definition of the model $p(a, A) \neq p(a, A)$. ■

A.12 Proof of Theorem 8

Proof. To prove (1) notice that, since we do not allow for ties, $p(a, A) \in (0, 1)$ if $a \in U^*$ then $\widetilde{W} \subseteq U^*$. Moreover if $a \in U^*$ then, given the full support assumption, $p(a, A) > 0$ for all $A \in \mathcal{D}$ then $a \in W^*$ as in Theorem 1.

(2) follows from Axiom 2. Notice that SARP does not requires complete data sets to guarantee the existence of a utility function that represents the revealed preference relation.

(3) follows from the definition of the model. Note that identification is only possible when surely revealed satisficing elements are available together in a menu. That is, we can identify the probability of a seen first than b in $A \in \mathcal{D}$, i.e. $\gamma_A(ar_Ab \quad \forall b \in (A \cap U^*)\backslash\{a\})$ if $a \in A \cap U^*$. Moreover notice that if this is the case, then $a, b \in \tilde{W}$. ■

A.13 Proof of Lemma 2

Proof. (I) If the data is generated by FSSM-RT then it satisfies SARP.

First notice that if $\tau(a) > 0$ then it follows under full support that $p(a, A) > 0$. Furthermore, by full support and the restrictions on the random threshold we observe also that $p(a, A) > 0$ when there is a $b \in A\backslash\{a\}$ such that $\tau(b) > 0$ only if $\tau(a) > 0$. Notice that if there is a $b \in A\backslash\{a\}$ with $\tau(b) > 0$ and $\tau(a) = 0$, then by monotonicity of the CDF F_{u^*} of the random threshold we know that it cannot be the case that $u(a) > u(b)$, thus $p(a, A) = 0$ (because $\tau(a) = 0$ implies the term $\mathbf{1}(u(a) > u(b) \; \forall \; b \in A\backslash\{a\}) = 0$) under the RT assumption).

To show that Axiom 2 holds, assume, by the way of contradiction, that (i) a is stochastically revealed preferred to b and (ii) b is stochastically strictly revealed preferred to a. From (ii), given that data admits a FSSM-RT we must have (by the full support assumption) that $\tau(a) = Pr(u(a) > u^*) = 0$ and $u(a) < u(b)$. If a is stochastically revealed preferred to b then there must exist a sequence of alternatives $c_1, ..c_N$ and choice sets $A_1, ..., A_{N-1}$, such that $c_1 = a$, $c_N = b$, $c_n, c_{n+1} \in A_n$ and $c_n \in C(A_n)$. If $\tau(b) = Pr(u(b) > u^*) > 0$ then it must be the case that $\tau(c_{N-1}) = Pr(u(c_{N-1}) > u^*) > 0$ (otherwise c_{N-1} could not be chosen with positive probability when c_N was available). Iterating on this argument implies that $\tau(a) = Pr(u(a) > u^*) > 0$.

If $\tau(b) = Pr(u(b) > u^*) = 0$ then this implies that $u(c_{N-1}) > u(b)$. If $Pr(u(c_n) > u^*) = 0$ for all n then iterating on this argument implies that $u(a) > u(b)$. Otherwise, the previous argument implies that $\tau(a) = Pr(u(a) > u^*) > 0$. Either provides a contradiction.

(II) A dataset that has a FSSM-RT representation satisfies the Deterministic no satisficing choice axiom.

Observe that $W^* = \{a \in X | \tau(a) = Pr(u(a) > u^*) > 0\}$ under full support, in that sense when $a \in X \backslash W^*$ and we have a menu A such that $a \in A$: (i) Either there is a $b \in W^* \cap A$ in which case $u(b) > u(a)$ because $\tau(b) > \tau(a) = 0$ only if $u(b) > u(a)$. This is because $Pr(u(b) > u^*) > 0 = Pr(u(a) > u^*)$ only if $u(b) > u(a)$ thus $p(a, A) = 0$. (ii) Or there is no $b \in W^* \cap A$ in which case $1 - P(A) = 1$, thus making the probability $p(a, A) = \mathbf{1}(u(a) > u(c) \; \forall \; c \in A)$ which is by definition either $p(a, A) = 0$ or $p(a, A) = 1$. ■

A.14 Proof of Theorem 9

Proof. First we prove that (1) implies (2).

If a dataset is generated by a FSSM with parameters $\{\{\gamma_A\}_{A \in \mathcal{D}}, u, \overline{u}^*\}$ then it can be generated by a FSSM-RT $\{\{\gamma_A\}_{A \in \mathcal{D}}, u^{RT}, \tau\}$ with $\tau(a) = Pr(u(a) > \overline{u}^*)$ where $\overline{u}^*$ is a constant random variable, such that $\tau(a) = 1$ for all elements in FSSM such that $u(a) > \overline{u}^*$ and $\tau(b) = 0$ for $u(b) < \overline{u}^*$, with the same utility $u^{RT} \equiv u$ and the same random search function $\gamma_A^{RT} \equiv \gamma_A$. In other words, FSSM is a special case of FSSM-RT with a constant threshold.

Now we prove that (2) implies (1).

If a dataset is generated by FSSM-RT by Lemma 2 it satisfies satisfies Deterministic no satisficing choice (Axiom 1), SARP (Axiom 2). Thus by Theorem 1 we can build an FSSM that generates the data. Thus (2) implies (1). ■

A.15 Proof of Lemma 3

Before proving Lemma 3 we prove auxiliary lemmata. Note that these will be also used in the proof for Theorem 10.

A.15.1 Preliminaries

Proof of Lemma 4

Lemma 4 *A weighed sum of totally monotonic mappings* $\hat{p}_i : X \times \mathcal{D} \mapsto [0, 1]$ *for* $i \in \{1, \cdots, I\}$ *with* $I \geq 1$ *an integer, evaluated at some* (a, A) *such that* $a \in A$ *and* $A \in \mathcal{D}$, $\hat{P}(a, A, \omega) = \sum_{i=1}^{I} \omega_i \hat{p}_i(a, A)$ *with* $\omega_i \geq 0$ *a non-negative weight also satisfies total monotonicity.*

Proof. In order to prove the total monotonicity condition for $\hat{P}(a, A, \omega)$, we first need to define the following function for all $i \in \{1, \cdots, I\}$ for each $A \in \mathcal{D}$ and $a \in A$:

$$f_i(a, A) = \sum_{D \in \mathbf{B}(A)} (-1)^{|D \setminus A|} \hat{p}_i(a, D)$$

where $\mathbf{B}(A)$ is the class of supersets of A (i.e., $\mathbf{B}(A) \equiv \{D \in \mathcal{D} | A \subseteq D\}$).

Now define $f(a, A) = \sum_{D \in \mathbf{B}(A)} (-1)^{|D \setminus A|} \hat{P}(a, A, \omega)$. Observe that $f(a, A) = \sum_{D \in \mathbf{B}(A)} (-1)^{|D \setminus A|} \sum_{i=1}^{I} \omega_i \hat{p}_i(a, A)$ by definition. We can interchange the first summation operator with the the second summation operator because the first does not depend on i, then $f(a, A) = \sum_{i=1}^{I} \omega_i \sum_{D \in \mathbf{B}(A)} (-1)^{|D \setminus A|} \hat{p}_i(a, A)$ which implies:

$$f(a, A) = \sum_{i=1}^{I} \omega_i f_i(a, A)$$

by the assumption that each of the mappings $\hat{p}_i$ is totally monotonic we have that $f_i(a, A) \geq 0$ for all $i \in \{1, \cdots, I\}$ thus establishing the result. ■

Proof of Lemma 5

Lemma 5 *The mapping* $\hat{P} : X \times \mathcal{D} \mapsto [0, 1]$ *defined by* $(a, A) \mapsto T(A)\mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\})$ *for a fixed* $a \in A$ *is totally monotone, where* $T(A) = \prod_{c \in A}(1 - \tau(c))$.

Before the proof of Lemma 5 we need the following preliminaries.

Remark 1 *It will be useful to notice that the FDSM-RT can be written in the following form:*

$$p(a, A) = \sum_{r_x \in R_X} p^{MM}_{r_A = r_X | A, \tau}(a, A)\gamma(r_X) + p^{MM}_{\tau}(o, A)\mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\})$$

where $p^{MM}_{r_A = r_X | A, \tau}(a, A) = \tau(a) \prod_{b \in A : b r_A a ; r_A = r_X | A}(1 - \tau(b))$ *is numerically equivalent to the MM probability of* $a \in A$ *being chosen and* $r_X | A$ *is the restriction of the ordering* r_X *to the set* A. *And* $p^{MM}_{\tau}(o, A) = \prod_{c \in A}(1 - \tau(c))$ *is numerical equivalent to the MM probability of the default alternative. Here there is no default alternative and we mean by* $p^{MM}(o, A)$ *the probability of no element in* A *being satisficing (keeping the MM notation for intuition).*

Definition 16 (Successive differences for mappings) *The successive differences for a mapping* $\hat{p} : X \times \mathcal{D} \mapsto [0, 1]$ *for a fixed* $a \in A$ *and* $A \in \mathcal{D}$ *and the probability* $\hat{p}(a, A)$, *is* *defined recursively as:*

$\Delta_{A_1}\hat{p}(a,A) = \hat{p}(a,A) - \hat{p}(a,A\cup A_1)$ *for* $A, A_1 \in \mathcal{D}$,

$\Delta_{A_n}\cdots\Delta_{A_1}\hat{p}(a,A) = \Delta_{A_{n-1}}\cdots\Delta_{A_1}\hat{p}(a,A) - \Delta_{A_{n-1}}\cdots\Delta_{A_1}\hat{p}(a,A\cup A_n)$ *for all* $n \geq 2$ *and for all* $A, A_1\cdots A_n \in \mathcal{D}$.

Definition 17 (Weakly increasing successive differences) *For a mapping* $\hat{p} : X \times \mathcal{D} \mapsto [0,1]$ *and a fixed* $a \in X$*, and all* $A \in \mathcal{D}$ *such that* $a \in A$ *and any* $\{A_i\}_{i=1}^n \in \mathcal{D}^n$ *the mapping* $\hat{p}$ *satisfies the weakly increasing successive difference property when the successive differences are non-negative* $\Delta_{A_n}\cdots\Delta_{A_1}\hat{p}(a,A) \geq 0$ *for all* $n \geq 1$.

Proof. Notice that any mapping $\hat{p} : X \times \mathcal{D} \mapsto [0,1]$ for a fixed (a,A) such that $a \in A$ for all $A \in \mathcal{D}$ is totally monotone if and only if it satisfies the weakly increasing successive differences (Molchanov (2005)).

Define $\hat{P}(a,A) = T(A)\mathbf{1}(u(a) > u(b)\forall b \in A\backslash\{a\})$ where $T(A) = \prod_{c\in A}(1-\tau(c))$ is totally monotone by MM (since $T(A) = p_{MM,\tau}(o,A)$ is numerically equivalent to the probability of a default choice in MM as explained in the remark) and thus satisfies for all $A \in \mathcal{D}$ such that $a \in A$ and any $\{A_i\}_{i=1}^n \in \mathcal{D}^n$ the weakly increasing successive differences $\Delta_{A_n}\cdots\Delta_{A_1}(1-P(A)) \geq 0$ for all $n \geq 1$, the same is true for $C_a(A) = \mathbf{1}(u(a) > u(b)\forall b \in A\backslash\{a\})$ such that $\Delta_{A_n}\cdots\Delta_{A_1}\mathbf{1}(u(a) > u(b)\forall b \in A\backslash\{a\}) \geq 0$.

Notice that

$$\begin{aligned}
\Delta_{A_1}\hat{P}(a,A) &= \hat{P}(a,A) - \hat{P}(a,A\cup A_1) \\
&= T(A)C_a(A) - T(A\cup A_1)C_a(A\cup A_1) \\
&= \begin{cases} 0 & C_a(A) = 0 \\ \Delta_{A_1}T(A) & C_a(A\cup A_1) = 1 \\ T(A) & C_a(A) = 1, C_a(A\cup A_1) = 0 \end{cases}
\end{aligned}$$

Also for the next difference $\Delta_{A_2}\Delta_{A_1}\hat{P}(a,A) = T(A)C_a(A) - T(A\cup A_1)C_a(A\cup A_1) - [T(A\cup A_2)C_a(A\cup A_2) - T(A\cup A_1\cup A_2)C_a(A\cup A_1\cup A_2)]$.

Now

$$\Delta_{A_2}\Delta_{A_1}\hat{P}(a,A) = \Delta_{A_1}\hat{P}(a,A) - \Delta_{A_1}\hat{P}(a,A\cup A_2)$$
$$= \begin{cases} 0 & C_a(A)=0 \\ \Delta_{A_1}T(A) & C_a(A\cup A_1)=1, C_a(A\cup A_2)=0 \\ T(A) & C_a(A)=1, C_a(A\cup A_1)=0, C_a(A\cup A_2)=0 \\ \Delta_{A_2}T(A) & C_a(A)=1, C_a(A\cup A_2)=1, C_a(A\cup A_1)=0 \\ \Delta_{A_2}\Delta_{A_1}T(A) & C_a(A\cup A_1\cup A_2)=1 \end{cases}.$$

Then we notice that in general the following properties hold:

(i) Null operator: $\Delta_{\emptyset}\Delta_{A_{n-1}}\cdots\Delta_{A_1}T(A)=0$,

(ii) Absorption: $\Delta_{A_{n-1}}\Delta_{A_{n-1}}\cdots\Delta_{A_1}T(A)=\Delta_{A_{n-1}}\cdots\Delta_{A_1}T(A)$.

$\Delta_{A_1}\hat{P}(a,A)\in\{\Delta_{X_1}T(A)\}$ for $X_1\in\{\emptyset,A_1\}$ and its value depends on the value of $C_a(A)$ and $C_a(A\cup A_1)$ but in any case we have

$\Delta_{X_1}T(A)\geq 0$ for all $X_1\in\{\emptyset,A_1\}$. Notice that $\Delta_{\emptyset}T(A)=0$.

Also, $\Delta_{A_2}\Delta_{A_1}\hat{P}(a,A) \in \{\Delta_{X_{2,2}}\Delta_{X_{2,1}}T(A)\}$ where $X_{2,1}\in\{\emptyset,A_1,A_2\}$ and $X_{2,2}\in\{\emptyset,A_1,A_2\}$ where the actual combination depends on the values of the vector $\{C_a(A\cup X_{2,1}\cup X_{2,2})\}_{X_{2,1}\in\{\emptyset,A_1,A_2\},X_{2,2}\in\{\emptyset,A_1,A_2\}}$ again in any case:

$\Delta_{X_{2,2}}\Delta_{X_{2,1}}T(A)\geq 0$ where $X_{2,1}\in\{\emptyset,A_1,A_2\}$ and $X_{2,2}\in\{\emptyset,A_1,A_2\}$. Notice that $\Delta_{\emptyset}\Delta_{X_{2,1}}T(A)=\Delta_{X_{2,1}}T(A)-\Delta_{X_{2,1}}T(A)=0$, $\Delta_{A_2}\Delta_{A_2}T(A)=\Delta_{A_2}T(A)-\Delta_{A_2}T(A\cup A_2)=\Delta_{A_2}T(A)$.

The induction hypothesis is:

$\Delta_{A_{n-1}}\cdots\Delta_{A_1}\hat{P}(a,A)\in\{\Delta_{X_{n-1,n-1}}\cdots\Delta_{X_{n-1,1}}T(A)\}$ for $X_{n-1,i}\in\{\emptyset,A_1,A_2,\cdots A_{n-1}\}$ for all $i\in\{1,\cdots,n-1\}$,

with $\Delta_{X_{n-1,n-1}}\cdots\Delta_{X_{n-1,1}}T(A)\geq 0$ for any $X_{n-1,i}\in\{\emptyset,A_1,A_2,\cdots A_{n-1}\}$ for all $i\in\{1,\cdots,n-1\}$.

We have to prove that

$$\Delta_{A_n} \cdots \Delta_{A_1} \hat{P}(a, A) \geq 0,$$

Notice that $\Delta_{A_n} \cdots \Delta_{A_1} \hat{P}(a, A) = \Delta_{A_{n-1}} \cdots \Delta_{A_1} \hat{P}(a, A) - \Delta_{A_{n-1}} \cdots \Delta_{A_1} \hat{P}(a, A \cup A_n)$ by definition.

Now by the induction step:

$\Delta_{A_n} \cdots \Delta_{A_1} \hat{P}(a, A) \in \{\Delta_{X_{n-1,n-1}} \cdots \Delta_{X_{n-1,1}} T(A) - \Delta_{X_{n-1,n-1}} \cdots \Delta_{X_{n-1,1}} T(A \cup A_n)\}$ for $X_{n-1,i} \in \{\emptyset, A_1, A_2, \cdots A_{n-1}\}$ for all $i \in \{1, \cdots, n-1\}$.

Finally by the definition of the difference operator:

$\Delta_{X_{n-1,n-1}} \cdots \Delta_{X_{n-1,1}} T(A) - \Delta_{X_{n-1,n-1}} \cdots \Delta_{X_{n-1,1}} T(A \cup A_n) = \Delta_{A_n} \Delta_{X_{n-1,n-1}} \cdots \Delta_{X_{n-1,1}} T(A)$.

We notice that $\Delta_{A_n} \Delta_{X_{n-1,n-1}} \cdots \Delta_{X_{n-1,1}} T(A) \geq 0$ is positive by total monotonicity of T, that preserves works for any combination of $X_{n-1,i} \in \{\emptyset, A_1, A_2, \cdots A_{n-1}\}$ for all $i \in \{1, \cdots, n-1\}$.

Thus $\Delta_{A_n} \cdots \Delta_{A_1} \hat{P}(a, A) \geq 0$ for any fixed $a \in A$ and all such $A \in \mathcal{D}$ and for all $\{A_i\}_{i=1}^n \in \mathcal{D}^n$ for all $n \geq 1$.

We conclude that $\hat{P}$ is a total monotone mapping. ■

A.15.2 Proof of Lemma 3

Proof. (I) If the data is generated by FDSM-RT then $p(a, A)$ satisfies total monotonicity for all $a \in A$ and all $A \subseteq X$.

We define the MM model for a fixed linear search r_A as $p^{MM}_{r_A,\tau}(a, A) = \tau(a) \prod_{b \in B_{r_A}(a)} (1 - \tau(b))$ and $p^{MM}_{\tau}(o, A) = \prod_{c \in A} (1 - \tau(c))$. Now notice that the FDSM-RT can be written as the average of the MM models, plus a correction for non satisficing cases.

$$p(a, A) = \sum_{r_x \in R_X} p^{MM}_{r_A = r_X | A, \tau}(a, A) \gamma(r_X) + p^{MM}_{\tau}(o, A) \mathbf{1}(u(a) > u(b) \; \forall \; b \in A \backslash \{a\})$$

It is evident by MM that $p^{MM}_{r_A,\tau}(a,A)$ satisfies total monotonicity for all $r_A \in R_A$ therefore by Lemma 4 we know that a summation (here a convex combination) of total monotonic mappings also satisfies total monotonocity, thus the term $\sum_{r_x \in R_X} p^{MM}_{r_A=r_X|A,\tau}(a,A)\gamma(r_X)$ satisfies total monotonicity.[14] The same argument is used for $p^{MM}_{\tau}(o,A)$ that by MM is also totally monotonic.

Also, notice that the indicator function $\mathbf{1}(u(a) > u(b)\ \forall b \in A\backslash\{a\})$ satisfies total monotonicity because it can be understood as a degenerate random utility distribution. We know also that $p^{MM}_{\tau}(o,A)\mathbf{1}(u(a) > u(b) \forall b \in A\backslash\{a\})$ is totally monotonic by Lemma 5.[15]

Since $p(a,A)$ is the summation of two totally monotonic mappings again by Lemma 4 we conclude that $p(a,A)$ satisfies total monotonicity.

(II) If the data is generated by FDSM-RT then it satisfies SARP. This follows from the fact that FDSM-RT is an special case of FSSM-RT and by Lemma 2 we know that it satisfies SARP.

(III) A dataset that has a FDSM-RT representation satisfies the Deterministic no satisficing choice axiom. This follows from the fact that FDSM-RT is an special case of FSSM-RT and by Lemma 2 we know that it satisfies the Deterministic no satisficing choice axiom. ■

A.16 Proof of Theorem 10

Proof. First we prove that (1) implies (2).

If a dataset is generated by a FDSM with parameters $\{\{\gamma_A\}_{A\in\mathcal{D}}, u, \overline{u}^*\}$ then it can be generated by a FDSM-RT $\{\{\gamma_A\}_{A\in\mathcal{D}}, u^{RT}, \tau\}$ with $\tau(a) = Pr(u(a) > \overline{u}^*)$ where $\overline{u}^*$ is a constant random variable, such that $\tau(a) = 1$ for all elements in FDSM such that $u(a) > \overline{u}^*$ and $\tau(b) = 0$ for $u(b) < \overline{u}^*$, with the same utility $u^{RT} \equiv u$ and the same random search function $\gamma^{RT}_A \equiv \gamma_A$. In other words, FDSM is a special case of FDSM-RT with a constant threshold.

[14] Notice that $\hat{p}(a,A) = \sum_{r_x \in R_X} p^{MM}_{r_A=r_X|A,\tau}(a,A)\gamma(r_X)$ satisfies total monotonicity however this term is a quasi-probability because it does not necessarily adds up to 1, so it is not a RUM in general. RUM is equivalent to total monotonicity, being non negative and adding up to 1.

[15] A non-constructive but equally valid argument is that $1 - p^{MM}_{\tau}(o,A)\mathbf{1}(u(a) > u(b)\ \forall b \in A\backslash\{a\})$ can be understood as the capacity of a random set that is the union of two other two independent random sets with capacities $1 - p^{MM}_{\tau}(o,A)$ and $(1 - \mathbf{1}(u(a) > u(b) \forall b \in A\backslash\{a\}))$ corresponding to the random sets $\hat{W} = \{x \in X : u(a) > u^*\}$ where u^* is the random threshold and $B(a) = \{b \in X : u(b) > u(a)\}$ that is the deterministic set of better than a items under the utility u. By Molchanov (2005) $p^{MM}_{\tau}(o,A)\mathbf{1}(u(a) > u(b)\ \forall b \in A\backslash\{a\})$ is totally monotone, and $p^{MM}_{\tau}(o,A)\mathbf{1}(u(a) > u(b)\ \forall b \in A\backslash\{a\}) = Pr(\hat{W} \cup B(a) \cap A = \emptyset)$ the probability that A does not have anything satisficing or better than a.

Now we prove that (2) implies (1).

If a dataset is generated by FDSM-RT by Lemma 3 it satisfies satisfies Deterministic no satisficing choice (Axiom 1), SARP (Axiom 2) and Total Monotonicity (Axiom 3). Thus by Theorem 3 we can build an FDSM that generates the data. Thus (2) implies (1). ∎

A.17 Proof of Theorem 11

In order to establish Theorem 11, we will establish an equivalent equivalence statement, with the help of the following property.

Definition 18 (Always satisficing random utility distribution) *A random utility distribution is always satisficing if for any item $x \in X$ such that for some $u \in \mathcal{U}$ with $\rho(u) > 0$, and $u(x) > u^*$ it follows that for any other $\hat{u} \in \mathcal{U}$ with $\rho(\hat{u}) > 0$ it also holds that $\hat{u}(x) > u^*$.*

The following are equivalent:

1. A complete stochastic choice dataset $(\mathcal{D}, p)$ can be generated by a RU-FDSM.
2. A complete stochastic choice dataset $(\mathcal{D}, p)$ satisfies Axiom 3 and Degeneracy (Axiom 4).
3. A complete stochastic choice dataset $(\mathcal{D}, p)$ can be generated by a RU-FDSM with Always satisficing random utility distribution.

Proof. First we prove that (1) implies (2).

First we recall that for a fixed $u \in \mathcal{U}$, $p_u^{FDSM}(a, A)$ satisfies Total Monotonicity (Axiom 3). By Lemma 4, the weighted average of totally monotonic mappings is also totally monotone $p(a, A) = \sum_{u \in \mathcal{U}} \rho(u) p_u^{FDSM}(a, A)$, thus RU-FDSM is totally monotone.

Second we notice that if $x \in X \backslash W^*$ it is never satisficing that is for all $u \in \mathcal{U}$ with $\rho(u) > 0$, $u(x) < u^*$, this means that for any fixed $u \in \mathcal{U}$, $p_u^{FDSM}(x, A) = 0$ if $A \cap W^* \neq \emptyset$, this in turn implies that $p(x, A) = \sum_{u \in \mathcal{U}} \rho(u) p_u^{FDSM}(x, A) = 0$. Thus degeneracy is established for RU-FDSM.

Now we prove that (2) implies (3).

With Claim 2 in hand, we define the following virtual dataset. We define the equivalence class $\sim= \{(a,b) \in X \times X : a \in X\backslash W^*, b \in X\backslash W^*\} \cup \{(c,c) \in X \times X : c \in W^*\}$ (it is symmetric, reflexive and transitive).

We define the set $X^\sim = X/\sim$ as the quotient space with respect to the equivalence relation. In words, we are "shrinking" all never satisficing elements to a singleton. We concentrate on the quotient set $X^\sim = X/\sim$, we define the canonical projection $j : X \mapsto X/\sim$ and its inverse mapping $j^{-1} : X/\sim \mapsto X$. We let $\mathcal{D}^\sim \equiv \{j(A)\}_{A\in\mathcal{D}}$ be the indexed set by $\mathcal{D}$. In particular define $p_\sim : X^\sim \times \mathcal{D}^\sim \mapsto [0,1]$ as $p(a^\sim, A^\sim) = \sum_{a\in j^{-1}(a^\sim)\cap A} p(a,A)$ for $A \in \mathcal{D}$ such that $j(A) = A^\sim$ and $a^\sim \in A^\sim$, this mapping is well defined. If Degeneracy (Axiom 4) holds it follows that the quotient dataset $\{p(a^\sim, A^\sim)\}_{a^\sim\in X^\sim, A^\sim\in\mathcal{D}^\sim}$ satisfies Deterministic not satisficing choice (Axiom 1) in $X^\sim$, because defining $W^{*\sim} = \{a^\sim \in X^\sim | p_\sim(a^\sim, A^\sim) > 0 \ \forall \ A^\sim \in \mathcal{D}^\sim\}$, we notice $X^\sim\backslash W^{\sim *}$ corresponds by construction to a singleton $\{z^\sim\} \equiv X^\sim\backslash W^{\sim *}$ when there is non-satisficing element (or empty if not) with probability $p(z^\sim, A^\sim) = \sum_{a\in j^{-1}(z)\cap A} p(a,A) \in \{0,1\}$ in all $A^\sim \in \mathcal{D}^\sim$. In particular, it is exactly 1 only when $A^\sim \equiv \{z^\sim\}$ and zero otherwise. Also notice that Degeneracy implies SARP in $\{p(a^\sim, A^\sim)\}_{a^\sim\in X^\sim, A^\sim\in\mathcal{D}^\sim}$, because all $x^\sim, y^\sim \in W^{*\sim}$ are declared stochastically revealed preferred to one another (i.e., stochastically revealed indifferent) and $x^\sim \in W^{*\sim}$ is always strictly revealed preferred to $z^\sim$ (or the non-satisficing singleton $\{z^\sim\} \equiv X^\sim\backslash W^{\sim *}$). Thus $C^\sim(A^\sim) = \{a^\sim \in A^\sim : p_\sim(a^\sim, A^\sim) > 0\}$ satisfies SARP. Also it is simple to see that the dataset in the quotient space $X^\sim$ also satisfies total monotonicity because $p(a^\sim, A^\sim) = \sum_{a\in j^{-1}(a^\sim)\cap A} p(a,A)$ for $A \in \mathcal{D}$ such that $j(A) = A^\sim$ and $a^\sim \in A^\sim$ is a sum of total monotonic mappings $p(a,A)$ by (2) Total Monotonicity (Axiom 3). By the theorem that characterizes the FDSM (Theorem 3) the dataset $\{p(a^\sim, A^\sim)\}_{a^\sim\in X^\sim, A^\sim\in\mathcal{D}^\sim}$ in the quotient space has a FDSM representation, thus we build a triple $(\overline{u}, u^*, \{\overline{\gamma}\}_{A^\sim\in\mathcal{D}^\sim})$, that generates the quotient dataset.

With this in hand we build the RU-FDSM in the actual dataset $\{p(a,A)\}_{a\in X, A\in\mathcal{D}}$. (i) We build a utility function $u : X \mapsto \mathbb{R}$, by the composition $u = \overline{u} \circ j$ where j is the canonical projection defined above. By construction $u(a) > u^*$ for all $a \in W^*$ and $u(b) < u^*$ for all $b \in X\backslash W^*$. We build now $(\rho, \mathcal{U})$, by making all elements in it have the following restriction $\hat{u} \in \mathcal{U}$ with $\rho(\hat{u}) > 0$ is such that $\hat{u}(a) = u(a)$ for all $a \in W^*$. For $b \in X\backslash W^*$ we use total monotonicity (Axiom 3) that holds for the restricted dataset $\{p(a,A)\}_{a\in X\backslash W^*, A\in\mathcal{D}, A\subseteq X\backslash W^*}$ to obtain a random utility by Falmagane (1978) defined on $X\backslash W^*$, we notice that under this $p(a,A) = \rho(\hat{u} : \hat{u}(a) > \hat{u}(b) \forall b \in A\backslash\{a\})$ for $a \in A \subseteq X\backslash W^*$, we fix an injective utility $\hat{u}_{X\backslash W^*} : X\backslash W^* \mapsto \mathbb{R}$ with $\rho(\hat{u}_{X\backslash W^*})$ compatible with $\rho(\hat{u}_{X\backslash W^*} : \hat{u}_{X\backslash W^*}(a) > \hat{u}_{X\backslash W^*}(b)\forall b \in A\backslash\{a\})$ with $A \subseteq X\backslash W^*$ (we omit the trivial construction). We extend $\hat{u}_{X\backslash W^*} : X\backslash W^* \mapsto \mathbb{R}$ to the grand set X, by defining $\hat{u} : X \mapsto \mathbb{R}$ using $\hat{u}(a) = u(a)$ for all $a \in W^*$ and $\hat{u}(b) = \hat{u}_{X\backslash W^*}$ for all $b \in X\backslash W^*$, this is an injective utility in X with mass $\rho(\hat{u}) = \rho(\hat{u}_{X\backslash W^*})$. We have built $(\rho, \mathcal{U})$ that has the Always satisficing random utility property.

(ii) The search probabilities are defined over the quotient set, we build search probabilities for the actual set X. $\overline{\gamma}$ defines a full support search distribution on each $A^{\sim} \in \mathcal{D}^{\sim}$. Now we define γ by the following algorithm: For each menu $A \in \mathcal{D}$, we obtain the menu $A^{\sim} \equiv j(A) \in \mathcal{D}^{\sim}$ in the quotient dataset, then take $R_{A^{\sim}}$ the support of $\overline{\gamma}_{A^{\sim}}$, now for any element $r_{A^{\sim}} \in R_{A^{\sim}}$ define the restriction $r_{A^{\sim}}|W^{\sim *}$. Now build the set of linear search orders on A $R_A(r_{A^{\sim}}|W^{\sim,*}) = \{r_A \in R_A : a, b \in W^*, ar_Ab \quad if \quad j(a)\ r_{A^{\sim}|W^{\sim *}}\ j(b)\}$. We assign to each element $r_A \in R_A(r_{A^{\sim}}|W^{\sim *})$, the probability $\gamma_A(r_A) = \sum_{r_{A^{\sim\prime}} \in R_{A^{\sim}}} \overline{\gamma}_{A^{\sim}}(r_{A^{\sim\prime}})\mathbf{1}(r_{A^{\sim\prime}}|W^{\sim *} = r_{A^{\sim}}|W^{\sim *})/|R_A(r_{A^{\sim}}|W^{\sim *})|$. This construction provides as with $\{\gamma_A\}_{A\in\mathcal{D}}$ that defines a Full Support random linear ordering on each $\mathcal{D}$ and by Total Monotonicity (Axiom 3) $\overline{\gamma}$ has the Fixed Distribution property and we build γ on the basis on it, extending it to the original dataset with a uniform rule, it follows that $\{\gamma_A\}_{A\in\mathcal{D}}$ has the Fixed Distribution property.

It is clear that for fixed $u \in \mathcal{U}$ constructed above with $\rho(u) > 0$, we have an FDSM with $(u, u^*, \{\gamma_A\}_{A\in\mathcal{D}}\})$ where $u, \{\gamma_A\}_{A\in\mathcal{D}}$ are constructed above and u^* is the same as the FDSM in the quotient space. Thus we have built a RU-FDSM $p(a, A) = \sum_{u\in\mathcal{U}} \rho(u)p_u^{FDSM}(a, A)$ with FDSM with $(u, u^*, \{\gamma_A\}_{A\in\mathcal{D}}\})$ with the always satisficing property for $(\rho, \mathcal{U})$. Finally, this generates the dataset $(p, \mathcal{D})$ by noticing that for $a \in W^*$ $p(a, A) = p_{\overline{u}}^{FDSM}(a, A)$ for any fixed $\overline{u} \in \mathcal{U}$ with $\rho(\overline{u}) > 0$ (thanks to the always satisficing property), so the fact that the constructed RU-FDSM (with always satisficing property) generates this probabilities follows from Theorem 3. For $a \in X\backslash W^*$, we have $p(a, A) = \mathbf{1}(\forall c \in A : \overline{u}(c) \leq u^*)\rho(u : u(a) > u(b)\forall b \in A\backslash\{a\})$ for any fixed $\overline{u} \in \mathcal{U}$ with $\rho(\overline{u}) > 0$ (again due to the always satisficing property), we have two cases $\mathbf{1}(\forall c \in A : \overline{u}(c) \leq u^*) = 1$, in which case the fact that the constructed RU-FDSM generates the data follows from Falmagane (1978) as these cases are equivalent to random utility in the restricted dataset defined in $X\backslash W^*$. The remaining case is $\mathbf{1}(\forall c \in A : \overline{u}(c) \leq u^*) = 0$ in which case we know that $\exists c \in A$ such that $c \in W^*$ and by Degenaracy (Axiom 4) we conclude that $p(a, A) = 0$ as it should be. Thus the constructed RU-FDSM generates the data.

Finally we prove (3) implies (1).

We notice that if complete stochastic choice dataset $(\mathcal{D}, p)$ can be generated by a RU-FDSM with Always satisficing random utility distribution, it is by definition a special case of RU-FDSM, with the additional Always satisficing restriction on $(\rho, \mathcal{U})$ thus the dataset $(\mathcal{D}, p)$ is also generated by a RU-FDSM. ■