Imprecise Data and Stochastic Choice*

Henrik Petri†

September 14, 2020

Abstract: We propose a framework of stochastic choice to accommodate situations where data on choices is imprecise/ambiguous. The primitive of our analysis is an imprecise data set. It contains information about the ordinal ranking of choice frequencies but not their exact values. Despite the limited information contained in an imprecise data set we argue that it can be fruitfully explored to draw inferences about choice. To this end we consider the testable implications of several models of stochastic choice, including the random utility model and subclasses of it, as well as the Luce model and close relatives. The testable implications of these models on imprecise data sets are in general strong. For some models, and as far as one is concerned about the testable implications of a model, going from exact to imprecise data is even immaterial.

Keywords: stochastic choice, random utility model, imprecise probability
JEL codes: C65, D11

*I thank Jay Lu, Paola Manzini, Marco Mariotti, Shasi Nandeibam and Efe A. Ok for valuable comments and encouragement.

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

1. Introduction

An important aim of discrete choice analysis is to describe and make predictions about human choice behavior. To this end, an array of theoretical models of choice have been developed. A key step in obtaining a better understanding of these models, is the characterization of the behavior consistent with them. I.e. determining their empirical content. The aim is to find a set of testable properties describing their behavior. The idea is that it should be possible to verify/test these properties by using observed data on choices. This exercise not only helps in understanding the scope and limitations of the models, but may also guide the empiricist in designing tests of the models.

Typically, the aforementioned models and their properties are studied in a quite idealized setting. It is assumed that an analyst/outside observer has access to a precise data set as encoded by a stochastic choice function (SCF). Formally, an SCF assigns *exact* probability distributions over alternatives to each menu in a collection of menus. However, in a number of situations quantifying exact choice frequencies is demanding (see example 1.1 -1.3 below). Indeed, in reality choice probabilities are *estimated* using subsamples of data and are in this sense always imprecise. An implication of this is that properties shown to describe a model using precise/exact data, may no longer do so with imprecise data. These properties hence hold *conditional on* the data being precise. Given impreciseness of data, basing a test on such a property seems questionable. Understanding how and if impreciseness of data impacts the observable/testable implications of models of choice thus seems important.

This paper proposes a framework to study the empirical content of stochastic choice models with imprecise data. To illustrate the type of situations with imprecise data we are interested in, we consider some examples:

Example 1.1. An analyst conducts a survey and asks a representative sample of 120 individuals to pick their favourite brand among chocolates $\{x, y, z\}$. Out of the 120 individuals only 100 individuals report their favourite brand and the corresponding choice frequencies are $\{61, 30, 9\}$. Given these choice frequencies it is not clear how to assign exact choice probabilities to the different brands. The probabilities depend on the individuals who did not participate in the survey. In contrast, the reported frequencies induce a likelihood ranking of alternatives: $x \succ y \succ z$. This ranking is *independent* of the 20 individuals not participating in the survey. $\lhd$

Example 1.2. An analyst conducts a survey and asks a representative sample of 130 individuals to pick their favourite brand among chocolates $\{x, y, z\}$. Out of the 130 individuals only 100 individuals report their favourite brand and the corresponding choice frequencies of the brands are $\{61, 30, 9\}$.The observed ordinal ranking of choice frequencies is $x \succ y \succ z$. In this case there is no complete ranking $\succ$ that is independent of the 30 individuals not participating in the survey. If all of them would choose z then the corresponding choice frequencies would be $\{61, 30, 39\}$, which would give a ranking $x \succ z \succ y$. Note, however, that the incomplete ranking declaring $x \succ y$, $x \succ z$ and y and z as incomparable is independent of this information. $\lhd$

Example 1.3. A grocery store keeps track of the number of individuals purchasing items from a set $\{x, y, z, u\}$. Suppose that the grocery store registers the following purchases of items: $\{45, 79, 41, 0\}$. Without further information on the number of individuals visiting the store and their respective purchases it is difficult to calculate exact choice probabilities. Some individuals may for instance buy bundles of several items when visiting the store. However, an analyst may still distinguish chosen alternatives (such as x, y, z) from never chosen alternatives (such as u). $\triangleleft$

The examples above share two common features:

1. The researcher observes choice frequencies as encoded by some SCF.

2. There is additional external information introducing uncertainty/ambiguity in the evaluation of choice frequencies.

Due to (2) calculating exact choice probabilities to use as input in the inference problem is difficult in each of these examples. Applying the standard stochastic choice framework is therefore problematic. To elaborate, consider example 1.1, it is not clear which SCF the researcher should use as input in her inference problem as it depends on the 20 individuals not participating in her survey. It would clearly be convenient if the researcher could study the induced ranking $\succ$ which is independent of this information. Similar reasoning applies to example 1.2 where it would be convenient if the researcher could directly study the incomplete relation $\succ$ declaring $x \succ y$, $x \succ z$ and y and z as incomparable. Or consider example 1.3 where studying a deterministic correspondence, declaring x, y, z as chosen, would make sense. Common to these examples is that data is extracted from an observed SCF by "forgetting" about exact choice frequencies and by considering relative likelihoods of alternatives in menus.

Formally, we propose to model imprecise data sets as a binary relation $\succsim$ on pairs (a, A) of choices $a \in A$ and menus A in a collection of menus $\mathscr{A}$. The interpretation of $(a, A) \succsim (b, B)$ being that a is as likely chosen in menu A as b is in B. To capture the idea that $\succsim$ is obtained from an observed SCF ρ by ignoring some of its information (as in the examples above), we impose on $\succsim$ a stochastic choice function ρ such that the ranking induced by ρ is "consistent" with $\succsim$. To accommodate examples such as example 1.2 and 1.3, we allow $\succsim$ to be incomplete.[1]

Ultimately, we are interested in understanding the behavior (empirical content) of stochastic choice models when data is imprecise. By modelling imprecise data as suggested above, it is quite clear that going from full data on choices (as encoded by an SCF) to imprecise data involves loss of information. Behaviors that were easily distinguishable using full data on choices, SCFs, may no longer be so with (certain types of) imprecise data. Given the less rich nature of imprecise data, it hence seems natural to ask (and these are the main queries addressed in the paper):

> *i) Are there models for which impreciseness is not an issue in the sense that imprecise data conveys as much about their underlying behavior as full data (an SCF) does?*
>
> *ii) For models that are sensitive to impreciseness, what can still be said about their behavior?*

To address query i), section 3 introduces our main concept of an ordinal model. An ordinal model is (in a specific sense) *immune* to impreciseness of data. Roughly speaking, it is

[1] Formally, we assume that ρ extends $\succsim$. That is for every pair (a, A) and (b, B) that are comparable according to $\succsim$ it must be that $\succsim$ and ρ agree about their ranking. I.e. we require that $\succsim$ is such that $(a, A) \succ (b, B)$ implies $\rho(a, A) > \rho(b, B)$ and $(a, A) \sim (b, B)$ implies $\rho(a, A) = \rho(b, B)$.

possible to fully describe the behavior of an ordinal model by using imprecise data. We illustrate this concept further below (see subsection 1.1 of the introduction). There is an abundance of ordinal models: the simple scalability model (Tversky, 1972), the additive perturbed utility model (APU) (Fudenberg, Iijima, and Strzalecki, 2015), the single crossing random utility model (Apesteguia, Ballester, and Lu, 2017), the dual random utility model (Manzini and Mariotti, 2018) and the random attention model (Cattaneo, Ma, Masatlioglu, and Suleymanov, 2020) are all such models. Furthermore, some models of stochastic choice require substantially less information in describing their behavior. To accommodate this, we introduce several natural weakenings of ordinality. The weakest notion of ordinality says that it is possible to fully describe the behavior of a model using deterministic data on choices (such as in example 1.3).

Not all models are ordinal, both the random utility model (RUM) and the Luce model fail to be ordinal. This does, however, not imply that the testable implications of these models, using imprecise data, are vacuous. To adress query ii), i.e. to study the empirical content of non-ordinal models with imprecise data, section 3 introduces the concept of the *ordinal closure* of a model. Roughly speaking, the ordinal closure of a model is the best description possible with imprecise data. I.e. it is not possible to distinguish the behavior of a model from its ordinal closure when data is imprecise. Impreciseness of data, is perhaps of the greatest concern for non ordinal models. As we will explain further below (see the discussion following the illustrative example in section 1.1), failing to acknowledge impreciseness of data may even cause a researcher to commit a type I error of rejecting a true null hypothesis of data being generated by a model. For such reasons, understanding the testable implications of non ordinal models is important.

To summarize, we identify two main classes of SCFs: ordinal and non-ordinal models. Ordinal models is our attempt to address query i) above, as their observable implications are unaffected by (certain types of) impreciseness. In an attempt to address query ii) we study the ordinal closures of various (non-ordinal) models of choice.

The remainder of the paper is devoted to studying versions of these notions and applying them to (well known) models of choice. We consider several notions of ordinality each corresponding to different types of imprecise data. Example 1.1 considers one particular type of imprecise data. In a sense, the impreciseness of that data set is "small". We also consider more extreme cases of impreciseness where less information on choices is available. As a particular example we consider data sets with only deterministic information on choices available (see example 1.3). Arguably, the degree of impreciseness of any data set "naturally occurring" in practice is in between these two in strength. I.e. it has more information on choices than a deterministic correspondence but less so than the full ranking induced by an SCF. Hence, as we discuss further in section 6.1, knowledge of the behavior of a model on these data sets is in many cases sufficient to infer their behavior on general imprecise data sets. For illustrative purposes we also consider two examples of data sets intermediate in strength of impreciseness. In the first case the researcher only observes how alternatives compare to each other (in terms of choice frequencies) within menus, whereas in the second case, the researcher observes how alternatives compare across menus. We give an exact description of these data sets in section 2.

In section 4 and 5, we consider specialized models of choice and their ordinal properties. Section 4 is confined to the classical random utility model and various subclasses of it. The classical random utility model is not ordinal. We characterize the ordinal closure of the random utility model. This problem is closely related to that of finding a representation of a subjective probability relation (see Kraft, Pratt, and Seidenberg (1959), Scott (1964), Fishburn (1969, 1986), Insua (1992), Alon and Lehrer (2014)). Our main axiom is cancellation, inspired by similar properties in subjective probability. This result thus links two hitherto largely unconnected literatures: one on stochastic/discrete choice and the other on subjective/imprecise probability.

Another main finding is that the observable implications of the random utility model are vacuous for a large collection of imprecise data sets. More precisely, the random utility model is not testable whenever the data set is such that it only contains information about the relative rankings of alternatives within menus. This result illustrates the rich behavior associated with a RUM and is in stark contrast to for example the Luce model, where the ranking of alternatives within menus needs to agree with a fixed utility function. Besides the classical random utility model, we consider the single crossing RUM of Apesteguia et al. (2017) and the dual random utility model of Manzini and Mariotti (2018). Interestingly, both of these models are ordinal. The dual random utility model is not only ordinal, but also deterministic. This means that it is possible to describe the behavior of this model using deterministic data on choices.

Section 5 considers ordinal properties of the Luce model and close relatives to it. More specifically we consider the classical Luce model, the simple scalability model (SSM) Tversky (1972), the APU model Fudenberg et al. (2015) and the gradual pairwise comparison rule (GPCR) Dutta (2019). Our main finding is that the SSM, APU model and GPCR are all ordinal. An implication is that it is possible to empirically differentiate between these models using imprecise data.

Section 6 contains a discussion of the framework adopted in the current paper. In particular section 6.1 discusses how to extend the findings of section 4 and 5 to general imprecise data sets. Section 6.2 discusses various types of imprecise data sets not covered by the present framework. We show how to apply our framework to classify properties of stochastic choice functions in section 6.3. In particular, and under special circumstances, the framework may be used to obtain characterizations of deterministic choice rules from characterizations of their corresponding stochastic choice rules and vice versa. We apply this technique to study the dual random utility model, obtaining a behavioral characterization of its deterministic counterpart, the top-and-the-top rule in Eliaz, Richter, and Rubinstein (2011).

The paper is concluded by a discussion of related literature in section 7.

1.1. An illustrative example

As a preview to our main ideas, we discuss a numerical example:

Example 1.4. An analyst observes choice frequencies from a population of 100 individuals from menus $A = \{a, b, c\}$ and $B = \{a, b\}$. The reported data looks as follows

	a	b	c
$\{a, b, c\}$	42	21	37
$\{a, b\}$	58	29	0

As is evident from the table the choice frequencies from the menus are not easily comparable. There are 100 observations from $\{a, b, c\}$, whereas there are only 87 observations from $\{a,b\}$.[2] Suppose that the researcher would like to test whether this data conforms to the Luce model, or equivalently the (Independence of Irrelevant Alternatives) IIA property: $\frac{\rho(a,A)}{\rho(b,A)} = \frac{\rho(a,B)}{\rho(b,B)}$ for all $a,b \in A \cap B$ and $A,B \in \mathscr{A}$. By inspecting data it is clear that $\frac{\rho(a,A)}{\rho(b,A)} = \frac{42}{21} = \frac{58}{29} = \frac{\rho(a,B)}{\rho(b,B)}$, suggesting that the IIA property holds. However, this conclusion ignores the issue of impreciseness of data due to the 13 non-participants in B. To see this consider the following table:

	a	b	c
$\{a,b,c\}$	42	21	37
$\{a,b\}$	$58+x$	$29+y$	0

For each positive x,y let $\rho_{x,y}$ be the SCF obtained from the data above. Note that x and y are unknown to the analyst and $x+y=13$. Further, note that for many values of x,y we have $\frac{\rho_{x,y}(a,A)}{\rho_{x,y}(b,A)} = \frac{42}{21} \neq \frac{58+x}{29+y} = \frac{\rho_{x,y}(a,B)}{\rho_{x,y}(b,B)}$. Due to the uncertainty pertaining to x,y it is not clear that we should attribute the behavior above to the Luce model. Depending on the choices of the non-participants it may or may not be consistent with the Luce model, but this is difficult to tell unless we elicit these choices as well. $\lhd$

As seen in example 1.4, the Luce model is sensitive to the choices of the non-participants and does, in this sense, not coop very well with imprecise data. In contrast, consider the APU model of Fudenberg, Iijima, and Strzalecki (2015). One can show that $\rho_{x,y}$ is an APU SCF for all $x,y \geq 0$ and $x+y=13$. Thus, *irrespective* of the choices of the non-participants, the data above is consistent with the APU model. In contrast to the Luce model, impreciseness seems to be less of an issue for the APU model. The reason behind this is, perhaps as anticipated, that the APU model is *ordinal*: transform the choice probabilities of an APU SCF using a strictly increasing (menu independent) map $v:[0,1] \to [0,1]$ and the transformed SCF will still be an APU. Roughly speaking, only ordinal information about choice probabilities is needed in describing its behavior. Note that the ordinal ranking induced by the choice probabilities in example 1.4 is independent of the non-participants x,y, i.e. for all positive x,y with $x+y=13$ we have:

$$\rho_{x,y}(a,A) \geq \rho_{x,y}(b,B) \Leftrightarrow \rho(a,A) \geq \rho(b,B). \tag{1}$$

Thus for each x,y, there is an increasing map $v_{x,y}:[0,1] \to [0,1]$ such that $\rho_{x,y} = v_{x,y} \circ \rho$. Now, since ρ is represented by a Luce SCF (it satisfies IIA) and since every Luce SCF is an APU, it follows by ordinality of the APU model that each $\rho_{x,y}$ is an APU. Quite strikingly, the behavior (empirical content) of the APU model is independent of impreciseness of data. As explained above, we will later show (see section 4 and 5) that this observation holds for many well known models of choice.

However, as already noted, not all models are ordinal. Example 1.4 shows that the Luce model fails to be ordinal. Whether or not the IIA property holds depends on the choices of the non-participants. Should we then reject the hypothesis of the data being generated by a Luce model? Although it is unclear whether the data in example 1.4 is generated by a Luce model, there is a sense in which it still is. The observed SCF ρ satisfies IIA and is

[2]The important property of this is example is that there are missing observations from at least one of the menus. It is possible to construct an example where equally many alternatives are missing from both menus A and B (which might be more realistic). For expositional purposes the current example is more convenient.

hence a Luce SCF. By equation (1) it then follows that each SCF $\rho_{x,y}$ agrees with the *ordinal ranking* of a Luce SCF. With imprecise data (as above), this is in a way the best description possible (assuming we want to explain the data using a Luce rule). I.e. it is not possible to distinguish the behavior of a Luce model from that of an increasing transformation of it when data is imprecise (as above). The ordinal closure of a model is defined as the set of SCFs being increasing transformations of/ordinally equivalent to some SCF in the model. The data above is thus consistent with an SCF in the ordinal closure of the Luce model.

To further illustrate this notion, suppose that a researcher would like to test the Luce model using imprecise data. Rejecting it based on an observed violation of the IIA property would, in light of the above discussion, not always be a correct decision. An observed violation of IIA may be due to impreciseness of data rather than actual choice behavior of individuals. Consider the following version of example 1.4.

Example 1.5. An analyst observes choice frequencies from a population of 100 individuals from menus $A = \{a, b, c\}$ and $B = \{a, b\}$. The reported data looks as follows

	a	b	c
$\{a, b, c\}$	42	21	37
$\{a, b\}$	58	34	0

$\triangleleft$

Effectively, we can think of the data above as resulting from the data in example 1.4 and from 5 of the 13 non-participants choosing b from $\{a, b\}$. I.e. there are now only 8 non-participants. Suppose that an analyst observes the data in example 1.5. Then she might be inclined to "reject" the Luce model based on an observed violation of the IIA property. Would this be a correct decision? Not necessarily. As noted above, the data in example 1.5 is in the ordinal closure of the Luce model (which follows using exactly the same reasoning as in example 1.4). Hence it *might* still be generated by a Luce SCF, but with imprecise data as above we cannot tell.

Thus, in order to test non ordinal models using imprecise data, it is instrumental to isolate the properties that these models satisfy independently of any impreciseness of data (or non-participants in the example above). Doing this is equivalent to studying the properties satisfied by their ordinal closures. Loosely, these are the set of properties of a model that are left *invariant* (unaffected) by increasing transformations of the data. By the examples above it follows that the IIA property is not such a property. However, every Luce SCF satisfy the property: $\rho(a, A) \geq \rho(b, A)$ if and only if $\rho(a, B) \geq \rho(b, B)$ and this property is preserved by increasing transformations of data, and hence in the ordinal closure of the Luce model. An observed violation of this property would, in the context of example 1.4 and 1.5, perhaps be more convincing evidence against the Luce model than an observed violation of the IIA property.

2. Imprecise data sets

Denote by X a finite set of alternatives. A subset $A \subseteq X$ is called a *menu*. We denote by $\mathscr{A}$ a collection of menus, i.e. $\mathscr{A} \subseteq 2^X \backslash \emptyset$. A *stochastic choice function* (SCF) is a function $\rho : X \times \mathscr{A} \mapsto [0,1]$ such that we have $\sum_{a \in A} \rho(a, A) = 1$ for all menus $A \subseteq X$ and $\rho(a, A) = 0$ for all $a \in X \backslash A$. This paper takes as primitive a binary relation $\succsim$ on $X \times \mathscr{A}$. The triplet

$(X, \mathscr{A}, \succsim)$ is called a *data set*. A data set is *total* if for all $(a, A), (b, B) \in X \times \mathscr{A}$: $(a, A) \succsim (b, B)$ or $(b, B) \succsim (a, A)$, it is *non-trivial* if there are $(a, A), (b, B) \in X \times \mathscr{A}$ with $(a, A) \succ (b, B)$. A data set, as encoded by an *arbitrary* binary relation, is in general very permissive. Not all data sets, can be properly interpreted as arising from choices of individuals. In order for a data set to be consistent with population choice, as in discrete choice analysis, more structure has to be imposed on it. To this end, a natural requirement is to impose on $\succsim$ a stochastic choice function ρ such that $\succsim$ agrees with the ordinal ranking induced by ρ. To allow for incomparability we assume that $\succsim$ uses a subset of the available information about ρ. Every stochastic choice function ρ induces a relation $\succsim_\rho$ on $X \times \mathscr{A}$ defined by: $(a, A) \succsim_\rho (b, B) \Leftrightarrow \rho(a, A) \geq \rho(b, B)$.

Definition 2.1. A data set $(X, \mathscr{A}, \succsim)$ is an *imprecise data set* if there is a stochastic choice function ρ such that $\succsim_\rho$ is an ordering extension[3] of $\succsim$. $\lhd$

Equivalently, a data set is an imprecise data set if there is a stochastic choice function ρ such that $(a, A) \succ (b, B)$ implies $\rho(a, A) > \rho(b, B)$ and $(a, A) \sim (b, B)$ implies $\rho(a, A) = \rho(b, B)$. Definition 2.1 is meant to capture the idea that an imprecise data set is obtained from an SCF by ignoring part of the information as encoded by an SCF. The researcher starts with an observed SCF ρ and then only considers the induced ordinal ranking $\succsim_\rho$ and pairs $(a, A), (b, B)$ that are comparable according to $\succsim$. Note also that imprecise data sets generalize stochastic choice functions along two dimensions by allowing choice probabilities to be both imprecise *and* incomparable.

Some important special cases of definition 2.1 are discussed next. The main part of the text will be centered around these examples. We choose to focus on the cases below mainly for two reasons: simplicity and tractability. We indicate how to extend the analysis to general imprecise data sets in section 6.1. As we discuss in section 6.1 focusing on these examples is in many cases without loss of generality as properties of models on these data sets may be used to infer properties on more general data sets. The first case below illustrates a situation where the researcher observes the full ranking induced by some SCF.

Definition 2.2. A data set with $\succsim = \succsim_\rho$ for some SCF ρ is called a *complete data set*. $\lhd$

A complete data set corresponds to a case of "pure" impreciseness in that the complete ordinal ranking induced by some SCF ρ is observed. Studying $\succsim_\rho$ could be useful in situations, like example 1.1 where an analyst has some degree of confidence in the ordinal ranking of alternatives, but is perhaps not convinced that observed choice frequencies, as encoded by ρ, reflect exact choice frequencies of the population. We will also consider some weakenings of $\succsim_\rho$ where less than the complete data set is observed.

Definition 2.3. Every SCF ρ induces a data set $(X, \mathscr{A}, \succsim_{\rho|X})$ defined by:

$$(a, A) \succsim_{\rho|X} (b, B) \Leftrightarrow \rho(a, A) \geq \rho(b, B) \text{ and } A = B$$

for all $(a, A), (b, B) \in X \times \mathscr{A}$. A data set with $\succsim = \succsim_{\rho|X}$ for some SCF ρ is called an *within menu data set*. $\lhd$

Intuitively, relation $\succsim_{\rho|X}$ captures within menu variation of ρ. This relation could be useful in situations where a researcher have reliable information about the relative ranking of

[3]A relation $\succsim'$ is an ordering extension of $\succsim$ if $\succsim \subseteq \succsim'$ and $\succ \subseteq \succ'$.

choice frequencies within menus, but where choice frequencies across menus are difficult to compare. As we will see later, there are models of stochastic choice whose empirical content could be fully described by using the limited information in the $\succsim_{\rho|X}$ ranking. Further, it is possible to distinguish between many models of (stochastic) choice using this information. The next example illustrates a scenario where data on choices is naturally encoded by a $\succsim_{\rho|X}$ relation.

Example 2.1. A researcher thinks that choices are governed by a stochastic choice function ρ^*. She conducts an experiment to test her hypothesis. However, she also believes that with small but positive probability some subjects in her experiment will choose a random option not conforming to the ρ^* process. This could be either because these subjects miss-interpret the choice task, or because they lack the incentives to reveal their true choices. The researcher hence observes an SCF ρ where:

$$\rho(a, A) = (1 - \alpha(A))\rho^*(a, A) + \alpha(A)\frac{1}{|A|}$$

for all $a \in A$ and $A \in \mathscr{A}$, where $\alpha : \mathscr{A} \to [0, 1]$ is an error function. Observing ρ could the researcher infer any (useful) properties about ρ^*? In this case the researcher observes the full $\succsim_{\rho^*|X}$ ranking, so it would clearly be convenient if the researcher could infer properties about ρ^* by studying $\succsim_{\rho^*|X}$. Note that the error term is menu dependent to reflect that subjects' incentives to reveal their true choices differ across menus. Subjects may for instance be less inclined to reveal their true choices when menus are large as opposed to small (as the cognitive costs to do so are higher). $\triangleleft$

Definition 2.4. Every SCF ρ induces a data set $(X, \mathscr{A}, \succsim_{\rho|\mathscr{A}})$ defined by:

$$(a, A) \succsim_{\rho|\mathscr{A}} (b, B) \Leftrightarrow \rho(a, A) \geq \rho(b, B) \text{ and } a = b$$

for all $(a, A), (b, B) \in X \times \mathscr{A}$. A data set with $\succsim = \succsim_{\rho|\mathscr{A}}$ for some SCF ρ is called an *across menu data set.* $\triangleleft$

In contrast, $\succsim_{\rho|\mathscr{A}}$ captures the across menu variation of stochastic choice models. This relation could be of use in situations where the researcher have imperfect information about the ranking of alternatives within menus. The following example illustrates a case where choice probabilities are comparable across (but not within) menus.

Example 2.2. An analyst observes choice frequencies from a population of 100 individuals from menus $A = \{a, b, c\}$ and $B = \{a, b\}$. The reported data looks as follows

	a	b	c
$\{a, b, c\}$	30	33	37
$\{a, b\}$	45	43	0

As is evident from the table the choice frequencies from the menus are not easily comparable. There are 100 observations from $\{a, b, c\}$, whereas there are only 88 observations from $\{a, b\}$. It is, of course, possible to transform the above choice frequencies to choice probabilities by dividing by the number of observations from each choice set. But this problematic due to the 12 "missing" observations from $\{a, b\}$. Consider the following table:

	a	b	c
$\{a,b,c\}$	30	33	37
$\{a,b\}$	$45+12-x$	$43+x$	0

Note that x is unknown to the analyst and $0 \leq x \leq 12$. If at least seven individuals not participating in $\{a,b\}$ would choose b, when given the choice between a and b, i.e. $x \geq 7$, then $\rho(b,ab) \geq \frac{43+7}{100} = \frac{45+12-7}{100} \geq \rho(a,ab)$. Due to the uncertainty pertaining to x it may hence be convenient for the analyst to leave exact choice probabilities unspecified and incomparable. $\triangleleft$

The next example reflects a case of extreme impreciseness. The researcher is only able to distinguish between alternatives based on whether they are chosen or not in a menu. Define $D_\rho(A) = \{a \in A : \rho(a,A) > 0\}$.

Definition 2.5. Every SCF ρ induces a data set $(X, \mathscr{A}, \succsim_\rho^{\mathscr{D}})$ defined by:

$$(a,A) \succsim_\rho^{\mathscr{D}} (b,B) \Leftrightarrow \mathbf{1}_{D_\rho(A)}(a) \geq \mathbf{1}_{D_\rho(B)}(b)$$

for all $(a,A),(b,B) \in X \times \mathscr{A}$. A data set with $\succsim = \succsim_\rho^{\mathscr{D}}$ for some SCF ρ is called a *deterministic data set.* $\triangleleft$

3. Ordinal properties of stochastic choice models

Our aim is to study the testable implications of various models of choice on imprecise data sets. Viewing an SCF ρ as an imprecise data set $\succsim$ the researcher "forgets about" exact choice frequencies and only considers properties of the ordinal ranking induced by ρ. It is therefore natural to wonder whether there are models of stochastic choice that are "fully described" by observing the limited information in an imprecise data set $\succsim$. And if not, by observing $\succsim$ (where $\succsim$ is "consistent" with ρ as in an imprecise data set) is it possible to infer any useful information about the background SCF ρ? This section thus investigates two related queries:

1. When is observing $\succsim$ equivalent to observing ρ? I.e. are there models of stochastic choice where $\succsim$ conveys as much information about the behavior of the underlying model as ρ does?

2. For models where observing $\succsim$ is not equivalent to observing ρ, what can still be said about their behavior? What is their empirical content?

To study these questions, we will consider various restrictions imposed on the impreciseness of $\succsim$, from the more restrictive (assuming that a researcher observes the full ordinal ranking $\succsim = \succsim_\rho$) to the less restrictive (assuming that only deterministic information about ρ is available). As explained in section 2, we will for tractability reasons focus our discussion around complete data sets $\succsim_\rho$, within menu data sets $\succsim_{\rho|X}$, across menu data sets $\succsim_{\rho|A}$ and deterministic data sets $\succsim_\rho^{\mathscr{D}}$. Section 6.1 provides a framework applicable to more general imprecise data sets. We show that considering the examples above is often without loss of generality, as knowledge of the behavior of models on these data sets may be used to infer their behavior on other imprecise data sets.

3.1. Ordinal models

Let $\mathtt{SCF}$ denote the set of all stochastic choice functions on $X \times \mathscr{A}$. A *model* is a subset $\mathtt{M} \subseteq \mathtt{SCF}$. In an attempt to address (1) above we introduce the concept of an *ordinal* model. Roughly speaking, a model is ordinal if the information provided by an imprecise data set $\succsim_\rho$ is sufficient to determine whether $\rho \in \mathtt{M}$ or not. Hence exact choice probabilities are not required in describing the behavior of the model.

Definition 3.1. A model $\mathtt{M}$ is *ordinal* if $\rho' \in \mathtt{M}$ and $\rho \in \mathtt{SCF}$ and

$$\rho(a,A) > \rho(b,B) \quad \Leftrightarrow \quad \rho'(a,A) > \rho'(b,B)$$

for all $(a,A),(b,B) \in X \times \mathscr{A}$ implies that $\rho \in \mathtt{M}$. ◁

The following proposition characterizes ordinal models.

Proposition 3.1. *The following statements are equivalent:*

1. $\mathtt{M}$ *is an ordinal model.*
2. *For all* $\rho' \in \mathtt{M}$ *and* $\rho \in \mathtt{SCF}$ *and* $\succsim_\rho = \succsim_{\rho'}$ *implies that* $\rho \in \mathtt{M}$.
3. *For all* $\rho \in \mathtt{SCF}$ *with* $\rho = v \circ \rho'$ *for some* $\rho' \in \mathtt{M}$ *and increasing (continuous) function* $v : [0,1] \to [0,1]$ *with* $v(0) = 0$ *we have* $\rho \in \mathtt{M}$.

Remark: If ρ and ρ' are such that $\succsim_\rho = \succsim_{\rho'}$ this not only requires the ranking of choice frequencies to be the same for ρ and ρ', it also requires their supports to be equal.[4] Thus, if a model $\mathtt{M}$ is ordinal then every SCF ρ having the same ordinal ranking *and* support as some $\rho' \in \mathtt{M}$ is in the model. The following example illustrates this feature of the definition further:

Example 3.1. Denote by $\mathtt{DET}$ the class of deterministic SCFs, i.e. $\rho \in \mathtt{DET}$ if and only if there for every $A \in \mathscr{A}$ is an x_A with $\rho(x_A, A) = 1$. $\mathtt{DET}$ is an ordinal model. To see this, let $\rho' \in \mathtt{DET}$ and $\rho \in \mathtt{SCF}$ and $\succsim_\rho = \succsim_{\rho'}$. Since $\succsim_\rho = \succsim_{\rho'}$, we have for all $a \in A$ and $A \in \mathscr{A}$: $\rho(a,A) > 0$ if and only if $\rho'(a,A) > 0$. This immediately gives $\rho \in \mathtt{DET}$. ◁

A proof of proposition 3.1 is in appendix A.4. Ordinal models are, in a specific sense, *immune* to impreciseness of data. Take an SCF ρ in an ordinal model $\mathtt{M}$ and apply any increasing transformation $v : [0,1] \to [0,1]$ to ρ and the new SCF $v \circ \rho$ is still in $\mathtt{M}$. As already discussed in example 1.4 of the introduction, the APU model of Fudenberg et al. (2015) is ordinal. Section 4 and 5 contain several examples of ordinal models. Let us see what ordinality of a model means in light of introductory example 1.1.

Example 1.1 (continuing from p. 2)**.** Recall that out of 120 individuals only 100 individuals report their favourite brand from the set $X = \{x, y, z\}$. The corresponding choice frequencies are $\{61, 30, 9\}$. After analyzing choice data, suppose the analyst concludes that ρ is in model $\mathtt{M}$.[5] If $\mathtt{M}$ is known to be ordinal, this means that the analyst's conclusion is robust/immune

[4] Too see this note that if $\succsim_\rho = \succsim_{\rho'}$ then $\rho(a,A) > \rho(b,B) \quad \Leftrightarrow \quad \rho'(a,A) > \rho'(b,B)$ for all $(a,A),(b,B) \in X \times \mathscr{A}$. This implies that $\rho(a,A) > \rho(b,B) \quad \Leftrightarrow \quad \rho'(a,A) > \rho'(b,B)$ for all $a \in A$ and $b \notin B$ and $A, B \in \mathscr{A}$, thus $\rho(a,A) > 0 \quad \Leftrightarrow \quad \rho'(a,A) > 0$ for all $a \in A$ and $A \in \mathscr{A}$.

[5] In this example there is only one menu X, so almost any conceivable model $\mathtt{M}$ is consistent with the data. The example generalizes straightforwardly to larger collections of menus.

to impreciseness of data caused by the non-participants in her survey. I.e. for all $u, v, w \geq 0$ with $u + v + w = 20$ we have that

$$\rho'(x, X) = \frac{61+u}{120} > \rho'(y, X) = \frac{30+v}{120} > \rho'(z, X) = \frac{9+w}{120},$$

so ρ' is an increasing transformation of the observed SCF ρ with

$$\rho(x, X) = \frac{61}{100} > \rho(y, X) = \frac{30}{100} > \rho(z, X) = \frac{9}{100}$$

and is hence still in $\mathtt{M}$. $\triangleleft$

Ordinal closure of models: As already mentioned, not all models are ordinal. For example the Luce and RUM model fails to be ordinal (see section 4 and 5 for further details). For these models, observing $\succsim_\rho$ is not equivalent to observing ρ in the sense described above. However, in many cases, $\succsim_\rho$ still provides useful information about the underlying population behavior as described by ρ. To understand the testable implications of non ordinal models we study their ordinal closures as defined below.

Definition 3.2. The *ordinal closure* of model $\mathtt{M}$, denoted $\mathscr{O}[\mathtt{M}]$, is the smallest ordinal model containing $\mathtt{M}$. $\triangleleft$

First note that the above definition is well founded since $\mathtt{SCF}$ is an ordinal model. Roughly speaking, the ordinal closure tells us everything there is to know about the ordinal properties of a model. For instance, if the ordinal closure of two models $\mathtt{M}$ and $\mathtt{M}'$ are different, then it is possible to empirically distinguish between the models using the $\succsim_\rho$ orderings. The following proposition characterizes the ordinal closure of a model.

Proposition 3.2. *The following statements are equivalent for an SCF ρ:*

1. *ρ is in the ordinal closure of $\mathtt{M}$, i.e. $\rho \in \mathscr{O}[\mathtt{M}]$*
2. *there is an SCF $\rho' \in \mathtt{M}$ such that $\succsim_\rho = \succsim_{\rho'}$.*
3. *there is a strictly increasing continuous function $v : [0,1] \to [0,1]$ with $v(0) = 0$ and an SCF $\rho' \in \mathtt{M}$ such that $\rho = v \circ \rho'$.*

By proposition 3.2 it follows that a model is ordinal if and only if it equals its ordinal closure, i.e. $\mathscr{O}[\mathtt{M}] = \mathtt{M}$. For non-ordinal models it hence follows that $\mathtt{M}$ is a strict subset of $\mathscr{O}[\mathtt{M}]$. Using this it is apparent why understanding the ordinal closures of models is important. Suppose that a researcher would like to test whether data is generated by model $\mathtt{M}$ and that data is imprecise. If the researcher nevertheless treats choice frequencies as being correctly described by an observed SCF ρ then the null hypothesis of data being generated by $\mathtt{M}$ would be rejected if $\rho \notin \mathtt{M}$. However, it may happen that $\rho \in \mathscr{O}[\mathtt{M}] \setminus \mathtt{M}$ and observing $\rho \in \mathscr{O}[\mathtt{M}] \setminus \mathtt{M}$ is not necessarily a violation of model $\mathtt{M}$. It may very well be that the underlying population behavior is correctly described by model $\mathtt{M}$ but impreciseness of data renders it impossible to tell whether $\rho \in \mathtt{M}$. With imprecise data the analyst can only tell whether $\rho \in \mathscr{O}[\mathtt{M}]$ or not. Thus, if the researcher rejects model $\mathtt{M}$ when observing $\rho \in \mathscr{O}[\mathtt{M}] \setminus \mathtt{M}$ she would sometimes reject a true null hypothesis of data being generated by model $\mathtt{M}$.

3.2. Within and across menu ordinal models

As explained above, a model is ordinal if it is possible to determine whether an SCF belongs to the model or not by using *the full* ordinal ranking of choice probabilities. However, for some models substantially less information is required in describing their behavior. The following two notions are natural variants of the notion of ordinality. The first, within menu ordinality, says that it is possible to determine model membership of an SCF by solely observing rankings of items within menus, i.e. the researcher observes an within menu data set $\succsim_{\rho|X}$ as in definition 2.3.

Definition 3.3. A model $\mathtt{M}$ is *within menu ordinal* if $\rho' \in \mathtt{M}$ and $\rho \in \mathtt{SCF}$ and

$$\rho(a,A) > \rho(b,A) \quad \Leftrightarrow \quad \rho'(a,A) > \rho'(b,A)$$

for all $(a,A),(b,A) \in X \times \mathscr{A}$ implies that $\rho \in \mathtt{M}$. Equivalently, a model is within menu ordinal if for all $\rho' \in \mathtt{M}$ and $\rho \in \mathtt{SCF}$ if $\succsim_{\rho|X} = \succsim_{\rho'|X}$ then $\rho \in \mathtt{M}$. $\lhd$

In contrast the next notion, across menu ordinality, requires the possibility to classify the behavior induced by SCFs using only information about menu rankings.

Definition 3.4. A model $\mathtt{M}$ is *across menu ordinal* if $\rho' \in \mathtt{M}$ and $\rho \in \mathtt{SCF}$ and

$$\rho(a,A) > \rho(a,B) \quad \Leftrightarrow \quad \rho'(a,A) > \rho'(a,B)$$

for all $(a,A),(a,B) \in X \times \mathscr{A}$ implies that $\rho \in \mathtt{M}$. Equivalently, a model is across menu ordinal if for all $\rho' \in \mathtt{M}$ and $\rho \in \mathtt{SCF}$ if $\succsim_{\rho|\mathscr{A}} = \succsim_{\rho'|\mathscr{A}}$ then $\rho \in \mathtt{M}$. $\lhd$

Not all models are within/across menu ordinal. Studying the testable implications of these models (assuming that $\succsim_{\rho|X}$ or $\succsim_{\rho|\mathscr{A}}$ is observed) is equivalent to studying their corresponding within/across menu closures:

Definition 3.5. The within menu closure of model $\mathtt{M}$, denoted $\mathscr{W}[\mathtt{M}]$, is the smallest within menu ordinal model containing $\mathtt{M}$. The across menu closure is defined similarly and is, with slight abuse of notation,[6] denoted $\mathscr{A}[\mathtt{M}]$. $\lhd$

Note that $\mathscr{O}[\mathtt{M}] \subseteq \mathscr{W}[\mathtt{M}]$ and $\mathscr{O}[\mathtt{M}] \subseteq \mathscr{A}[\mathtt{M}]$ for all models $\mathtt{M}$. If a model is within/across menu ordinal it hence follows that the model is ordinal. Roughly speaking, this is because if it is possible to describe the behavior of a model using an imprecise data set then it should be possible to do so using any imprecise data set containing more information on choices as well.

3.3. Deterministic models

The strongest notion of ordinality that we will consider essentially says that the properties satisfied by a model are deterministic. Differently put, such a model is (very) insensitive to the correct specification of choice probabilities.

Definition 3.6. A model $\mathtt{M}$ is a *deterministic model* if $\rho' \in \mathtt{M}$ and $\rho \in \mathtt{SCF}$ and

$$\rho(a,A) > 0 \quad \Leftrightarrow \quad \rho'(a,A) > 0$$

[6] We also use $\mathscr{A}$ to denote the collection of menus on X.

for all $(a, A) \in X \times \mathscr{A}$ implies that $\rho \in \mathtt{M}$. Equivalently, a model is a deterministic model if for all $\rho' \in \mathtt{M}$ and $\rho \in \mathtt{SCF}$ and $\succsim_\rho^{\mathscr{D}} = \succsim_{\rho'}^{\mathscr{D}}$ implies that $\rho \in \mathtt{M}$. ◁

Thus, if a model is deterministic and ρ is an SCF of a certain model, then every SCF with support equal to ρ belongs to the model. Deterministic models represent a case of extreme impreciseness, in that it is possible to pin down model membership although data is (very) imprecise. A deterministic *model* should not be confused with a deterministic stochastic choice function. The former refers to a property satisfied by a class of SCFs, whereas the latter refers to a property satisfied by a single SCF. To elaborate, recall that a deterministic SCF ρ is such that there for every $A \in \mathscr{A}$ is an x_A with $\rho(x_A, A) = 1$. As we will explain later, the dual Random Utility Model of Manzini and Mariotti (2018) is a deterministic model. But it is clear that not every SCF with a dRUM representation is deterministic.

Definition 3.7. The deterministic closure of model $\mathtt{M}$, denoted $\mathscr{D}[\mathtt{M}]$, is the smallest deterministic model containing $\mathtt{M}$. ◁

4. Ordinality of random utility models

We investigate ordinal properties of the random utility model (and close relatives to it). As we will see, a notable property of the random utility model is that it is not ordinal. There are however interesting subclasses of RUMs that are ordinal.

4.1. Classical random utility model

Let $\mathscr{P}$ denote the set of strict linear orders on X. For all $A \subseteq X$ and $a \in A$ denote by $\mathscr{P}(a, A) = \{P \in \mathscr{P} : \max(P, A) = a\}$. A stochastic choice function ρ has a *random utility representation* if there is a measure μ on $\mathscr{P}$ such that $\rho(a, A) = \mu(\mathscr{P}(a, A))$ for all $a \in A$ and $A \in \mathscr{A}$. The *random utility model* is the class of SCFs with a random utility representation and is denoted by $\mathtt{RUM}$.

Proposition 4.1. *The random utility model is not an ordinal model.*

The proposition above follows by the next example.[7]

Example 4.1. Let $X = \{a, b, c, d\}$. First note that any RUM must satisfy the following inequality:

$$\rho(a, ab) - \rho(a, abc) \geq \rho(a, abd) - \rho(a, abcd) \tag{2}$$

We construct a RUM ρ and a stochastic choice function ρ' such that $\succsim_{\rho'} = \succsim_\rho$ and such that ρ' violates the previous inequality. Let linear orders P_1, P_2, P_3 be defined as in the table below.

P_1	P_2	P_3
c	c	d
a	d	a
b	b	b
d	a	c

[7] Example 4.1 requires $|X| > 3$. The random utility model is characterized by monotonicity when $|X| \leq 3$. The set of SCFs satisfying monotonicity is an ordinal model.

Let p_1, p_2, p_3 be numbers such that $0 < p_1 < p_2 < p_3 < 1$ and $p_1 + p_2 + p_3 = 1$. Define a probability measure μ on $\mathscr{P}$ by $\mu(P_i) = p_i$ for all $i \in \{1,2,3\}$. Now we note that:

$$\rho(a, ab) = p_1 + p_3 \tag{3}$$
$$\rho(a, abc) = p_3 \tag{4}$$
$$\rho(a, abd) = p_1 \tag{5}$$
$$\rho(a, abcd) = 0 \tag{6}$$

Thus the inequality in equation (2) is binding. Further, note that there is only one $(b, B) \in X \times \mathscr{A}$ such that $\rho(b, B) = p_1 + p_3$ and this is $(b, B) = (a, ab)$ and similarly there is only one (b, B) such that $\rho(b, B) = p_2$ and this is (b, ab). Let

$$\varepsilon_1 = \min_{(b,B) \in X \times \mathscr{A} \setminus (a,ab)} |\rho(a, ab) - \rho(b, B)|,$$

$$\varepsilon_2 = \min_{(b,B) \in X \times \mathscr{A} \setminus (b,ab)} |\rho(b, ab) - \rho(b, B)|$$

and let $0 < \varepsilon < \frac{\min\{\varepsilon_1, \varepsilon_2\}}{2}$. Define a new stochastic choice function ρ' by $\rho'(a, A) = \rho(a, A)$ for all $(a, A) \in X \times \mathscr{A} \setminus \{(a, ab), (b, ab)\}$ and $\rho'(a, ab) = \rho(a, ab) - \varepsilon$ and $\rho'(b, ab) = \rho(b, ab) + \varepsilon$. It is now straightforward to check that $\rho(a, A) > \rho(b, B) \Leftrightarrow \rho'(a, A) > \rho'(b, B)$ for all $(a, A), (b, B) \in X \times \mathscr{A}$. But note that

$$\rho'(a, ab) - \rho'(a, abc) = p_1 + p_3 - \varepsilon - p_3 = p_1 - \varepsilon < p_1 = \rho'(a, abd) - \rho'(a, abcd).$$

It follows that ρ' fails to satisfy equation (2) and hence cannot have a random utility representation. $\lhd$

Since RUM is not an ordinal model, it may be of interest to know exactly what the ordinal properties of it are, i.e. given an imprecise data set $\succsim_\rho$ what testable restrictions (if any) does the random utility model impose on $\succsim_\rho$?

Characterizing ordinal closure of RUM: The main property is called cancellation and is similar to cancellation properties appearing in the literature on subjective probability Kraft, Pratt, and Seidenberg (1959). Call a pair of sequences $(a_1, A_1), \ldots, (a_n, A_n)$ and $(b_1, B_1), \ldots, (b_n, B_n)$ *admissible* if there are natural numbers $(k_i)_{i=1}^n$ s.t. $\sum_{i=1}^n k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \le \sum_{i=1}^n k_i \mathbf{1}_{\mathscr{P}(b_i, B_i)}(P)$ for all $P \in \mathscr{P}$.

> **Cancellation (C):** There is no admissible sequence such that $(a_i, A_i) \succsim (b_i, B_i)$ for all $i \in \{1, \ldots, n\}$ and $(a_n, A_n) \succ (b_n, B_n)$.

To understand cancellation, consider some of its implications. Cancellation implies the well-known monotonicity/regularity property requiring that for all $a \in X$ and $A, B \in \mathscr{A}$: if $A \subseteq B$ then $\rho(a, A) \ge \rho(b, B)$.

Monotonicity alone is however not sufficient to characterize $\mathscr{O}[\texttt{RUM}]$.[8] Finding a "simple" behavioral characterization of imprecise data sets described by random utility models would clearly be desirable. But this is a challenging problem, related to that of finding a behavioral characterization of the random utility model (which to date is open).

[8] It is straightforward to verify that the following additivity property holds for any such data set: For all $A, B, C \subseteq X$ with $A \supseteq B$ if $\rho(a, A) \ge \rho(a, B)$ then not $\rho(a, B \cup C) > \rho(a, A \cup C)$.

Theorem 4.2. *A data set* $(X, \mathscr{A}, \succsim)$ *satisfies totality and cancellation if and only if there is a measure* μ *on preferences* $\mathscr{P}$ *such that*

$$(a, A) \succsim (b, B) \quad \Leftrightarrow \quad \mu(\mathscr{P}(a, A)) \geq \mu(\mathscr{P}(b, B))$$

for all $(a, A), (b, B) \in X \times \mathscr{A}$.

Note that theorem 4.2 holds for general data sets, i.e. there is a priori no restriction at all on the binary relation $\succsim$. The proof of theorem 4.2 uses a rational version of Farkas' lemma (see appendix A.1 for a statement and proof). A main technical challenge in proving theorem 4.2 is that the collection of sets $\mathscr{P}(a, A)$ is not an algebra of sets and hence theorem 4.2 does not follow as a direct corollary to similar results in subjective probability.[9] Say that a stochastic choice function ρ satisfies cancellation if $\succsim_\rho$ satisfies cancellation. Denote by `Cancellation` the set of SCFs satisfying cancellation. Since $\succsim_\rho$ is total we have the following corollary to theorem 4.2.

Corollary 4.3. *A stochastic choice function* ρ *satisfies cancellation if and only if* $\rho \in \mathscr{O}[\texttt{RUM}]$. *That is,* $\mathscr{O}[\texttt{RUM}] = \texttt{Cancellation}$.

Within/across menu properties of RUM: The following result shows that there for every stochastic choice function ρ is a RUM μ such that the ranking of alternatives induced by the RUM coincides with the ranking of ρ in in every menu A. Note that our representation below only takes variation of choice frequencies *within* menus into account. This result illustrates the rich behavior consistent with a RUM. Indeed, rank the alternatives of any given menu in any way you like and there is a RUM inducing exactly the same ranking of alternatives! This is in quite stark contrast to for example the Luce model, where the ranking of alternatives within menus need to be consistent with a fixed utility function. The main result of this section reads:

Theorem 4.4. *For every stochastic choice function* ρ *there is a RUM* μ *such that*

$$\rho(a, A) \geq \rho(b, A) \;\Leftrightarrow\; \mu(\mathscr{P}(a, A)) \geq \mu(\mathscr{P}(b, A))$$

for all $a, b \in A$ *and* $A \in \mathscr{A}$.

Remark: Theorem 4.4 does not imply that $\mathscr{W}[\texttt{RUM}] = \texttt{SCF}$. The reason is that the above result "ignores" deterministic information about ρ as we require $a, b \in A$. For an SCF to lie in the within menu closure of `RUM` it is required that there is a RUM μ such that $\rho(a, A) \geq \rho(b, A) \;\Leftrightarrow\; \mu(\mathscr{P}(a, A)) \geq \mu(\mathscr{P}(b, A))$ for all $a, b \in X$ and $A \in \mathscr{A}$ (i.e. it may happen that $a, b \notin A$).

This result is perhaps not surprising given that the RUM offers quite great flexibility in that any (strict) preference relation may be part of the support of a RUM. However, a striking

[9] To see this, first note that a subjective probability relation is defined on an algebra of sets. Given a data set $(X, \mathscr{A}, \succsim)$, a natural proof strategy is to define a derived relation $\succsim_0$ on subsets of the form $\mathscr{P}(a, A)$ by $\mathscr{P}(a, A) \succsim_0 \mathscr{P}(b, B)$ if and only if $(a, A) \succsim (b, B)$. The proof of the theorem would be finished if we could show that $\succsim_0$ satisfies the axioms characterizing subjective probability. However, this is not possible since the collection of sets $\mathscr{A} = \{R \subseteq \mathscr{P} : R = \mathscr{P}(a, A)$ for some $(a, A)\}$ is not an algebra of sets. Thus $\succsim_0$ is not defined on an algebra of subsets. In order to apply any of the existing representation results in subjective probability we would have to somehow extend the relation $\succsim_0$ to an algebra of sets including $\mathscr{A}$ and this seems difficult.

consequence of theorem 4.4 is that there is no relation whatsoever between the rankings that a RUM induces on different menus. For instance, it follows as a corollary to theorem 4.4 that there is a RUM such that $\rho(a,A) > \rho(b,A)$ for all $A \supseteq \{a,b\}$ with $|A| > 2$, but with $\rho(b,ab) > \rho(a,ab)$. As already noted above, this behavior distinguishes the RUM from other models of stochastic choice.

The across menu properties of the random utility model are, in contrast to its within menu properties, far from being vacous. Call a pair of sequences $(a_1,A_1),\dots,(a_n,A_n)$ and $(b_1,B_1),\dots,(b_n,B_n)$ *menu admissible* if (1) $a_i = b_i$ for all i, (2) there are natural numbers $(k_i)_{i=1}^n$ such that $\sum_{i=1}^n k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \leq \sum_{i=1}^n k_i \mathbf{1}_{\mathscr{P}(b_i,B_i)}(P)$ for all $P \in \mathscr{P}$.

> **Across menu cancellation:** There is no menu admissible sequence such that $\rho(a_i,A_i) \geq \rho(b_i,B_i)$ for all $i \in \{1,\dots,n\}$ and $\rho(a_n,A_n) > \rho(b_n,B_n)$.

Denote by $\mathtt{AC}$ the set of SCFs satisfying across menu cancellation.

Proposition 4.5. *A stochastic choice function ρ satisfies across menu cancellation if and only if $\rho \in \mathscr{A}[\mathtt{RUM}]$. That is, $\mathscr{A}[\mathtt{RUM}] = \mathtt{AC}$.*

We conjecture that $\mathscr{O}[\mathtt{RUM}] = \mathtt{AC} = \mathscr{A}[\mathtt{RUM}]$. I.e. for the RUM, across menu data is as rich as complete (ordinal) data. In light of theorem 4.4 this would not be very surprising.

Deterministic properties of RUM Since $\mathtt{RUM}$ is not an ordinal model, it follows that $\mathtt{RUM}$ is not deterministic. Thus, the deterministic closure of $\mathtt{RUM}$, $\mathscr{D}[\mathtt{RUM}]$ is a strict superset of $\mathtt{RUM}$. We next provide a characterization of $\mathscr{D}[\mathtt{RUM}]$. The following properties are straightforward analogues of the corresponding deterministic properties.

> **Stochastic Pseudo WARP:** for all $A,B \in \mathscr{A}$, if for all $a \in A$: $\rho(a,A) > 0$ implies $a \in B$ then if $a \in A$ and $\rho(a,B) > 0$ it follows that $\rho(a,A) > 0$.

Denote by $\mathtt{SPWARP}$ the collection of SCFs satisfying stochastic pseudo WARP. We have the following result:

Proposition 4.6. *We have $\mathscr{D}[\mathtt{RUM}] \subseteq \mathtt{SPWARP}$. Moreover, if $\mathscr{A}$ is the collection of all non-empty menus then $\mathscr{D}[\mathtt{RUM}] = \mathtt{SPWARP}$.*

Proof. It is straightforward to check that if $\rho \in \mathscr{D}[\mathtt{RUM}]$ then ρ satisfies stochastic pseudo WARP. Suppose that ρ satisfies stochastic pseudo WARP and $\mathscr{A} = 2^X \setminus \emptyset$. Consider the deterministic correspondence D_ρ defined by $D_\rho(A) = \{a \in A : \rho(a,A) > 0\}$. Then it is straightforward to check that D_ρ satisfies the following property: for all $A,B \in \mathscr{A}$: if $D_\rho(A) \subseteq B$ then $D_\rho(B) \cap A \subseteq D_\rho(A)$. Thus by Manzini and Mariotti (2015, Prop. 6, p.253) there is a collection of linear orders $P_1,\dots,P_n$ on X such that for all $A \in \mathscr{A}$: $D_\rho(A) = \bigcup_{i=1}^n \max(P_i,A)$. Now let ρ' be a RUM with support $P_1,\dots,P_n$ and $\mu(P_i) = \frac{1}{n}$ for all $i \in \{1,\dots,n\}$. Then $D_\rho(A) = D_{\rho'}(A)$ for all $A \in \mathscr{A}$. □

4.2. Single crossing random utility models

In a recent, very interesting paper, Apesteguia et al. (2017) introduce the single crossing random utility model (SCRUM). Formally, let $\succ$ be a linear order. A stochastic choice function ρ is a single crossing random utility model, if there is a measure μ on $\mathscr{P}$ and an ordering

$\{P_1,...,P_T\}$ of the support of μ such that if $x \succ y$ and $s > t$ then xP_ty implies xP_sy. I.e. P_s is more "aligned" with $\succ$ than P_t. Denote by $\texttt{SCRUM}$ the class of SCRUM stochastic choice functions. Clearly $\texttt{SCRUM} \subset \texttt{RUM}$. Since $\texttt{RUM}$ is not an ordinal model, one may expect the same to be true for $\texttt{SCRUM}$. This is however not the case. As the following proposition shows $\texttt{SCRUM}$ is an ordinal model:

Proposition 4.7. $\texttt{SCRUM}$ *is an across menu ordinal model (and hence an ordinal model).*

Proof. Let $\rho \in \texttt{SCRUM}$ and let $\rho' \in \texttt{SCF}$ with $\succsim_{\rho|\mathscr{A}} = \succsim_{\rho'|\mathscr{A}}$. Apesteguia et al. (2017) characterize SCRUMs using monotonicity/regularity and a property called *centrality*: If $x \succ y \succ z$ and $\rho(y,\{x,y,z\}) > 0$, then $\rho(x,\{x,y\}) = \rho(x,\{x,y,z\})$ and $\rho(z,\{y,z\}) = \rho(z,\{x,y,z\})$. Hence it suffices to show that ρ' satisfies monotonicity and centrality. It is immediate that ρ' satisfies monotonicity. Let $x \succ y \succ z$ and $\rho'(y,\{x,y,z\}) > 0$. Then, since $\rho'(y,\{x,z\}) = 0$ it follows that $\rho'(y,\{x,y,z\}) > \rho'(y,\{x,z\})$. Hence $\rho(y,\{x,y,z\}) > \rho(y,\{x,z\}) = 0$. Since ρ satisfies centrality we have $\rho(x,\{x,y\}) = \rho(x,\{x,y,z\})$ and $\rho(z,\{y,z\}) = \rho(z,\{x,y,z\})$ and using that $\succsim_{\rho|\mathscr{A}} = \succsim_{\rho'|\mathscr{A}}$ it immediately follows that $\rho'(x,\{x,y\}) = \rho'(x,\{x,y,z\})$ and $\rho'(z,\{y,z\}) = \rho'(z,\{x,y,z\})$. $\square$

Thus, in contrast to the random utility model, $\texttt{SCRUM}$ coops quite well with across menu data (or any imprecise data set with richer information on choices, see proposition 6.3). However, $\texttt{SCRUM}$ is not an within menu ordinal model (and hence not a deterministic model).

Proposition 4.8. $\texttt{SCRUM}$ *is not an within menu ordinal model.*

Proposition 4.8 follows by example 4.2 below.

Example 4.2. To see that $\texttt{SCRUM}$ is not within menu ordinal. Let $a \succ b \succ c$. Define P_1 by cP_1bP_1a and P_2 by bP_2aP_2c. Then P_1 and P_2 are single crossing w.r.t. $\succ$. Define a SCRUM ρ by $\rho(a,A) = \frac{1}{3}\mathbf{1}\{a = \max(A,P_1)\} + \frac{2}{3}\mathbf{1}\{a = \max(A,P_2)\}$. Note that $\rho(c,\{a,b,c\}) = \rho(c,\{b,c\}) = \frac{1}{3}$ and $\rho(b,\{a,b,c\}) = \rho(b,\{b,c\}) = \frac{2}{3}$. Define an SCF ρ' by $\rho'(b,B) = \rho(b,B)$ for all $b \in B \subseteq \{a,b,c\}$ with $B \neq \{b,c\}$. Set $1-\rho'(c,bc) = \rho'(b,bc) = \frac{3}{4}$. It is then clear that $\rho' \in \mathscr{W}[\texttt{SCRUM}]$. However $\rho'(c,bc) = \frac{1}{4} < \frac{1}{3} = \rho'(c,abc)$, so ρ' is not monotone. Hence $\rho' \notin \texttt{SCRUM}$. $\triangleleft$

Deterministic properties of SCRUM: Call a choice correspondence $c : \mathscr{A} \to 2^X \setminus \emptyset$ a deterministic SCRUM if there is a linear order $\succ$ on X and a collection of linear orders $\{P_1,...,P_T\}$ satisfying the single crossing property such that $c(A) = \bigcup_{i\in\{1,...,T\}} \max(P_i, A)$ for all $A \in \mathscr{A}$. Let $\texttt{detSCRUM}$ denote the set of all SCFs such that D_ρ is a deterministic SCRUM. The following proposition is then immediate:

Proposition 4.9. $\mathscr{D}[\texttt{SCRUM}] = \texttt{detSCRUM}$.

The class of deterministic SCRUMs is characterized in Costa, Ramos, and Riella (2020). As explained there, the properties characterizing deterministic SCRUMs are not obtained by straightforwardly rewriting the properties of SCRUM in terms of the support of a stochastic choice function. The present framework sheds some light on why this is the case. The reason being that SCRUM is not a deterministic model.

4.3. Dual random utility model

The dual Random Utility Model (dRUM) is a variation of the classical random utility model, where utility depends on at most two states and where choice probabilities are allowed to be menu dependent. The class of dRUMs was introduced by Manzini and Mariotti (2018).[10] Formally, a stochastic choice function ρ is a dual Random Utility Model (dRUM) if there is a function $\alpha : 2^X \setminus \emptyset \to (0,1)$ and there are linear orders P_1 and P_2 such that for all $a \in A$ and $A \subseteq X$:

$$\rho(a, A) = \alpha(A)\mathbf{1}\{a = \max(A, P_1)\} + (1 - \alpha(A))\,\mathbf{1}\{a = \max(A, P_2)\}.$$

If there is $\alpha \in (0,1)$ with $\alpha(A) = \alpha$ for all $A \in \mathscr{A}$ then ρ is an independent dRUM (idRUM). Denote the class of (i)dRUMs by $\texttt{(i)dRUM}$. Every idRUM is a RUM, but the class of dRUMs is not necessarily a subset of the class of RUMs, due to the menu dependent choice probabilities $\alpha(A)$. The following proposition shows that only deterministic (hence ordinal) information is needed in determining whether an SCF is a dRUM or not.

Proposition 4.10. $\texttt{dRUM}$ *is a deterministic model.*

Proof. Let ρ be an SCF s.t. there is a function $\alpha : 2^X \setminus \emptyset \to (0,1)$ and there are linear orders P_1 and P_2 such that for all $a \in A$ and $A \subseteq X$: $\rho(a, A) = \alpha(A)\mathbf{1}\{a = \max(A, P_1)\} + (1 - \alpha(A))\,\mathbf{1}\{a = \max(A, P_2)\}$. Let $\rho' \in \texttt{SCF}$ with $D_{\rho'} = D_\rho$. Define $\beta(A) = \rho'(\max(P_1, A), A)$ for all $A \in \mathscr{A}$. Then $\rho'(a, A) = \beta(A)\mathbf{1}\{a = \max(A, P_1)\} + \big(1 - \beta(A)\big)\,\mathbf{1}\{a = \max(A, P_2)\}$. Thus $\rho' \in \texttt{dRUM}$, as we wanted to show. □

Since $\texttt{dRUM}$ is a deterministic model, it follows as a corollary that $\texttt{dRUM}$ is an within/across menu ordinal model as well. In contrast to the situation for dRUMs, much more information about choice probabilities is needed in order to pin down the behavior of idRUMs. As the following proposition shows idRUMs are ordinal, but not necessarily across/within menu ordinal (and hence not deterministic). The intuition behind this result is quite simple as idRUMs put harsh restrictions on choice probabilities in that they are required to be constant across menus.

Proposition 4.11. $\texttt{idRUM}$ *is an ordinal model.* $\texttt{idRUM}$ *is not an across/within menu ordinal model.*

Proof. Let ρ be an SCF s.t. there is an $\alpha \in (0,1)$ and there are linear orders P_1 and P_2 such that for all $a \in A$ and $A \subseteq X$: $\rho(a, A) = \alpha\mathbf{1}\{a = \max(A, P_1)\} + (1 - \alpha)\,\mathbf{1}\{a = \max(A, P_2)\}$. Let $\rho' \in \texttt{SCF}$ and $\rho' = \nu \circ \rho$ where $\nu : [0,1] \to [0,1]$ is a strictly increasing function s.t. $\nu(0) = 0$. Then by checking a few cases (depending on whether $a = \max(A, P_1)$ or $a = \max(A, P_2)$ or both) it is readily verified that $\rho'(a, A) = \nu(\alpha)\mathbf{1}\{a = \max(A, P_1)\} + (1 - \nu(\alpha))\,\mathbf{1}\{a = \max(A, P_2)\}$ for all $a \in A$ and $A \subseteq X$. The proof that $\texttt{idRUM}$ is not within menu ordinal follows by example 4.2. Showing that $\texttt{idRUM}$ is not across menu ordinal is also demonstrated by example and is available from the author upon request. □

There is a quite illuminating characterization of the across menu closure of $\texttt{idRUM}$. One can show that the across menu closure of $\texttt{idRUM}$ exactly coincides with the class of SCFs satisfying the original axioms characterizing idRUMs in Manzini and Mariotti (2018) (but later shown insufficient in Manzini et al. (2019)). A proof of this result is available upon request and is part of ongoing work. Not very surprisingly, idRUMs are indistinguishable from dRUMs using deterministic data.

[10] see also Manzini, Mariotti, and Petri (2019).

Proposition 4.12. $\mathscr{D}[\texttt{idRUM}] = \texttt{dRUM}$.

4.4. Summary

The following table summarizes the findings of this section. Perhaps interestingly, it is possible to differentiate between all the models considered in this section by using the full ordinal ranking $\succsim_\rho$. I.e. their ordinal closures are mutually distinct. This should be quite good news to the researcher: it suggests that a lot can be said about the behavior of these models even though exact choice frequencies are not available. Indeed, even the deterministic closures of all of these models are mutually distinct. Thus it would (in theory) be possible to differentiate between the models using very imprecise (i.e. deterministic) data.

Closure	RUM	SCRUM	dRUM
$\mathscr{O}$	Cancellation	SCRUM	dRUM
$\mathscr{W}$	$\subseteq$ SPWARP	$\neq$ SCF	dRUM
$\mathscr{A}$	AC	SCRUM	dRUM
$\mathscr{D}$	$\subseteq$ SPWARP	detSCRUM	dRUM

5. Ordinality of Luce models and relatives

In this section we will look at ordinal properties of the classical Luce model and close relatives to it.

5.1. Luce model

A positive stochastic choice function[11] ρ has a Luce representation if there is a non-negative utility function $u : X \to \mathbb{R}$ such that:

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

for all $a \in A$ and $A \in \mathscr{A}$. This model is due to Luce (1959). Denote by $\texttt{Luce}$ the set of stochastic choice functions with a Luce representation. It is fairly well known that the Luce model is characterized by *Independence of Irrelevant Alternatives*, stating that: $\frac{\rho(a,A)}{\rho(b,A)} = \frac{\rho(a,B)}{\rho(b,B)}$ for all $a, b \in A \cap B$. By this property it is immediate that the Luce model is not ordinal.

Proposition 5.1. *The Luce model is not an ordinal model.*

Proof. We exhibit SCFs ρ and ρ' such that $\rho \in \texttt{Luce}$, $\rho' = v \circ \rho$ and $\rho' \in \texttt{SCF} \setminus \texttt{Luce}$. Let $X = \{a_1, a_2, a_3\}$ and $\mathscr{A} = \{\{a_1, a_2, a_3\}, \{a_1, a_2\}\}$ and let ρ be a Luce model with utility function u defined by $u(a_1) = 1$ and $u(a_{i+1}) = u(a_i) + \frac{1}{10}$ for all $i \in \{1, 2\}$. Then $\rho(a_1, \{a_1, a_2\}) = \frac{10}{21}$. Define a stochastic choice function ρ' such that $\rho'(a_1, \{a_1, a_2\}) = \frac{100}{201}$ and $\rho'(a_i, \{a_1, a_2, a_3\}) = \rho(a_i, \{a_1, a_2, a_3\})$ for all $i \in \{1, 2, 3\}$ then ρ' is ordinally equivalent to ρ. But it is clear that ρ' is not a Luce rule as $\frac{\rho'(a_1,\{a_1,a_2\})}{\rho'(a_2,\{a_1,a_2\})} = \frac{100}{101} > \frac{10}{11} = \frac{\rho(a_1,\{a_1,a_2\})}{\rho(a_2,\{a_1,a_2\})} = \frac{\rho(a_1,\{a_1,a_2,a_3\})}{\rho(a_2,\{a_1,a_2,a_3\})} = \frac{\rho'(a_1,\{a_1,a_2,a_3\})}{\rho'(a_2,\{a_1,a_2,a_3\})}$. □

[11] An SCF ρ is positive if $\rho(a, A) > 0$ for all $a \in A \in \mathscr{A}$.

Within/across menu properties of Luce model: In order to facilitate the discussion of the within menu closure of the Luce model and other models considered in this section, we introduce the following rationality concept for stochastic choice functions.

Definition 5.1. Call a stochastic choice function ρ *ordinally rational* if there is a utility function $u : X \to \mathbb{R}$ such that:

$$u(a) \geq u(b) \;\Leftrightarrow\; \rho(a, A) \geq \rho(b, A)$$

for all $a, b \in A$ and for all $A \in \mathscr{A}$. Denote by $\mathtt{OR}$ the class/model of stochastic choice functions that are ordinally rational. ◁

Proposition 5.2. $\mathscr{W}[\mathtt{Luce}] = \mathtt{OR}$.

Proof. If $\rho \in \mathscr{W}[\mathtt{Luce}]$ there is a Luce model with utility function $u : X \to \mathbb{R}$ such that:

$$\rho(a, A) \geq \rho(b, A) \;\Leftrightarrow\; \frac{u(a)}{\sum_{c \in A} u(c)} \geq \frac{u(b)}{\sum_{c \in A} u(c)} \;\Leftrightarrow\; u(a) \geq u(b).$$

By the equivalences above it hence follows that $\rho \in \mathtt{OR}$. If $\rho \in \mathtt{OR}$ then there is a utility function $u : X \to \mathbb{R}$ such that for all $A \in \mathscr{A}$: $(a, A) \succsim (b, A)$ if and only if $u(a) \geq u(b)$. Using the second equivalence above the claim then follows. □

The problem of characterizing the across menu closure of the Luce model is closely related to establishing conditions such that a subjective probability relation is representable by a probability measure. The across menu properties of the Luce model are non-vacous, i.e. $\mathscr{A}[\mathtt{Luce}] \neq \mathtt{SCF}$. In particular, note that every SCF in the across menu closure of the Luce model must satisfy the following property: $\rho(a, A) \geq \rho(a, B)$ if and only if $\rho(b, A) \geq \rho(b, B)$ for all $a, b \in A \cap B$ and for all $A, B \in \mathscr{A}$.

Deterministic properties of the Luce model: Denote by $\mathtt{Pos}$ the collection of positive stochastic choice functions. The following proposition is then immediate:

Proposition 5.3. $\mathscr{D}[\mathtt{Luce}] = \mathtt{Pos}$.

5.2. Simple scalability

The simple scalability model (SSM) is formally a generalization of the Luce model. It was first characterized in Tversky (1972). [12] To define it, we need some further notation. Define a relation $\geq_m$ on $\mathbb{R}^m$ by $x \geq_m y$ if $x_1 \geq y_1$ and $x_i \leq y_i$ for all $i \in \{2, ..., m\}$. Similarly, $>_m$ is defined by $x >_m y$ if $x \neq y$ and $x \geq_m y$.

Definition 5.2. A stochastic choice function ρ has a simple scalability representation if there is a scale $u : X \to \mathbb{R}$ and for each $2 \leq k \leq n$ a function F_k such that for each $A = \{a_1, ..., a_k\}$ with $|A| = k$ it holds that

$$\rho(a_1, A) = F_k(u(a_1), ..., u(a_k)). \tag{7}$$

Further, each F_k is strictly $\geq_k$ increasing, meaning that: $F_k(x) \geq F_k(y)$ if $x \geq_m y$ and $F_k(x) > F_k(y)$ if $x >_m y$. ◁

[12] Suppes et al. (1989) corrects an issue with the definition of the simple scalability model presented in Tversky (1972).

Intuitively, a stochastic choice model is SSM if the alternatives in A can be scaled by a utility function u such that any choice probability $\rho(a,A)$ is a function $F_{|A|}$ of the scale values. Denote the class of stochastic choice functions with a simple scalability representation by $\mathtt{SSM}$. Already from the definition of SSM, due to the functions F_k, it is quite clear that $\mathtt{SSM}$ is an ordinal model.

Proposition 5.4. *Simple scalability is an ordinal model.*

Proof. It suffices to show that $\mathscr{O}[\mathtt{SSM}] \subseteq \mathtt{SSM}$. Let $\rho \in \mathscr{O}[\mathtt{SSM}]$ then it follows by proposition 3.2 that there is a $\rho' \in \mathtt{SSM}$ and a strictly increasing function $v : [0,1] \to [0,1]$ with $v(0) = 0$ and $\rho = v \circ \rho'$. Since $\rho' \in \mathtt{SSM}$ there is a scale u and there are strictly $\geq_k$ increasing functions F'_k for $2 \leq k \leq n$ such that equation (7) holds. Thus for each $A = \{a_1, ..., a_k\}$ with $|A| = k$ it holds that

$$\rho(a_1, A) = v \circ \rho'(a_1, A) = v \circ F'_k(u(a_1), ..., u(a_k)).$$

To conclude, it suffices to show that $F_k = v \circ F'_k$ is strictly $\geq_k$ increasing for all $2 \leq k \leq n$. But this is immediate since v is strictly increasing. □

Within/across menu properties of SSM: The within menu closure of a full support SSM is indistinguishable from the within menu closure of the Luce model. An implication of this result is that these models are indistinguishable using data sets with less information on choices than an within menu data set $(X, \succsim_{\rho|X}, \mathscr{A})$.

Proposition 5.5. $\mathscr{W}[\mathtt{SSM}] = \mathtt{OR}$.

Proof. We show that $\mathscr{W}[\mathtt{SSM}] = \mathtt{OR}$. Since every Luce model is an SSM it follows that $\mathtt{OR} = \mathscr{W}[\mathtt{Luce}] \subseteq \mathscr{W}[\mathtt{SSM}]$. It thus suffices to show that $\mathscr{W}[\mathtt{SSM}] \subseteq \mathtt{OR}$. Let $\rho \in \mathscr{W}[\mathtt{SSM}]$ then there is a $\rho' \in \mathtt{SSM}$ such that $\rho(a,A) \geq \rho(b,A) \Leftrightarrow \rho'(a,A) \geq \rho'(b,A)$ for all $a, b \in A$ and $A \in \mathscr{A}$. Since ρ' is SSM it satisfies the following property: $\rho(a,A) \geq \rho(b,A)$ if and only if $\rho(a,B) \geq \rho(b,B)$ for all $a, b \in A \cap B$ and $A, B \in \mathscr{A}$. Thus $\rho \in \mathtt{OR}$. □

A straightforward consequence of simple scalability is the following (across menu) property, for all $a \in A$ and $b \in B$ and $c, d \in X \setminus (A \cup B)$:

$$\rho(a, A \cup \{c\}) \geq \rho(a, A \cup \{d\}) \quad \Leftrightarrow \quad \rho(b, B \cup \{c\}) \geq \rho(b, B \cup \{d\}).$$

The across menu properties of the simple scalability model are hence non-vacous since every SCF ρ in the across menu closure of $\mathtt{SSM}$ must satisfy this property as well. I.e. we have $\mathscr{A}[\mathtt{SSM}] \neq \mathtt{SCF}$.

Remark: Many of the popular binary stochastic choice models are ordinal.[13] A binary stochastic choice function is a function $P : X \times X \to [0,1]$ such that $P(a,b) + P(b,a) = 1$ for all $a, b \in X$. A binary stochastic choice function is *Fechnerian* if there is a utility function $u : X \to \mathbb{R}$ and non-decreasing function $F : \mathbb{R} \to \mathbb{R}$ for all $x \in \mathbb{R}$ such that $P(a,b) = F(u(a) - u(b))$ for all $a, b \in X$. It follows as a corollary to proposition 3.2 that Fechnerian models are ordinal.

[13] I thank Jay Lu for bringing this to my attention.

5.3. Additive perturbed utility model

Fudenberg, Iijima, and Strzalecki (2015) introduce and characterize the additive perturbed utility (APU) model. For simplicity of analysis, we will throughout this section assume that $\rho(a, A) > 0$ for all $a \in A$ and $A \in \mathscr{A}$. A function c is a *cost function* if $c : [0,1] \to \mathbb{R} \cup \{\infty\}$ is strictly convex and C^1 over $(0,1)$ and $\lim_{q\to 0} c'(q) = -\infty$. A stochastic function has an APU representation if:

$$\{\rho(a, A)\} = \underset{p\in\Delta(A)}{\arg\max} \sum_{a\in A} \left[u(a)p(a) - c(p(a))\right].$$

Denote by $\texttt{APU}$ the class of SCFs with an APU representation. From its definition it is not obvious that such models are ordinal. However, this is immediate once we appeal to the main characterization result in Fudenberg, Iijima, and Strzalecki (2015). They characterize APUs using a property called acyclicity. Call a pair of sequences $(a_1, A_1), \ldots, (a_n, A_n)$ and $(b_1, B_1), \ldots, (b_n, B_n)$ such that $(a_1, \ldots, a_n)$ is a permutation of $(b_1, \ldots, b_n)$ and $(A_1, \ldots, A_n)$ is a permutation of $(B_1, \ldots, B_n)$ admissible.

> **Acyclicity:** There is no admissible sequence such that $\rho(a_i, A_i) \geq \rho(b_i, B_i)$ for all $i \in \{1, \ldots, n-1\}$ and $\rho(a_n, A_n) > \rho(b_n, B_n)$.

It is clear that if a stochastic choice function ρ satisfies acyclicity and if ρ' is an increasing transformation of ρ then ρ' satisfies acyclicity. Thus we have:

Proposition 5.6. *The APU model is an ordinal model.*

Within/across menu properties of APU model: The following proposition shows that the within menu closure of the APU model is indistinguishable from that of the SSM and Luce model.

Proposition 5.7. $\mathscr{W}[\texttt{APU}] = \texttt{OR}$.

Proof. Since $\texttt{Luce} \subseteq \texttt{APU}$ it follows that $\mathscr{W}[\texttt{APU}] \supseteq \mathscr{W}[\texttt{Luce}] = \texttt{OR}$. Let $\rho \in \mathscr{W}[\texttt{APU}]$ then there is a $\rho' \in \texttt{APU}$ such that $\rho(a, A) \geq \rho(b, A) \Leftrightarrow \rho'(a, A) \geq \rho'(b, A)$ for all $a, b \in A$ and $A \in \mathscr{A}$. Since ρ' is an APU model it satisfies acyclicity. We next note that acyclicity implies that $\rho(a, A) \geq \rho(b, A)$ if and only if $\rho(a, B) \geq \rho(b, B)$ for all $a, b \in A \cap B$ and $A, B \in \mathscr{A}$. The result follows. $\square$

The across menu properties of the APU model are non-vacous. In other words $\mathscr{A}[\texttt{APU}] \neq \texttt{SCF}$. This follows since every SCF in the across menu closure of the APU model satisfies a property called menu acyclicity from Fudenberg et al. (2015).[14]

Deterministic properties of APU model: Since ρ is assumed to be strictly positive it follows that the APU model is indistinguishable from the Luce model based on deterministic data. Note, however, that Fudenberg et al. (2015) characterize a version of APU allowing for non positive choice probabilities. Characterizing the deterministic properties of that model is left for future research.

[14]Define a relation $\succsim_m$ on menus by $A \succsim_m B$ if there is an $a \in A \cap B$ with $\rho(a, A) \geq \rho(a, B)$. Menu acyclicity requires there to be no sequence of menus $A_1, \ldots, A_k$ such that we have: $A_1 \succsim_m A_2 \succsim_m \ldots \succsim_m A_k \succ_m A_1$.

5.4. Gradual pairwise comparision rule

The gradual pairwise comparison rule (GPCR) considered in Dutta (2019) is formally a generalization of simple scalability. Similarly, to simple scalability and the APU model it is characterized using a set of axioms only impacting the ordinal ranking of choice probabilities. Informally, 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 $\mathscr{X} = \{X_1, ..., X_I\}$ be an ordered partition of P. Define sets recursively as follows:

$$\begin{aligned} M_0^{\mathscr{X}}(A) &= A \\ M_i^{\mathscr{X}}(A) = \{a \in M_{i-1}^{\mathscr{X}}(A) : \forall b \in M_{i-1}^{\mathscr{X}}(A), (b,a) &\notin X_i\} \\ \text{for all } 1 \leq i &\leq I \end{aligned}$$

Let $X_0 = \emptyset$. Since P is a strict preference relation there is for every $A \in \mathscr{A}$ an $I(A) \leq I$ such that $|M_{I(A)}^{\mathscr{X}}(A)| = 1$. A stopping function is a function $\pi : (\mathscr{X} \cup \{X_0\}) \times \mathscr{A} \to [0,1]$ such that $\sum_{i \in I} \pi(X_i, A) + \pi(X_0, A) = 1$ and $\pi(X_{I(A)}, A) > 0$ for all $A \in \mathscr{A}$. The GPCR can now be defined formally as follows:

$$\rho(a, A) = \begin{cases} \sum_{\{i : a \in M_i^{\mathscr{X}}(A)\}} \frac{\pi(X_i, A)}{|M_i^{\mathscr{X}}(A)|} & \text{if } a \in A \\ 0 & \text{otherwise} \end{cases}$$

The model composed of all SCFs with an GPCR representation is denoted by $\texttt{GPCR}$. The GPCR model can be shown to be within menu ordinal and this singles out the GPCR model among other models in this section.

Proposition 5.8. *The GPCR model is an within menu ordinal model (and hence an ordinal model).*

Proof. By Theorem 2 in (Dutta, 2019, p.11) it suffices to show that every GPCR satisfies *unique best*, i.e. that there for every $A \in \mathscr{A}$ is an $a \in A$ such that $\rho(a, A) > \rho(b, A)$ for all $b \in A \setminus \{a\}$. But this fact immediately follows by noting that for each $A \in \mathscr{A}$ the unique $a \in M_{I(A)}^{\mathscr{X}}(A)$ satisfies $\rho(a, A) > \rho(b, A)$ for all $b \in A \setminus \{a\}$. Since ρ satisfies unique best it follows by (Dutta, 2019, Thm. 2, p.11) that $\texttt{GPCR}$ is an within menu ordinal model. $\square$

The deterministic properties of $\texttt{GPCR}$ are more difficult to establish. Call a choice correspondence $c : \mathscr{A} \to 2^X \setminus \emptyset$ a deterministic GPCR if there for every $A \in \mathscr{A}$ is an $J(A) \in \{1, ..., I\}$ such that $c(A) = M_{J(A)}^{\mathscr{X}}(A)$. Denote the class of SCF's ρ such that D_ρ is a deterministic GPCR by $\texttt{detGPCR}$. It follows that:

Proposition 5.9. $\mathscr{D}[\texttt{GPCR}] = \texttt{detGPCR}$.

A characterization of the class of deterministic GPCRs is closely related to that of characterizing sequential rationalizability settled in Manzini and Mariotti (2012) (see also Manzini and Mariotti (2007)).

5.5. Random attention models

There is by now a large literature focusing on stochastic choice models where individuals only consider a subset of all available options and then choose a preference maximizing alternative within the subset. Notable in this regard are the papers by Manzini and Mariotti (2014); Brady and Rehbeck (2016); Cattaneo, Ma, Masatlioglu, and Suleymanov (2020). The models in Manzini and Mariotti (2014); Brady and Rehbeck (2016) fail to be ordinal for much the same reason as the Luce model fails to be ordinal. However, the recent random attention model (RAM) by Cattaneo, Ma, Masatlioglu, and Suleymanov (2020) can indeed be shown to be ordinal. To introduce it we need some further notation.

Definition 5.3. A function $\mu : \mathscr{A} \times \mathscr{A} \to [0,1]$ is called an *attention rule* if for all $A \in \mathscr{A}$: $\mu(B|A) \geq 0$ for all $B \subseteq A$, $\mu(B|A) = 0$ otherwise, and $\sum_{B\subseteq A} \mu(B|A) = 1$. A *monotonic attention rule* is such that $\mu(B|A) \leq \mu(B|A \setminus \{a\})$ for all $a \in A \setminus B$. ◁

We can now formally define SCFs with a random attention representation as follows.

Definition 5.4. A stochastic choice function ρ has a random attention representation if there is a monotonic attention rule μ and a linear order $\succ$ on X such that

$$\rho(a, A) = \sum_{B\subseteq A} \mathbf{1}\{a = \max(B, \succ)\} \cdot \mu(B|A)$$

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

The random attention model is the set of SCFs with a random attention representation and is denoted by $\mathtt{RAM}$. The next result follows by noting that the main axiom characterizing RAMs in Cattaneo et al. (2020) is an acyclicity condition on choice probabilities, and this condition is "preserved" by increasing transformations of SCFs.

Proposition 5.10. *The random attention model is an across menu ordinal model.*

Proof. Define a relation P by aPb if and only if there is an $A \in \mathscr{A}$ with $a, b \in A$ and $\rho(a, A) > \rho(a, A \setminus \{b\})$. Cattaneo et al. (2020) show that ρ has a random attention representation if and only if P is acyclic. Let $\rho \in \mathtt{RAM}$ and $\rho' \in \mathtt{SCF}$ be s.t. $\rho(a, A) > \rho(a, B)$ if and only if $\rho'(a, A) > \rho'(a, B)$. Let P and P' be relations corresponding to ρ and ρ'. Then it follows that aPb if and only if $aP'b$. Hence P' is acyclic. Thus, by Cattaneo, Ma, Masatlioglu, and Suleymanov (2020, Thm. 2) it follows that $\rho' \in \mathtt{RAM}$. As we wanted to show. □

5.6. Summary

As the following table illustrates, the $\mathtt{SSM}$, $\mathtt{APU}$, $\mathtt{GPCR}$ and $\mathtt{RAM}$ models are ordinal. Hence it is possible to empirically differentiate between these models using the $\succsim_\rho$ data set. The ordinal closure of the Luce model is a strict subset of $\mathtt{APU}$ so it is possible to distinguish between the Luce model and the other models of this section. In contrast to the previous section, it is not possible to differentiate between these models using within menu data. The Luce, SSM and APU model are all indistinguishable based on such data. This hence suggests that it should be possible to see the difference between these models using across menu data. The GPCR model is notable in that it is within menu ordinal, and hence distinguishable from the other models based on within menu data.

Closure	Luce	SSM	APU	GPCR	RAM
$\mathscr{O}$	$\subset$ APU	SSM	APU	GPCR	RAM
$\mathscr{W}$	OR	OR	OR	GPCR	$*$
$\mathscr{A}$	$\neq$ SCF	$\neq$ SCF	$\neq$ SCF	$*$	RAM
$\mathscr{D}$	Pos	$\neq$ SCF	Pos	detGPCR	$*$

6. Discussion

6.1. A general framework

In this section we discuss the testable implications of stochastic choice models on general imprecise data sets. A purpose of the following discussion is to show that knowledge of the behavior of models on the data sets $\succsim_\rho, \succsim_{\rho|X}, \succsim_{\rho|\mathscr{A}}$ and $\succsim_\rho^{\mathscr{D}}$ may be used to infer properties on general imprecise data sets.

Recall that a data set is an imprecise data set if $\succsim_\rho$ is an ordering extension of $\succsim$. We now give an equivalent definition of imprecise data sets, which will allow us to extend the ordinality notions of section 3 to a more general setting. Define a *comparability relation* to be a symmetric[15] binary relation C on $X \times \mathscr{A}$. Intuitively, $(a, A)C(b, B)$ means that the researcher is able to compare the choice-menu pair (a, A) to (b, B). Clearly if (a, A) is comparable to (b, B) then (b, B) should be comparable to (a, A) and therefore symmetry is imposed on C.

Proposition 6.1. *A data set $(X, \mathscr{A}, \succsim)$ is an imprecise data set if and only if there is a stochastic choice function ρ and a comparability relation C such that $\succsim = \succsim_\rho \cap C$.*

The above characterization of imprecise data sets (whose proof is omitted and available upon request) suggests the following extension of the notion of ordinality in section 3.

Definition 6.1. Let C be a comparability relation. A model $\mathtt{M}$ is *ordinal w.r.t. C* if $\rho' \in \mathtt{M}$ and $\rho \in \mathtt{SCF}$ and

$$\rho(a, A) > \rho(b, B) \quad \Leftrightarrow \quad \rho'(a, A) > \rho'(b, B)$$

for all $(a, A), (b, B) \in X \times \mathscr{A}$ with $(a, A)C(b, B)$ implies that $\rho \in \mathtt{M}$. Equivalently, a model is ordinal w.r.t. C if for all $\rho' \in \mathtt{M}$ and $\rho \in \mathtt{SCF}$ if $\succsim_\rho \cap C = \succsim_{\rho'} \cap C$ then $\rho \in \mathtt{M}$. $\lhd$

By suitably defining comparability relations C it can be shown that all notions of ordinality in section 3 are special cases of the above definition.[16] The ordinal closure of a model w.r.t. a comparability relation C may be defined analogously. Denote the ordinal closure of a model w.r.t. to C by $\mathscr{O}_C[\mathtt{M}]$. We then have the following result:

Proposition 6.2. *Let C and C' be comparability relations with $C \supseteq C'$ then for all models $\mathtt{M}, \mathtt{M}'$: if $\mathscr{O}_C[\mathtt{M}] = \mathscr{O}_C[\mathtt{M}']$ then $\mathscr{O}_{C'}[\mathtt{M}] = \mathscr{O}_{C'}[\mathtt{M}']$.*

The above proposition shows that much can be learnt about the behavior of models on (general) imprecise data sets by just studying their behavior on data sets $\succsim_\rho, \succsim_{\rho|X}, \succsim_{\rho|\mathscr{A}}$ and $\succsim_\rho^{\mathscr{D}}$. To illustrate, recall that it is possible to distinguish the RUM from various subclasses of it using deterministic data. As a corollary to proposition 6.2 it follows that it is possible to distinguish between these models using any imprecise data set $\succsim = \succsim_\rho \cap C \supseteq \succsim_\rho^{\mathscr{D}}$. As another example, recall that the Luce, SSM and APU model are indistinguishable based on within menu data. It hence follows as a corollary to proposition 6.2 that these models are indistinguishable using any imprecise data set $\succsim$ with $\succsim = \succsim_\rho \cap C \subseteq \succsim_{\rho|X}$. The following proposition is also immediate:

Proposition 6.3. *Let C and C' be comparability relations with $C \subseteq C'$ then $\mathscr{O}_C[\mathtt{M}] \supseteq \mathscr{O}_{C'}[\mathtt{M}]$.*

[15] A binary relation $\succsim$ on X is symmetric if for all $x, y \in X$: $x \succsim y$ implies $y \succsim x$.

[16] For instance, defining W by $(a, A)W(b, B)$ if and only if $A = B$, it follows that a model is ordinal w.r.t. W if and only if it is within menu ordinal.

In particular the above proposition implies that if $C \subseteq C'$ and $\mathscr{O}_C[\mathtt{M}] = \mathtt{M}$ then $\mathscr{O}_{C'}[\mathtt{M}] = \mathtt{M}$. As a corollary it follows that the dual random utility model is ordinal w.r.t. C for any comparability relation C. Similarly, it follows that $\mathscr{O}_C[\mathtt{RUM}] \neq \mathtt{RUM}$ and $\mathscr{O}_C[\mathtt{Luce}] \neq \mathtt{Luce}$ for all comparability relations C.

6.2. Other types of data sets

The notion of an imprecise data set may be generalized in many plausible directions not covered by definition 2.1. An example that stands out is that of a modal data set, i.e. the researcher only have information on modal (most frequent) choices in menus available. Formally, the researcher observes a modal choice correspondence $M_\rho : \mathscr{A} \to 2^X \setminus \emptyset$ such that $M_\rho(A) = \{a \in A : \rho(a, A) = \max_{b \in A} \rho(b, A)\}$ where ρ is an SCF.[17]

A modal data set is in general very sparse. To illustrate this consider the random utility model. It follows as a corollary to theorem 4.4 that there for every choice correspondence C is a RUM μ such that the chosen set $C(A)$ equals the set of alternatives in A chosen with maximum probability according to μ.[18] Differently put, this means that the RUM has no testable implications on modal data sets.

Proposition 6.4. *For every choice correspondence $C : \mathscr{A} \to 2^X \setminus \emptyset$ there is a probability measure μ (with rational values) on $\mathscr{P}$ such that*

$$C(A) = \{a \in A : \mu(\mathscr{P}(a, A)) = \max_{b \in A} \mu(\mathscr{P}(b, A))\}$$

for all $A \in \mathscr{A}$.

Proof. Define a stochastic choice function for all $a \in X$ and $A \in \mathscr{A}$ by: $\rho(a, A) = \frac{1}{|C(A)|}$ if $a \in C(A)$ and $\rho(b, A) = 0$ if $b \notin C(A)$. By theorem 4.4 there is a RUM μ such that

$$\rho(a, A) \geq \rho(b, A) \Leftrightarrow \mu(\mathscr{P}(a, A)) \geq \mu(\mathscr{P}(b, A)).$$

In particular, this implies that

$$\{a \in A : \rho(a, A) = \max_{b \in A} \rho(b, A)\} = \{a \in A : \mu(\mathscr{P}(a, A)) = \max_{b \in A} \mu(\mathscr{P}(b, A))\}.$$

Since $C(A) = \{a \in A : \rho(a, A) = \max_{b \in A} \rho(b, A)\}$ the claim follows. □

6.3. Classifying properties of SCFs

The framework of the previous sections may be used to classify properties of stochastic choice models. We identify with each property P a subset of SCFs satisfying this property. As an example, the well known monotonicity/regularity property, requiring that $\rho(a, A) \geq \rho(a, B)$ for all $A \subseteq B$ and $a \in A$, is identified with the subset of SCFs satisfying monotonicity. In this way, properties and models (as defined in the previous section) are in one-to-one

[17] These correspondences are also studied in Fishburn (1978).

[18] A similar result is obtained in Dogan and Yildiz (2020). They show that every choice correspondence is plurality rationalizable. Indeed, proposition 6.4 follows as a corollary to Dogan and Yildiz (2020, Prop. 1, p.13). Using a rational version of Farkas' lemma, we may assume that μ is rational valued. Hence proposition 1 in Dogan and Yildiz (2020) also follows as a corollary to our proposition 6.4.

correspondence with each other. Each property corresponds to a subset $\mathtt{P} \subseteq \mathtt{SCF}$ as follows: $\mathtt{P} := \{\rho \in \mathtt{SCF} : \rho \text{ satisfies property P}\}$.

Definition 6.2. A property P is an ordinal (across/within menu ordinal, deterministic) property if the model $\mathtt{P}$ is an ordinal (across/within menu ordinal, deterministic) model. ◁

Viewed in this way, a set of properties characterize a model if and only if their intersection equals the model.

Definition 6.3. We say that a model $\mathtt{M}$ is characterized by using a set of properties $\mathtt{P}_1, ..., \mathtt{P}_n$ if

$$\mathtt{M} = \bigcap_{i=1}^{k} \mathtt{P}_i.$$

◁

Every ordinal (across/within menu ordinal, deterministic) model is in a trivial way characterized using ordinal properties, take $\mathtt{P} = \mathtt{M}$. Conversly, if a model is characterized using a set of ordinal (across/within menu ordinal, deterministic) properties $\mathtt{P}_1, ..., \mathtt{P}_k$ then, since the intersection of ordinal (across/within menu ordinal, deterministic) properties is ordinal (across/within menu ordinal, deterministic), it follows that $\mathtt{M}$ is an ordinal model. We have the following observation:

Proposition 6.5. *A model $\mathtt{M}$ is characterized by a collection of ordinal (across/within menu ordinal, deterministic) properties $\mathtt{P}_1, ..., \mathtt{P}_n$ if and only if $\mathtt{M}$ is ordinal (across/within menu ordinal, deterministic).*

Thus, if an analyst knows that a model is ordinal, then this suggests that there is a characterization of the model entirely in terms of ordinal properties. Say that a researcher wishes to take a model to data and test it. If it is known that a model could be expressed using only ordinal properties, then this is convenient for the analyst, as less precise information about choice probabilities is required in testing the properties of the model. As an example, consider the dual Random Utility Model which is a deterministic model. Due to dRUM being deterministic there ought to be a characterization of it using deterministic properties. This is indeed correct, as the properties characterizing dRUM in Manzini and Mariotti (2018) can all be shown to be deterministic properties.

6.4. Relationship between stochastic and ordinal properties

The above dicussion suggests that if a model is ordinal then it should be possible to give a characterization of it using ordinal properties. Hence it should be possible to describe its properties entirely in terms of the $\succsim$ ranking. Similarly, if a model is a deterministic model then it should be possible to describe its properties using sloley deterministic information (i.e. choice correspondences). In this section, we show that this is indeed the case. We show that properties satisfied by a stochastic choice model may, under special circumstances, be translated to properties satisfied by the corresponding deterministic choice model. We apply the result to the dual random utility model, thereby obtaining a new characterization of the top-and-the-top procedure in Eliaz, Richter, and Rubinstein (2011). The analysis of this section hence suggests that these two choice models are equivalent in the sense that only deterministic information is needed to describe the behavior of the dual random utility model.

Let $\mathscr{C}$ denote the set of choice functions on X. Generic subsets of $\mathscr{C}$ are denoted $\mathscr{S}, \mathscr{S}_1, ..., \mathscr{S}_m$. For each $\mathscr{S} \subseteq \mathscr{C}$ let $\mathtt{SCF}(\mathscr{S}) = \{\rho \in \mathtt{SCF} : D_\rho \in \mathscr{S}\}$. The following proposition is immediate and hence stated without proof.

Proposition 6.6. *A model* $\mathtt{M}$ *is deterministic if and only if there is a subset of choice functions* $\mathscr{S} \subseteq \mathscr{C}$ *such that* $\mathtt{SCF}(\mathscr{S}) = \mathtt{M}$.

The next proposition shows that a deterministic model $\mathtt{SCF}(\mathscr{S})$ is characterized by deterministic properties $\mathtt{SCF}(\mathscr{S}_1)$ and $\mathtt{SCF}(\mathscr{S}_1)$ if and only if $\mathscr{S}$ is characterized by $\mathscr{S}_1$ and $\mathscr{S}_2$.

Proposition 6.7. $\mathtt{SCF}(\mathscr{S}) = \mathtt{SCF}(\mathscr{S}_1) \cap \mathtt{SCF}(\mathscr{S}_2)$ *if and only if* $\mathscr{S} = \mathscr{S}_1 \cap \mathscr{S}_2$.

Proof. It is clear that $\mathtt{SCF}(\mathscr{S}) \subseteq \mathtt{SCF}(\mathscr{S}_i)$ if and only if $\mathscr{S} \subseteq \mathscr{S}_i$. Further, we have $\mathtt{SCF}(\mathscr{S}) \supseteq \mathtt{SCF}(\mathscr{S}_1) \cap \mathtt{SCF}(\mathscr{S}_2)$ if and only if $\mathscr{S} \supseteq \mathscr{S}_1 \cap \mathscr{S}_2$. To see this, assume that $\mathtt{SCF}(\mathscr{S}) \supseteq \mathtt{SCF}(\mathscr{S}_1) \cap \mathtt{SCF}(\mathscr{S}_2)$ and $c \in \mathscr{S}_1 \cap \mathscr{S}_2$. Then $\mathtt{SCF}(\{c\}) \subseteq \mathtt{SCF}(\mathscr{S}_1) \cap \mathtt{SCF}(\mathscr{S}_2) \subseteq \mathtt{SCF}(\mathscr{S})$ implying that $c \in \mathscr{S}$. Conversly, if $\mathscr{S}_1 \cap \mathscr{S}_2 \subseteq \mathscr{S}$ and $\rho \in \mathtt{SCF}(\mathscr{S}_1) \cap \mathtt{SCF}(\mathscr{S}_2)$ then $D_\rho \in \mathscr{S}_1$ and $D_\rho \in \mathscr{S}_2$ and hence $D_\rho \in \mathscr{S}_1 \cap \mathscr{S}_2$ implying that $\rho \in \mathtt{SCF}(\mathscr{S}_1 \cap \mathscr{S}_2)$. □

Application to dual Random Utility Model: It is quite instructive to apply the above proposition to the dual random utility model. Let $\mathscr{S}$ be the set of choice functions such that $c(A) = \max(P_1, A) \cup \max(P_2, A)$ for all $A \in \mathscr{A}$. Then it is readily verified that $\mathtt{dRUM} = \mathtt{SCF}(\mathscr{S})$. Say that b impacts a in A if $a \in c(A)$ and $a \notin c(A \cup \{b\})$ or $\{a\} = c(A)$ and $\{a\} \subset c(A \cup \{b\})$. Consider the following properties (imposed on choice functions):

Deterministic Monotonicity: If $a \in B \subseteq A$ and $a \notin c(B)$ then $a \notin c(A)$ and if $\{a\} = c(A)$ then $\{a\} = c(B)$.

Deterministic Contraction Consistency: If b impacts a in A and $B \subseteq A$, then b impacts a in B.

Deterministic Impact Consistency: If b does not impact a in A for all $a \in A$ then $b \notin c(A \cup \{b\})$.

Proposition 6.8. *A choice function c satisfies deterministic monotonicity, deterministic contraction consistency and deterministic impact consistency if and only if there are linear orders* P_1 *and* P_2 *on X such that*

$$c(A) = \max(P_1, A) \cup \max(P_2, A)$$

for all $A \in \mathscr{A}$.

Proof. It is straightforward to check that the axioms above are the deterministic analogues of the properties considerd in Manzini and Mariotti (2018). The proposition then follows by applying proposition 6.7. I.e. let $\mathscr{S}_1, \mathscr{S}_2, \mathscr{S}_2$ be the set of deterministic choice functions satisfying deterministic monotonicity, deterministic contraction consistency and deterministic impact consistency. Let further $\mathtt{MM}, \mathtt{MCC}, \mathtt{MIC}$ be the sets of SCFs satisfying modal monotonicity, modal contraction consistency and modal impact consistency (these properties are stated formally in Manzini and Mariotti (2018)). Then it is easy to verify that $\mathtt{MM} = \mathtt{SCF}(\mathscr{S}_1)$,

$\texttt{MCC} = \texttt{SCF}(\mathscr{S}_2)$ and $\texttt{MIC} = \texttt{SCF}(\mathscr{S}_3)$. Thus, since $\texttt{dRUM} = \texttt{MM} \cap \texttt{MCC} \cap \texttt{MIC}$ by (Manzini and Mariotti, 2018, Thm. 2), it follows by repeated applications of proposition 6.7 that $\mathscr{S} = \mathscr{S}_1 \cap \mathscr{S}_2 \cap \mathscr{S}_3$. □

6.5. Characterization of imprecise data sets

In this section we provide a characterization of imprecise data sets that may be of some independent interest. As a corollary we obtain a characterization of complete data sets $(X, \mathscr{A}, \succsim)$. The main property is a version of cancellation in section 4. To state it we need some further notation. Let a collection $\mathscr{A}$ of menus be given (as above). A choice function is a function $c : \mathscr{A} \to X$ such that $c(A) \in A$ for all $A \in \mathscr{A}$.[19] Let $\mathscr{C}$ denote the collection of all choice functions on X and $\mathscr{A}$. For all $A \subseteq X$ and $a \in A$ denote by $\mathscr{C}(a, A) = \{c \in \mathscr{C} : c(A) = a\}$. Call a pair of sequences $(a_1, A_1), \ldots, (a_n, A_n)$ and $(b_1, B_1), \ldots, (b_n, B_n)$ *admissible* if (1) $a_i \in A_i$ and $b_i \in B_i$ for all i, (2) there are natural numbers $(k_i)_{i=1}^n$ such that $\sum_{i=1}^n k_i \mathbf{1}_{\mathscr{C}(a_i, A_i)}(c) \leq \sum_{i=1}^n k_i \mathbf{1}_{\mathscr{C}(b_i, B_i)}(c)$ for all $c \in \mathscr{C}$.

> **Weak cancellation:** There is no admissible sequence such that $(a_i, A_i) \succsim (b_i, B_i)$ for all $i \in \{1, \ldots, n\}$ and $(a_n, A_n) \succ (b_n, B_n)$.

Theorem 6.9. *A non-trivial data set $(X, \mathscr{A}, \succsim)$ satisfies weak cancellation if and only if there is a stochastic choice function ρ and a comparability relation C such that $\succsim = \succsim_\rho \cap C$.*

The proof of theorem 6.9 is in the Appendix and follows the same line of reasoning used to characterize the ordinal closure of the random utility model. Theorem 6.9 can be viewed as providing a consistency check on part of the analyst. If $(X, \mathscr{A}, \succsim)$ is an imprecise data set then it must satisfy cancellation. A corollary is that not every data set $(X, \mathscr{A}, \succsim)$ qualifies as an imprecise data set. Using theorem 6.9 we obtain the following characterization of complete data sets:

Corollary 6.10. *A data set $(X, \mathscr{A}, \succsim)$ satisfies weak cancellation and totality if and only if there is a stochastic choice function ρ such that*

$$(a, A) \succsim (b, B) \Leftrightarrow \rho(a, A) \geq \rho(b, B)$$

for all $(a, A), (b, B) \in X \times \mathscr{A}$.

6.6. Identification on imprecise data sets

So far we have not discussed the important issue of identification of models when data is imprecise. In general we expect impreciseness of data to complicate the identification exercise. Even for models whose parameters are unique given observations encoded by an SCF, we do not expect the same level of uniqueness to hold with imprecise data. To illustrate, consider the single crossing random utility model of Apesteguia et al. (2017). A desirable feature of this model is that its parameters are unique, i.e. given observed choices conforming to the model it is possible to uniquely pin down its parameters. However, for the deterministic version of their model, studied in Costa et al. (2020), uniqueness is lost. Some level of uniqueness is still possible with imprecise data. Consider for example the Luce model (or

[19] Note that $|c(A)| = 1$ for all $A \in \mathscr{A}$.

any of its relatives in section 5). In the standard setting, where an SCF is observed, the utility function in a Luce representation is unique up to a positive scalar. That is, if u and u' are utility functions generating the same stochastic choices, then $u = \alpha u'$ for some $\alpha > 0$. Now suppose instead that an within menu data set $(X, \mathscr{A}, \succsim_{\rho|X})$ is observed and that there are two different Luce rules with utility functions u and u' generating these choices. Although the previous uniqueness result fails, it is still the case that u and u' are ordinally equivalent, i.e. $u(a) > u(b)$ if and only if $u'(a) > u'(b)$. Similar results can be shown to hold for other models considered in this paper. It is for instance straightforward to verify that the preference relation P in a GPCR rule is unique (given within menu data on choices). We conjecture that issues of non-uniqueness are less severe for ordinal models (w.r.t. some comparability relation C) compared non-ordinal models. We leave the topic of identification of models in the presence of imprecise data to future research.

7. Related literature

Our paper is related to a literature investigating issues pertaining to limited information and data availability in stochastic/discrete choice. Notable in this regard are the papers by Fishburn (1978), Ok and Tserenjigmid (2020), Balakrishnan, Ok, and Ortoleva (2020), De Clippel and Rozen (2020) and Ahn, Echenique, and Saito (2018). Ahn et al. (2018) studies an axiomatic characterization of the Luce rule assuming that only data on average choices from menus is available. De Clippel and Rozen (2020) consider issues of limited data availability in testing theories of bounded rationality.

Closely related is the work by Fishburn (1978), Ok and Tserenjigmid (2020) and especially Balakrishnan, Ok, and Ortoleva (2020). Balakrishnan, Ok, and Ortoleva (2020) study a method to elicit choice correspondences from repeated observations on choices. The first part of their paper motivates this method by an axiomatic characterization. The second part then develops statistical techniques that can be used to implement the method with real world, finite sample, data. Hence, similarly to us, Balakrishnan et al. (2020) tackles the issue of impreciseness of data (due to its finite sample properties), but from a different angel. They provide a method to compute an imprecise data set/correspondence from actual observations on choices, whereas this paper takes an imprecise data set as given and then asks what can be inferred about the behavior of various models of choice.

Fishburn (1978) and Ok and Tserenjigmid (2020) study deterministic implications of stochastic choice models. I.e. they associate to each stochastic choice function a deterministic choice correspondence and then study the relationship between the properties satisfied by the SCF and those satisfied by the correspondence. Ok and Tserenjigmid (2020) are interested in identifying when a stochastic choice model can be thought of as arising from either indifference, indecisiveness or experimentation. Consider an individual with a possibly incomplete preference relation over alternatives. Whenever indifferent or indecisive, she randomizes between alternatives resulting in stochastic choice behavior. In repeated choices, such an individual may also occasionally "experiment", by choosing suboptimal alternatives, and this behavior will appear stochastic to an outsider observer. Ok and Tserenjigmid (2020) investigate under what conditions well-known models of stochastic choice can be thought of as arising from such behavior. In this paper we also study deterministic implications of stochastic choice models. However, our focus is different from Ok and Tserenjigmid (2020). We are interested in the testable implications of stochastic choice models assuming that deterministic data is observed, whereas they are focusing on the rationality properties of the induced correspondences.

There are other papers considering the observable implications of models on the ordinal ranking of choice probabilities. As mentioned in section 5, Tversky (1972), Fudenberg, Iijima, and Strzalecki (2015) and Dutta (2019) consider versions of the Luce model, each being characterized by properties impacting the ordinal ranking of choice probabilities. Although related, the papers differ from the current paper in several ways. Importantly, the primitive in all of these papers is a stochastic choice function whereas we consider an inference problem posed for general imprecise data sets. An earlier working paper version of Fudenberg et al. (2015) considers a ranking $\succsim$ defined on $X \times \mathscr{A}$. They also note that their acyclicity axiom resembles cancellation properties in subjective probability. However, they do not pursue this direction further.

There are also several papers acknowledging the deterministic implications of stochastic choice models. Aguiar, Boccardi, and Dean (2016); Aguiar and Kimya (2019) consider stochastic versions of Simon's (Simon, 1955) satisficing procedure. The menu dependent versions of their models are characterized by deterministic properties. Using the framework of the current paper, it is straightforward to show that these models are deterministic and hence satisfy definition 3.6.

Related are also the papers Apesteguia and Ballester (2016) and Apesteguia and Ballester (2020). Similarly to us, they study *models* of stochastic choice and various properties satisfied by these models. However, they differ from the current paper in scope. Apesteguia and Ballester (2020) proposes a goodness of fit measure for stochastic choice models, which is the largest fraction of data consistent with the model. Apesteguia and Ballester (2016) studies the existence representative agents of stochastic choice models $\mathtt{M}$. The existence of a representative agent is equivalent to a model $\mathtt{M}$ being a convex subset of $\mathtt{SCF}$. In contrast we focus on the ordinality of stochastic choice models and the testable implications of models on imprecise data sets.

This paper, and in particular our characterization of RUMs in section 4, relates to a literature investigating the testable implications of versions of the random utility model. Notable contributions in this regard are Gul and Pesendorfer (2006), Apesteguia, Ballester, and Lu (2017), Manzini and Mariotti (2018) and Frick, Iijima, and Strzalecki (2019). Our paper, and especially our results showing that the within menu and modal properties of RUM are vacuous, relates to a number of papers establishing "anything goes" results for choice functions. Kalai, Rubinstein, and Spiegler (2002) and Bossert and Sprumont (2013) obtain such results for deterministic choice functions, whereas Manzini and Mariotti (2014), Li and Tang (2017), Dogan and Yildiz (2020),Saito (2018) and Tserenjigmid and Kovach (2019) obtain similar results in the stochastic choice setup.

A. Proofs

A.1. Rational version of Farkas' lemma

The proof of theorem 4.2 and theorem 6.9 are applications of a rational version of Farkas' Lemma. The lemma below follows as a corollary to theorem 1.6.1. in Stoer and Witzgall (2012) (see also Echenique and Saito (2015) who applies this result in a different context). Alternatively, the result can be derived from theorem 3.2 in Fishburn (1973) (see also Chambers and Echenique (2014) for an application of this result).

Lemma A.1. *Let A be an $n\times m$ matrix, B an $n\times l$ matrix and C be an $n\times k$ matrix all with rational entries. Then exactly one of the following statements is true.*

1. *There are $x\in\mathbb{Q}^m$, $y\in\mathbb{Q}^l$ and $z\in\mathbb{Q}^k$ such that $Ax+By+Cz=0$, $x\geq 0$, $y\geq 0$ and $z>0$.*
2. *There is an $x\in\mathbb{Q}^n$ such that $A^Tx\geq 0$, $B^Tx\geq 0$ and $C^Tx\gg 0$.*

A.2. Proof of theorem 4.2 and theorem 6.9

To prove theorem 4.2 and theorem 6.9 we will show a more general result from which both results follow as corollaries. Let $\mathscr{S}$ denote a subset of the collection of choice functions on $\mathscr{A}$, i.e. $\mathscr{S}\subseteq\mathscr{C}$. For all $A\subseteq X$ and $a\in A$ denote by $\mathscr{S}(a,A)=\{c\in\mathscr{S}: c(A)=a\}$. Call a pair of sequences $(a_1,A_1),\dots,(a_n,A_n)$ and $(b_1,B_1),\dots,(b_n,B_n)$ $\mathscr{S}$- *admissible* if (1) $a_i\in A_i$ and $b_i\in B_i$ for all i, (2) there are natural numbers $(k_i)_{i=1}^n$ such that $\sum_{i=1}^n k_i\mathbf{1}_{\mathscr{S}(a_i,A_i)}(c)\leq\sum_{i=1}^n k_i\mathbf{1}_{\mathscr{S}(b_i,B_i)}(c)$ for all $c\in\mathscr{S}$.

> $\mathscr{S}$**-Cancellation (C):** There is no $\mathscr{S}$- admissible sequence such that $(a_i,A_i)\succsim(b_i,B_i)$ for all $i\in\{1,\dots,n\}$ and $(a_n,A_n)\succ(b_n,B_n)$.

Theorem A.2. *A non-trivial data set $(X,\mathscr{A},\succsim)$ satisfies $\mathscr{S}$- cancellation if and only if there is a measure μ on $\mathscr{S}$ such that*

$$(a,A)\succ(b,B)\;\Rightarrow\;\mu(\mathscr{S}(a,A))>\mu(\mathscr{S}(b,B)) \tag{8}$$
$$(a,A)\succsim(b,B)\;\Rightarrow\;\mu(\mathscr{S}(a,A))\geq\mu(\mathscr{S}(b,B)) \tag{9}$$

for all $(a,A),(b,B)\in X\times\mathscr{A}$.

Proof. Let $(X,\mathscr{A},\succsim)$ be a imprecise data set such that $\mathscr{S}$-cancellation and non-triviality holds. For all $(a,A),(b,B)\in X\times\mathscr{A}$ with $(a,A)\succsim(b,B)$ define a vector $\mathbf{x}[(a,A),(b,B)]=\mathbf{1}_{\mathscr{S}(a,A)}-\mathbf{1}_{\mathscr{S}(b,B)}$. Define a matrix A by letting the columns correspond to the vectors $\mathbf{x}[(a,A),(b,B)]$. Let B be the identity matrix, i.e. $B=I$. Define for all $(a,A),(b,B)\in X\times\mathscr{A}$ with $(a,A)\succ(b,B)$ a vector $\mathbf{y}[(a,A),(b,B)]=\mathbf{1}_{\mathscr{S}(a,A)}-\mathbf{1}_{\mathscr{S}(b,B)}$. Define a matrix C by letting its columns correspond to the vectors $\mathbf{y}[(a,A),(b,B)]$. Since $(X,\mathscr{A},\succsim)$ is non-trivial there is at least one such vector $\mathbf{y}[(a,A),(b,B)]$.

STEP 1: We first claim that there are no $x\in\mathbb{Q}^m$, $y\in\mathbb{Q}^l$ and $z\in\mathbb{Q}^k$ such that $Ax+By+Cz=0$, $x\geq 0$, $y\geq 0$ and $z>0$.

Suppose that there are $x\in\mathbb{Q}^m$, $y\in\mathbb{Q}^l$ and $z\in\mathbb{Q}^k$ such that $Ax+By+Cz=0$, $x\geq 0$, $y\geq 0$ and $z>0$. Hence there are positive rational numbers $x_1,\dots x_m$ $y_1,\dots,y_l$ and positive $z_1,\dots,z_k$ with $z_j>0$ for some $j\geq 1$ such that

$$\sum_{i=1}^n x_i\mathbf{x}[(a_i,A_i),(b_i,B_i)]+\sum_{i=1}^l y_i\mathbf{e}_i+\sum_{i=1}^k z_i\mathbf{y}[(c_i,C_i),(d_i,D_i)]=0,$$

where $(a_i,A_i)\succsim(b_i,B_i)$ for all $i\in\{1,\dots,m\}$ and $(c_i,C_i)\succ(d_i,D_i)$ for all $i\in\{1,\dots,k\}$.

Multiplying the expression above with least common denominators it follows that there are integers $k_1,...,k_m$ $s_1,...,s_l$ and $r_1,...,r_k$ such that

$$\sum_{i=1}^{n} k_i \mathbf{x}[(a_i,A_i),(b_i,B_i)] + \sum_{i=1}^{l} s_i \mathbf{e}_i + \sum_{i=1}^{k} r_i \mathbf{y}[(c_i,C_i),(d_i,D_i)] = 0,$$

and this further implies that

$$\sum_{i=1}^{n} k_i \mathbf{x}[(a_i,A_i),(b_i,B_i)] + \sum_{i=1}^{k} r_i \mathbf{y}[(c_i,C_i),(d_i,D_i)] \leq$$

$$\sum_{i=1}^{n} k_i \mathbf{x}[(a_i,A_i),(b_i,B_i)] + \sum_{i=1}^{l} s_i \mathbf{e}_i + \sum_{i=1}^{k} r_i \mathbf{y}[(c_i,C_i),(d_i,D_i)] = 0$$

The latter inequality holds if and only if

$$\sum_{i=1}^{n} k_i \left[\mathbf{1}_{\mathscr{S}(a_i,A_i)} - \mathbf{1}_{\mathscr{S}(b_i,B_i)}\right] + \sum_{i=1}^{k} r_i \left[\mathbf{1}_{\mathscr{S}(c_i,C_i)} - \mathbf{1}_{\mathscr{S}(d_i,D_i)}\right] \leq 0.$$

But $\mathscr{S}$-cancellation now implies that $(d_i,D_i) \succsim (c_i,C_i)$ for all $i \in \{1,...,k\}$. A contradiction. By lemma A.1 it follows that there is an $x \in \mathbb{Q}^n$ such that $A^T x \geq 0$, $B^T x \geq 0$ and $C^T x \gg 0$.

STEP 2: Define a set function by:

$$\mu(R) = \frac{x \cdot \mathbf{1}_R}{x \cdot \mathbf{1}_{\mathscr{S}}}$$

for all $R \subseteq \mathscr{S}$. We next show that μ is a probability measure. To do this we need to check that $\mu(\cdot)$ is a positive, finitely additive measure and that $\mu(\mathscr{S}) = 1$. It is clear that $\mu(\mathscr{S}) = \frac{x\cdot\mathbf{1}_{\mathscr{S}}}{x\cdot\mathbf{1}_{\mathscr{S}}} = 1$. Additivity follows since for all disjoint $R,S \subseteq \mathscr{S}$ we have $\mu(R\cup S) = \frac{x\cdot\mathbf{1}_{R\cup S}}{x\cdot\mathbf{1}_{\mathscr{S}}} = \frac{x\cdot[\mathbf{1}_R+\mathbf{1}_S]}{x\cdot\mathbf{1}_{\mathscr{S}}} = \frac{x\cdot\mathbf{1}_R}{x\cdot\mathbf{1}_{\mathscr{S}}} + \frac{x\cdot\mathbf{1}_S}{x\cdot\mathbf{1}_{\mathscr{S}}} = \mu(R)+\mu(S)$. Positivity follows by noting that $B^T x \geq 0$ implies that $x \geq 0$. Using this observation and that the inner product of two positive vectors is a positive scalar it follows that $\mu(R) = \frac{x\cdot\mathbf{1}_R}{x\cdot\mathbf{1}_{\mathscr{S}}} \geq 0$ for all $R \subseteq \mathscr{S}$. Thus μ is a probability measure.

STEP 3: For all $(a,A),(b,B) \in X \times \mathscr{A}$: If $(a,A) \succsim (b,B)$ then $\mu(\mathscr{S}(a,A)) \geq \mu(\mathscr{S}(b,B))$ and if $(a,A) \succ (b,B)$ then $\mu(\mathscr{S}(a,A)) > \mu(\mathscr{S}(b,B))$.

If $(a,A) \succsim (b,B)$ then since $A^T x \geq 0$ it follows that $x\cdot\mathbf{x}[(a,A),(b,B)] \geq 0$ and hence $x\cdot(\mathbf{1}_{\mathscr{S}(a,A)} - \mathbf{1}_{\mathscr{S}(b,B)}) \geq 0$ implying that $\mu(\mathscr{S}(a,A)) \geq \mu(\mathscr{S}(b,B))$. If $(a,A) \succ (b,B)$ then since $C^T x \gg 0$ it follows that $x\cdot\mathbf{x}[(a,A),(b,B)] > 0$ and hence $x\cdot(\mathbf{1}_{\mathscr{S}(a,A)} - \mathbf{1}_{\mathscr{S}(b,B)}) > 0$. Thus $\mu(\mathscr{S}(a,A)) > \mu(\mathscr{S}(b,B))$.

Conversely, let $(X,\mathscr{A},\succsim)$ be a triple such that equation (8)-(9) holds. We show that $\mathscr{S}$-cancellation holds. Let $(a_1,A_1),...,(a_n,A_n)$ and $(b_1,B_1),...,(b_n,B_n)$ be such that $(a_i,A_i) \succsim (b_i,B_i)$ for all $i \in \{1,...,n\}$ and assume that there is a sequence of natural numbers $(k_i)_{i=1}^n$ such that

$$\sum_{i=1}^{n} k_i \mathbf{1}_{\mathscr{S}(a_i,A_i)}(c) \leq \sum_{i=1}^{n} k_i \mathbf{1}_{\mathscr{S}(b_i,B_i)}(c)$$

for all $c \in \mathscr{S}$. Then taking expectations it follows that

$$\sum_{i=1}^{n} k_i \mu(\mathscr{S}(a_i, A_i)) = E_\mu \left[\sum_{i=1}^{n} k_i \mathbf{1}_{\mathscr{S}(a_i,A_i)} \right] \leq E_\mu \left[\sum_{i=1}^{n} k_i \mathbf{1}_{\mathscr{S}(b_i,B_i)} \right] = \sum_{i=1}^{n} k_i \mu(\mathscr{S}(b_i, B_i)). \tag{10}$$

Note that $(a_i, A_i) \succsim (b_i, B_i)$ implies that $\mu(\mathscr{S}(a_i, A_i)) \geq \mu(\mathscr{S}(b_i, B_i))$ for all $i \in \{1, ..., n\}$. If $(a_n, A_n) \succ (b_n, B_n)$ then $\mu(\mathscr{S}(a_n, A_n)) > \mu(\mathscr{S}(b_n, B_n))$. It then follows that $\sum_{i=1}^{n} k_i \mu(\mathscr{S}(a_i, A_i)) > \sum_{i=1}^{n} k_i \mu(\mathscr{S}(b_i, B_i))$. A contradiction to equation (10). □

Theorem 4.2 follows by noticing that $\mathscr{P}$-cancellation is equivalent to cancellation. To see how theorem 6.9 follows from theorem A.2 note that weak cancellation is equivalent to $\mathscr{C}$-cancellation. Hence weak cancellation holds if and only if there is a measure μ on $\mathscr{C}$ such that

$$\begin{aligned}
(a, A) \succ (b, B) \;&\Rightarrow\; \mu(\mathscr{C}(a, A)) > \mu(\mathscr{C}(b, B)) \\
(a, A) \succsim (b, B) \;&\Rightarrow\; \mu(\mathscr{C}(a, A)) \geq \mu(\mathscr{C}(b, B))
\end{aligned}$$

for all $(a, A), (b, B) \in X \times \mathscr{A}$. But every measure μ on $\mathscr{C}$ gives an SCF defined by $\rho(a, A) = \mu(\mathscr{C}(a, A))$ for all $a \in A$ and $A \in \mathscr{A}$. Conversely, for every SCF ρ there is a measure μ on $\mathscr{C}$ such that $\rho(a, A) = \mu(\mathscr{C}(a, A))$ for all $a \in A$ and $A \in \mathscr{A}$. Theorem 6.9 thus follows.

A.3. Proof of theorem 4.4

In the following proofs we will use I as a shorthand for the set $\{1, ..., n\}$.

Lemma A.3. *Let $A \subseteq X$. If it holds that*

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A)}(P) \leq \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A)}(P)$$

for all $P \in \mathscr{P}$ then there is a $P \in \mathscr{P}$ with

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A)}(P) = \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A)}(P).$$

Proof. Since $\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A)}(P) \leq \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A)}(P)$ for all $P \in \mathscr{P}$ it suffices to show that there is a $P \in \mathscr{P}$ with $\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A)}(P) \geq \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A)}(P)$. To show this we note that $\sum_{\{i \in I : a_i = b\}} k_i - \sum_{\{i \in I : b_i = b\}} k_i \geq 0$ for some $b \in \{a_1, ..., a_n\}$. If not, then $\sum_{i \in I} k_i = \sum_{a \in \{a_1, ..., a_n\}} \sum_{\{i \in I : a_i = a\}} k_i < \sum_{a \in \{a_1, ..., a_n\}} \sum_{\{i \in I : b_i = a\}} k_i \leq \sum_{i \in I} k_i$. A contradiction. Let b be such that $\sum_{\{i \in I : a_i = b\}} k_i - \sum_{\{i \in I : b_i = b\}} k_i \geq 0$.Then if $P \in \mathscr{P}(b, X)$ it follows that

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A)}(P) = \sum_{\{i \in I : a_i = b\}} k_i \geq \sum_{\{i \in I : b_i = b\}} k_i = \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A)}(P).$$

The claim follows. □

Lemma A.4. *If it holds that*

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \leq \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}$ then there is a $P \in \mathscr{P}$ with

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) = \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P).$$

Proof. The proof is by induction on the number of distinct sets in $A_1,...,A_n$. The base case $A_1 = ... = A_n = A$ for some $A \subseteq X$ follows directly by lemma A.3. As induction hypothesis assume that the claim is true for all sequences $A_1,...,A_n$ with $k-1$ distinct sets. Assume that $A_1,...,A_n$ is a sequence with k distinct sets such that

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \leq \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P \in \mathscr{P}$. Let $b \in \arg\max_{a\in\{a_1,...,a_n\}} \left(\sum_{\{i\in I: a_i=a\}} k_i - \sum_{\{i\in I: b_i=a\}} k_i\right)$ then we have that $\sum_{\{i\in I:a_i=b\}} k_i - \sum_{\{i\in I:b_i=b\}} k_i \geq 0$. If not, then $\sum_{i\in I} k_i = \sum_{a\in\{a_1,...,a_n\}} \sum_{\{i\in I:a_i=a\}} k_i < \sum_{a\in\{a_1,...,a_n\}} \sum_{\{i\in I:b_i=a\}} k_i \leq \sum_{i\in I} k_i$. A contradiction.

Thus $\sum_{\{i\in I:a_i=b\}} k_i - \sum_{\{i\in I:b_i=b\}} k_i \geq 0$. Let $I(2) = \{i \in I : b \notin A_i\}$ and note that $\sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) - \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) \geq 0$ for all $P \in \mathscr{P}(b,X)$. Further, by assumption, it holds that

$$\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) + \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \leq \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) + \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P \in \mathscr{P}$. Rearranging the above expression we obtain

$$\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) - \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) \leq \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) - \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \leq 0$$

for all $P \in \mathscr{P}(b,X)$. This in particular implies that $\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) - \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) \leq 0$ for all $P \in \mathscr{P}$. Since $(A_i)_{i\in I(2)}$ is a sequence with $k-1$ distinct sets it follows by the induction hypothesis that there is a $P \in \mathscr{P}$ with $\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) = \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$. Since $b \notin A_i$ for any $i \in I(2)$ we w.l.o.g. assume that $P \in \mathscr{P}(b,X)$. Using this we obtain that

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) = \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) + \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \geq$$

$$\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) + \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) = \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P).$$

The claim follows.

□

Lemma A.5. *If it holds that*

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \leq \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P \in \mathscr{P}$ then it follows that

$$\sum_{\{i\in I:a_i=b\}} k_i - \sum_{\{i\in I:b_i=b\}} k_i = 0$$

for all $b \in X$. This further implies that $\{a_1,...,a_n\} = \{b_1,...,b_n\}$.

Proof. Let

$$b \in \underset{a \in \{a_1,\dots,a_n\}}{\arg\max} \left(\sum_{\{i \in I : a_i = a\}} k_i - \sum_{\{i \in I : b_i = a\}} k_i \right)$$

then we have that $\sum_{\{i \in I : a_i = b\}} k_i - \sum_{\{i \in I : b_i = b\}} k_i \geq 0$.

CASE 1: Assume that $\sum_{\{i \in I : a_i = b\}} k_i - \sum_{\{i \in I : b_i = b\}} k_i > 0$.

Let $I(2) = I \backslash \{i \in I : b \in A_i\}$. Then it follows that $\sum_{i \in I \backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P) - \sum_{i \in I \backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) < 0$ for all $P \in \mathscr{P}(b, X)$. Further, by assumption, it holds that

$$\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) + \sum_{i \in I \backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \leq \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P) + \sum_{i \in I \backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}$. Rearranging the above expression we obtain

$$\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) - \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P) \leq \sum_{i \in I \backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P) - \sum_{i \in I \backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) < 0$$

for all $P \in \mathscr{P}(b, X)$. This in particular implies that $\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) - \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P) \leq 0$ for all $P \in \mathscr{P}$. It folllows that there is a $P \in \mathscr{P}$ with $\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) = \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$. Since $b \notin A_i$ for any $i \in I(2)$ we w.l.o.g. assume that $P \in \mathscr{P}(b, X)$. Using this we obtain that

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) = \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) + \sum_{i \in I \backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) >$$

$$\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P) + \sum_{i \in I \backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P) = \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P).$$

A contradiciton.

CASE 2: If $\sum_{\{i \in I : a_i = b\}} k_i - \sum_{\{i \in I : b_i = b\}} k_i = 0$ then $\sum_{\{i \in I : a_i = a\}} k_i - \sum_{\{i \in I : b_i = a\}} k_i = 0$ for all $a \in \{a_1, \dots, a_n\}$. If not, then

$$\sum_{i \in I} k_i = \sum_{a \in \{a_1,\dots,a_n\}} \sum_{\{i \in I : a_i = a\}} k_i < \sum_{a \in \{a_1,\dots,a_n\}} \sum_{\{i \in I : b_i = a\}} k_i \leq \sum_{i \in I} k_i.$$

Finally, we note that $\{a_1, \dots, a_n\} = \{b_1, \dots, b_n\}$, since if $b \in \{a_1, \dots, a_n\} \backslash \{b_1, \dots, b_n\}$ then $\sum_{\{i \in I : a_i = b\}} k_i - \sum_{\{i \in I : b_i = b\}} k_i > 0$ and if $b \in \{b_1, \dots, b_n\} \setminus \{a_1, \dots, a_n\}$ then $\sum_{\{i \in I : a_i = b\}} k_i - \sum_{\{i \in I : b_i = b\}} k_i < 0$. A contradiction again.

□

Lemma A.6. *Let $A \subseteq X$. If it holds that*

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A)}(P) \leq \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A)}(P)$$

for all $P \in \mathscr{P}$ then

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A)}(P) = \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A)}(P).$$

for all $P \in \mathscr{P}$.

Proof. This follows immediately by lemma A.5. □

Lemma A.7. *Let $A_1, \dots, A_n \subseteq X$. If it holds that*

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \le \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}$ then

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) = \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}$.

Proof. The claim is proven by induction on the number of distinct sets in the sequence $A_1, \dots, A_n$ (i.e. on the cardinality of $\{A_1, \dots, A_n\}$). The base case follows by lemma A.7 above. Assume that the claim is true for all sequences with $k-1$ distinct sets in $A_1, \dots, A_n$. Assume that $A_1, \dots, A_n$ is a sequence with k distinct sets such that

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \le \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}$. By lemma A.5 it follows that

$$\sum_{\{i \in I : a_i = b\}} k_i - \sum_{\{i \in I : b_i = b\}} k_i = 0$$

for all $b \in X$ and moreover $\{a_1, \dots, a_n\} = \{b_1, \dots, b_n\}$.

There are $i, j \in \{1, \dots, n\}$ with $A_i \setminus A_j \neq \emptyset$. Let $b \in A_i \setminus A_j$ and consider the set of orders $P \in \mathscr{P}$ with b as their top element in X, i.e. $\mathscr{P}(b, X)$.

Let $I(2) = \{i \in I(1) : b \notin A_i\}$. If $b \notin \{a_1, \dots, a_n\}$ then $\sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) = 0 = \sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$. If $b \in \{a_1, \dots, a_n\}$ then $\sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) = \sum_{\{i \in I : a_i = b\}} k_i = \sum_{\{i \in I : b_i = b\}} k_i = \sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$. It hence follows that

$$\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \le \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}(b, X)$.

But since $b \notin A_i$ for any $i \in I(2)$ this implies that

$$\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \le \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}$.

Since $(A_i)_{i \in I(2)}$ is a sequence with $k-1$ distinct sets it follows by the induction hypotheis that $\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) = \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$ for all $P \in \mathscr{P}$.

This further implies that

$$\sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \le \sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}$. Since there are less than $k-1$ sets in the sequence $(A_i)_{i \in I \setminus I(2)}$ it follows by the induction hypothesis that $\sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) = \sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$ for all $P \in \mathscr{P}$. This proves the claim. □

Finally, we are ready for the proof of theorem 4.4.

Proof. Define a relation $\succsim$ on $X \times \mathscr{A}$ by $(a, A) \succsim (b, B)$ if and only if $\rho(a, A) \geq \rho(b, B)$ and $A = B$. We check that $\succsim$ satisfies $\mathscr{P}$-cancellation. Let $(a_1, A_1), \ldots, (a_n, A_n)$ and $(b_1, B_1), \ldots, (b_n, B_n)$ be such that $(a_i, A_i) \succsim (b_i, B_i)$ for all $i \in \{1, \ldots, n\}$ and assume that :

$$\sum_{i=1}^{n} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \leq \sum_{i=1}^{n} k_i \mathbf{1}_{\mathscr{P}(b_i, B_i)}(P)$$

for all $P \in \mathscr{P}$.

Since $(a_i, A_i) \succsim (b_i, B_i)$ for all $i \in \{1, \ldots, n\}$ it follows that $A_i = B_i$ for all $i \in \{1, \ldots, n\}$ by definition of $\succsim$. We prove the claim by induction on the number of distinct sets in the sequence $A_1, \ldots, A_n$. The base case is clear. Assume that the claim holds when there are $k-1$ distinct sets in the sequence. Let $A_1, \ldots, A_n$ be a sequence with k distinct sets. Hence there are $i, j \in \{1, \ldots, n\}$ with $A_i \setminus A_j \neq \emptyset$. Let $b \in A_i \setminus A_j$ and consider the set of orders $P \in \mathscr{P}$ with b as their top element in X, i.e. $\mathscr{P}(b, X)$. Let $I(2) = \{i \in I : b \notin A_i\}$ and note that

$$\sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) - \sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P) = \sum_{\{i \in I : a_i = b\}} k_i - \sum_{\{i \in I : b_i = b\}} k_i = 0$$

for all $P \in \mathscr{P}(b, X)$, where the first equality follows since $P \in \mathscr{P}(b, X)$ and the second equality follows by lemma A.5. Using this, it then follows that

$$\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \leq \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}(b, X)$. But since $b \notin A_i$ for any $i \in I(2)$ we futher have that

$$\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \leq \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}$. If $n \in I(2)$ the claim follows by induction hypothesis. Otherwise, note that lemma A.7 implies that $\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) = \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$ for all $P \in \mathscr{P}$. This further implies that

$$\sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \leq \sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}$. Since there are less than $k-1$ sets in the sequence $(A_i)_{i \in I \setminus I(2)}$ it follows by the induction hypothesis that $(b_n, A_n) \succsim (a_n, A_n)$. As we wanted to show. □

A.4. Proof of proposition 3.2 and prosposition 3.1

We first prove proposition 3.2.

Proof. 3.2 It is clear that 3 implies 2. We show that 2 implies 3. By assumption there is an SCF $\rho' \in \mathtt{M}$ such that:

$$\rho(a, A) \geq \rho(b, B) \;\Leftrightarrow\; \rho'(a, A) \geq \rho'(b, B)$$

for all $(a, A), (b, B) \in X \times \mathscr{A}$. Using this relation it is easily verified that there is a strictly increasing function $v : [0, 1] \to [0, 1]$ with $v(0) = 0$ and such that

$$\rho(a, A) = v \circ \rho'(a, A)$$

for all $a \in A$ and $A \subseteq X$. To see this, first define v on the range

$$Y = \{y \in [0,1] : y = \rho'(a, A) \text{ for some } (a, A) \in X \times \mathscr{A}\}$$

by $v(y) = \rho(a, A)$ if $y = \rho'(a, A)$ (where (a, A) is some representative of all (a, A) with $y = \rho'(a, A)$). This function is clearly increasing and since Y is finite it may be extended to an increasing function on $[0,1]$. Further $v(0) = 0$ since if $x \notin A$ then $v(0) = v(\rho'(x, A)) = \rho(x, A) = 0$ (note that $\rho(x, A) = 0$ for all $x \notin A$ follows by definition of a stochastic choice function).

We next show the equivalence of 1 and 3. Suppose that $\rho \in \mathscr{O}[\mathtt{M}]$. Consider the model

$$\mathtt{A} = \{\rho \in \mathtt{SCF} : \rho = v \circ \rho', \rho' \in \mathtt{M}, \text{ increasing } v : [0,1] \to [0,1], v(0) = 0\}.$$

Model $\mathtt{A}$ is clearly ordinal, hence $\mathscr{O}[\mathtt{M}] \subseteq \mathtt{A}$ and 3 follows. Conversely, if there is a strictly increasing function $v : [0,1] \to [0,1]$ with $v(0) = 0$ and an SCF $\rho' \in \mathtt{M}$ such that $\rho = v \circ \rho'$ then since $\rho' \in \mathtt{M} \subseteq \mathscr{O}[\mathtt{M}]$ it follows that $\rho = v \circ \rho' \in \mathscr{O}[\mathtt{M}]$. $\square$

Proposition 3.1 follows from proposition 3.2 by noting that a model is ordinal if and only if $\mathscr{O}[\mathtt{M}] = \mathtt{M}$.

References

V. H. Aguiar and M. Kimya. Adaptive stochastic search. *Journal of Mathematical Economics*, 81:74–83, 2019.

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

D. S. Ahn, F. Echenique, and K. Saito. On path independent stochastic choice. *Theoretical Economics*, 13(1):61–85, 2018.

S. Alon and E. Lehrer. Subjective multi-prior probability: A representation of a partial likelihood relation. *Journal of Economic Theory*, 151:476–492, 2014.

J. Apesteguia and M. A. Ballester. Stochastic representative agent. Technical report, 2016.

J. Apesteguia and M. A. Ballester. Separating predicted randomness from residual behavior. *Journal of the European Economic Association*, 2020.

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

N. Balakrishnan, E. A. Ok, and P. Ortoleva. Computational choice theory. 2020.

W. Bossert and Y. Sprumont. Every choice function is backwards-induction rationalizable. *Econometrica*, 81(6):2521–2534, 2013.

R. L. Brady and J. Rehbeck. Menu-dependent stochastic feasibility. *Econometrica*, 84(3):1203–1223, 2016.

M. D. Cattaneo, X. Ma, Y. Masatlioglu, and E. Suleymanov. A random attention model. *Journal of Political Economy*, 128(7):000–000, 2020.

C. P. Chambers and F. Echenique. On the consistency of data with bargaining theories. *Theoretical Economics*, 9(1):137–162, 2014.

M. Costa, P. H. Ramos, and G. Riella. Single-crossing choice correspondences. *Social Choice and Welfare*, 54(1):69–86, 2020.

G. De Clippel and K. Rozen. Bounded rationality and limited datasets. *Theoretical Economics, forthcoming*, 2020.

S. Dogan and K. Yildiz. Every choice function is pro-con rationalizable. *Available at SSRN 3085542*, 2020.

R. Dutta. Gradual pairwise comparison and stochastic choice. *Theoretical Economics, forthcoming*, 2019.

F. Echenique and K. Saito. Savage in the market. *Econometrica*, 83(4):1467–1495, 2015.

K. Eliaz, M. Richter, and A. Rubinstein. Choosing the two finalists. *Economic Theory*, 46(2): 211–219, 2011.

P. C. Fishburn. Weak qualitative probability on finite sets. *The Annals of Mathematical Statistics*, 40(6):2118–2126, 1969.

P. C. Fishburn. *The theory of social choice*. Princeton University Press, 1973.

P. C. Fishburn. Choice probabilities and choice functions. *Journal of Mathematical Psychology*, 18(3):205–219, 1978.

P. C. Fishburn. The axioms of subjective probability. *Statistical Science*, pages 335–345, 1986.

M. Frick, R. Iijima, and T. Strzalecki. Dynamic random utility. *Econometrica*, 87(6):1941–2002, 2019.

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

F. Gul and W. Pesendorfer. Random expected utility. *Econometrica*, 74(1):121–146, 2006.

D. R. Insua. On the foundations of decision making under partial information. *Theory and decision*, 33(1):83–100, 1992.

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

C. H. Kraft, J. W. Pratt, and A. Seidenberg. Intuitive probability on finite sets. *The Annals of Mathematical Statistics*, 30(2):408–419, 1959.

J. Li and R. Tang. Every random choice rule is backwards-induction rationalizable. *Games and Economic Behavior*, 104:563–567, 2017.

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

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

P. Manzini and M. Mariotti. State dependent choice. *Social Choice and Welfare*, 45(2):239–268, 2015.

P. Manzini and M. Mariotti. Dual random utility maximisation. *Journal of Economic Theory*, 177:162–182, 2018.

P. Manzini, M. Mariotti, and H. Petri. Corrigendum to “Dual random utility maximisation”[J. Econ. Theory 177 (2018) 162–182]. *Journal of Economic Theory*, 184:104944, 2019.

E. A. Ok and G. Tserenjigmid. Indifference, indecisiveness, experimentation and stochastic choice. 2020.

K. Saito. Axiomatizations of the mixed logit model. 2018.

D. Scott. Measurement structures and linear inequalities. *Journal of mathematical psychology*, 1(2):233–247, 1964.

H. A. Simon. A behavioral model of rational choice. *The quarterly journal of economics*, 69 (1):99–118, 1955.

J. Stoer and C. Witzgall. *Convexity and optimization in finite dimensions I*, volume 163. Springer Science & Business Media, 2012.

P. Suppes, D. H. Krantz, R. D. Luce, and A. Tversky. *Foundations of measurement: Geometrical, threshold, and probabilistic representations*, volume 2. 1989.

G. Tserenjigmid and M. Kovach. Behavioral Foundations of Nested Stochastic Choice and Nested Logit. 2019.

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