RANDOM (SEQUENTIAL) MULTIVALUED CHOICE*

HENRIK PETRI

Abstract. Multivalued choice rules, or choice correspondences, are effective means to describe a wide array of behaviors conforming to and deviating from rationality. Modeling heterogeneity in such behaviors is problematic as it gives rise to serious identification issues. We introduce a model of stochastic choice that is general enough to explain a high degree of heterogeneity in multivalued choice, yet with good identification properties. We provide two main applications. First, we study satisficing behavior in the spirit of Simon (1955). Second, we introduce a dynamic version of our model and use it to study long run behavior of individuals.

1. Introduction

The analysis of this paper rests on two important and, to various extents, empirically validated theses about human choice behavior.

(1) Individual behaviors are fundamentally *multivalued* and hence described by choice correspondences.
(2) Individuals display differences in their behaviors, i.e. there is *heterogeneity* in choice behaviors.

As for (1), multivalued - as opposed to single valued - choice rules are ubiquitous in describing human choice behavior. Whether dealing with behaviors conforming to or deviating from rationality, some behaviors are necessarily expressed as properties of a choice correspondence. For instance, modeling the behavior of a rational, utility maximizing, individual requires a choice correspondence in describing the potential indifference classes of the individual. Similarly, the behavior of an indecisive individual is intrinsically multivalued, and a choice correspondence is needed to describe the sets of alternatives that cause indecisiveness. What is more, distinguishing between these two modes of choice is virtually impossible with single valued choice functions. Choice correspondences are also essential in describing many behaviors outside the rational paradigm,[1] such as the behavior of individuals who only pay attention to a subset of alternatives (Masatlioglu, Nakajima, and Ozbay, 2012), use a

First version: February, 2022. *This version:* July, 2022. I thank Victor Aguiar, Jose Apesteguia, Miguel Ballester, Daniele Caliari and Georgios Gerasimou for their valuable comments.

Department of Economics, University of Bath, henrik@petri.se.

[1]Abundant evidence, as well as introspection, suggests that individuals may fail to adhere to the basic tenets of rationality (see e.g. Tversky (1969), Huber, Payne, and Puto (1982), Loomes, Starmer, and Sugden (1991), Harbaugh, Krause, and Berry (2001),Echenique, Lee, and Shum (2011),Apesteguia and Ballester (2015)).

sequence of criteria to narrow down choice (Manzini and Mariotti, 2007), or use a set of rationales/criteria to justify choice (Cherepanov, Feddersen, and Sandroni, 2013).

Concerning (2) there is pervasive evidence supporting the thesis that individuals display differences in their tastes/preferences and other behavioral traits affecting choice.[2] Modeling such choice heterogeneity poses a number of challenges for the researcher. Particularly pressing is the non-uniqueness or identification issue caused by the the large number of parameters in such models. This means that there are multiple distinct (i.e. non-unique) sets of parameters inducing exactly the same observable behavior. Roughly speaking, with many types there are many parameters compared to the number of observations in the data. Non uniqueness of the parameters then arises for much the same reason as it does in an equation system with fewer equations than unknowns. This non uniqueness or lack of identifiability of a models underlying parameters is problematic for several reasons, hindering important tasks such as parameter estimation, comparative statics and out-of-sample predictions.

To further appreciate the identification issue in models of choice heterogeneity, note that even apparently simple models of choice heterogeneity fail to be identified. An example is the popular random utility model. It posits a probability distribution on a set of preference maximizing types. The choice probability of an alternative is the sum of probabilities of types having the alternative as their top choice. The random utility model is thus only able to explain heterogeneity in tastes/preferences. Despite the limited degree of heterogeneity tolerated by the random utility model, it is well-known to suffer from identifiability issues. Even the simplest version of the random utility model, with only two distinct preference maximizing types, is unidentified.[3]

In this paper we study a generalization of the random utility model that allows for heterogeneity in choice types described by multivalued choice rules. The model thus posits a probability distribution on a set of choice correspondences. Our main result is on identification of the underlying type distribution. Since our proposed model has a larger number of heterogeneous types, and contains the random utility model as a special case, the identification issue is even more pressing. To appreciate the extent of the issue, note that there are more than *twenty million* different choice correspondences on a collection of menus consisting of a set of four alternatives and its subsets.[4] Compare this figure to the number of preference orderings (24) and it is easy to see (given the identification issues already posed

[2]There is evidence from diverse fields such as finance (Curcuru, Heaton, Lucas, and Moore, 2010) on portfolio choice, discrete choice (Train, 1998) on anglers' choices of fishing site, on commuting mode choice (Greene and Hensher, 2010) and from marketing (Draganska and Klapper, 2011) on the role of advertising on choice set heterogeneity.

[3]See e.g. Fishburn (1998) and Turansick (2021) for a discussion. Given the identifiability issues of the random utility model, Turansick (2021) discusses necessary and sufficient conditions on choice data such that it is identified.

[4]More precisely, there are $(2^4-1)(2^3-1)^4(2^2-1)^6 = 26,254,935$ choice correspondences on a menu collection consisting of a set of four elements and its subsets.

by the random utility model) that identification within our model is challenging. Due to this we focus on a special case of the general model. To ensure identification within our model we impose two important, yet flexible, assumptions.

As a first crucial assumption, we impose a sequential condition on the collection of types in our model. Informally, a collection of choice types is sequential if they can be ordered such that choice sets of higher types are subsets of those of lower types. Examples of models that satisfy this condition are abundant (several examples are in section 2) and include models where an individual sequentially narrows down choices by using a sequence of criteria. Later survivor sets are then subsets of earlier sets. A wealth of models in the literature have this structure. Manzini and Mariotti (2007) introduce and characterize a model of two stage sequential choice.[5] Their model is later generalized in several interesting directions. Notably, both Manzini and Mariotti (2012) and Apesteguia and Ballester (2013) discuss characterizations of multi-stage versions of the baseline model in Manzini and Mariotti (2007). A closely related model is also the gradual pairwise comparison rule of Dutta (2020). Another example is provided within the theory of incomplete preferences (see e.g. Aumann (1962), Eliaz and Ok (2006) and Galaabaatar and Karni (2013)). Consider an indecisive individual with an incomplete preference relation and suppose that there is heterogeneity in the degree of decisiveness of the individual (i.e. in the number of pairs of alternatives that she can compare). This behavior can be modeled using a sequential collection of choice correspondences, where choice sets corresponding to states of mind where the individual is more decisive are subsets of those where the individual is less decisive.

Our second assumption is an exogenous/observed tie-breaking or reference rule. To explain this assumption let us first explain the role of a reference/tie-breaking rule in our setting. We assume that the outside observer/researcher/analyst has access to standard stochastic choice data, which is a collection probability distributions on *single valued* choices. Whether elicited through experiments or empirically observed, this is the type of data most common in practice.[6] Since types in our model are described by choice correspondences that assign *sets* of alternatives to menus, and data appears as a probability distribution on single valued choices, our model posits that individuals use a reference/tie-breaking rule (depending on the behavior studied) to arrive at a final (single) choice. We discuss several examples of reference rules below.

[5] Dutta and Horan (2015) discuss the important issue of identification in the two stage sequential choice model.

[6] An alternative approach would be to consider a probability distribution over sets as an observed primitive. Only a few recent papers propose methods to elicit (deterministic) choice correspondences directly using (non-) forced multivalued choice tasks (see e.g. Agranov and Ortoleva (2017), Costa-Gomes, Cueva, and Gerasimou (2021), Bouacida (2021), Gerasimou (2021) and references therein). However, these papers are mainly situated in a deterministic (non-repeated) setting. To the best of our knowledge experimental or empirical data on *repeated* multivalued choices is rare. See Balakrishnan et al. (2021) for an elaboration on closely related issues.

The exogeneity assumption on the reference rule means that the researcher observes the reference rule used by individuals. This is a plausible assumption in many applications.[7] In a dynamic choice environment individuals may look at past periods' choices to inform their current choices and choose alternatives based on their popularity in previous periods (we will study this sort of behavior in section 6, where we introduce a dynamic version of our model). If the researcher has access to time series data then each periods' choices may serve as a reference rule for next period's choices and is hence observed by the researcher. In an online environment individuals may use a common set of criteria such a price in ascending/descending order, discount or popularity to choose among considered alternatives. When indifferent between alternatives, it is conceivable that individuals uniformly randomize to break ties. Or, another class of examples, is when the reference rule is induced by a strict reference order reflecting the ingrained characteristics of alternatives. Such reference orders arise in many economically relevant situations and we refer to section 4 for a further discussion.

To summarize, our model, the *Random Sequential Model* (RSM) posits a type distribution μ on a set of choice types/correspondences $(C_i)_{i \in I}$ satisfying our sequential condition,[8] such that the probability of a being chosen from menu A is

$$\rho(a, A) = \sum_{i \in I} \mu_i \pi(a, C_i(A)), \tag{1}$$

where π is an exogenous reference rule. Note that $C_i(A)$ is the multiset of choices of type i. The reference rule π is used to choose a final single alternative a from $C_i(A)$ with probability $\pi(a, C_i(A))$.

Our main result is on identification of the underlying type distribution. As explained above, we assume observability/exogeneity of the reference rule. We find characterizing conditions on the reference rule such that every stochastic choice function has a unique representation with respect to the reference rule. The main property is a weak version of regularity. Regularity, roughly speaking requires alternatives in a set to be chosen with positive probability whenever they are chosen with positive probability in a superset (a set containing it). A corollary to our identification result is that the underlying type distribution is unique if the reference rule is regular. A host of different reference rules satisfy regularity. Examples include random reference orders (which are probability distributions on reference orders) logit/Luce reference rules and uniform reference rules (formally a special case of Luce rules). The weak regularity property is also consistent with highly irregular choice patterns. For instance, any reference rule induced by a deterministic single valued choice function, satisfy this condition. After

[7]In other contexts (such as when the reference rule corresponds to a behavioral trait of the decision maker) it needs to be inferred from choice data and we discuss this issue further in section 8.1

[8]Mathematically, this condition can be expressed as $C_1(A) \supseteq ... \supseteq C_{|I|}(A)$ for all menus A.

establishing our main identification result, we apply it in section 4 to study various examples of reference rules.

To illustrate the versatility of the proposed framework we discuss three main applications. In section 5 we use the framework to study more specific sequential rules. In particular we provide characterizations of stochastic versions of Simon's satisficing procedure.[9] An agent using a satisficing procedure considers all alternatives with satisfaction exceeding some random (possibly menu dependent) threshold. A reference rule is then used to discriminate between alternatives exceeding the threshold. Such rules have a sequential structure as sets of alternatives with satisfaction levels exceeding higher thresholds are subsets of those with lower thresholds. Prior literature has mainly focused on deterministic choice (e.g. Aleskerov et al. (2007), Manzini et al. (2013) and Frick (2016)) and cases where the reference rule is uniform. We extend these results by considering an array of empirically relevant tie-breaking/reference rules. The main property is an acyclicity condition imposed on pairs of alternatives. This condition can be checked by using the algorithmic construction in the proof of our main representation result. We also discuss examples of tie-breaking rules where this condition is particularly easy to check, such as when the reference rule is a logit rule.

A second main application of our framework is in section 6 where we introduce a dynamic version of our model and use it to study long run behavior of individuals. In the dynamic setting there is a particularly illuminating interpretation of the reference rule as a proxy for previous periods' choices. In line with a large literature on social learning (see e.g. Banerjee (1992), Becker (1991), Ellison and Fudenberg (1993), Acemoglu et al. (2011) and Che and Hörner (2018)) we assume that individuals use information on past periods' choices (i.e. the choices of others) to inform todays' choices. In particular, they sort alternatives based on their popularity in previous periods and then choose alternatives with probability proportional to their popularity. As an example, consider an individual who is deciding on a pair of new sneakers using an online shopping service. She is indifferent between several pairs of shoes. To break ties, she sorts them based on their popularity (in previous periods) and chooses each alternative with probability proportional to their popularity. This sort of popularity based updating behavior is not peculiar to the preceding example. Indeed, mounting evidence suggests that individuals are more likely to choose alternatives that appear higher up in (popularity) lists.[10]

Our aim is to use the dynamic model to study long run behavior of individuals. If each individual in a population of individuals uses a popularity based updating rule, will their

[9]See Simon (1955). A stochastic version of this model is also studied in Dardanoni et al. (2022).

[10]For recent evidence see Joachims et al. (2005) who find evidence on "trust bias" which leads to more clicks on links that are ranked higher up by Google's search engine.

long run behavior converge? And, if so, what will it converge to?[11] We identify sufficient conditions on the type distribution such that long run behavior converges. Further, we show that it converges to the behavior of the highest type in the population. We show that a range of relevant behaviors are consistent with our sufficient condition, such as when each individual in the population is a satisficer (Simon, 1955), a checklist user (Mandler, Manzini, and Mariotti, 2012) or a rationalizer (Cherepanov, Feddersen, and Sandroni, 2013). It follows that long run behavior converges to the behavior of the highest type in each of these models.

A final application is in section 7 where we discuss an extension of the progressive random choice model of Filiz-Ozbay and Masatlioglu (2022). As mentioned, their model is defined as a probability distribution over single valued choice functions where the collection of choice functions in the support of the distribution satisfy a property called progressiveness. Roughly speaking, a collection of choice functions is progressive with respect to a strict reference order if they can be ordered in such a way that choice functions higher up in the ordering are more aligned with the reference order than choice functions lower down. We extend their framework to a case where the reference rule is an *arbitrary* single valued choice function. In particular, it follows as corollary to our main identification result that this model is identified.

Section 8 concludes the paper with a discussion of extensions and special cases of our framework. Subsection 8.1 relaxes the observability assumption on the reference rule and discusses a sufficient condition under which both the reference rule and underlying type distribution are revealed by choice data. Subsection 8.2 closes with a short discussion on the relation between our work and the works of Apesteguia et al. (2017) and Filiz-Ozbay and Masatlioglu (2022). In particular, we show that both of their models can be viewed as special cases of our framework.

1.1. **Related literature.** This paper is related to a literature that aims to address issues of identification in models with choice heterogeneity. With our sequential condition we join a recent literature that uses sorted type spaces to obtain identification in mixture models. In particular, our sequential condition generalizes both the single-crossing condition in Apesteguia, Ballester, and Lu (2017) and progressiveness in Filiz-Ozbay and Masatlioglu (2022). We elaborate on the relation to Apesteguia et al. (2017) and Filiz-Ozbay and Masatlioglu (2022) in subsection 8.2. Our paper is also related to Apesteguia and Ballester (2021) who explore menu restrictions in characterizations of the random utility model with ordered types and domains. Their objective is to meet empirical requirements by working with arbitrary domains and they illustrate the practicality of their results by applying them to empirically and experimentally important menu collections such as those with menus composed of lotteries from the Marschak-Machina triangle. Similar to Apesteguia and Ballester

[11] Similar questions are, independently from us, explored in Analytis, Cerigioni, Gelastopoulos, and Stojic (2022).

(2021) our main results on identification apply to arbitrary menu collections. In addition, our results allow for possibly multivalued and non-rational types.[12]

There is also a burgeoning literature that has made good progress to achieve identification in models with heterogeneity across both preferences and cognition (such as attention and mood). Notable contributions are Goeree (2008), Abaluck and Adams-Prassl (2021) and Barseghyan, Coughlin, Molinari, and Teitelbaum (2021). Recently, Dardanoni et al. (2022) study identification in models with heterogeneity across preferences and cognition, assuming that the analyst has access to mixture choice data (which is richer than traditional stochastic choice data in that it records correlation among alternatives). In contrast to Dardanoni et al. (2022), we work with traditional stochastic choice data.[13] In our main identification result we assume observability of the reference rule and obtain identification of the underlying type distribution. Later, we relax this requirement and impose a condition called binariness in section 4. Roughly speaking, it requires all individuals to pay full attention in binary menus. Similarly to the aforementioned studies, it allows us to obtain identification of both preference parameters and the underlying type distribution.

2. Model

We denote by X a finite set of alternatives. A nonempty subset $A \subseteq X$ is called a *menu*. Let $\mathcal{A}$ denote the collection of all nonempty subsets of X.[14] A stochastic choice function (SCF) is a function $\rho : X \times \mathcal{A} \mapsto [0,1]$ such that for all menus $A \in \mathcal{A}$: $\sum_{a \in A} \rho(a, A) = 1$ and $\rho(a, A) = 0$ for all $a \in X \setminus A$. The quantity $\rho(a, A)$ can either be interpreted as the fraction of times (in repeated trials) that an individual chooses alternative a from menu A, or as the fraction of individuals in a population that chooses a from A. The object ρ is the empirical primitive in our exercise and is observed by an outside observer/researcher/analyst. The aim of the researcher is to use ρ to infer properties about underlying behavior.

2.1. Choice correspondences and reference/tie-breaking rules.

To state our main model, we will need some further preliminaries. The behavior of each type/individual in our model is described by a choice correspondence. Formally, a *choice correspondence* is a function $C : \mathcal{A} \to 2^X \setminus \emptyset$ such that $C(A) \subseteq A$ for all $A \in \mathcal{A}$. Single valued choice correspondences, i.e. choice correspondences C such that $|C(A)| = 1$ for all $A \in \mathcal{A}$ are called *choice functions*. Choice functions are denoted with lower case c.

[12] For similar considerations as in Apesteguia and Ballester (2021) see also Petri (2021) who characterizes the single-crossing random utility model restricted to the collection of binary menus.

[13] Recently, Kashaev and Aguiar (2021) obtain identification results in a model with random attention and preference heterogeneity. Similar to us they only need standard stochastic choice data.

[14] Many of our results hold for arbitrary menu collections. We will indicate whenever this is the case.

In contrast to choice functions, choice correspondences are generally not univalent (single valued). Nevertheless, our observed primitive (an SCF) associates to each menu A a probability distribution on single choices a in A. To arrive at a final single choice, we will assume that each type/individual uses a common reference/tie-breaking rule.

Definition 2.1. A *reference/tie-breaking rule* is a stochastic choice function $\pi : X \times 2^X \setminus \emptyset \rightarrow [0, 1]$. A *positive reference rule* is a reference rule such that $\pi(a, A) > 0$ for all $a \in A$ and $A \subseteq X$. ◁

An individual described by a choice correspondence C will thus, in a given menu A, first select a subset of alternatives $C(A)$ and then randomize and choose alternatives $a \in C(A)$ with probability $\pi(a, C(A))$. To start with, we will assume that π is exogenous and observed by the researcher. We will later relax this requirement. Ideally, the reference rule should reflect application specific features of the studied problem. For instance, in a model with non strict preferences/utility function, it seems conceivable that individuals resolve indifferences by using a uniform rule. Whereas, in a dynamic setting, individuals may use information on past choices to inform today's choices (we will study this case further in section 6). In other situations, other reference rules might be more appropriate.

2.2. **Random Sequential Model.** We consider stochatic choices that are goverened by a probability distribution μ on choice correspondences in $\mathcal{C}$. As explained in the introduction, we impose a sequential condition on the set of correspondences in the support of the measure μ.

Definition 2.2. A collection of distinct choice correspondences $(C_i)_{i \in I}$ is called *sequential* if they can be ordered such that $C_1(A) \supseteq ... \supseteq C_{|I|}(A)$ for all $A \in \mathcal{A}$. ◁

We will give several examples of collections of correspondences that satisfy the sequential condition below. We are now ready to introduce the main model studied in this paper.

Definition 2.3. A stochastic choice function ρ is called a *Random Sequential Model* (RSM) w.r.t. a reference rule π if there is a finite set $I = \{1, ..., n\}$, a sequential collection of choice correspondences $\mathcal{C} = (C_i)_{i \in I}$, and a probability measure μ with support on $\mathcal{C}$ such that

(1) $\rho(a, A) = \sum_{i \in I} \mu_i \pi(a, C_i(A))$ for all $a \in A$ and $A \in \mathcal{A}$ and

(2) $C_i(A) \subseteq \{a \in A : \rho(a, A) > 0\}$ for all $i \in I$ and $A \in \mathcal{A}$.

◁

Remark: If either π or ρ is positive, then condition (2) is implied by condition (1) and is hence superflous. This condition is imposed in order to obtain identification of the underlying parameters of an RSM. To see why this condition is necessary for identification we refer the reader to example B.1 in the Appendix.

The model operates in two stages. First, a choice correspondence $C_i \in \mathcal{C}$ is drawn from $\mathcal{C}$ with probability μ_i and a subset of alternatives $C_i(A)$ is chosen from a given menu A. In a second stage, to arrive at a unique final choice in A type i then uses the common reference or tie-breaking function π and chooses each alternative $a \in A$ with probability $\pi(a, C_i(A))$.

An RSM restricts the support of μ to be sequential. We will next give several examples of sequential correspondences to illustrate their prevalence in choice. Informally, we may think of the set $C_i(A) \subseteq A$ as the alternatives that are cognitively feasible to individual i in menu A. For example, $C_i(A)$ could be interpreted as the alternatives that individual i pays attention to in menu A. Sequentialness of the correspondences $(C_i)_{i\in I}$ then means that individuals can be ordered based on the severity of their cognitive constraints (or their attention). We may thus assign a degree of attentiveness to each individual i based on their rank according to this ordering. Alternatively, we may interpret sequentialness as a property imposed on a set of correspondences of a single individual. Higher choice correspondences then correspond to states of mind where the cognitive constraints of the decision maker are more severe or to later survivor sets in a sequential elimination process employed by the decision maker. Sequentialness is closely related to the progressiveness property in Filiz-Ozbay and Masatlioglu (2022), as well as the single-crossing property in Apesteguia et al. (2017). In particular, both of these models can be seen as special cases of our model (see section 7 for further elaboration).

Our sequential condition imposes significant structure on the representation. For a broad range of (exogenous) reference rules π, we show (in section 3) that any RSM representation w.r.t. π is unique. At the same time, sequentialness is a permissive condition and has a clear interpretation in an array of relevant and well studied choice models. We next provide several examples.

Example 2.4. (*Sequential rationalizable choice (Manzini and Mariotti, 2007)*). There is a sequence of asymmetric relations $P_1, ..., P_n$ such that $C_0(A) = A$ and $C_k(A) = \{x \in C_{k-1}(A) : yP_kx \text{ for no } y \in C_{k-1}(A)\}$. In words, the decision maker first eliminates all alternatives that are sub optimal according to P_1. Among remaining alternatives she then eliminates alternatives that are sub optimal according to P_2 . She continues in a similar fashion until all criteria P_i have been looked at. ◁

Example 2.5. (*The gradual pairwise comparison rule (GPCR) (Dutta, 2020)*). A decision maker using a GPCR has a rational preference relation P on X, but she only considers pairwise comparisons according to P gradually, with the interpretation that easier comparisons precede more difficult ones. Let $\mathcal{X} = \{X_1, ..., X_n\}$ be an ordered partition of P. Define a collection $(C_i)_{i\in\{1,\dots,n\}}$ recursively as follows. Set $C_1(A) = A$ and for all $1 < k \le n$ set:

$$C_k(A) = \{a \in C_{k-1}(A) : \forall b \in C_{k-1}(A), (b, a) \notin X_k\}.$$

By construction $(C_i)_{i\in\{1,\dots,n\}}$ is a sequential collection of choice correspondences. ◁

Example 2.6. (*Satisficing behavior (Simon, 1955)*). For each $A \in \mathcal{A}$, let $\delta_1(A), \dots, \delta_n(A)$ be real numbers such that $\delta_1(A) \le \delta_2(A) \le \dots \le \delta_n(A)$ and let $u : X \to \mathbb{R}$ be a utility function. Define $C_i(A) = \{a \in A : u(a) \ge \delta_i(A)\}$. I.e. each individual i only considers alternatives in menu A that are above the threshold $\delta_i(A)$. Individuals share a common utility u but differ in their level of satisfaction. ◁

Example 2.7. (*Checklist users (Mandler, Manzini, and Mariotti, 2012)*) A checklist user has a finite collection of properties in mind and proceeds through the list sequentially. At each step of the process, alternatives lacking the considered property are eliminated. The agent proceeds in this way until all properties have been checked. Formally, a checklist P is a finite set of properties $P_1, \dots, P_n$, where each property P_i is a subset of X. Define $C_1(A) = A$ and for all $1 < k \le n$

$$C_k(A) = \begin{cases} P_k \cap C_{k-1}(A) & \text{if } P_k \cap C_{k-1}(A) \neq \emptyset \\ C_{k-1}(A) & \text{otherwise.} \end{cases}$$

Clearly, $(C_i)_{i\in\{1,\dots,n\}}$ is a sequential collection of correspondences. ◁

Example 2.8. (*Shades of indecisiveness/indifference*) The decision maker is indecisive and has an incomplete preference relation $\succsim$ on X. The degree of indecisiveness is stochastic. This is formalized by a collection of transitive and reflexive relations $\succsim_0, \dots, \succsim_n$ on X such that for each $k, l \in \{1, \dots, n\}$ with $k < l$ we have that $a \succ_k b$ implies that $a \succ_l b$. Define $C_k(A) = \{x \in A : y \succ_k x \text{ for no } y \in A\}$ for all $A \in \mathcal{A}$. Then $(C_i)_{i\in\{1,\dots,n\}}$ is a sequential collection of correspondences. A special case of this model is when each $\succsim_i$ is a weak order and the sizes of the indifference classes are stochastic. ◁

Example 2.9. (*Multiple rationales*) Individuals have multiple rationales in mind and only considers alternatives that can be justified by some rationale.[15] Formally, there is a collection of preferences/rationales $\succ_1, \dots, \succ_n$ and each agent considers all alternatives $a \in A$ that are maximal according to some rationale $\succ_i$. Thus, there is for each individual $k \in I$ a subset of rationales $I(k) \subseteq \{1, \dots, n\}$ such that $C_k(A) = \bigcup_{j\in I(k)} \max(A, \succ_j)$. If higher types use fewer rationales, so that $I(k) \subseteq I(l)$ if $k > l$, then the collection of choice correspondences $(C_i)_{i\in\{1,\dots,n\}}$ used by these individuals is sequential. ◁

[15] See e.g. Kalai et al. (2002) and Cherepanov et al. (2013) who study models of rationalization.

2.3. **Representation.** We are now ready to state our first result.

Theorem 2.10. *Let $\pi : X \times \mathcal{A} \to [0,1]$ be a reference rule. Then every stochastic choice function ρ has a random sequential representation w.r.t. π.*

Remark: As is clear by inspection of the proof of theorem 2.10 no domain assumption is needed for it to go through. I.e. it holds for arbitrary menu collections $\mathcal{A}$. This is convenient for experimental design as it allows the researcher to choose the domain most suitable for their particular experiment.

There are two main takeaways from theorem 2.10. First, fixing an exogenous reference rule π theorem 2.10 says that any stochastic choice function has an RSM representation w.r.t. this particular reference rule. This shows that our model has high explanatory power. Second, according to theorem 2.10 any given stochastic choice function has a multitude of different RSM representations, one for each reference rule π. This explains the role of the exogeneity assumption on the reference rule, as without it the model would not be identified. In the next section we will show that the underlying type distribution is identified assuming an exogenous reference rule. We will later (section 8.1) also discuss cases where both the reference rule and the underlying type distribution are revealed by data.

The proof of theorem 2.10 is constructive and delegated to appendix A.1. Our algorithmic construction is different from and extends the construction in Filiz-Ozbay and Masatlioglu (2022).[16] The algorithm constructs both the probability measure μ as well as the correspondences in its support recursively. In contrast the construction of μ in Filiz-Ozbay and Masatlioglu (2022) calculates μ from the cumulative probabilities on lower contour sets of an exogenous order $\succ$. Their construction thus depends on the existence of an exogenous linear order $\succ$. Since we consider general (stochastic) reference rules we cannot straightforwardly adapt their algorithm to our setting.

We next give a brief proof outline and description of the algorithm. The first choice correspondence C_1 is, for each menu, set equal to the set of alternatives chosen with positive probability in the menu. The probability mass assigned to C_1 is then equal to the minimum of the ratio $\frac{\rho(a,A)}{\pi(a,C_1(A))}$ among all alternatives a (chosen with positive probability) and all menus A in the menu collection. Let a^* and A^* denote the alternative a^* and menu A^* that achieve this minimum value. With this choice of μ_1 it is clear that the defining formula for an RSM will hold by construction for a^* and A^*. Moreover, it assures that not too much probability is deducted from the other choice probabilities in that $\rho(a,A) \geq \mu_1 \pi(a, C_1(A))$ for all alternatives a and menus A. The next step of the construction defines correspondence C_2 to be equal to the support of the function $\rho(a,A) - \mu_1 \pi(a, C_1(A))$, i.e. the alternatives

[16] A similar algorithm is also considered in a prior working paper version of Dardanoni et al. (2022) where they use it to prove a representation theorem for non-reversing random choice models.

with positive probability that remains after deducting $\mu_1\pi(a, C_1(A))$ from $\rho(a, A)$. Similarly to μ_1, the probability mass μ_2 assigned to C_2 is then set equal to the minimum of the ratio $\frac{\rho(a,A)-\mu_1\pi(a,C_1(A))}{\pi(a,C_2(A))}$ among all alternatives a in $C_2(A)$ and all menus A. One can also show that this choice of μ_2 assures that the defining formula for an RSM will hold exactly for the alternatives that achieve the minimum ratio, and that not too much probability is deducted from the other choice probabilities. A similar construction then defines C_k and μ_k from C_{k-1} and μ_{k-1}. The remaining part of the proof is devoted to showing that the algorithm stops, that each step of the algorithm is well-defined and that the constructed probabilities and choice correspondences gives an RSM representation of the stochastic choice function ρ.

3. Identification

In this section we aim to address the issue of identification in the Random Sequential Model. More precisely, given data on choices when are the parameters of the representation unique? Such uniqueness properties are important for a variety of reasons, but perhaps most notably for comparative statistics, out-of-sample prediction and estimation. Consider a policy maker, who would like to analyze the effects of a certain policy on individual choices. This cannot be done unambiguously if the model is unidentified. We establish a general identification result that holds for a large class of RSMs. As a corollary, it follows that the class of progressive random choice models in Filiz-Ozbay and Masatlioglu (2022) and the class of single-crossing random utility models in Apesteguia et al. (2017) are identified. Moreover, any RSM with a (strictly) positive reference rule is identified.

3.1. **Regular reference rules.** In line with our discussion above (paragraph following theorem 2.10), to ensure identification, we fix an exogenous/observed reference rule π. The plausibility of this assumption was discussed in the introduction. The aim of this subsection is to show that we obtain identification in the RSM model for a large class of exogenous reference rules. A first observation is that there are reference rules π such that not every stochastic choice function admits a unique representations w.r.t. π (see example C.1). This section establishes a simple (and weak) sufficient condition on the reference rule to ensure uniqueness. We will assume that the reference rule π satisfies the following condition.

Definition 3.1. A refererence rule is *regular* if for all $a \in B \subseteq A$: if $\pi(a, A) > 0$ then $\pi(a, B) > 0$. ◁

Regularity requires a to be chosen with positive probability in a small set B whenever it is chosen with positive probability in a larger set A containing B. This is a stochastic analogue of the classic Chernoff property from deterministic choice and is a characterizing property of single valued choice functions maximized by rational preference orders. If we allow for multivalued or stochastic choice this is a fairly weak condition and is consistent with behaviors deviating from the rational paradigm (see section 4 for examples). The following result shows that the Random Sequential Model is identified if π satisfies this condition.

Theorem 3.2. *Let π be a regular reference rule. Then every stochastic choice function ρ has a unique RSM representation with respect to π.*

Remark: Similarly to our representation theorem 2.10, no domain assumption is needed for theorem 3.2 to go through. It holds for arbitrary collections of menus $\mathcal{A}$.

Theorem 3.2 already includes many cases of reference rules such as uniform reference rules, random reference orders and logit reference rules (a more thorough discussion follows in section 4 on applications). However, there are plausible reference rules that fail to satisfy regularity.[17] Perhaps surprisingly, it is possible to extend our identification result to cover reference rules that admit quite wild regularity violations. The aim of the next subsection is to extend theorem 3.2 as far as possible to include such cases.

3.2. **Characterization.** The previous subsection shows that regularity is sufficient for identification in the Random Sequential Model. Is it also necessary? I.e., if every stochastic choice function has a unique RSM representation w.r.t. a reference rule π, does this imply that the reference rule is regular? Fortunately, this is not the case, it turns out that a large class of non-regular reference rules allows identification in our model. We will next find necessary and sufficient conditions on π such that every stochastic choice function has a unique RSM representation w.r.t. π. Our main characterizing property is a weak version of regularity and is stated below.

Definition 3.3. A reference rule π is *weakly regular* if for all $a \in B \subseteq A$ with $\pi(b, A) > 0$ for all $b \in A \setminus B$: if $\pi(a, A) > 0$ then $\pi(a, B) > 0$. $\triangleleft$

Weak regularity restricts regularity to all pairs of sets $B \subseteq A$ where alternatives belonging to A but not to B are chosen with strictly positive probability. This property is consistent with a wide range of behaviors. Clearly any, regular choice function is weakly regular. Weak regularity is also consistent with behaviors that can be considered highly irregular. For instance, any stochastic choice function induced by a deterministic choice function satisfies weak regularity. A concrete example of such a rule is if there are multiple reference orderings $\succ_1, ..., \succ_n$ and the choice function chooses an alternative that is maximal according to one of these orderings (as in the choice by multiple rationales model of Kalai, Rubinstein, and Spiegler (2002)). Such reference orderings are commonplace in applications. In online shopping environments alternatives can often be sorted based on several attribute orders such as price and popularity/customer rating. We will give more examples in section 4 where we discuss applications.

[17] Subsection 4.3 contains examples.

Theorem 3.4. *Let π be reference rule. Then the following statements are equivalent:*

(1) π satisfies weak regularity.

(2) Every stochastic choice function ρ has a unique RSM representation w.r.t. π.

The proof of theorem 3.4 follows the steps of the algorithmic construction in the proof of theorem 2.10. Weak regularity is needed to ensure that there is a unique "path" from each step of the algorithm to the next. To show necessity of weak regularity we show its contrapositive, i.e. that if π fails to satisfy weak regularity then it is possible to construct an SCF with a non unique RSM representation. Example C.1 in the appendix illustrates the necessity of weak regularity in theorem 3.4.

4. Specific reference rules

We next discuss special cases of our main model obtained by considering different examples of reference rules π. More specifically, and in increasing order or specialization, we consider cases where π is a random reference order (i.e. a probability distribution on strict reference orders), a Luce/logit reference rule and a deterministic rule induced by a choice function.

4.1. Random reference order.

In this section we consider a class of reference rules π that are (menu dependent) probability distributions on collections of strict reference orders. A reference rule π is *random reference order* if there are reference orders $\succ_1,...,\succ_n$ and for each menu $A \in \mathcal{A}$ corresponding probabilities $\alpha(A) = (\alpha_1(A), ..., \alpha_n(A))$[18] such that

$$\pi(a, A) = \sum_{i=1}^{n} \alpha_i(A)\mathbf{1}\{a = \max(A, \succ_i)\}$$

for all $a \in A$ and $A \in \mathcal{A}$.

An RSM with a random reference order can be interpreted as a model where each choice type uses a common set of reference orders to break ties among considered alternatives $C_i(A)$, and where each type entertains a particular reference order $\succ_k$ with common probability $\alpha_k(A)$. Consider an online shopping environment, customers can then sort alternatives based on a common set of attributes such as price in ascending/descending order, customer rating, discount and so on. We can model this by letting each order $\succ_i$ correspond to such an attribute.

It follows as a direct corollary to theorem 2.10 that every stochastic choice function has an RSM representation w.r.t. a fixed (exogenous) random reference order π. Since, every random reference order π satisfies regularity it follows by corollary 3.2 that any ρ has a unique representation w.r.t. π.

[18] I.e. where $\alpha_i(A) > 0$ for all $i \geq 1$ and $\sum_{i=1}^{n} \alpha_i(A) = 1$.

Proposition 4.1. *Let π be a random reference order. Every stochastic choice function ρ has an RSM representation w.r.t. π. Moreover, the representation is unique*

4.1.1. *Dual random reference order.* A special case of the rule above is when there are only two reference orders $\succ_1$ and $\succ_2$ that are randomly entertained by individuals. We call this a dual random reference order. As an example, consider an online shopping environment where customers can sort alternatives in ascending or descending order based on attributes such as price. This can be modeled as a dual random reference order with support on a linear order $\succ$ (reflecting the attribute) and its reverse order $\succ'$ defined by $a \succ' b$ if and only if $b \succ a$. The following proposition follows as a direct corollary to theorem 2.10 and proposition 4.1.

Proposition 4.2. *Let π be a dual random reference order. Every stochastic choice function ρ has an RSM representation w.r.t. π. Moreover, the representation is unique.*

4.2. **Logit reference rule.** In this section we consider random sequential models with logit tie-breaking/reference rules. A reference rule is a *logit reference rule* if there is a utility function $u : X \to (0, \infty)$ such that

$$\pi(a, A) = \frac{u(a)}{\sum_{b \in A} u(b)}$$

for all $a \in A$ and $A \in \mathcal{A}$. This can be interpreted as a case where individuals use a common reference order agreeing with u, but where they occasionally make misstakes and choose suboptimal alternatives (or, alternatively, where they deliberately experiment and try suboptimal alternatives) and where the probability of choosing a suboptimal alternative is proportional to its utility. The following proposition follows as a corollary to theorem 2.10 and 3.4 (since π is positive).

Proposition 4.3. *Let π be a logit reference rule. Every stochastic choice function ρ has an RSM representation w.r.t. π. Moreover, the representation is unique.*

4.3. **Deterministic reference rule.** In this section we consider a deterministic reference rule. We assume that there is a (single valued) choice function c defined on the subsets of X such that the reference rule takes the following form

$$\pi_c(a, A) = \begin{cases} 1 & \text{if } a = c(A) \\ 0 & \text{otherwise.} \end{cases}$$

We interpret c as a common observed choice rule. An important example is when c is dictated by a collection of exogenous reference/attribute orders $\succ_1, ..., \succ_n$ such that $c(A)$ is $\succ_i$ maximal for some i.[19] Consider for instance an online environment where alternatives can be ordered based on attributes such as price in ascending/descending order and/or popularity. A special

[19] This is the choice by multiple rationales model in Kalai et al. (2002).

case is when c is maximized by a single reference order $\succ$, i.e. $c(A) = \max(\succ, A)$ for all $A \in \mathcal{A}$. Such reference rules will be denoted by $\pi_\succ$. As we will see in subsection 8.2 the Random Sequential Model w.r.t. $\pi_\succ$ is closely related to the models considered in Apesteguia et al. (2017) and Filiz-Ozbay and Masatlioglu (2022).

The following proposition shows that any stochastic choice function has a unique RSM representation w.r.t. π_c. This is perhaps a bit surprising given the irregular choice patterns consistent with a general single valued choice rule c.

Proposition 4.4. *Let π_c be a deterministic reference rule. Every stochastic choice function ρ has an RSM representation w.r.t. π_c. Moreover, the representation is unique.*

Proof. It is clear that π_c is weakly regular. This is because if $B \subseteq A$ and $b \in A \setminus B$ with $\pi_c(b, A) > 0$ then $\pi_c(b, A) = 1$. Hence $\pi_c(a, A) = 0$ for all $a \in B$, so weak regularity vacuously holds. □

5. Satisficing Behavior

So far we have studied RSMs without any restrictions on the choice functions in their support. We mentioned several plausible classes of such choice functions in the introduction, each consistent with different types of boundedly rational behavior. One prominent and well studied type of behavior is satisficing. In satisficing a decision maker has a (random) threshold utility and only considers alternatives with utility exceeding the threshold. A model along these lines was first conceived in Simon (1955) and has since then attracted wide attention in a large literature on behavioral economics and decision theory.[20] In this section we provide a characterization of a satisficing model in the spirit of Simon (1955).

Definition 5.1. A collection $(C_i)_{i \in I}$ is *consistent with satisficing behavior* if there is:

i) a common (utility) function $v : X \to (0, \infty)$ and

ii) for each $i \in I$ a threshold function $\delta_i : \mathcal{A} \to (0, \infty)$

such that for all $A \in \mathcal{A}$ and for all $i \in I$: $C_i(A) = \{a \in A : v(a) > \delta_i(A)\}$. ◁

In words, a collection $(C_i)_{i \in I}$ is consistent with satisficing behavior if there is a common utility v such that each type i only considers alternatives with utility exceeding a (possibly) type-dependent threshold $\delta_i(A)$. Heterogeneity in this model is due to the individual (type dependent) thresholds δ_i. I.e. types differ in their satisfaction *levels* but have a common satisfaction utility v.

Our characterization result is in terms of an acyclicity condition imposed on a preference relation revealed from stochastic choice data. We next define this relation. Fix a weakly

[20] See Tyson (2008), Caplin et al. (2011), Manzini et al. (2013) and Aguiar et al. (2016) for some recent explorations.

regular reference rule π (we impose weak regularity to ensure uniqueness of the revealed preference relation). Let ρ be a stochastic choice function and $(C_i^\pi)_{i\in I}$ the unique collection of sequential choice correspondences in an RSM representation of ρ w.r.t. the reference rule π. We define a binary relation $\mathtt{P}^\pi$ on X by $a\mathtt{P}^\pi b$ if and only if $a \in C_i^\pi(A)$ and $b \in A \setminus C_i^\pi(A)$ for some $i \in I$ and $A \in \mathcal{A}$. In words, a is related to b if there is a menu A, with both a and b available, and some type $i \in I$, such that a is chosen by type i, but b is not. Note that $a\mathtt{P}^\pi b$ reveals a strict preference for a over b, since then $a \in C_i(A)$ and $b \in A \setminus C_i(A)$ implies $v(a) > \delta_i(A) \geq v(b)$. The relation $\mathtt{P}^\pi$ is well-defined since weak regularity of π implies that every SCF ρ has a unique RSM representation w.r.t. π (follows by theorem 2.10 and theorem 3.2). Our acyclicity condition is related to acyclicity conditions used to characterize deterministic versions of the satisficing model (see e.g. Aleskerov et al. (2007), Manzini et al. (2013) and Frick (2016)).

Proposition 5.2. *Let π be a weakly regular reference rule. A stochastic choice function ρ has a satisficing representation*[21] *w.r.t. π if and only if $\mathtt{P}^\pi$ is acyclic.*

Proof. Assume that ρ has a satisficing representation w.r.t. π and that there are thresholds $\delta_1, ..., \delta_m$ and a function $v : X \to (0, \infty)$ such that definition 5.1 holds. We first show that $\mathtt{P}^\pi$ is acyclic. To show this it suffices to show that $a\mathtt{P}^\pi b$ implies $v(a) > v(b)$. Assume that $a\mathtt{P}^\pi b$. Then there is an $i \in I$ with $a \in C_i(A)$ and $b \notin C_i(A)$ and it follows that $v(a) > \delta_i(A) \geq v(b)$.

Next, we assume that $\mathtt{P}^\pi$ is acyclic (the relation $\mathtt{P}^\pi$ is well-defined due to weak regularity of π). Let $v : X \to (0, \infty)$ be a representation of a linear extension of $\mathtt{P}^\pi$, i.e. v is such that $a\mathtt{P}^\pi b$ implies $v(a) > v(b)$. Let $C_1, ..., C_n$ be the unique sequential collection of choice correspondences in an RSM representation of ρ as per theorem 3.2. We will show that it is possible to define $\delta_i : \mathcal{A} \to \mathbb{R}$ for each $i \in I$ such that $C_i(A) = \{a \in A : v(a) > \delta_i(A)\}$. Let $i \in I$. For all $A \in \mathcal{A}$ define $\delta_i(A) = \arg\max_{a \in A\setminus C_i(A)} v(a)$. Let $a \in A$ with $v(a) > \delta_i(A)$. Then $v(a) > \delta_i(A) \geq v(b)$ for all $b \in A \setminus C_i(A)$ and hence it follows that $a \in C_i(A)$. On the other hand, if $a \in C_i(A)$ then $a\mathtt{P}^\pi b$ for all $b \in A \setminus C_i(A)$, so $v(a) > v(b)$ for all $b \in A \setminus C_i(A)$, which in turn implies that $v(a) > \delta_i(A)$. The claim follows. □

5.1. **Logit satisficing.** The relation $\mathtt{P}^\pi$ is, for certain reference rules, quite cumbersome to compute. To compute it we first need to execute the algorithm in the construction of an RSM representation. We next study a case where the relation $\mathtt{P}^\pi$ takes a particularly simple form, namely when π belongs to the class of logit functions. If the reference rule π has a logit representation with utility function $u : X \to (0, \infty)$ then we may define a relation $\mathtt{P}^u$ by $a\mathtt{P}^u b$ if and only if $\rho(a, A)u(b) > \rho(b, A)u(a)$ for some $A \supseteq \{a, b\}$. The next proposition

[21]In the sequel, we will say that a stochastic choice function ρ has a satisficing representation w.r.t. a reference rule π, if ρ has an RSM representation w.r.t. π and if the collection $(C_i)_{i\in I}$ in the representation is consistent with satisficing behavior.

shows that acyclicity of $\mathtt{P}^\pi$ is equivalent to acyclicity of $\mathtt{P}^u$ which in turn is equivalent to a satisficing represenation of ρ (with a logit reference rule).

Proposition 5.3. *An SCF ρ has a satisficing representation w.r.t. a logit reference rule (with utility $u : X \to (0, \infty)$) if and only if $\mathtt{P}^u$ is acyclic.*

The intuition behind proposition 5.3 is simple. The standard Luce/logit model is characterized by independence of irrelevant alternatives (IIA) requiring that $\frac{\rho(a,A)}{\rho(b,A)} = \frac{u(a)}{u(b)}$ for all $a, b \in A$ and for all menus $A \subseteq X$. Our acyclicity condition is based on the idea that an IIA violation at a, b, meaning that $\frac{\rho(a,A)}{\rho(b,A)} > \frac{u(a)}{u(b)}$, reveals that a is a more satisfying alternative than b (i.e. that $v(a) > v(b)$). Hence we require acyclicity of this relation.

Proof. It suffices to show that for all $a, b \in X$: $a\mathtt{P}^u b$ if and only if $a\mathtt{P}^\pi b$. If $a\mathtt{P}^u b$ then $\rho(a, A)u(b) > \rho(b, A)u(a)$ for some $A \supseteq \{a, b\}$. Assume, by contradiction, that for all $i \in I$: $a \in C_i(A)$ implies $b \in C_i(A)$. Then $\frac{\mathbf{1}\{a\in C_i(A)\}}{\sum_{c\in C_i(A)} u(c)} \leq \frac{\mathbf{1}\{b\in C_i(A)\}}{\sum_{c\in C_i(A)} u(c)}$ for all $i \in I$, which in turn implies that

$$\rho(a, A)u(b) = \sum_{i\in I} \mu_i \frac{u(a)\mathbf{1}\{a \in C_i(A)\}}{\sum_{c\in C_i(A)} u(c)} u(b) = u(a)u(b) \sum_{i\in I} \mu_i \frac{\mathbf{1}\{a \in C_i(A)\}}{\sum_{c\in C_i(A)} u(c)} \leq$$

$$\leq u(a)u(b) \sum_{i\in I} \mu_i \frac{\mathbf{1}\{b \in C_i(A)\}}{\sum_{c\in C_i(A)} u(c)} = \sum_{i\in I} \mu_i \frac{u(b)\mathbf{1}\{b \in C_i(A)\}}{\sum_{c\in C_i(A)} u(c)} u(a) = \rho(b, A)u(a).$$

This is a contradiction. It thus follows that there is an $i \in I$ with $a \in C_i^\pi(A)$ and $b \in A \setminus C_i^\pi(A)$. Hence $a\mathtt{P}^\pi b$. Conversely, assume that $a\mathtt{P}^\pi b$. Since π is positive there is an index $i(a)$ such that $a \in C_i^\pi(A)$ if and only if $i \leq i(a)$ and an index $i(b)$ such that $b \in C_i^\pi(A)$ if and only if $i \leq i(b)$. Further, $i(a) > i(b)$. Hence

$$\rho(a, A)u(b) = \sum_{i\in I} \mu_i \frac{u(a)\mathbf{1}\{a \in C_i(A)\}}{\sum_{c\in C_i(A)} u(c)} u(b) = u(a)u(b) \sum_{i\leq i(a)} \mu_i \frac{1}{\sum_{c\in C_i(A)} u(c)} >$$

$$> u(a)u(b) \sum_{i\leq i(b)} \mu_i \frac{1}{\sum_{c\in C_i(A)} u(c)} = \sum_{i\in I} \mu_i \frac{u(b)\mathbf{1}\{b \in C_i(A)\}}{\sum_{c\in C_i(A)} u(c)} u(a) = \rho(b, A)u(a).$$

The claim follows. □

Of some independent interest is perhaps the case where the reference rule takes the form of a uniform rule. This corresponds to a special case of the logit rule where the logit utilities of all alternatives are equal. To characterize it, define a relation $\mathtt{P}$ by $a\mathtt{P}b$ if and only if $\rho(a, A) > \rho(b, A)$.

Corollary 5.4. *A stochastic choice function ρ has a satisficing representation w.r.t. a uniform tie-breaking rule if and only if $\mathtt{P}$ is acyclic.*

Proof. Let $u : X \to (0, \infty)$ be such that $u(a) = u(b)$ for all $a, b \in X$. Then for all $a, b \in X$: $a\mathtt{P}^u b$ if and only if $a\mathtt{P}b$. Hence $\mathtt{P}^u$ is acyclic and it follows by proposition 5.3 that ρ has a satisficing representation w.r.t. $u : X \to (0, \infty)$. But since $\frac{u(a)}{\sum_{b \in C_i(A)} u(b)} = \frac{1}{|C_i(A)|}$ for all $i \in I$, it follows that ρ has a satisficing representation with uniform tie breaking. The converse is immediate. □

6. Dynamic choice

Our model may be tailored to study dynamic choice. In line with a large literature on social learning, we assume that individuals use information on past periods' choices to inform current decisions. In this setting we assume that the outside observer/analyst observes a sequence or time series of stochastic choices. We call a sequence $(\rho_t)_{t=0}^{\infty}$, where each ρ_t is a stochastic choice function, a *dynamic stochastic choice function*.

Definition 6.1. A dynamic stochastic choice function $(\rho_t)_{t=0}^{\infty}$ is governed by a *Dynamic Random Sequential Model* (DRSM) if there for each $t \in \mathbb{N}$ is a finite set $I(t) = \{1, ..., n(t)\}$, a collection of sequential choice correspondences $\mathcal{C}^t = (C_i^t)_{i \in I(t)}$, and a probability measure μ^t on $\mathcal{C}^t$ such that

$$\rho_t(a, A) = \sum_{i \in I(t)} \mu_i^t \rho_{t-1}(a, C_i^t(A)) \tag{2}$$

for all $a \in A$ and $A \in \mathcal{A}$. ◁

An interpretation of the model is as follows. There is a population of agents $I(t)$ whose deterministic choices are described by choice correspondences C_i^t in period t. These agents may deem several alternatives as choosable in A since $C_i^t(A)$ is not necessarily singleton. The model assumes that each agent resolves any ambiguity, due to non singleton $C_i^t(A)$, by using information on past period's choices ρ_{t-1}. I.e. past choices ρ_{t-1} are used as a reference for today's choices ρ_t. The following proposition is a direct corollary to theorem 2.10.

Corollary 6.2. *Every dynamic stochastic choice function $(\rho_t)_{t=0}^{\infty}$ has a DRSM representation.*

Proof. The result follows by applying theorem 2.10 to each pair (ρ_t, ρ_{t+1}). I.e. let ρ_t be the reference rule in theorem 2.10. □

Moreover, if each ρ_t is weakly regular, then it is possible to uniquely recover the parameters of the DRSM by using dynamic stochastic choices. This follows as a direct corollary to theorem 3.4.

Corollary 6.3. *Assume that ρ_t is weakly regular for all $t \geq 0$. Then $(\rho_t)_{t=0}^{\infty}$ has a unique DRSM representation.*

6.1. Long run behavior.

The model above can be used to study long run behavior of individuals. However, without any further assumptions imposed on the model, the long run behavior may be very erratic. This is because any dynamic stochastic choice function has a

DRSM representation, so any type of long run behavior is consistent with the model. In this section we will therefore assume that there is a *fixed* population of agents who update their choices using information on past choices as in equation (2). More formally, we will assume that there is a population of individuals, consisting of n different types $I = \{1, ..., n\}$, and that the fraction of each type in the population is constant (time-invariant) and equals μ_i. We assume that each individual uses a correspondence C_i to narrow down choices, and then uses information on past choices to choose among considered options $C_i(A)$ in A, i.e. for each period t aggregate choices satisfy $\rho_t(a, A) = \sum_{i \in I} \mu_i \rho_{t-1}(a, C_i(A))$ for all $a \in A$ and $A \in \mathcal{A}$. This is the same equation as in (2) except that we have dropped the time superscripts on μ and the $C_i's$ due to the assumption of a fixed/constant population. We will impose one final assumption, and this is idempotence:

Definition 6.4. A collection of choice correspondences $(C_i)_{i \in I}$ is *idempotent* if $C_j \circ C_i(A) = C_i \circ C_j(A) = C_j(A)$ for all types $i, j \in I$ with $i \leq j$. ◁

Idempotence is a quite weak assumption and is satisfied by a range of different models. Examples of models that satisfy this assumption are the shades of indecisiveness model, the multiple rationales model and the checklist model introduced in example 2.6 - 2.9. Another example is the satisficing model in section 5 with menu independent thresholds, i.e. threshold functions $\delta_i : \mathcal{A} \to \mathbb{R}$ such that $\delta_i(A) = \delta_i$ for all $A \in \mathcal{A}$. The following lemma, whose proof is in the Appendix, proves that these models are idempotent.

Lemma 6.5. *Let $(C)_{i \in I}$ have a shades of indecisiveness representation, a multiple rationales representation, a checklist representation or an independent satisficing representation. Then $(C_i)_{i \in I}$ is idempotent.*

We will next study the long run behavior of individuals. Suppose that each individual uses information on past choices to inform current choice, i.e. their choices are in line with the model described above. Will long run behavior converge? And, if so, what will long run behavior converge to? For instance, will individuals be better off in the long run? The following proposition provides a partial answer to these questions and shows that idempotence is a sufficient condition for convergence of long run behavior.

Proposition 6.6. *Let $(C_i)_{i \in I}$ be an idempotent collection of choice functions. The long run choice distribution converges to the initial distribution of the highest type, i.e. for all $a \in A$ and $A \in \mathcal{A}$:*

$$\lim_{t \to \infty} \rho_t(a, A) = \rho_0(a, C_n(A)).$$

In the context of the shades of indecisiveness model proposition 6.6 says that the long run behavior converges to the initial distribution of the least indecisive type. Suppose that there is a non-negliable probability that all indecisiveness is resolved, i.e. that choice correspondence

C_n maximizes a linear preference, then by proposition 6.6 long run behavior will converge to deterministic preference/utility maximization. This provides a testable hypothesis of the model and suggests that long run behavior gets less heterogeneous (individual behaviors more aligned with each other) as time increases.

The intuition behind proposition 6.6 is quite simple. To explain it, consider a special case with two types, high (H) and low (L), where each type uses a satisficing procedure $C_H(A) = \{a \in A : u(a) > \delta_H\}$ and $C_L(A) = \{a \in A : u(a) > \delta_L\}$ and where $\delta_H > \delta_L$. High and low types will look at prior periods choices to inform current choice. I.e. they will consider alternatives in $C_L(A)$ ($C_H(A)$) and use prior periods' choices ρ_1 to break ties. In particular, the choice probability of $a \in A$ in period 2 is $\rho_2(a, A) = \mu_L \rho_1(a, C_L(A)) + \mu_H \rho_1(a, C_H(A))$. Expanding and using the idempotence assumption we can express this probability as:

$$\rho_2(a, A) = \mu_L^2 \rho_0(a, C_L(A)) + (1 - \mu_L^2)\rho_0(a, C_H(A))$$

In particular, low types will inevitably put larger probability on alternatives with satisfaction level exceeding δ_H and are hence indistinguishable from high types (this is because they will observe high types choosing such alternatives in the previous period). In contrast, high types will keep choosing alternatives above their satisfaction threshold δ_H. In rough terms, each period probability will thus be "shifted" from outside $C_L(A) \setminus C_H(A)$ to $C_H(A)$, while no probability will "escape" from $C_H(A)$. Thus, in the long run, a gradually higher probability mass will be put on C_H such that in the limit all mass will be on C_H. By induction/iteration, one can show that

$$\rho_n(a, A) = \mu_L^n \rho_0(a, C_L(A)) + (1 - \mu_L^n)\rho_0(a, C_H(A)),$$

implying the statement for the case of two types (since $\mu_L < 1$ implies that $\mu_L^n \to 0$ as $n \to \infty$). The proof of the general n type case follows a similar logic, but relies on the (global) convergence of a simple first order linear difference equation. See appendix A.4 for further details.

7. General progressive choice model

In this section we apply our framework to study extensions of the single-crossing random utility model (SCRUM) in Apesteguia et al. (2017) and the progressive random choice model in Filiz-Ozbay and Masatlioglu (2022). Both of these papers consider representations w.r.t. an exogenous linear reference order $\succ$. We extend their results to allow arbitrary exogenous choice functions. Examples of such reference rules were provided in subsection 4.3. An archetypical example being online environments where alternatives can be ordered based on a set of different criteria, such as price in ascending or descending order. Let $\mathbb{C}$ denote the set of all (single valued) choice functions.

Definition 7.1. A collection of choice functions $\mathcal{C} \subseteq \mathbb{C}$ is progressive w.r.t. an exogenous choice function c if $\mathcal{C}$ can be ordered $\{c_1, ..., c_n\}$ such that $c(\{c_i(A), ..., c_n(A)\}) = c_i(A)$ for all $i \in \{1, ..., n\}$. ◁

A stochastic choice function has a progressive random choice representation w.r.t. c if there is a measure μ on $\mathbb{C}$ such that the support of μ is progressive w.r.t. c and $\rho(a, A) = \mu(\mathbb{C}(a, A))$ for all $a \in A$ and $A \in \mathcal{A}$, where $\mathbb{C}(a, A) = \{c \in \mathbb{C} : c(A) = a\}$.

Proposition 7.2. *Every stochastic choice function ρ has a progressive representation w.r.t. c. Moreover, the representation is unique.*

The proof of proposition 7.2 is provided in appendix A.5. The proof that every ρ has a progressive representation w.r.t. c is straightforward and follows by adapting the algorithm in the proof of theorem 2.10 or by using a similar algorithmic construction as in Filiz-Ozbay and Masatlioglu (2022). Uniqueness does not follow from the proof in Filiz-Ozbay and Masatlioglu (2022). Note, however, that corollary 4.4 implies that any RSM representation w.r.t. π_c is unique. Given a PRC representation of ρ with support $c_1, ..., c_n$ and type distribution μ define a sequential collection of choice correspondences $(C_i)_{i \in I}$ by $C_i(A) = \cup_{j=i}^{n} c_j(A)$. One can then show that ρ has an RSM representation (w.r.t. π_c) using this collection of choice correspondences and the same type distribution μ. Uniqueness of the PRC representation then follows by corollary 4.4.

Proposition 7.2 suggests an extension of the single-crossing random utility model to a setting with an arbitrary exogenous choice function. However, the following result shows that there are limits to how far this result can be extended.

Proposition 7.3. *Let ρ be a positive stochastic choice function and c a choice function. If ρ is progressive w.r.t. c and the type distribution μ has support consisting of rational choice functions then there is a linear order $\succ$ s.t. $c(A) = \max(A, \succ)$ for all $A \in \mathcal{A}$.*

Proof. Since c is single valued, it suffices to show that c satisfies Sen's α (for a proof see (Sen, 1971, proposition T.6,p. 313)) , i.e. if $a \in B \subseteq A$ and $a = c(A)$ then $a = c(B)$. Let $a = c(A)$ then $a = c(c_1(A), ..., c_n(A)) = c_1(A)$ (due to the full support assumption). But since c_1 is rational and $a = c_1(A)$ it follows that $a = c_1(B)$. Hence $c(B) = c(c_1(B), ..., c_n(B)) = c_1(B) = a$. □

8. Discussion

We first discuss a sufficient condition under which both the reference rule and underlying type distribution are revealed by choice data. We then provide a more in depth discussion of the relation to Apesteguia et al. (2017) and Filiz-Ozbay and Masatlioglu (2022).

8.1. **Identification of reference rule *and* type distribution.** In order to achieve identification of the underlying type distribution we assume observability of the reference rule π. Although a plausible assumption, there are instances where π may not be directly observed by the researcher, such as when it corresponds to a decision rule/preference. We will next discuss a sufficient condition for identification of both the reference rule π and the type distribution μ in the context of the examples of reference rules discussed in section 4.

Definition 8.1. A collection of choice correspondences $(C_i)_{i \in I}$ is called *binary* if $C_i(\{x, y\}) = \{x, y\}$ for all $x, y \in X$ and for all $i \in I$. $\triangleleft$

This assumption implies that each type considers both alternatives in each binary menu. If we interpret the choice correspondence C_i as the cognitive bounds of type i, this assumption means that individuals are the least cognitively constrained in binary menus (which are the non singleton menus of smallest cardinality).[22] This, arguably, makes sense due to the fewer number of alternatives in such sets.

Proposition 8.2. *Let ρ be a positive stochastic choice function. If ρ has a binary RSM representation w.r.t. a dual random reference order then the parameters of ρ are unique.*

Proof. Since ρ is positive when restricted to binary subsets and ρ is a binary RSM it follows that $\rho(x, xy) = \pi(x, xy)$ for all $x, y \in X$. Thus it suffices to show that if π is positive on binary subsets then it is represented by a unique dual random reference order. The proof is straightforward (albeit a bit tedious) and available from the author upon request. $\square$

Our next example is when π is a logit rule. Then we can interpret an RSM rule ρ as a rule where each type considers a set of alternatives $C_i(A)$ and then chooses final alternatives with probability proportional to a common utility u. If we assume that ρ satisfies a positivity constraint, we recover "full" identification of the parameters of the model.

Proposition 8.3. *Let ρ be a positive stochastic choice function. If ρ has a binary RSM representation w.r.t. a logit reference rule π then the parameters of ρ are unique.*[23]

Proof. Suppose that ρ is represented by possibly different parameter triples $\langle (C_i)_{i \in I}, \mu, u \rangle$ and $\langle (C'_i)_{i \in I}, \mu', u' \rangle$. By binariness it follows that $C_i(\{a, b\}) = \{a, b\} = C'_i(\{a, b\})$ for all $i \in I$ and hence we have that $\frac{u(a)}{u(a)+u(b)} = \rho(a, ab) = \frac{u'(a)}{u'(a)+u'(b)}$. Thus, by uniqueness of the Luce/logit model when restricted to binary subsets it follows that $u = \beta u'$ for some $\beta > 0$. The uniqueness of $(C_i)_{i \in I}, \mu$ then follows by theorem 3.2. $\square$

[22] A similar condition is used in a characterization of the focal Luce model in Kovach and Tserenjigmid (2020).

[23] In particular, the utility function u is unique up to an affine transformation. See proof for details.

8.2. **Relation to progressive random choice and single-crossing choice.** This section discusses further relationships to Apesteguia et al. (2017) and Filiz-Ozbay and Masatlioglu (2022). As mentioned before, our framework may be viewed as an extension of both of these works, allowing random reference rules. I.e. their models are "isomorphic" to a special case of RSM with a reference rule induced by a strict reference order.

To show that the progressive random choice model (and hence the single-crossing random utility model) is isomorphic to a special case of our model. Recall from subsection 4.3 that the reference rule $\pi_\succ$ is defined as follows:

$$\pi_\succ(a, A) = \begin{cases} 1 & \text{if } a = \min(\succ, A) \\ 0 & \text{otherwise.} \end{cases}$$

Our next proposition, shows that an RSM representation w.r.t. this reference rule is closely related to progressiveness in Filiz-Ozbay and Masatlioglu (2022).

Proposition 8.4. *A stochastic choice function ρ has an RSM representation with type distribution μ and sequential collection $(C_i)_{i\in I}$ w.r.t. reference rule $\pi_\succ$ if and only if it has a PRC representation w.r.t. $\succ$ with the same type distribution μ and with support $(c_i)_{i\in I}$ such that $c_i(A) = \min(\succ, C_i(A))$.*

Proof. Let ρ be a stochastic choice function. Suppose that μ and $(C_k)_{k=1}^n$ are the parameters of an RSM representation of ρ w.r.t. the reference rule $\pi_\succ$. Then $\rho(a, A) = \sum_{i\in I}\mu_i\pi_\succ(a, C_i(A)) = \sum_{i\in I}\mu_i\mathbf{1}\{a = \min(\succ, C_i(A))\}$. For all $i \in I$ define $c_i(A) = \min(\succ, C_i(A))$ for all $A \in \mathcal{A}$. Then, since $C_1(A) \supseteq ... \supseteq C_m(A)$ it follows that $c_1(A) \precsim .... \precsim c_m(A)$. Further, $\rho(a, A) = \mu(\mathcal{C}(a, A))$ so ρ has a progressive representation with support $\{c'_1, ..., c'_m\}$. By theorem 1 in Filiz-Ozbay and Masatlioglu (2022) ρ has a unique progressive representation w.r.t. $\succ$ and the proposition follows. □

The model in Filiz-Ozbay and Masatlioglu (2022) enjoys high explanatory power in that any stochastic choice function can be represented by the PRC model. Ultimately, however, the PRC model is a model of single valued choice (as it posits a probability distribution on single valued choices). Some multivalued behaviors are therefore (and arguably) not that well-explained within the PRC framework. As an example consider a decision maker who has a weak order $\succsim$ on alternatives but where the indifference classes of ρ are stochastic. This is a special case of the shades of indecisiveness model discussed in example 2.8, where each $\succsim_i$ in the representation is a weak order (and with uniform tie-breaking).

Definition 8.5. A stochastic choice function ρ has a *shades of indifference* representation if there is a collection of preferences $\succsim_1, ..., \succsim_n$ such that $a \succ_i b$ implies $a \succ_j b$ for all $i < j$ and

$$\rho(a, A) = \sum_{i \in I} \mu_i \frac{\mathbf{1}\{a \in \max(\succsim_i, A))\}}{|\max(\succsim_i, A)|},$$

for all $a \in A$ and $A \in \mathcal{A}$. ◁

It is straightforward to show that the model above is a random utility model and hence consistent with population rationality.[24] However, as the next result shows there is no reference order $\succ$ such that the behavior of this model is described by a SCRUM. In particular, since every stochastic choice function can be explained by the PRC model, this implies that a PRC representation of this behavior has to assign positive probability to a "non-rational" choice function.

Proposition 8.6. *Let ρ have a shades of indifference representation (with an indifference class of at least three alternatives).*[25] *Then there is no reference order $\succ$ such that ρ is a SCRUM w.r.t. $\succ$.*

Proof. Let $\succ$ be a reference order and suppose that ρ is SCRUM w.r.t. $\succ$. Consider a set $\{a, b, c\}$ of three elements such that $a \sim_1 b \sim_1 c$. W.l.o.g. assume that $a \succ b \succ c$ (the other cases are symmetric). Since ρ is a SCRUM and $\rho(b, \{a, b, c\}) > 0$ it follows that $\rho(a, \{a, b, c\}) = \rho(a, \{a, b\})$. But, note that $\frac{1}{|\max(\succsim_1, \{a,b,c\})|} = \frac{1}{3} < \frac{1}{2} = \frac{1}{|\max(\succsim_1, \{a,b\})|}$. Hence, it follows that $\rho(a, \{a, b, c\}) = \sum_{i \in I} \mu_i \frac{\mathbf{1}\{a \in \max(\succ_i, \{a,b,c\})\}}{|\max(\succsim_i, \{a,b,c\})|} < \sum_{i \in I} \mu_i \frac{\mathbf{1}\{a \in \max(\succ_i, \{a,b\})\}}{|\max(\succsim_i, \{a,b\})|} = \rho(a, \{a, b\})$. This is a contradiction to the centrality axiom characterizing SCRUM (see Apesteguia et al. (2017)). □

Appendix A. Proofs omitted from main text

A.1. **Proof of theorem 2.10.** It will be convenient to allow division by zero in the proof below. We will use the convention that $\frac{x}{0} = \infty$ for all positive real numbers x.

Proof. Let $\rho : X \times \mathcal{A} \mapsto [0, 1]$ be a stochastic choice function and π a reference rule. We will find an RSM representation of ρ by using a recursive construction.

STEP 1: Let $\hat{\rho}_1 = \rho$. For each $A \in \mathcal{A}$ define a correspondence C_1 by setting $C_1(A) = \{a \in A : \hat{\rho}_1(a, A) > 0\}$. Set

$$\mu(C_1) = \min_{A \in \mathcal{A}} \min_{a \in C_1(A)} \frac{\hat{\rho}_1(a, A)}{\pi(a, C_1(A))}$$

Note that $\mu(C_1) < \infty$, since for all $A \in \mathcal{A}$ we have $\pi(a, C_1(A)) > 0$ for some $a \in A$.

[24] A proof is available from the author upon request.

[25] If all indifference classes of $\succsim_1$ are less than or equal to two, then one can show that ρ is indeed a SCRUM.

STEP 2: Define a function $\hat{\rho}_2 : X \times \mathcal{A} \mapsto [0,1]$ by setting

$$\hat{\rho}_2(a,A) = \begin{cases} \hat{\rho}_1(a,A) - \mu(C_1)\pi(a, C_1(A)) & \text{if } a \in C_1(A) \\ \hat{\rho}_1(a,A) & \text{otherwise} \end{cases} \tag{3}$$

Set $C_2(A) = \{a \in A : \hat{\rho}_2(a,A) > 0\}$. If $C_2(A) \neq \emptyset$ for some $A \in \mathcal{A}$ set

$$\mu(C_2) = \min_{A \in \mathcal{A}} \min_{a \in C_2(A)} \frac{\hat{\rho}_2(a,A)}{\pi(a, C_2(A))},$$

(note that $\mu(C_2) < \infty$ by similar reasoning as above) and continue to next step. Otherwise, stop. Assume that C_i, $\mu(C_i)$ and $\hat{\rho}_i$ for all $i \leq k$ has already been defined.

STEP $k+1$: Define a function $\hat{\rho}_{k+1} : X \times \mathcal{A} \mapsto [0,1]$ by setting

$$\hat{\rho}_{k+1}(a,A) = \begin{cases} \hat{\rho}_k(a,A) - \mu(C_k)\pi(a, C_k(A)) & \text{if } a \in C_k(A) \\ \hat{\rho}_k(a,A) & \text{otherwise} \end{cases} \tag{4}$$

Note that $\hat{\rho}_{k+1}(a,A) \geq 0$, since $\frac{\hat{\rho}_k(a,A)}{\pi(a,C_k(A))} \geq \mu(C_k)$ for all $A \in \mathcal{A}$ by equation (5). Set $C_{k+1}(A) = \{a \in A : \hat{\rho}_{k+1}(a,A) > 0\}$ for all $A \in \mathcal{A}$. If $C_{k+1}(A) \neq \emptyset$ for some $A \in \mathcal{A}$, set

$$\mu(C_{k+1}) = \min_{A \in \mathcal{A}} \min_{a \in C_{k+1}(A)} \frac{\hat{\rho}_{k+1}(a,A)}{\pi(a, C_{k+1}(A))} \tag{5}$$

and note that $\mu(C_{k+1}) < \infty$ and continue to next step. Otherwise, i.e. if $C_{k+1}(A) = \emptyset$ for all $A \in \mathcal{A}$, stop.

We will first show that the algorithm above stops, i.e. that there is a smallest number $k \in \mathbb{N}$ such that $C_k(A) = \emptyset$ for all $A \in \mathcal{A}$. But this follows since the universal set X is finite and since $C_{k+1}(A) \subseteq C_k(A)$ for all $A \in \mathcal{A}$ and $C_{k+1}(B) \subset C_k(B)$ for some $B \in \mathcal{A}$.

We next show that the algorithm above indeed gives an RSM representation of ρ. Let $\mu_1, ..., \mu_m$ and $C_1, ..., C_m$ be obtained from the algorithm where $m+1$ is the step at which the algorithm stops.

We first show that the $C_1, ..., C_m$ are non-empty valued choice correspondences. I.e. that $C_i(A) \neq \emptyset$ for all $A \in \mathcal{A}$ and $i \in \{1, ..., m\}$. Let $k \in \{1, ...m\}$. By construction (of the algorithm) we know that $C_k(A) \neq \emptyset$ for some $A \in \mathcal{A}$ (otherwise we would have a contradiction to $m+1$ being the smallest number at which the algorithm stops).

To show that $C_k(A) \neq \emptyset$ for all $A \in \mathcal{A}$, it suffices to note that for all $A \in \mathcal{A}$:

$$\sum_{a\in A}\hat\rho_m(a,A)=\sum_{a\in A\setminus C_m(A)}\hat\rho_m(a,A)+\sum_{a\in C_m(A)}\hat\rho_m(a,A)=$$
$$=\sum_{a\in A\setminus C_m(A)}\hat\rho_{m-1}(a,A)+\sum_{a\in C_m(A)}(\hat\rho_{m-1}(a,A)-\mu(C_m)\pi(a,C_m(A)))=$$
$$=\sum_{a\in A}\hat\rho_{m-1}(a,A)-\mu(C_m)=...=\sum_{a\in A}\rho(a,A)-\sum_{i\in\{1,\dots,m\}}\mu(C_i)=$$
$$=1-\sum_{i\in\{1,\dots,m\}}\mu(C_m)=...=\sum_{b\in B}\hat\rho_m(b,B).$$

By the equalities established above it then follows that if $C_m(A)\neq\emptyset$ and $C_m(B)=\emptyset$ for some $B\in\mathcal{A}$, then $\sum_{a\in A}\hat\rho_m(a,A)>0=\sum_{b\in B}\hat\rho_m(b,B)=0$. A contradiction.

We next show that μ is indeed a probability measure. To show this, it suffices to show that $\sum_{i\in I}\mu(C_i)=1$ and that μ is positive, i.e. that $\mu(C_i)\geq 0$ for all $i\in I$. Positivity of μ follows by equation (5) and since $\hat\rho_k(a,A)\geq 0$ for all $k\geq 1$. To see that probabilities sum to one, note that it follows by the recursion in equation (4) that $\hat\rho_{k+1}(a,\{a\})=\hat\rho_k(a,\{a\})-\mu(C_{k+1})\pi(a,C_{k+1}(\{a\}))=\hat\rho_k(a,\{a\})-\mu(C_{k+1})$ for all $k\geq 1$. Using this it then follows that

$$\sum_{i=1}^{m}\mu(C_i)=\mu(C_1)+...+\mu(C_n)=$$
$$=(\rho(a,\{a\})-\hat\rho_1(a,\{a\}))+...+(\hat\rho_{m-1}(a,\{a\})-\hat\rho_m(a,\{a\}))=\rho(a,\{a\})=1.$$

Thus μ is a probability measure on $\mathcal{C}$.

We finally show that the defining formula (1) for an RSM holds, i.e. that

$$\rho(a,A)=\sum_{i\in I}\mu_i\pi(a,C_i(A))$$

for all $a\in A$ and $A\in\mathcal{A}$.

Let $a\in A$ and $A\in\mathcal{A}$ with $\rho(a,A)>0$. By reasoning above there is a smallest k such that $a\notin C_{k+1}(A)$ and note that $\hat\rho_k(a,A)=\mu(C_k)\pi(a,C_k(A))$ for this k. It hence follows that

$$\sum_{i\in I}\mu_i\pi(a,C_i(A))=\sum_{i=1}^{k}\mu(C_i)\pi(a,C_i(A))=\mu(C_k)\pi(a,C_k(A))+\sum_{i=1}^{k-1}\mu(C_i)\pi(a,C_i(A))=$$
$$=\hat\rho_k(a,A)+\sum_{i=1}^{k-1}(\hat\rho_i(a,A)-\hat\rho_{i+1}(a,A))=\hat\rho_1(a,A)=\rho(a,A).$$

Finally, if $\rho(a,A)=0$ then $a\notin\{a\in A:\rho(a,A)>0\}=C_1(A)$ and since $C_i(A)\subseteq C_1(A)$ it follows that $a\notin C_i(A)$ for any $i\geq 1$ implying that $\sum_{i\in I}\mu_i\pi(a,C_i(A))=0$. □

A.2. **Proof of theorem 3.4.**

Proof. We fist show that (1) implies (2). Assume that π satisfies weak regularity. Let ρ be an SCF. We will show that ρ has a unique RSM representation. Assume that $\rho(a, A) = \sum_{i \in I} \mu_i \pi(a, C_i(A))$ and $C_i(A) \subseteq \{a \in A : \rho(a, A) > 0\}$ for all $A \in \mathcal{A}$, where $(C_i)_{i \in I}$ is a sequential collection of choice correspondences.

STEP 1: We first show that $\pi(a, C_k(A)) > 0$ for all $a \in C_k(A) \setminus C_{k+1}(A)$ for all $k \geq 1$ with $C_k(A) \neq C_{k+1}(A)$.

As base case, we show that $\pi(a, C_1(A)) > 0$ for all $a \in C_1(A) \setminus C_2(A)$ with $C_1(A) \neq C_2(A)$. Let $a \in C_1(A) \setminus C_2(A)$. If $\pi(a, C_1(A)) = 0$ then $\rho(a, A) = 0$. A contradiction. As induction hypothesis, assume that $\pi(a, C_k(A)) > 0$ for all $a \in C_k(A) \setminus C_{k+1}(A)$ with $C_k(A) \neq C_{k+1}(A)$ for all $k < m$. Let $a \in C_m(A) \setminus C_{m+1}(A)$. Assume, by contradiction, that $\pi(a, C_m(A)) = 0$. If $\pi(a, C_i(A)) = 0$ for all $i < m$ then $\rho(a, A) = 0$. A contradiction. Hence $\pi(a, C_i(A)) > 0$ for some $i < m$ where $C_i(A) \neq C_{i+1}(A)$. By induction hypothesis it follows that $\pi(b, C_i(A)) > 0$ for all $b \in C_i(A) \setminus C_{i+1}(A)$. Since $\pi(a, C_i(A)) > 0$ it follows that $\pi(a, C_i(A)) > 0$ and since π satisfies weak regularity it follows that $\pi(a, C_m(A)) > 0$. As we wanted to show.

STEP 2: This step entails showing uniqueness of the representation.

Assume that ρ has two RSM representations with parameters $\mu, (C_i)_{i \in I}$ and $\mu', (C'_i)_{i \in I'}$. We first note that $C_1(A) = \{a \in A : \rho(a, A) > 0\} = C'_1(A)$. By assumption it holds that $C_1(A) \subseteq \{a \in A : \rho(a, A) > 0\}$ and clearly if $\rho(a, A) > 0$ then $a \in C_1(A)$ (since otherwise $a \notin C_i(A)$ for all $i \geq 1$, a contradiction). Next, we show that

$$\mu_1 = \min_{A \in \mathcal{A}} \min_{a \in C_1(A)} \frac{\rho(a, A)}{\pi(a, C_1(A))}.$$

To see this, note that $\mu_1 \leq \frac{\rho(a,A)}{\pi(a,C_1(A))}$ for all $a \in A$ and $A \in \mathcal{A}$. Further, since $C_1(A) \neq C_2(A)$ for some $A \in \mathcal{A}$ there is an $a \in C_1(A) \setminus C_i(A)$ for all $i > 1$. Hence, it follows that $\pi(a, C_j(A)) = 0$ for all $j > 1$ and we have $\rho(a, A) = \mu_1 \pi(a, C_1(A))$. As induction hypothesis, assume that $\mu_i = \mu'_i$ and $C_i = C'_i$ for all $i \leq k$. We divide the proof that the statement holds for step $k + 1$ into two cases:

Case 1: We show that $C_{k+1} = C'_{k+1}$. Assume not, then there is an $A \in \mathcal{A}$ with $C_{k+1}(A) \neq C'_{k+1}(A)$. Assume w.l.o.g. that there is an $a \in C_{k+1}(A) \setminus C'_{k+1}(A)$ (the other case being symmetric). First, note that $\pi(a, C'_i(A)) > 0$ for some $i < k+1$. If not, then $\pi(a, C'_i(A)) = 0$ for all $i < k + 1$ and $a \notin C'_j(A)$ for all $j \geq k + 1$ implies that $\rho(a, A) = 0$. By induction hypothesis $C_i = C'_i$ and thus it follows that $\pi(a, C_i(A)) > 0$ for some $i \leq k + 1$. Let

$i^* \le k+1$ be the largest index with $\pi(a, C_{i^*}(A)) > 0$. We claim that $i^* = k+1$. Assume that $i^* < k+1$ (otherwise $\pi(a, C_{k+1}(A)) = \pi(a, C_{i^*}(A)) > 0$ and we are done). Further, note that $C_{i^*}(A) \neq C_{i^*+1}(A)$ since otherwise $\pi(a, C_{i^*}(A)) > 0$ implies $\pi(a, C_{i^*+1}(A)) = \pi(a, C_{i^*}(A)) > 0$, contradicting the maximality of i^*. By STEP 1 above it follows that $\pi(b, C_{i^*}(A)) > 0$ for all $b \in C_{i^*}(A) \setminus C_{i^*+1}(A)$ and since $\pi(a, C_{i^*}(A)) > 0$ and π is weakly regular it follows that $\pi(a, C_{i^*+1}(A)) > 0$ and hence $\pi(a, C_{i^*+1}(A)) > 0$. A contradiction again. Thus $\pi(a, C_{k+1}(A)) > 0$. It follows that $\rho(a, A) - \sum_{i=1}^{k} \mu_i \pi(a, C_i(A)) \ge \mu_{k+1}\pi(a, C_{k+1}(A)) > 0$ and using the induction hypothesis we have that $\rho(a, A) - \sum_{i=1}^{k} \mu_i' \pi(a, C_i'(A)) > 0$. But, since $a \notin C_j'(A)$ for all $j \ge k+1$, we must have $\rho(a, A) - \sum_{i=1}^{k} \mu_i' \pi(a, C_i'(A)) \le 0$. A contradiction.

Case 2: We next have to show that

$$\mu_{k+1} = \min_{A \in \mathcal{A}} \min_{a \in C_{k+1}(A)} \frac{\rho(a, A) - \sum_{i=1}^{k} \mu_i \pi(a, C_i(A))}{\pi(a, C_{k+1}(A))}.$$

There are two subcases:

Case 2a: If $k+1 = n$ then clearly $\mu_n \pi(a, C_n(A)) = \rho(a, A) - \sum_{i=1}^{n-1} \mu_i \pi(a, C_i(A))$ and the claim follows.

Case 2b: If $k+1 < n$ then it is clear that $\mu_{k+1}\pi(a, C_{k+1}(A)) \le \rho(a, A) - \sum_{i=1}^{k} \mu_i \pi(a, C_i(A))$ for all $a \in A$ and $A \in \mathcal{A}$. Further, since $C_{k+1}(A) \neq C_{k+2}(A)$ for some $A \in \mathcal{A}$, there is an $a \in A$ with $a \in C_{k+1}(A) \setminus C_j(A)$ for all $j > k+1$. Further, since $a \in C_{k+1}(A) \setminus C_{k+2}(A)$ it follows by STEP 1 that $\pi(a, C_{k+1}(A)) > 0$. Since $\pi(a, C_{k+1}(A)) > 0$ we have $\rho(a, A) = \sum_{i=1}^{k} \mu_i \pi(a, C_i(A)) + \mu_{k+1}\pi(a, C_{k+1}(A))$, so $\mu_{k+1} = \frac{\rho(a,A) - \sum_{i=1}^{k} \mu_i \pi(a, C_i(A))}{\pi(a, C_{k+1}(A))}$. This finishes the proof of the theorem.

We next show that (2) implies (1). Suppose that π fails to satisfy weak regularity. Then there are $A, B \subseteq X$ with $\pi(b, A) > 0$ for all $b \in A \setminus B$ such that $a \in B \subset A$ but $\pi(a, A) > 0$ and $\pi(a, B) = 0$. Since $X \in \mathcal{A}$, let ρ be a stochastic choice function such that $\rho(a, X) > 0$ if and only if $a \in A$. More specifically, define ρ s.t. $\frac{\rho(a,X)}{\pi(a,A)} = \frac{\rho(b,X)}{\pi(b,A)} < \frac{\rho(c,X)}{\pi(c,A)}$ for all $c \in B \setminus \{a\}$ and $b \in A \setminus B$. To obtain two distinct representations of ρ, carry out the algorithm used to prove theorem 2.10 to obtain an RSM representation $C_1, ..., C_k, C_{k+1}, ..., C_n$. Next, carry out the algorithm until the first step k where $C_{k+1}(X) \neq C_1(X)$ and note that we have $C_{k+1}(X) = B \setminus \{a\}$. Instead of continuing with C_{k+1} define $C_{k+1}'(X) = B$ at step $k+1$ (this is possible since $\pi(a, B) = 0$) and then continue the algorithm as before to obtain an RSM representation with $C_1, ..., C_k, C_{k+1}', ..., C_m'$. Since $C_{k+1}' \neq C_{k+1}$ we have two distinct representations. This finishes the proof that (2) implies (1). $\square$

Since regularity implies weak regularity theorem 3.2 follows as a direct corollary to theorem 3.4.

A.3. **Proof of lemma 6.5.**

Proof. Satisficing with menu independent thresholds: Idempotence is straightforward. Let $i < j$. Then $C_j(C_i(A)) = \{a \in C_i(A) : u(a) > \delta_j\} = \{a \in A : u(a) > \delta_j\} = C_j(A)$, where the second equality follows since $\delta_j > \delta_i$. Similarly, it follows that $C_i(C_j(A)) = \{a \in C_j(A) : u(a) > \delta_i\} = \{a \in A : u(a) > \delta_j\} = C_j(A)$.

Shades of indecisiveness: So suppose that $(C_i)_{i\in I}$ has a shades of indecisiveness representation. I.e. the collection $(C_i)_{i\in I}$ satisfies $C_k(A) = \{x \in A : y \succ_k x \text{ for no } y \in A\}$ for all $A \in \mathcal{A}$, where $\succsim_1, ..., \succsim_I$ on X are such that for each $k, l \in \{1, ..., |I||\}$ with $k < l$ we have that $a \succ_k b$ implies that $a \succ_l b$. We prove idempotence. Let $i < j$. Then $C_i(C_j(A)) = \{x \in C_j(A) : y \succ_i x \text{ for no } y \in C_j(A)\} \subseteq C_j(A)$. Assume that $x \in C_j(A)$ and that $y \succ_i x$ for some $y \in C_j(A)$. Then $y \succ_i x$ implies $y \succ_j x$, which yields a contradiction to $x \in C_j(A)$. We next show that $C_j(C_i(A)) = C_j(A)$. Let $x \in C_j(A)$ then $y \succ_j x$ for no $y \in A$. Hence $y \succ_j x$ for no $y \in C_i(A)$, so $C_j(A) \subseteq C_j(C_i(A))$. Conversley, assume that $x \notin C_j(A)$ then $y \succ_j x$ for some $y \in A$. We claim that $z \succ_j x$ for some $z \in C_j(A)$. If $y \in C_j(A)$ we are done. Otherwise, there is an $y' \succ_j y$ and since $y \succ_j x$ we have $y' \succ_j x$. If $y' \in C_j(A)$ we are done again. If not, continue as before. Since A is finite the claim follows. But since $z \succ_j x$ and $z \in C_j(A) \subseteq C_i(A)$ it follows that $x \notin C_j(C_i(A))$. As we wanted to show.

Multiple rationales: We show that the sequential multiple rationales model is idempotent. Let $\succ_1, ..., \succ_n$ be a collection of preferences. We first show that $C_j \circ C_i = C_i$ where $i < j$ and $C_j(A) = \bigcup_{k\in I(j)} \max(A, \succ_k)$ and $C_i(A) = \bigcup_{k\in I(i)} \max(A, \succ_k)$ and $I(j) \subseteq I(i)$. If $x \in C_j(A)$ then there is a $k \in I(j)$ with $x \succ_k y$ for all $y \in A$. Since $k \in I(j) \subseteq I(i)$ and $x \succ^k y$ for all $y \in C_i(A)$, we have that $x \in C_j(C_i(A))$. Conversely, if $x \notin C_j(A)$ then for each $k \in I(j)$ we have $y_k = \max(A, \succ_k) \succ_k x$. Since $y_k \in C_j(A) \subseteq C_i(A)$, this implies that $x \notin C_j(C_i(A))$.

We next show that $C_i \circ C_j = C_j$. Clearly $C_i(C_j(A)) \subseteq C_j(A)$ since by construction $C_i(A) \subseteq A$ for all $A \in \mathcal{A}$. Let $x \in C_j(A)$ then there is a $k \in I(j)$ with $x \succ_k y$ for all $y \in A$. Since $k \in I(j) \subseteq I(i)$ and $x \succ_k y$ for all $y \in C_j(A)$, it follows that $x \in C_i(C_j(A))$.

Checklist user: This model is a special case of the shades of indecisiveness model. Indeed, as shown in Mandler et al. (2012) a checklist user acts as if maximizing a weak order on alternatives. Idempotence thus follows from idempotence of the shades of indecisiveness model. □

A.4. **Proof of proposition 6.6.** The following lemma on the convergence of a first order linear difference equation is used in the proof of proposition 6.6.

Lemma A.1. *Let $x_0 \in \mathbb{R}$ and let $(x_t)_{t=1}^{\infty}$ and $(b_t)_{t=0}^{\infty}$ be sequences such that $x_{t+1} = ax_t + b_t$ and $|b_t| < 1$ for all $t \geq 0$ and $a < 1$. Then $(x_t)_{t=0}^{\infty}$ is convergent.*

Proof. Using an induction argument it is straightforward to show that

$$x_t = a^t x_0 + \sum_{k=0}^{t} b_k a^{t-k-2}.$$

First note that $|\sum_{k=0}^{t} b_k a^{t-k-2}| \leq \sum_{k=0}^{t} a^{t-k-2} \leq \sum_{k=0}^{\infty} a^{t-k-2}$. Since each term of the series $\sum_{k=0}^{t} b_k a^{t-k-2}$ is non negative and bounded by the real number $\sum_{k=0}^{\infty} a^{t-k-2}$ it follows that the series is convergent. Clearly, the sequence $a^t x_0$ is also convergent. Thus, since x_t is the sum of two convergent series it follows that x_t is convergent. $\square$

We are now ready to present our proof of proposition 6.6.

Proof. Define $\mu_i(0) = \mu_i$ for all $i \in \{1, ..., n\}$ and

$$\mu_i(t+1) = \mu_1\mu_i(t) + \mu_2\mu_i(t) + ... + \mu_{i-1}\mu_i(t) + \mu_i\left[\mu_1(t) + ... + \mu_i(t)\right]$$

for all $t \geq 1$ and $i \in \{1, ..., n\}$. We claim that for all $t \geq 0$ we have $\rho_t(a, A) = \sum_{i=1}^{n} \mu_i(t)\rho_0(a, C_i(A))$. The base case $t = 1$ holds by assumption. Suppose the claim is true for all $k \leq t$. Then

$$\rho_{t+1}(a, A) = \sum_{i=1}^{n} \mu_i \rho_t(a, C_i(A)) = \sum_{i=1}^{n} \mu_i \left[\sum_{j=1}^{n} \mu_j(t)\rho_0(a, C_j \circ C_i(A))\right] =$$

$$= \sum_{k=1}^{n} \left[\sum_{j=1}^{k-1} \mu_j \mu_k(t) + \mu_k \sum_{i=1}^{k} u_i(t)\right] \rho_0(a, C_k(A)) = \sum_{i=1}^{n} \mu_i(t+1)\rho_0(a, C_i(A)).$$

The second equality follows by the induction assumption. The third equality follows since $C_j \circ C_i(A) = C_k(A)$ if and only if $i, j \leq k$ and $i = k$ or $j = k$. The last equality follows by construction of the sequences $\mu_i(t)$.

We next show that $\lim_{t\to\infty} \mu_i(t) = 0$ for all $i \in \{1, ..., n-1\}$. First note that $\mu_1(t+1) = \mu_1\mu_1(t)$ for all $t \geq 1$, which gives the explicit solution $\mu_1(t) = (\mu_1)^t$ and since $\mu_1 \in (0, 1)$ it follows that $\lim_{t\to\infty} \mu_1(t) = 0$. Assume that $\lim_{t\to\infty} \mu_i(t) = 0$ for all $i \in \{1, ..., k\}$ where $k < n - 1$. We claim that $\lim_{t\to\infty} \mu_{k+1}(t) = 0$. Since $\mu_{k+1}(t+1) = [\mu_1 + ... + \mu_{k+1}]\mu_{k+1}(t) + \mu_{k+1}\left[\mu_1(t) + ... + \mu_k(t)\right]$, it follows by lemma A.1 that $\lim_{t\to\infty} \mu_{k+1}(t) = \mu^*$ for some $\mu^* \in \mathbb{R}$. We claim that $\mu^* = 0$. To see this, note that

$$\begin{aligned}\mu^* - (\mu_1 + ... + \mu_k + \mu_{k+1})\,\mu^* &= \lim_{t\to\infty}\left[\mu_{k+1}(t+1) - (\mu_1 + ... + \mu_k + \mu_{k+1})\,\mu_{k+1}(t)\right] \\ &= \lim_{t\to\infty} \mu_{k+1}\left[\mu_1(t) + ... + \mu_k(t)\right] = 0.\end{aligned}$$

Hence, since $\mu_1 + ... + \mu_k + \mu_{k+1} < 1$ it follows that $\mu^* = 0$. Finally, since $\sum_{i=1}^n \mu_i(t) = 1$ for all $t \geq 1$, it follows that $\lim_{t\to\infty} \mu_n(t) = 1$. Thus

$$\lim_{t\to\infty} \rho_t(a, A) = \lim_{t\to\infty} \sum_{i=1}^{n} \mu_i(t)\rho_0(a, C_i(A)) = \rho_0(a, C_n(A)),$$

as we wanted to show. □

A.5. **Proof of proposition 7.2.**

Proof. Assume by contradiction that ρ has two distinct PRC representations μ on $\{c_1, ..., c_n\}$ and μ' on $\{c'_1, ..., c'_m\}$. Define $C_i(A) = \cup_{j=i}^n c_j(A)$ and $C'_i(A) = \cup_{j=i}^m c'_j(A)$ for all $i \geq 1$. Then $\rho(a, A) = \sum_{i=1}^n \mu_i \pi_c(a, C_i(A)) = \sum_{i=1}^m \mu'_i \pi_c(a, C'_i(A))$. Moreover, $C_1(A) \subseteq S^\rho(A)$ and $C'_1(A) \subseteq S^\rho(A)$ for all $A \in \mathcal{A}$. It follows by corollary 4.4 that $\mu = \mu'$, $n = m$ and that $C_i = C'_i$ for all $i \in \{1, ..., n\}$. We prove by induction that $c_i = c'_i$ for all $i \in \{1, ..., n\}$.

Base case: Clearly $c_n = C_n = C'_n = c'_n$. If $c_{n-1}(A) = c_n(A)$ then $C_{n-1}(A) = C_n(A)$ and $C'_{n-1}(A) = C_{n-1}(A) = C_n(A) = C'_n(A)$. Thus $c'_{n-1}(A) = c'_n(A)$. If $c_{n-1}(A) \neq c_n(A)$, then $c_{n-1}(A) = C_{n-1}(A) \setminus C_n(A) = C'_{n-1}(A) \setminus C'_n(A) = c_{n-1}(A)$.

Assume as induction hypothesis that $c_i = c'_i$ for all $i \geq k$ and $c_{i-1}(A) = C_{i-1}(A) \setminus C_i(A)$ if $C_{i-1}(A) \neq C_i(A)$ and $c_{i-1}(A) = c_i(A)$ otherwise (assume similarly that $c'_{i-1}(A) = C'_{i-1}(A) \setminus C'_i(A)$ if $C'_{i-1}(A) \neq C'_i(A)$ and $c'_{i-1}(A) = c'_i(A)$ otherwise).

Proof of induction step: We claim that $c_{k-1} = c'_{k-1}$. Let $A \in \mathcal{A}$. If $c_{k-1}(A) = c_k(A)$ then $C'_{k-1}(A) = c'_{k-1}(A) \cup ... \cup c'_n(A) = c_{k-1}(A) \cup ... \cup c_n(A) = c_k(A) \cup ... \cup c_n(A) = C_k(A) = C'_k(A)$. Hence $c'_k(A) = c'_{k-1}(A)$. If $c_{k-1}(A) \neq c_k(A)$ then $c_{k-1}(A) = C_{k-1}(A) \setminus C_k(A) = C'_{k-1}(A) \setminus C'_k(A) = c'_{k-1}(A)$. □

Appendix B. Remark 1: Example and proof

The following example shows that condition (2) in definition 2.3 is necessary for identification in the Random Sequential Model.

Example B.1. Consider a universal choice set of three alternatives $X = \{a, b, c\}$ and $a \succ b \succ c$. Assume that $C_1(A) = C_2(A)$ for all $A \neq X$ and $C_1(X) = X$ and $C_2(X) = \{c\}$. Define an SCF ρ by

$$\rho(a, A) = \frac{1}{2}\mathbf{1}\{a = \max(\succ, C_1(A))\} + \frac{1}{2}\mathbf{1}\{a = \max(\succ, C_2(A))\}$$

for all $a \in A$ and $A \subseteq X$. Note that $C_1(X) = X \neq \{a, c\} = \{x \in X : \rho(x, X) > 0\}$, so ρ violates condition (2) of an RSM.

We next show that there is a distinct RSM representation ρ. Define an SCF ρ' by

$$\rho'(a, A) = \frac{1}{2}\mathbf{1}\{a = \max(\succ, C_1'(A))\} + \frac{1}{2}\mathbf{1}\{a = \max(\succ, C_2'(A))\}$$

for all $a \in A$ and $A \subseteq X$, where $C_1'(X) = \{a, c\}$ and $C_1'(B) = B$ for all $B \neq X$ and $C_2' = C_2$.

It is then straightforward to verify that $\rho = \rho'$, but the representation is not unique since $C_1'(X) = \{b, c\} \neq X = C_1(X)$. $\triangleleft$

The next result shows that condition (2) in definition 2.3 is implied by condition (1), and hence superfluous, when either the reference rule π or the stochastic choice function ρ is positive.

Proposition B.2. *If ρ satisfies condition (1) of an RSM, and if either π or ρ is positive, then $C_i(A) \subseteq \{a \in A : \rho(a, A) > 0\}$ for all $A \in \mathcal{A}$ and $i \in I$.*

Proof. If ρ is positive the result is immediate. Next, assume that π is positive. If suffices to show that $C_1(A) \subseteq \{a \in A : \rho(a, A) > 0\}$ for all $A \in \mathcal{A}$. Let $a \in C_1(A)$ then $\pi(a, C_1(A)) > 0$ and since ρ has an RSM representation it follows by the defining equation of an RSM (condition 1) that $\rho(a, A) > 0$. $\square$

Appendix C. Example C.1

Example C.1. Consider a choice set of three alternatives $X = \{a, b, c\}$. Suppose that the reference rule π is such that $\pi(b, \{b, c\}) = 0$ and $\pi(a, A) = \frac{1}{|A|}$ for all $a \in A$ and $A \neq \{b, c\}$. Note, that π fails to satisfy weak regularity, since $\pi(a, A) > 0$ and $\pi(b, A) > 0$, but $\pi(b, A \setminus a) = 0$.

Define an SCF ρ by

$$\rho(a, A) = \mu_1 \pi(a, C_1(A)) + \mu_2 \pi(a, C_2(A))$$

for all $a \in A$ and $A \subseteq X$, where $C_1(X) = X$ and $C_2(X) = \{b, c\}$ and $C_1(B) = C_2(B) = B$ for all $B \neq X$ and $\mu_1 = \mu_2 = \frac{1}{2}$.

Define an SCF ρ' by

$$\rho'(a, A) = \mu_1' \pi(a, C_1'(A)) + \mu_2' \pi(a, C_2'(A))$$

for all $a \in A$ and $A \subseteq X$, where $C_1'(X) = X$ and $C_2'(X) = \{c\}$ and $C_1(B) = C_2(B) = B$ for all $B \neq X$ and $\mu_1' = \mu_2' = \frac{1}{2}$.

It is then straightforward to verify that $\rho = \rho'$, but the representation is not unique since $C_2'(X) = \{c\} \neq \{b, c\} = C_2(X)$. $\triangleleft$

References

J. Abaluck and A. Adams-Prassl. What do consumers consider before they choose? Identification from asymmetric demand responses. *The Quarterly Journal of Economics*, 136(3): 1611–1663, 2021.

D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236, 2011.

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

V. H. Aguiar, M. J. Boccardi, and M. Dean. Satisficing and stochastic choice. *Journal of Economic Theory*, 166:445–482, 2016.

A. Ahumada and L. Ülkü. Luce rule with limited consideration. *Mathematical Social Sciences*, 93:52–56, 2018.

F. Aleskerov, D. Bouyssou, and B. Monjardet. *Utility maximization, choice and preference*, volume 16. Springer Science & Business Media, 2007.

P. P. Analytis, F. Cerigioni, A. Gelastopoulos, and H. Stojic. Sequential choice and self-reinforcing rankings. Technical report, 2022.

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

J. Apesteguia and M. A. Ballester. A measure of rationality and welfare. *Journal of Political Economy*, 123(6):1278–1310, 2015.

J. Apesteguia and M. A. Ballester. Random utility models with ordered types and domains. 2021.

J. Apesteguia, M. A. Ballester, and J. Lu. Single-crossing random utility models. *Econometrica*, 85(2):661–674, 2017.

R. J. Aumann. Utility theory without the completeness axiom. *Econometrica*, pages 445–462, 1962.

N. Balakrishnan, E. A. Ok, and P. Ortoleva. Inferential choice theory. 2021.

A. V. Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107 (3):797–817, 1992.

L. Barseghyan, M. Coughlin, F. Molinari, and J. C. Teitelbaum. Heterogeneous choice sets and preferences. *Econometrica*, 89(5):2015–2048, 2021.

G. S. Becker. A note on restaurant pricing and other examples of social influences on price. *Journal of political economy*, 99(5):1109–1116, 1991.

E. Bouacida. Identifying choice correspondences. 2021.

A. Caplin, M. Dean, and D. Martin. Search and satisficing. *American Economic Review*, 101 (7):2899–2922, 2011.

Y.-K. Che and J. Hörner. Recommender systems as mechanisms for social learning. *The Quarterly Journal of Economics*, 133(2):871–925, 2018.

V. Cherepanov, T. Feddersen, and A. Sandroni. Rationalization. *Theoretical Economics*, 8 (3):775–800, 2013.

M. Costa-Gomes, C. Cueva, and G. Gerasimou. Choice, deferral and consistency. *Quantitative Economics, forthcoming*, 2021.

S. Curcuru, J. Heaton, D. Lucas, and D. Moore. Heterogeneity and portfolio choice: Theory and evidence. In *Handbook of financial econometrics: Tools and techniques*, pages 337–382. Elsevier, 2010.

V. Dardanoni, P. Manzini, M. Mariotti, H. Petri, and C. J. Tyson. Mixture choice data: Revealing preferences and cognition. *Journal of Political Economy, forthcoming*, 2022.

M. Draganska and D. Klapper. Choice set heterogeneity and the role of advertising: An analysis with micro and macro data. *Journal of Marketing Research*, 48(4):653–669, 2011.

R. Dutta. Gradual pairwise comparison and stochastic choice. *Theoretical Economics*, 15(4): 1335–1364, 2020.

R. Dutta and S. Horan. Inferring rationales from choice: Identification for rational shortlist methods. *American Economic Journal: Microeconomics*, 7(4):179–201, 2015.

F. Echenique and K. Saito. General luce model. *Economic Theory*, 68(4):811–826, 2019.

F. Echenique, S. Lee, and M. Shum. The money pump as a measure of revealed preference violations. *Journal of Political Economy*, 119(6):1201–1223, 2011.

F. Echenique, K. Saito, and G. Tserenjigmid. The perception-adjusted luce model. *Mathematical Social Sciences*, 93:67–76, 2018.

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

G. Ellison and D. Fudenberg. Rules of thumb for social learning. *Journal of Political Economy*, 101(4):612–643, 1993.

E. Filiz-Ozbay and Y. Masatlioglu. Progressive random choice. *Journal of Political Economy, forthcoming*, 2022.

P. C. Fishburn. Stochastic utility. In S. Barbera, P. J. Hammond, and C. Seidl, editors, *Handbook of utility theory*, pages 273–318. Kluwer Dordrecht, 1998.

M. Frick. Monotone threshold representations. *Theoretical Economics*, 11(3):757–772, 2016.

D. Fudenberg, R. Iijima, and T. Strzalecki. Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409, 2015.

T. Galaabaatar and E. Karni. Subjective expected utility with incomplete preferences. *Econometrica*, 81(1):255–284, 2013.

G. Gerasimou. Model-rich approaches to eliciting weak preferences: Evidence from a multi-valued choice experiment. *Working Paper*, 2021.

M. S. Goeree. Limited information and advertising in the us personal computer industry. *Econometrica*, 76(5):1017–1074, 2008.

W. H. Greene and D. A. Hensher. Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation*, 37(3):413–428, 2010.

W. T. Harbaugh, K. Krause, and T. R. Berry. Garp for kids: On the development of rational choice behavior. *American Economic Review*, 91(5):1539–1545, 2001.

S. Horan. Stochastic semi-orders. *Journal of Economic Theory*, 192:105171, 2021.

J. Huber, J. W. Payne, and C. Puto. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research*, 9(1): 90–98, 1982.

T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting click-through data as implicit feedback. In *Acm Sigir Forum*, volume 51, pages 4–11. Acm New York, NY, USA, 2005.

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

N. Kashaev and V. H. Aguiar. A random attention and utility model. *Working Paper*, 2021.

M. Kovach and G. Tserenjigmid. The focal luce model. *American Economic Journal: Microeconomics, forthcoming*, 2020.

G. Loomes, C. Starmer, and R. Sugden. Observing violations of transitivity by experimental methods. *Econometrica*, pages 425–439, 1991.

R. D. Luce. *Individual choice behavior: A theoretical analysis*. 1959.

M. Mandler, P. Manzini, and M. Mariotti. A million answers to twenty questions: Choosing by checklist. *Journal of Economic Theory*, 147(1):71–92, 2012.

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

P. Manzini and M. Mariotti. Choice by lexicographic semiorders. *Theoretical Economics*, 7 (1):1–23, 2012.

P. Manzini, M. Mariotti, and C. J. Tyson. Two-stage threshold representations. *Theoretical Economics*, 8(3):875–882, 2013.

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

H. Petri. Binary single-crossing random utility models. *Working Paper*, 2021.

A. K. Sen. Choice functions and revealed preference. *The Review of Economic Studies*, 38 (3):307–317, 1971.

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

K. E. Train. Recreation demand models with taste differences over people. *Land economics*, pages 230–239, 1998.

C. Turansick. Identification in the random utility model. *Working Paper*, 2021.

A. Tversky. Intransitivity of preferences. *Psychological review*, 76(1):31, 1969.

A. Tversky. Choice by elimination. *Journal of Mathematical Psychology*, 9(4):341–367, 1972.

C. J. Tyson. Cognitive constraints, contraction consistency, and the satisficing criterion. *Journal of Economic Theory*, 138(1):51–70, 2008.

Department of Economics, University of Bath, United Kingdom, BA2 7AY
Email address: henrik@petri.se