

Likelihood Regions: An Axiomatic Approach*

Fan Wang[†], Itzhak Gilboa[‡], and Stefania Minardi[§]

October 31, 2024

Abstract

We consider a reasoner who selects a set of distributions given a database of observations. A likelihood region is monotonic with respect to the likelihood function. We provide axiomatic foundations for such a selection rule. Starting with an abstract set of theories, we propose conditions on choice correspondences (across different databases) for which there exists a statistical model such that the choice correspondence is a likelihood region relative to that model, for the general case and for the case of a fixed likelihood-ratio threshold. We interpret the results as supporting the notion of likelihood regions for the selection of theories.

1 Introduction

Consider a government who needs to make decisions in novel situations, where scientific data exist, but do not suffice to pin down a single probability distribution. For example, the government has to decide how to tax CO_2 emission. Scientists may agree that emission contributes to global warming, but have

*We thank Fuad Aleskerov, Yoav Benjamini, Mira Frick, Paola Manzini, Marco Mariotti, Isaac Meilijson, Efe Ok, Fernando Payro Chew, Marcus Pivato, Saharon Rosset, and Gilles Stoltz for comments and references. Gilboa gratefully acknowledges ISF Grants 1077/17 and 1443/20, the Investissements d’Avenir ANR -11- IDEX-0003 / Labex ECODEC No. ANR -11-LABX-0047, the AXA Chair for Decision Sciences at HEC and the Foerder Institute at Tel-Aviv University.

[†]ESSEC, Singapore. b00786349@essec.edu

[‡]HEC, Paris, and Tel-Aviv University. tzachigilboa@gmail.com

[§]HEC, Paris. minardi@hec.fr

different estimates of the temperature distribution 30 years hence, given different levels of emission.¹ Or, facing a new pandemic, the government has to determine its vaccination policy. Medical evidence accumulates, but, at present, it doesn't suffice to have an agreed-upon estimate of the probability of various complications.² In such cases the decision maker needs to act without having sufficient data to single out a point estimate of the underlying probability distribution. Further, she does not have any reliable data on which to base a prior probability over the varying theories. A decision maker who is accountable to a wide, heterogenous public may wish to stick to estimation that is as objective as possible, selecting a set of theories without specifying a Bayesian posterior over them.

Likelihood regions The problem is clearly statistical in nature, and, indeed, several approaches have been proposed in the literature to deal with it. Specifically, it stands to reason to select distributions that obtain a certain likelihood threshold. Specifically, assume that, given a database of observations, D , and a collection of possible distributions, F , the reasoner has to select a subset of distributions for consideration, to be denoted $C_D(F)$. Likelihood regions are monotonic in the likelihood function. That is, denoting the likelihood function by $L(\cdot|D)$, for any two distributions f, g in F , if $L(f|D) \geq L(g|D)$, then, if g is included in the set $C_D(F)$, so should be f . (Equivalently, if f is excluded from the set, so should be g .) A function $C_D(F)$ that satisfies this property for all databases D is referred to as *monotonic* with respect to L . The sets it generates, $C_D(F)$, are called *likelihood regions* with respect to the statistical model (that is, the set of distributions).

A likelihood region $C_D(F)$ thus includes the top-tier likelihood distributions: there exists some $\gamma \in [0, 1]$ so that

¹The role of uncertainty has received considerable attention in climate policy. See, e.g., Pindyck (2007), Millner, Dietz, and Heal (2013), Heal and Millner (2014), Berger, Emmerling and Tavoni (2017), and Berger and Marinacci (2020).

²See, e.g., Manski (2018, 2019) and Berger et al. (2021) on how to incorporate uncertainty in policy making and, in particular, in public health contexts.

$$C_D(F) = \left\{ f \in F \mid L(f|D) \geq \gamma \sup_{g \in F} L(g|D) \right\} \quad (1)$$

or

$$C_D(F) = \left\{ f \in F \mid L(f|D) > \gamma \sup_{g \in F} L(g|D) \right\} \quad (2)$$

In the case of a finite set of distributions, the threshold γ is typically not unique and it can be chosen to obtain a representation according to (1) or to (2). More generally, the intuitive likelihood-monotonicity condition does not imply that either of these representations should hold for all databases D .³

The choice of the threshold γ reflects the reasoner’s degree of prudence, where a more conservative approach would be associated with a lower γ . When γ is very low, the reasoner hardly learns anything from the data, considering as relevant many possible distributions. When γ is close to 1, the reasoner tends to dismiss theories quite easily. The choice of γ thus leaves room for subjective judgment. Note, however, that this is a statistician’s choice, as are the level of confidence for confidence sets and the level of significance for hypothesis tests. A conservative statistical choice, of a low γ , may end up justifying highly non-conservative decisions (for example, if the decision maker maximizes the maximal expected utility over the set of “relevant” distributions).

Overview of results In this paper we axiomatize choice correspondences that can be rationalized as likelihood regions for some statistical model. We consider a set of abstract “theories” F , where a theory $f \in F$ is no more than a symbol. We use the term “theory” to highlight the fact that, for the time being, it is not associated with any distribution over observations, or any other data, numerical or otherwise. The association of distributions to theories will be derived from the reasoner’s choice correspondence C . That is, in the spirit of the revealed preferences paradigm, we are interested in conditions on the function C under which the reasoner selects theories *as if* she had a statistical

³See Epstein and Schneider (2007), who use such sets to select the distributions that are to be updated in a Bayesian fashion.

model in mind, and she picked likelihood regions relative to this statistical model.

More specifically, we assume that the reasoner observes $x \in X$. A database is a collection of observations, and we implicitly assume that their order does not matter (in the spirit of de Finetti's exchangeability). Thus, a database D is simply a counting function, indicating how many times each x was observed. Given such a (finite) D , we assume that the reasoner can select theories out of various subsets of F : for $A \subset F$, $C_D(A) \subset A$ are the theories that the reasoner considers relevant. We propose conditions on the choice function $C_D(A)$, relating its values for different sets (A) and different databases (D), under which there exists a statistical model relative to which $C_D(A)$ is a likelihood region for all D, A . That is, we derive, for each observation $x \in X$ and each theory $f \in F$, a number $p(x|f) \in [0, 1]$, interpreted as the conditional probability of observation x given theory f , and we show that, for each D, A , $C_D(A)$ is monotonic with respect to the associated likelihood function $L(\cdot|D)$.

Monotonicity of $C_D(A)$ with respect to a likelihood function $L(\cdot|D)$ allows the threshold γ to depend both on the choice set (A) and the database (D). It is natural to wonder when such a threshold is independent of these. We therefore suggest additional axioms that yield such a result. Specifically, these axioms yield a *quasi-representation* of the choice correspondences so that there exists a single $\gamma \in (0, 1]$ such that, for every $D \in \mathbb{D}$ and every $A \subset F$, and every $f \in A$

$$\begin{aligned} L(f|D) > \gamma \sup_{g \in A} L(g|D) &\Rightarrow f \in C_D(A) \\ L(f|D) < \gamma \sup_{g \in A} L(g|D) &\Rightarrow f \notin C_D(A) \end{aligned}$$

Observe that a quasi-representation does not specify the status of theories whose likelihood value is precisely at $\gamma \sup$. A (precise) *representation* would take the form

$$L(f|D) \geq \gamma \sup_{g \in A} L(g|D) \Leftrightarrow f \in C_D(A)$$

for every $D \in \mathbb{D}$ and every $A \subset F$. As we show below, such a precise representation can be obtained under additional topological structure and axioms.

Descriptive vs. normative interpretations Our results can be interpreted both descriptively and normatively. A descriptive⁴ interpretation would suggest that if a reasoner – a person, an organization, a group of scientists or statisticians – selects theories out of available sets while satisfying our axioms, her choice can be represented by a statistical model, with respect to which she selects likelihood regions. According to this interpretation, the reasoner need not be explicitly or even consciously using a statistical model. The model is silent on the mental, computational, or organizational processes that lead to theory selection. As long as the selection is made in accordance with the axioms, such an underlying statistical model exists, and it can (in principle) be estimated from past choice data for prediction of future choices.⁵ Adopting a normative interpretation, our results may convince a reasoner that she would like to select sets of theories that are likelihood regions. Further, the axioms and related analysis may assist the reasoner in eliciting a statistical model that best captures her beliefs.

The results presented here derive a statistical model from presumably-observable choices, without assuming that such a model is explicitly given. However, if a statistical model is given, it should surely not be ignored. To the contrary, we would expect the model of the reasoner to coincide with the objectively-given one, and our results would only serve as an additional argument for the selection of likelihood regions, whether the model is used descriptively or normatively. The main import of the model is, however, for the case of theories that are not structured in the mold of a statistical model. Taking a normative interpretation, the model may convince a reasoner that it makes sense to adopt a statistical model and choose likelihood regions with respect to this model. Admittedly, an additional source of subjectivity is introduced in this case: different reasoners may assign different conditional probabilities to the theories at hand. Yet, agreement over the general framework may reduce divergence of opinions. In a descriptive interpretation, the same argument

⁴We do not distinguish here between “descriptive” and “positive”, nor between “normative” and “prescriptive”.

⁵The statistical model we derive is unique only up to certain transformations, as will be specified in the formal statement of the main theorem.

would apply to an economist who models the reasoner: the axiomatization might make it more reasonable to assume that reasoners engage into theory selection as if they had a statistical model in mind.⁶

Organization The rest of this paper is organized as follows. Section 2 is devoted to choice-theoretic foundations, where we fix the database D and impose conditions on choice correspondences that guarantee that they be monotonic with respect to some order. Our conditions provide a new characterization of choice correspondences that correspond to semiorders (Luce, 1956). Section 3 embeds the choice-theoretic setup in the database context, and imposes additional conditions on $C_D(A)$, focusing on the way it changes across databases. Combining the conditions, we derive a statistical model relative to which $C_D(A)$ is a likelihood region for all D, A . The main result of this section also discusses uniqueness of the statistical model and it is preceded by a lemma that states a precise equivalence result. Next, Section 4 provides the axiomatization of the quasi-representation with a fixed threshold for all subsets and databases. A precise representation is obtained in Section 5, where the set of theories is assumed to be endowed with a topology, allowing us to state additional conditions. We provide a comparative statics result on degrees of prudence. Section 6 concludes with a discussion.

2 Choice-Theoretic Foundations

In this section we focus on a given database D , and suppress it from the notation. A few words of introduction might be useful to place the analysis in the context of choice theory.

We are interested in choice correspondences that are monotonic with respect to a function L , that is, functions $C(A)$ for which, for every $A \subset F$

⁶The situation is analogous to axiomatizations of subjective probability, which are not supposed to suggest that objective probabilities be ignored.

there exists $\gamma \in (0, 1]$ such that,

$$C(A) = \left\{ f \in A \mid L(f) \geq (>)\gamma \sup_{g \in A} L(g) \right\}$$

In order to derive the function L from choice data, we will impose additional conditions, which will be equivalent to the existence of a monotonic transformation of L , for which γ is independent of A . For a given database D , we may assume, without loss of generality, that the transformation is the identity. Thus we are interested in choice correspondences that can be represented by a function $u : F \rightarrow \mathbb{R}$ and a constant $\Delta \geq 0$ such that

$$C(A) = \left\{ f \in A \mid u(f) \geq (>)\max_{g \in A} u(g) - \Delta \right\} \quad (3)$$

(taking $u = \log(L)$ and $\Delta = \log\left(\frac{1}{\gamma}\right)$). While our motivation is the selection of theories given databases of observations, choice correspondences of this nature appear in other set-ups as well. In particular, decision makers or players in a game might have a utility function u , and only ε -maximize it. This might result from bounded rationality, unmodeled computation costs, or limited ability to discern differences in utility. Specifically, based on the literature in psychophysics (dating back to Weber, 1834, Fechner, 1868), we may consider a strict preference relation P that is a semiorder according to Luce (1956). In a finite setup, we may think of semiorders as relations P that can be represented by a function $u : F \rightarrow \mathbb{R}$ and a constant $\Delta \geq 0$ such that, for all f, g ,

$$gPf \iff u(g) - u(f) > \Delta. \quad (4)$$

Clearly, the representation (3) can be thought of as the set of P -undominated elements in the set A .

Related literature There is a rather rich literature on choice correspondences that can be represented as P -undominated elements for various partial orders P , and for semiorders in particular. Jamison and Lau (1973) presented the first result regarding semiorders; Fishburn (1975) corrected the stated version and added conditions for the more general case of P being an interval

order.⁷ In their setup a binary relation P was assumed, and the choice correspondence $C(A)$ was defined as the set of P -undominated elements from each subset A . Schwartz (1976) referred to the latter as the “Binary Choice Property (BICH)”, and offered several characterization results, covering the cases of semiorders and interval orders, where BICH is assumed in one set of conditions but not in the other. Aagaev and Aleskerov (1993) provided conditions for C to be determined by an interval order P , and Aleskerov, Bouyssou, and Monjardet (2007) dealt with the case of semiorders as well, complementing the axiomatization by a “ \mathcal{PO} ” condition, which guarantees that the choice correspondence can indeed be derived from a partial order (see Theorem 3.8 on p. 92). van Rooij (2008) deals with a similar problem, where there are two correspondences, one selecting the “good” and the other – the “bad” alternatives from each set A . Manzini and Mariotti (2012) characterize choice correspondences that can be represented as the selection of lexicographically-undominated elements according to some sequence of semiorders (but the sequence may well be longer than 1). Frick (2016) axiomatizes choice correspondences that have monotone threshold representations, that is, representations in which the threshold (Δ) can depend on the set (A). Importantly, she also provides a characterization for the case of a fixed Δ (Lemma 1(i), p. 765). Aguilera and Ulku (2017) axiomatize interval orders in which the threshold may depend on the menu, and, specifically, when this dependence is monotone with respect to set inclusion.

However, in order to relate the choice correspondences across different databases, we need a more explicit reference to the underlying weak order associated with the semiorder in question (this is the weak order which corresponds to the likelihood function we intend to derive). Given that, we find a different set of axioms to be more intuitive (though this is clearly a matter of taste). We thus provide here a different characterization of choice correspondences that can be represented as P -undominated elements for a semiorder P .

⁷An interval order can be thought of as a binary relation for which a representation as above exists, where Δ is allowed to depend on x or on y . See a formal definition and more details in the Appendix.

A by-product of our axiomatization is the numerical representation of these choice correspondences – and the associated semiorders – for sets of arbitrary cardinality. Beja and Gilboa (1992) provide conditions for such a representation, including separability. While these conditions can be translated to the language of choice correspondences, we find the resulting axioms rather cumbersome. In this paper the need for special conditions for infinite sets is obviated, because the numerical representation will be based on the context of the databases. As in Gilboa and Schmeidler (2003), most of the mathematical work (such as the use of separating hyperplane theorems) is done in the space of “contexts” (the different databases) rather than the alternatives themselves.

2.1 Axiomatic Characterization

We now turn to the formal analysis. Let F be a set. For any $A \subset F$, let $C(A) \subset A$ be a selection from A . If A is finite and nonempty, $C(A)$ is assumed to be nonempty as well. We wish to consider choice correspondences that select “top-tier elements” of each set A . Consider the standard axioms,

A1 Sen’s α : If $f \in B \subset A$ and $f \in C(A)$, then $f \in C(B)$.

Sen’s β : If $f, g \in C(A)$, $A \subset B$, and $g \in C(B)$, then $f \in C(B)$.

Sen’s α is a natural axiom for our purposes (and thus received the signifier A1): if f is among the top-tier elements in a larger set, it should also be among the top-tier ones in any subset thereof (that still contains f). By contrast, Sen’s β is too strong for our purposes: it is possible that both f, g are among the top-tier elements in a smaller set (A), but when additional elements are considered ($B \setminus A$), which may be better than both, the inferiority of f becomes apparent, while g remains “good enough” to be selected. In a standard choice setting, this might appear to be a result of bounded rationality, or at least (in the case of semiorders) of bounded perception abilities. By contrast, in our motivating example the objects to be ranked are probability distributions, and, bearing in mind that the data one observes are inherently random, the inclusion of theories that are not among the maximizers of the

likelihood function is viewed as a conscious, rational decision of a reasoner, akin to a statistician who constructs a confidence interval around a point estimator.

Be that as it may, the type of choice correspondences considered here need not satisfy Sen’s β and thus not the IIA in general. Specifically, choosing options that are close enough to the maximizers makes the choice procedure context-dependent. We note that, under Sen’s α , for all A, B , we have $C(A \cup B) \subset C(A) \cup C(B)$, so that this axiom restricts the choice set of the union “from above”, saying it cannot be too large. In a sense, Sen’s β could be viewed as requiring that it not be too small either. While we do not require this axiom, we have a comparable one that guarantees that the union operation doesn’t make too many elements drop out of the chosen set:

A2 Union: If $f \in C(A)$ and $g \in C(B)$, then $f \in C(A \cup B)$ or $g \in C(A \cup B)$.

Axiom A2 thus deals with expansion of a set, say from A to $A \cup B$. Clearly, the addition of new elements may make a highly-ranked element, f , less highly-ranked in the new, larger set. Intuitively, this would be the case if among the new elements there are some that are more highly-ranked than f . But in that case, those that were sufficiently highly-ranked in B should also remain so in the union. That is, the union of two sets can’t simultaneously dethrone two elements that were chosen in each of the sets separately. It can be easily verified that, in the presence of Sen’s α , Sen’s β implies Union.⁸

We next define a binary relation over elements of F that is intended to indicate strict preference.

Definition: For $f, g \in F$, $f \succ g$ if one of the following holds:

- (i) there exists a set A such that $f, g \in A$, $f \in C(A)$ and $g \notin C(A)$;
- (ii) there exists a set B such that $f, g \notin B$, $z \in B$, and $z \in C(B \cup \{g\})$, while $z \notin C(B \cup \{f\})$.

We can refer to clause (i) of the definition as *directly* revealed preference, and to (ii) – as *indirectly* revealed preference. Indeed, clause (i) deals with a

⁸Clearly, Union can be written as (for all nonempty A, B) $C(A) \subset C(A \cup B)$ or $C(B) \subset C(A \cup B)$.

choice from a set that contains both alternatives, and thus involves a direct comparison between them (among other elements). By contrast, clause (ii) involves choices in different contexts, none of which contains both alternatives. Yet, comparing choices in these two contexts indicates a higher ranking of f as compared to g : f is sufficiently highly-ranked to make another alternative, z , appear less attractive, while g isn't: adding g to B leaves z among the top-rated, whereas adding f to B dethrones z .

We can now state an additional axiom:

A3 Monotonicity: If $f, g \in A$, $g \in C(A)$ and $f \succ g$, then $f \in C(A)$.

Axiom A3 says that, if g is ranked sufficiently highly to be selected out of A , and f has been revealed to be ranked above g in another context, f should also be considered sufficiently high-ranking to be selected out of A .

The representation result can now be stated:

Proposition 1 If F is finite, a choice correspondence C on F satisfies A1-A3 iff there exist a function $u : F \rightarrow \mathbb{R}$ and $\Delta \geq 0$ such that, for all $A \subset F$,

$$\begin{aligned} C(A) &= \{f \in A \mid \nexists g \in A, \quad u(g) - u(f) > \Delta\} \\ &= \left\{ f \in A \mid u(f) \geq \max_{g \in A} u(g) - \Delta \right\} \end{aligned}$$

For the case of an infinite set F we need an additional axiom:

A4 Continuity: Assume that $(A_\tau)_{\tau \in \Upsilon}$ is an increasing chain ($A_\tau \subset A_{\tau'} \subset F$ for $\tau < \tau'$ where $<$ is a linear order on Υ). If $f \in C(A_\tau)$ for all $\tau \in \Upsilon$, then $f \in C(\cup_{\tau \in \Upsilon} A_\tau)$.

Axiom A4 states that an element f that is sufficiently highly ranked to be included in the choice from any set A_τ should also be sufficiently highly ranked in the limit (the union of the increasing chain). Alternatively, if f is not at the top of $\cup_{\tau \in \Upsilon} A_\tau$, there should be some large enough τ for which f will already be ruled out when only A_τ is considered. Equipped with this axiom we can formulate the next result:

Theorem 1 *A choice correspondence C on F satisfies A1-A4 iff there exist a semiorder P on F , such that, for all $A \subset F$,*

$$C(A) = \{ f \in A \mid \nexists g \in A, \quad gPf \}$$

For the sake of completeness we also state a similar result for the more general class of interval orders (originally defined in Fishburn, 1970, 1985) by dropping Monotonicity:

Proposition 2 *A choice correspondence C on F satisfies A1, A2, and A4 iff there exists an interval order P on F such that, for all $A \subset F$,*

$$C(A) = \{ f \in A \mid \nexists g \in A, \quad gPf \}$$

Interval orders are not sufficiently structured to obtain the result we seek, because they do not induce a well-defined weak order.⁹ As in the case of a semiorder, if there are only finitely many alternatives, Continuity holds vacuously and can thus be dropped from the statement of the proposition.

3 Likelihood Regions: A Preliminary Representation

In the rest of the paper, we integrate the choice-theoretic setup into the database context and study how the selection of theories changes with the evidence provided by different databases of observations.

Database Context Let X be the set of (types of) *observations*. The set of *databases* is defined as

$$\mathbb{D} \equiv \{ D \mid D : X \rightarrow \mathbb{Z}_+, \quad \sum_{x \in X} D(x) < \infty \}.$$

⁹For the sake of intuition, an interval order can be thought of by its representation: an alternative f is associated with an interval of values $[b(f), e(f)]$ with $b(f) \leq e(f)$, and gPf iff $b(g) > e(f)$. In the case $b(f) < b(g) < e(g) < e(f)$ neither P nor any binary relation we derive from it will rank f and g .

A database $D \in \mathbb{D}$ is interpreted as a counter vector, where $D(x)$ counts how many observations of type x appear in the database represented by D . Algebraic operations on \mathbb{D} are performed pointwise. Thus, for $D, D' \in \mathbb{D}$ and $k \geq 0$, $D + D' \in \mathbb{D}$, and $kD \in \mathbb{D}$ are well-defined. Similarly, the inequality $D \geq D'$ is read pointwise. For $D \in \mathbb{D}$ we define the support to be $\text{supp}D = \{x \in X \mid D(x) > 0\}$.

Let F be a set of *distributions*. For concreteness, one may bear in mind a parametric problem, where the distribution of a random variable is known up to finitely many parameters (but allow only for a limited precision in the measurement of the parameters). In such a setup all distributions are deemed to have the same level of complexity and, intuitively, the only criterion for evaluating them is the degree to which they explain the observations.¹⁰ However, in the formal analysis that follows the set of distributions may be any set of theories. At this point this is an abstract set, and its statistical meaning – specifically, the distribution that each $f \in F$ induces on the observations X – will be derived from the reasoner’s choices of elements of F given databases $D \in \mathbb{D}$.

Both X and F may be finite or not. To avoid trivial cases, we assume that each has at least two elements. Next, we assume that, for each $D \in \mathbb{D}$, and for each nonempty $A \subset F$, there exists $C_D(A) \subset A$ interpreted as the set of distributions in A that are considered relevant for decision making given the database D . Specifically, we may assume that there are observations that completely refute some theories, and that a set A consists of the theories that have not been refuted by the data.¹¹ If A is nonempty and finite, we assume that $C(A)$ is also nonempty.¹²

Basic Axioms We impose the following assumptions:

¹⁰In the Discussion we describe a more general model, in which the distributions may vary in terms of simplicity or some other criteria, such as subjective prior beliefs.

¹¹We thank Marcus Pivato for this interpretation.

¹²Observe that in our setup there is no loss of generality in assuming $C_D(A) \neq \emptyset$ even for an infinite A : the likelihood function is bounded from above by 1. Thus, the set $C(A) = \{f \in A \mid L(f) \geq \gamma \sup_{g \in A} L(g)\}$ will be empty only if the $\max_{g \in A} L(g)$ is not obtained and $\gamma = 1$.

P1 Choice: For every $D \in \mathbb{D}$, C_D on F satisfies Sen's α , Union, Monotonicity, and Continuity.

Given the analysis in Section 2, we know that for every $D \in \mathbb{D}$ there is a semiorder P_D on F such that, for every $A \subset F$, $C_D(A) \subset A$ consists precisely of the P_D -undominated distributions in A . Let \succsim_D be the associated weak order on F , with \succ_D and \sim_D denoting its asymmetric and symmetric parts, respectively.¹³

P2 Combination: For every $D, D' \in \mathbb{D}$ and every $f, g \in F$, $f \succsim_D g$ ($f \succ_D g$) and $f \succsim_{D'} g$ imply $f \succsim_{D+D'} g$ ($f \succ_{D+D'} g$).

The Combination axiom deals with the union of databases, and it states that, if each of two databases, D and D' , indicates that distribution f is at least as likely as g , the union of the two cannot suggest the reversed ranking. Observe that, D, D' being counter vectors, the union of the databases is captured by the addition of these vectors ($D + D'$), corresponding to the union of disjoint sets of observations. The axiom is stated for weak orders ($f \succsim_D g$ and $f \succsim_{D'} g$) but it is strengthened to require that, if one of them is strict ($f \succ_D g$), the same applies to the unified database $D + D'$.

The next axiom is Archimedean in nature:

P3 Archimedeanity: For every $D, D' \in \mathbb{D}$ and every $f, g \in F$, if $f \succ_D g$, then there exists $l \in \mathbb{N}$ such that $f \succ_{lD+D'} g$.

P3 assumes that a database D makes a distribution f more likely than another distribution g . It then requires that, even if we start with a database D' that favors g over f , sufficiently many repetitions of D will eventually overwhelm the evidence in D' .

We need the following assumption, which may be viewed as richness of potential observations relative to the distributions:

P4 Diversity: For every list (f, g, h, k) of distinct elements of F there exists $D \in \mathbb{D}$ such that $f \succ_D g \succ_D h \succ_D k$. If $|F| < 4$, then for any strict ordering of the elements of F there exists $D \in \mathbb{D}$ such that \succ_D is that ordering.

¹³It is well-known that a semiorder induces a weak order (see Luce, 1956, Definition 1 and Theorem 1).

Threshold-dependent quasi-representation To state the precise characterization result, we first define a matrix $v : F \times X \rightarrow \mathbb{R}$ to be *diversified* if there are no elements $f, g, h, k \in F$ with $g, h, k \neq f$ and $\lambda, \mu, \theta \in \mathbb{R}$ with $\lambda + \mu + \theta = 1$ such that $v(f, \cdot) \leq \lambda v(g, \cdot) + \mu v(h, \cdot) + \theta v(k, \cdot)$. That is, v is diversified if no row in v is dominated by an affine combination of three (or fewer) other rows.

Lemma 1 Let there be given X, F , and choice correspondences $\{C_D\}_{D \in \mathbb{D}}$. Then the following two statements are equivalent:

- (i) $\{C_D\}_{D \in \mathbb{D}}$ satisfy P1-P4;
- (ii) There is a diversified matrix $v : F \times X \rightarrow \mathbb{R}$ such that, for every $D \in \mathbb{D}$, there exists a semiorder P_D such that, for all A ,

$$C_D(A) = \{f \in A \mid \nexists g \in A, \quad g P_D f\}$$

and \succsim_D , its associated weak order, is represented by $U_D(\cdot) \equiv \sum_{x \in X} D(x)v(\cdot, x)$.¹⁴

The lemma combines the result of the previous section with the main result of Gilboa and Schmeidler (2003).

Finally, we impose another richness assumption, guaranteeing a maximal element for each singleton database. Specifically, for $x \in X$, let $D_x \in \mathbb{D}$ satisfy $D_x(x) = 1$ and $D_x(y) = 0$ for $y \neq x$. We require

P5 Maximality: For every $x \in X$, \succsim_{D_x} has a maximal element (in F).

We can now state our characterization of threshold-dependent likelihood regions.

Proposition 3 Let there be given X, F , and choice correspondences $\{C_D\}_{D \in \mathbb{D}}$ that satisfy P1-P5. There exist, for each $x \in X$ and $f \in F$, a number

¹⁴That is, for every $f, g \in F$, $f \succsim_D g$ iff

$$U_D(f) \geq U_D(g)$$

for

$$U_D(f) = \sum_{x \in X} D(x)v(f, x)$$

$p(x|f) \in (0, 1]$, such that, for every $D \in \mathbb{D}$ and every $A \subset F$, $C_D(A)$ is a likelihood region relative to the likelihood function

$$L(f|D) = \prod_{\{x \in X \mid D(x) > 0\}} [p(x|f)]^{D(x)}$$

Further, the conditional probabilities $(p(x|f))_{x,f}$ are unique up to multiplication of $(p(x|\cdot))_{x,\cdot}$ by $\lambda_x > 0$ and raising $(p(\cdot|f))_{\cdot,f}$ to a positive power.

Proposition 3 is a simple implication of Lemma 1 combined with Maximality. Note that it provides a quasi-representation of likelihood regions with thresholds that depend both on the choice set and the database. The next sections will build upon this basic result by developing a precise representation and by deriving a unique threshold for all choice sets and all databases.

4 A Fixed Threshold

In this section we consider additional axioms, which would guarantee that all likelihood regions are defined by the same threshold γ . In the following, for any database $D \in \mathbb{D}$ let P_D be the semiorder associated with $C_D(\cdot)$, namely, xP_Dy iff $C_D(\{x, y\}) = \{x\}$.

P6 Database Monotonicity:

- (i) For every $f, g \in F$, and every $D, D' \in \mathbb{D}$, if fP_Dg and $f \succsim_{D'} g$, then $fP_{(D+D')}g$.
- (ii) For every $f, g \in F$, and every $D, D' \in \mathbb{D}$, if $\neg(fP_Dg)$ and $g \succsim_{D'} f$, then $\neg(fP_{(D+D')}g)$.

Part (i) of the axiom says that if, given database D , f is sufficiently more highly ranked than g as to exclude the latter from any set containing f , and if according to another database D' f is more highly ranked than g , then the union of the databases should not only make f more highly ranked than g , but also retain the conclusion based on D , namely that g should be excluded from the choice set. Observe that, if $g \succ_{D'} f$ were the case, the additional information in D' may weaken the evidence in favor of f contained in D , and may even reverse the ordering. But the condition $f \succsim_{D'} g$ rules out this

possibility. Part (ii) of the axiom follows a very similar logic: starting with $\neg(fP_Dg)$, we observe that the evidence in D does not suffice to exclude g based on the presence of f . (Indeed, it is even possible that the opposite happens, or, at least that g is more highly ranked than f given D .) If we have additional evidence, D' , that, in and of itself makes g more highly ranked than f , we should certainly not exclude g in favor of f given the combined database.

We will also need another Archimedean axiom:

P7 P-Archimedeanity: For every $f, g \in F$, and every $D \in \mathbb{D}$, if $f \succ_D g$, then for every $D' \in \mathbb{D}$ there exists $l \in \mathbb{N}$ such that $fP_{lD+D'}g$.

The Archimedean axiom P3 guarantees that sufficiently many repetitions of a database D , which provides strict evidence in favor of f as compared to g , would overwhelm the evidence in any other database D' . The current axiom strengthens this requirement and demands that, with sufficiently many repetitions of D , the evidence in the combined database in favor of f would suffice to exclude g from the chosen set. Note that the axiom is necessary for a representation with a fixed threshold: when a database is replicated, the log-likelihood function for any two (non-trivial) theories tends to $-\infty$, and so does the difference between them. (Equivalently, the likelihood ratio of the lower-likelihood theory to the higher one tends to zero.) Hence, for sufficiently many repetitions, this difference will be large enough to exceed the jnd Δ .

Finally, we impose another richness axiom:

P8 Richness: For every distinct $f, g, h \in F$, there exist distinct $x, y, z \in X$ such that $\{\succsim_D\}_{\{D|\text{supp}(D)=\{x,y,z\}\}}$ equals the six permutations on (f, g, h) . If $|F| = 2$, the corresponding condition holds for some $x, y \in X$.

Axiom P8 requires that, for every three theories, $f, g, h \in F$, one can find observations $x, y, z \in X$ such that, when considering databases consisting of positive appearances of these observations, but only these, one obtains all the linear orderings on (f, g, h) , but only those. The fact that all six permutations can be obtained is similar to the Diversity axiom P4. The main new requirement is that, when considering only the three observations x, y, z , there can be no non-trivial equivalences between any pair in (f, g, h) . In the

language of Lemma 1, this is equivalent to the requirement that the matrix $[v(f, \cdot) - v(g, \cdot), v(f, \cdot) - v(h, \cdot)]$ will not consist only of numbers that are rational relative to each other. This condition will be naturally satisfied if the matrix v denotes the log-likelihood values for continuous distributions on some space \mathbb{R}^k . Indeed, in this case, the set of triples x, y, z that do not satisfy the non-equivalence condition is of measure zero in the respective space (\mathbb{R}^{3k}).

With these we can now state

Theorem 2 *Let there be given X, F , and choice correspondences $\{C_D\}_{D \in \mathbb{D}}$ that satisfy P1-P8.¹⁵ There exist, for each $x \in X$ and $f \in F$, a number $p(x|f) \in (0, 1]$, and there exists a unique constant $\gamma \in (0, 1]$ such that, for every $D \in \mathbb{D}$ and every $A \subset F$, $C_D(A)$ is a likelihood region relative to the likelihood function*

$$L(f|D) = \prod_{\{x \in X \mid D(x) > 0\}} [p(x|f)]^{D(x)}$$

and the threshold γ . That is, for every $D \in \mathbb{D}$ and every $A \subset F$, and every $f \in A$

$$\begin{aligned} L(f|D) > \gamma \sup_{g \in A} L(g|D) &\Rightarrow f \in C_D(A) \\ L(f|D) < \gamma \sup_{g \in A} L(g|D) &\Rightarrow f \notin C_D(A) \end{aligned}$$

Further, the conditional probabilities $(p(x|f))_{x,f}$ are unique up to multiplication of $(p(x|\cdot))_{x,\cdot}$ by $\lambda_x > 0$ and raising $(p(\cdot|\cdot))_{\cdot,\cdot}$ to a positive power.

5 Continuous Representations

The representation provided by Theorem 2 classifies theories whose likelihood value, $L(f|D)$, is strictly above, as well as strictly below, the threshold, $\gamma \sup_{g \in A} L(g|D)$. They remain silent on the theories that are equal to that threshold. We know that, for every $D \in \mathbb{D}$ and for every $A \subset F$, all the theories that have $L(f|D) = \gamma \sup_{g \in A} L(g|D)$ have to be treated identically

¹⁵As mentioned above, P7 implies P3. Hence, “P1-P8” can be replaced by “P1,P2,P4-P8”.

– either all of them are in $C_D(A)$ or all of them are not. But, as we vary D, A , both cases are possible. It is therefore of interest to ask, what would it take, in terms of the presumably-observable choice correspondences $\{C_D\}_{D \in \mathbb{D}}$, to obtain a representation by

$$L(f|D) \geq \gamma \sup_{g \in A} L(g|D) \quad \Leftrightarrow \quad f \in C_D(A)$$

for every $D \in \mathbb{D}$ and every $A \subset F$, and every $f \in A$ (or, alternatively, the same equivalence with a strict inequality).

One way to obtain such a representation would be to demand that the sets $C_D(A)$ be closed in an underlying topology on F , and then make sure that the derived likelihood values are continuous with respect to this topology. Specifically, assume that F is a separable connected topological space. This would indeed be the case in most relevant examples of sets of distributions: in fact, if f and g are distributions, so will be $\alpha f + (1 - \alpha)g$ for all $\alpha \in [0, 1]$, and it is natural to define a topology on F relative to which the interval connecting two distributions is a path. Recall, however, that we start out with an abstract set F , not yet endowed with numbers that can be interpreted as conditional probabilities or densities. Thus, we assume the additional structure of the topology on F , add the closedness assumption, and proceed to make sure that the numbers derived from Theorem 2 are continuous relative to the topology assumed. Formally, we add three axioms/conditions, stated in the language of the topology on F :

P9 Closedness: For every $D \in \mathbb{D}$ and for every $A \subset F$, $C_D(A)$ is closed (in F).

P10 F -Continuity: For every $D \in \mathbb{D}$, the relation \succsim_D is continuous (on F).

Finally, we also need another richness assumption. To state it, let us denote, for \succsim_D and $f \in F$, f 's \succsim_D -equivalence class by $E_D(f)$, that is

$$E_D(f) = \{g \in F \mid g \sim_D f\}$$

We impose

P11 Connected richness: For every $x \in X$ and $f \in F$ there exists $y \in X$ such that (i) $E_{\mathbf{1}_{\{x\}}}(f)$ contains a non-trivial connected subspace of F which is not included in $E_{\mathbf{1}_{\{y\}}}(f)$ (that is, $g \succ_{\mathbf{1}_{\{y\}}} g'$ for some $g, g' \in E_{\mathbf{1}_{\{x\}}}(f)$) and (ii) there are $g, h \in F$ such that $g \succ_{\mathbf{1}_{\{x\}}} f \succ_{\mathbf{1}_{\{x\}}} h$ and $h \succ_{\mathbf{1}_{\{y\}}} f \succ_{\mathbf{1}_{\{y\}}} g$.

We can now state

Theorem 3 *Let there be given X, F , and choice correspondences $\{C_D\}_{D \in \mathbb{D}}$ that satisfy P1-P11. There exist, for each $x \in X$, a continuous function $p(x|\cdot) : F \rightarrow (0, 1]$, and there exists a unique constant $\gamma \in (0, 1]$ such that, for every $D \in \mathbb{D}$ and every $A \subset F$,*

$$L(f|D) \geq \gamma \sup_{g \in A} L(g|D) \quad \Leftrightarrow \quad f \in C_D(A)$$

where

$$L(f|D) = \prod_{\{x \in X \mid D(x) > 0\}} [p(x|f)]^{D(x)}$$

Further, the conditional probabilities $(p(x|f))_{x,f}$ are unique up to multiplication of $(p(x|\cdot))_{x,\cdot}$ by $\lambda_x > 0$ and raising $(p(\cdot|\cdot))_{\cdot,\cdot}$ to a positive power.

5.1 Comparative Prudence

The precise characterization achieved with the continuous model allows us to develop a comparative analysis of the effects of the threshold γ on the selection of theories. In this subsection we consider choice correspondences that admit a characterization as in Theorem 3. We refer to such choice correspondences as *continuous likelihood regions* and denote their representation by the pair (p, γ) .

Consider two reasoners with continuous likelihood regions given by $\{C_D^1\}_{D \in \mathbb{D}}$ and $\{C_D^2\}_{D \in \mathbb{D}}$. We say that $\{C_D^1\}_{D \in \mathbb{D}}$ is *comparable* to $\{C_D^2\}_{D \in \mathbb{D}}$ if, for every $D \in \mathbb{D}$ and every $f, g \in F$,

$$f \succ_D^1 g \Leftrightarrow f \succ_D^2 g$$

The notion of comparability requires that the two reasoners rank theories in the same way, conditional on having access to the same data. It has a simple characterization, as shown next.

Observation 1 *Let $\{C_D^1\}_{D \in \mathbb{D}}$ and $\{C_D^2\}_{D \in \mathbb{D}}$ be two choice correspondences with continuous likelihood regions representations given by (p_1, γ_1) and (p_2, γ_2) , respectively. Then, the following conditions are equivalent:*

(i) $\{C_D^1\}_{D \in \mathbb{D}}$ is comparable to $\{C_D^2\}_{D \in \mathbb{D}}$;

(ii) For every $x \in X$, $p_1(x|\cdot) = \lambda_x p_2(x|\cdot)^\alpha$ for some $\lambda_x > 0$ and $\alpha > 0$.

Behavioral differences between two comparable reasoners can therefore be attributed solely to differing attitudes towards retaining/dismissing possible distributions, as captured by a single parameter, the threshold γ , which measures the reasoner's degree of prudence used to generate likelihood regions. The next result characterizes what it means for one reasoner to be more prudent than another.

Observation 2 *Let $\{C_D^1\}_{D \in \mathbb{D}}$ and $\{C_D^2\}_{D \in \mathbb{D}}$ be two choice correspondences with continuous likelihood regions representations given by (p_1, γ_1) and (p_2, γ_2) , respectively. Assume that $\{C_D^1\}_{D \in \mathbb{D}}$ is comparable to $\{C_D^2\}_{D \in \mathbb{D}}$. Then, $\gamma_1 \leq \gamma_2$ iff, for every $D \in \mathbb{D}$ and every $A \subset F$, $C_D^1(A) \supseteq C_D^2(A)$.*

Observation 2 provides a notion of comparative prudence: larger likelihood regions are characterized by lower threshold levels. Hence, the more prudent a reasoner is, the less she will dismiss possible theories. Observation 2 formally shows that this more conservative approach is characterized by lower γ . (The proof is straightforward, given the previous analysis, and, therefore, omitted.)

6 Discussion

6.1 Simplicity

Comparing distributions according to their likelihood function is a very sensible approach when the distributions are a priori on equal footing. But it can be a rather poor idea if they differ in meaningful ways ex ante. In particular, it is well known that, if theories vary in complexity, maximization of the like-

likelihood function typically results in overfitting. Thus, one would wish to select theories in ways that would trade off likelihood and simplicity, as in the AIC and BIC criteria (Akaike, 1974, Schwarz, 1978).

Gilboa and Schmeidler (2011) offer an axiomatic approach to this problem, deriving a representation of a binary relation over theories by the log-likelihood function with an additional additive term (that depends on the theory but not on the data). This functional form includes the AIC, where the additive term depends only on the number of the parameters used by the theory. The axiomatization can be adapted to our context: one would have to re-state the axioms in that paper for the preference orders \succsim_D obtained from the choice correspondence $C_D(A)$. However, further analysis is called for to axiomatize other rules, such as theory selection based on the BIC (where the additive term depends also on the number of observations in the database).

6.2 Vague Theories

As stated in the introduction, the main significance of our result is likely to be for situations in which the problem is not formulated in terms of a statistical model to begin with. For example, in geopolitical setups one may consider theories such as “Fundamentalism is on the rise” or “Countries are becoming more democratic”. These theories should be tested in light of evidence, and, indeed, it seems that people do assign more credence to theories that have predicted past events better than others. Our result may be read as a suggestion to make this process more precise, by quantifying the theories (a priori) in terms of a statistical model, and then calculating their likelihood function.

However, this idea might appear somewhat unrealistic: can a reasoner specify, ex-ante, what is the conditional probability of, say, a military coup, given a theory which is as vague as “Fundamentalism is on the rise”? It seems that many theories in the social sciences are hard to quantify so precisely. This difficulty suggests that one may wish to consider a more general model, in which theories are less restrictive. For instance, a theory need not specify a single probability for any observation, but rather provide a range of probabilities.

We believe that there is a need for more general models for these cases.

However, we mention in passing that a particular, simple generalization does not call for a new model: if a “vague theory” is conceived of as a set of possible distributions, we may treat each and every one of these as a separate theory in our model. Thus, each specific theory within a vague theory would be judged according to its own likelihood function, and the vague theory would be implicitly judged by the highest likelihood of its constituents.

6.3 Bayesianism

Selection of theories can also be done by an appeal to a (typically subjective) probabilistic belief over the theories. Indeed, a Bayesian would have a prior over theories, and would update it to generate a posterior. Viewed from the Bayesian perspective, the selection of a subset of theories lacks nuance: theories aren’t dichotomously divided into “reasonable” and “unreasonable”. Rather, the reasoner has probabilistic beliefs about the correct theory, which are naturally measured on a continuous scale. Moreover, this approach relies on formidable foundations, such as Savage’s (1954) derivation of subjective expected utility maximization.

Despite the elegance and coherence of the Bayesian approach, which is probably unmatched by any other method of inference and reasoning, and despite the stunning beauty and depth of Savage’s result, we find that the Bayesian approach is too demanding in many situations of interest. Reasoners and scientists may find it difficult to assign (single) probabilities to theories about wars and financial crises, the evolution of the planet and of its inhabitants. Coming up with *some* probabilities, just in order to use the Bayesian machinery, often seems arbitrary. In many situations it appears that it would be more rational to admit that we do not know all the relevant probabilities, rather than to pretend that we do. The present paper attempts to make a contribution to rational selection of theories for those reasoners who find the Bayesian model too restrictive.

Appendix: Proofs and Related Analysis

We divide this appendix into two: the analysis that relates to choice functions in general, and then – to choice correspondences given databases.

7 Choice-Theoretic Analysis

In order to prove the sufficiency of the axioms in Proposition 1, Theorem 1, and Proposition 2, several auxiliary lemmas will be useful.

7.1 Preliminary lemmas

Assume first that C is a choice correspondence satisfying A1, A2, and A4.

Lemma 2 For all $A, B \subset F$, we have $C(A \cup B) \subset C(A) \cup C(B)$.

Proof: Consider $f \in C(A \cup B)$. If $f \in A$, then by Sen's α , $f \in C(A)$, and if $f \in B$, then by the same logic, $f \in C(B)$. Thus, $f \in C(A) \cup C(B)$ (and, in fact, if $f \in A \cap B$ then we also have $f \in C(A) \cap C(B)$). \square

Note that the proof only uses Sen's α .

Lemma 3 For all $A \subset F$ and $f \in A$, if for all $g \in A$ we have $f \in C(\{f, g\})$, then $f \in C(A)$.

Proof: Consider first the case of a finite set A . Suppose that $A = \{f, g_1, g_2, \dots, g_k\}$. By induction on $i \leq k$, one observes that $f \in C(\{f, g_1, g_2, \dots, g_i\})$ using Union and the fact that $f \in C(\{f, g_{i+1}\})$ for $0 \leq i < k$.

In case A is infinite, we need to use both Union and Continuity. Using Zorn's lemma, consider the set of sets $B \subset A (\subset F)$ such that $(f \in B \text{ and } f \in C(B))$, ordered by set inclusion. Continuity guarantees that every increasing chain in this poset, $(B_\tau)_{\tau \in \Upsilon}$ ($B_\tau \subset B_{\tau'} \subset F$ for $\tau < \tau'$) has an upper bound, namely $\cup_{\tau \in \Upsilon} B_\tau$, with $f \in C(\cup_{\tau \in \Upsilon} B_\tau)$. The lemma implies that there exists a maximal element B^* in the poset, so that $f \in C(B^*)$. If $B^* = A$, we are done. Otherwise, there exists $g \in A \setminus B^*$. We know that $f \in C(\{f, g\})$ and, using

Union again, this implies that $f \in C(B^* \cup \{g\})$ so that B^* is not a maximal element. This contradiction implies that $B^* = A$ and $f \in C(A)$ as required. \square

Note that the proof only uses Union and Continuity.

The following lemma summarizes the previous two and suggests that, under the axioms A1, A2, A4, binary comparisons contain all relevant information.

Lemma 4 For all $A \subset F$ and $f \in A$, $f \in C(A)$ iff for all $g \in A$ we have $f \in C(\{f, g\})$.

Proof: Lemma 3 implies the “if” direction. As for the “only if”, assume that $f \in C(A)$. If, for some $g \in A$ we have $f \notin C(\{f, g\})$ then, by Sen’s α , f cannot be chosen out of the superset, A , of $\{f, g\}$. \square

Note that the proof above only uses Sen’s α , Union, and Continuity (for the infinite case).

Lemma 5 : If there exists a set $B \subset F$ such that $f, g \notin B$, $f \in C(B \cup \{f\})$ and $g \notin C(B \cup \{g\})$, then $f \succ g$.

Proof: Consider $D = B \cup \{f, g\}$. Because we may write $D = [B \cup \{f\}] \cup \{g\}$, with $f \in C(B \cup \{f\})$ and $g \in C(\{g\})$, Union implies that $f \in C(D)$ or $g \in C(D)$. However, $g \notin C(B \cup \{g\})$ implies, using Sen’s α , that $g \notin C(D)$. Thus $f \in C(D)$ and $f \succ g$ follows. \square

Let us now assume, throughout the rest of this subsection, that C also satisfies A3 (Monotonicity).

Lemma 6 : \succ is irreflexive, asymmetric, and transitive.

Proof: The fact that \succ is irreflexive, that is, that $f \succ f$ can’t hold for any f , is obvious from the definition. We proceed to prove asymmetry and transitivity.

Claim: \succ is asymmetric.

Proof: Assume, to the contrary, that $f \succ g$ and $g \succ f$ for some f, g . Distinguish between two cases:

Case 1: $f \succ g$ or $g \succ f$ by clause (i). Suppose, w.l.o.g., that it is $f \succ g$. Then there exists A such that $f, g \in A$, $f \in C(A)$ and $g \notin C(A)$. However, Monotonicity says that, because $f \in C(A)$ and $g \succ f$, we must also have $g \in C(A)$, a contradiction.

Case 2: Both $f \succ g$ and $g \succ f$ hold (only) by clause (ii). In this case, there exist two sets, B and E , such that $f, g \notin B, E$ and two elements $h \in B$ and $d \in E$ such that $(f \succ g)$ [$h \notin C(B \cup \{f\})$ and $h \in C(B \cup \{g\})$] as well as $(g \succ f)$ [$d \in C(E \cup \{f\})$ and $d \notin C(E \cup \{g\})$]. Consider $D = B \cup E \cup \{f, g\}$. Because $D = [B \cup \{g\}] \cup [E \cup \{f\}]$ with $h \in C(B \cup \{g\})$ and $d \in C(E \cup \{f\})$, Union implies that $h \in C(D)$ or $d \in C(D)$. However, $h \notin C(B \cup \{f\})$ rules out the former (by Sen's α), and, similarly, $d \notin C(E \cup \{g\})$ rules out the latter, a contradiction.

Claim: \succ is transitive.

Proof: Let there be given f, g, h with $f \succ g$ and $g \succ h$. Observe that f, g, h have to be distinct because \succ is irreflexive and asymmetric. Distinguish among the following cases:

Case 1: $f \succ g$ because of clause (i). That is, there exists a set A such that $f, g \in A$ and $f \in C(A)$ while $g \notin C(A)$. If $h \in A$ then $g \notin C(A)$ implies $h \notin C(A)$ (by Monotonicity) and $f \succ h$ follows (by clause (i)). If $h \notin A$ consider $D = A \cup \{h\}$ with $f, g, h \in D$. By Union, at least one of f ($\in C(A)$) and h ($\in C(\{h\})$) is in $C(D)$. But $h \in C(D)$ would imply (by Monotonicity) $g \in C(D)$ and (by Sen's α) $g \in C(A)$, a contradiction. Hence $h \notin C(D)$ and $f \in C(D)$, implying $f \succ h$ again.

Case 2: $g \succ h$ because of clause (i). That is, there exists a set A such that $g, h \in A$ and $g \in C(A)$ while $h \notin C(A)$. If $f \in A$ we have $f \in C(A)$ by Monotonicity, and $f \succ h$ follows. Otherwise, consider $D = A \cup \{f\}$ with $f, g, h \in D$. By Union, at least one of f ($\in C(\{f\})$) and g ($\in C(A)$) is in $C(D)$. Yet, $g \in C(D)$ implies $f \in C(D)$ by Monotonicity. In both cases, then, $f \in C(D)$ while $h \notin C(D)$ (because $h \notin C(A)$). Thus $f \succ h$.

Case 3: Both $f \succ g$ and $g \succ h$ hold (only) by clause (ii). In this case, there exist two sets, B and E , $f, g \notin B$ and $g, h \notin E$, and two elements $d \in B$ and $t \in E$ such that $(f \succ g)$ [$d \notin C(B \cup \{f\})$ and $d \in C(B \cup \{g\})$] as well as

$(g \succ h) [t \notin C(E \cup \{g\}) \text{ and } t \in C(E \cup \{h\})]$. Consider $D = B \cup E \cup \{g, h\}$. As $D = [B \cup \{g\}] \cup [E \cup \{h\}]$, Union implies that $d \in C(D)$ or $t \in C(D)$. However, $t \notin C(E \cup \{g\})$ rules out (by Sen's α) the latter, and we have $d \in C(D)$. Hence we also have $d \in C(B \cup E \cup \{h\})$. (In particular, this also means that $f \notin E$, because $d \notin C(B \cup \{f\})$.) However, $d \notin C(B \cup \{f\})$ implies $d \notin C(B \cup E \cup \{f\})$ and $f \succ h$ follows (by clause (ii)). \square

It is natural to define

Definition 1 For $f, g \in F$, $f \sim g$ if neither $f \succ g$ nor $g \succ f$.

Lemma 7 : \sim is an equivalence relation.

Proof: Because \succ is irreflexive, \sim is reflexive. Symmetry follows from the definition of \sim . Thus we only have to prove transitivity. Assume that $f \sim g$, $g \sim h$. If $f \sim h$ doesn't hold, we have a \succ relation between f and h , without loss of generality, $f \succ h$. There are two possibilities to consider:

Case 1: $f \succ h$ because of clause (i). That is, there exists a set A such that $f, h \in A$ and $f \in C(A)$ while $h \notin C(A)$. If $g \in A$, then we either have $g \in C(A)$, in which case $g \succ h$ (by definition of \succ), or $g \notin C(A)$, which implies that $f \succ g$ (again, by definition). In both cases we obtain a contradiction.

Assume, then, that $g \notin A$ and consider $B = A \cup \{g\}$. Observe that $h \notin C(B)$ by Sen's α . If $f \in C(B)$, proceed as above (to show that $f \succ g$ or $g \succ h$ must hold). If $f \notin C(B)$, then, by Union, we must have $g \in C(B)$ (as $C(\{g\}) = \{g\}$). But then $g \succ f$ follows – again, a contradiction.

Case 2: $f \succ h$ because of clause (ii). That is, there exists a set B such that $f, h \notin B$, and d such that $d \notin C(B \cup \{f\})$ and $d \in C(B \cup \{h\})$. If $g \notin B$, we consider $B \cup \{g\}$. If $d \in C(B \cup \{g\})$, we have $f \succ g$, and if $d \notin C(B \cup \{g\})$ – $g \succ h$ follows (in both cases, by clause (ii) of the definition of \succ) – a contradiction.

Consider, then, the sub-case in which $g \in B$. Define $B' = B \setminus \{g\}$ and consider $B' \cup \{f, h\}$. If $d \notin C(B' \cup \{f, h\})$, then, since $d \in C(B' \cup \{g, h\})$ ($=C(B \cup \{h\})$) we get $f \succ g$ (for the set $B' \cup \{h\}$). If, however, $d \in$

$C(B' \cup \{f, h\})$ then, since $d \notin C(B' \cup \{f, g\}) (=C(B \cup \{f\}))$ we get $g \succ h$ (for the set $B' \cup \{f\}$). Thus, in both cases we obtain a contradiction. \square

We now define a binary relation P that is to be interpreted as “sufficiently more highly ranked than”:

Definition 2 For $f, g \in F$, fPg if $g \notin C(\{f, g\})$.

Observe that, for $f \neq g$, $g \notin C(\{f, g\})$ means that $C(\{f, g\}) = \{f\}$. As stated, the definition also applies to the case $f = g$, in which case it means that P is irreflexive. Evidently, P is also asymmetric.

Considering $A = \{f, g\}$, we observe that fPg implies $f \succ g$. The converse need not hold, as the difference in ranking between f and g might only be revealed in the presence of other alternatives. We also define

Definition 3 For $f, g \in F$, fIg if neither fPg nor gPf .

Note that I is symmetric by its definition, and that it is reflexive because P is irreflexive. However, I need not be transitive. We now wish to show that the relation P and its associated I satisfy Luce’s axioms for a semiorder (Luce, 1956). Specifically, using the notation of concatenation of binary relations, we wish to show that $PIP \subset P$ and $PPI \subset P$.¹⁶ We first observe that both PI and IP imply the relation \succ .

Lemma 8 : For all f, g, h if $(fPg$ and $gIh)$ or $(fIg$ and $gPh)$, then $f \succ h$.

Proof: If $g = h$, fPg means fPh and $f \succ h$ follows. Otherwise, f, g, h are distinct ($f \neq g$ because P is irreflexive and $f \neq h$ because fPg but hIg).

Assume first that fPg and gIh . Thus, $C(\{f, g\}) = \{f\}$ and $C(\{g, h\}) = \{g, h\}$. Considering $B = \{g\}$, we have $g \in C(B \cup \{h\})$ but $g \notin C(B \cup \{f\})$ and $f \succ h$ follows by clause (ii) of the definition of \succ .

Assume next that fIg and gPh . Then, $C(\{f, g\}) = \{f, g\}$ and $C(\{g, h\}) = \{g\}$. Setting $B = \{g\}$ again, we have $f \in C(B \cup \{f\})$ but $h \notin C(B \cup \{h\})$ and $f \succ h$ follows from Lemma 5. \square

¹⁶The concatenation is defined by an existential quantifier: xR_1R_2y if there exists a z such that xR_1zR_2y . It is used inductively, so that, for example, $xPIP_y$ means that there are z, w such that xPz , zIw , wPy

We now proceed to prove that (P, I) satisfy Luce's axioms.

Lemma 9 : For all $f, g, h, d \in F$, if fPg , gIh , and hPd , then fPd .

Proof: Assume that fPg , gIh , and hPd . Because each of the relations P, IP, PI was proven to imply \succ , and because \succ is irreflexive and transitive, we have $f \succ g$, $f \succ h$, $g \succ d$, and $h \succ d$. Also, $d = f$ would imply $dPIh$ and $d \succ h$ while we have $h \succ d$. Hence we conclude that among f, g, h, d we have at least three distinct elements, where the only possible equality is $g = h$. We allow for this possibility in the sequel (with the obvious notational convention that $\{g, h\} = \{g\}$ if $g = h$).

We have $C(\{f, g\}) = \{f\}$, $C(\{g, h\}) = \{g, h\}$, $C(\{h, d\}) = \{h\}$ and need to show that $C(\{f, d\}) = \{f\}$. Assume, to the contrary, that $d \in C(\{f, d\})$. Since $g \in C(\{g, h\})$ we can apply Union to conclude that at least one of d, g is in $C(\{f, g, h, d\})$. However, $d \in C(\{f, g, h, d\})$ would imply $d \in C(\{h, d\})$ and $g \in C(\{f, g, h, d\})$ would imply $g \in C(\{f, g\})$, both of which are known to be false. \square

Observe that Lemma 9 also proves that P is transitive (corresponding to the case $g = h$ in the proof above). We now turn to the second axiom:

Lemma 10 : For all f, g, h, d , if fPg , gPh , and hId , then fPd .

Proof: Assume that fPg , gPh , and hId . As above, we have $f \succ g$, $g \succ h$, $f \succ h$. Further, as P is transitive, we have fPh and thus both $fPI d$ and $gPI d$ so that $f, g \succ d$. It follows that among f, g, h, d we have at least three distinct elements, where the only possible equality is $h = d$. However, in this case we have fPd (because fPh), so we may turn to the case in which all four elements are distinct.

We have $C(\{f, g\}) = \{f\}$, $C(\{g, h\}) = \{g\}$, $C(\{h, d\}) = \{h, d\}$ and need to show that $C(\{f, d\}) = \{f\}$. Assume, to the contrary, that $d \in C(\{f, d\})$. Consider $B = \{f\}$ and observe that $g \notin C(B \cup \{g\})$ while $d \in C(B \cup \{d\})$ and thus, by Lemma 5, $d \succ g$. However, we already established that $g \succ d$, a contradiction. \square

Next, we wish to show that the semiorder P captures all the relevant information in the choice correspondence C . To this end, we remind the reader that, if P is a semiorder, then $Q^* \equiv PI \cup IP$ is the strict part of a weak order. That is, Q^* is irreflexive, asymmetric, and transitive, where $Q^0 \equiv (Q^* \cup Q^{*-1})^c$ is an equivalence relation. Thus, $Q = Q^* \cup Q^0$ is the weak order whose asymmetric part is Q^* and whose symmetric part is Q^0 . It is well known (and easy to verify) that, with the above definitions, $QP \subset P$ and $PQ \subset P$.

We wish to show that $(Q^*, Q^0) = (\succ, \sim)$, that is, that the relations (Q^*, Q^0) , defined in terms of the binary relation P , are identical to the relations (\succ, \sim) defined in terms of the choice correspondence C .

Lemma 11 : $(Q^*, Q^0) = (\succ, \sim)$.

Proof: We have already shown in Lemma 8 that $Q^* \subset \succ$, which also means that $\sim \subset Q^0$ and we now wish to show the converse, that is, that $f \succ g$ implies fQ^*g (and thus fQ^0g would imply $f \sim g$). Assume, then, that $f \succ g$. We have two possibilities to consider, depending on the clause of the definition of \succ by which the relation holds:

Case 1: There exists a set A such that $f, g \in A$, $f \in C(A)$ and $g \notin C(A)$.

By Lemma 3, there exists $h \in A$ such that $g \notin C(\{g, h\})$ (or else the lemma implies that $g \in C(A)$). Thus, hPg . If, however, $f \notin C(\{f, h\})$ then $f \in C(A)$ would be impossible. Thus, $f \in C(\{f, h\})$. This leaves us with two possibilities: fIh (if $C(\{f, h\}) = \{f, h\}$), in which case $fIPg$ and fQ^*g , or fPh (if $C(\{f, h\}) = \{f\}$), which implies fPg and fQ^*g .

Case 2: There exists a set B such that $f, g \notin B$, and $h \in C(B \cup \{g\})$ (in particular, $h \in B$) while $h \notin C(B \cup \{f\})$. By Lemma 4, $h \in C(B \cup \{g\})$ means that, for each $d \in B \cup \{g\}$, we cannot have dPh . In particular, we have gIh or hPg . By the same lemma, $h \notin C(B \cup \{f\})$ implies that there exists an element t in $B \cup \{f\}$ such that tPh , and since this element isn't in B , it has to be f . That is, fPh . Thus we have $fPIg$ (in case gIh) or $fPPg$, hence fPg (in case hPg). In both cases, fQ^*g follows. \square

7.2 Proof of Theorem 1 and of Proposition 1

Proof of Theorem 1:

The “only if” has been proved above: we started with a choice correspondence C that satisfied A1-A4, and defined a semiorder P associated with it, namely fPg iff $g \notin C(\{f, g\})$. Lemma 4 has shown that, for every set A , and every element $f \in A$, $f \in C(A)$ iff $f \in C(\{f, g\})$ for every $g \in A$, that is, iff for every $g \in A$ it is not the case that gPf .

To see the converse, assume that P is a semiorder and that $C(A) = \{f \in A \mid \nexists g \in A, gPf\}$ for every A . It is immediate that C satisfies Sen’s α , because, if $B \subset A$, an element that isn’t P -dominated by any element in A can’t be P -dominated by any element in B .

To see that Union holds, assume that $f \in C(A)$ and $g \in C(B)$. Let h be a Q -maximizer in A and let d be a Q -maximizer in B . Assume w.l.o.g. that hQd . We know that $\neg hPf$ and hence f is not P -dominated by any element in B (because tPf for some $t \in B$ would imply dPf and then hPf). Thus, f is not P -dominated by any element in A or in B and $f \in C(A \cup B)$ follows. (Naturally, $g \in C(A \cup B)$ follows from dQh .)

We now turn to prove Monotonicity. Assume that $f, g \in A$, $g \in C(A)$ and $f \succ g$. We argue that fQ^*g (where $Q^* = PI \cup IP$ is the strict part of the weak order defined by the semiorder P). Again, there are two possibilities to consider:

Case 1: $f \succ g$ because of clause (i). That is, there exists a set D such that $f, g \in D$ and $f \in C(D)$ while $g \notin C(D)$. In this case there exists $h \in D$ such that hPg (as g is P -dominated in D) but we do not have hPf (as f is P -undominated in D). Then we have fPh (in which case $fPhPg$ and then fPg) or fIh (and $fIhPg$). In both cases, fQ^*g .

Case 2: $f \succ g$ because of clause (ii). That is, there exists a set B such that $f, g \notin B$, and h such that $h \notin C(B \cup \{f\})$ and $h \in C(B \cup \{g\})$. The latter implies that we have neither gPh nor dPh for any $d \in B$. However, $h \notin C(B \cup \{f\})$ implies that there exists a $t \in C(B \cup \{f\})$ such that tPh , and this t can only be f . Thus, fPh . Since $\neg gPh$ we have either hPg and then $fPhPg$ implying fPg , or hIg , and then $fPhIg$. In both cases, fQ^*g .

Given that fQ^*g and $g \in C(A)$, we argue that $f \in C(A)$ has to hold. Indeed, if not, there exists $h \in A$ such that hPf . But in this case we would have $hPfQ^*g$ and therefore hPg , contrary to $g \in C(A)$. Thus f isn't P -dominated by any element in A , and $f \in C(A)$ follows.

Finally, we note that C satisfies Continuity: let there be given an increasing chain $(A_\tau)_{\tau \in \Upsilon}$ with $f \in C(A_\tau)$ for all $\tau \in \Upsilon$. If $f \in C(\cup_{\tau \in \Upsilon} A_\tau)$ does not hold, there exists $g \in \cup_{\tau \in \Upsilon} A_\tau$ that satisfies gPf . But then there exists $\tau \in \Upsilon$ with $g \in A_\tau$ contradicting $f \in C(A_\tau)$. \square

Proof of Proposition 1:

Observe that, when F is finite, A4 (Continuity) isn't needed for the proof of sufficiency in Theorem 1. Hence, A1-A3 imply that C is defined by a semiorder P . We now need only cite a representation result (see Luce, 1956, and Beja and Gilboa, 1992), saying that P is a semiorder on a finite set F iff there exist a function $u : F \rightarrow \mathbb{R}$ and $\Delta \geq 0$ such that, for all f, g ,

$$gPf \iff u(g) - u(f) > \Delta.$$

\square

7.3 Proof of Proposition 2

We begin with the “if” part. Let there be an interval order P on F . Define C by $C(A) = \{f \in A \mid \nexists g \in A, gPf\}$. Observe, first, that $C(A) \neq \emptyset$ for all $A \neq \emptyset$, because P cannot have cycles. To see that it satisfies Sen's α , let $f \in B \subset A$ and $f \in C(A)$. Clearly, there exists no $g \in B \subset A$ such that gPf and thus $f \in C(B)$. Next, consider Union. Assume that $f \in C(A)$ and $g \in C(B)$. We need to show that $f \in C(A \cup B)$ or $g \in C(A \cup B)$. If this is not the case, then $f \notin C(A \cup B)$ implies that there exists $h \in C(A \cup B)$ such that hPf . As $f \in C(A)$ it has to be the case that $h \in B \setminus A$. Similarly, $g \notin C(A \cup B)$ implies that there exists $d \in A \setminus B$ such that dPg . Further, we know that $\neg hPg$ (by optimality of g in B) and (similarly) $\neg dPf$. If gPh then, by transitivity of P , gPf , and with dPg we also get dPf , a contradiction. Hence $\neg hPg$ and $\neg gPh$, that is gIh . However, $dPgIhPf$ implies dPf , again,

a contradiction. Finally, C satisfies Continuity as in the case of a semiorder (and, indeed, as in the case of a function C that is defined by selecting P -undominated elements for any binary relation P).

As for the “only if” part, assume that C satisfies Sen’s α , Union, and Continuity. Define fPg by $g \notin C(\{f, g\})$. By Lemma 4, because C satisfies Sen’s α , Union, and Continuity,

$$C(A) = \{f \in A \mid \forall g \in A, f \in C(\{f, g\})\}$$

holds for all A . Given that, for $g \in A$, $f \in C(\{f, g\})$ means $\neg gPf$, we have $C(A) = \{f \in A \mid \nexists g \in A, gPf\}$.

We now wish to prove that this relation P is indeed an interval order. The fact that it is irreflexive is obvious from the definition. We thus need to show that $PIP \subset P$. Let there be given f, g, h, d such that $fPgIhPd$ and we need to show that fPd . Assume not. Consider $A = \{f, d\}$ and $B = \{g, h\}$. Because $\neg fPd$, we have $d \in C(A)$. We know that gIh so that $g \in C(B)$. Hence at least one of d or g is in $C(A \cup B) = C(\{f, g, h, d\})$. But $d \in C(\{f, g, h, d\})$ is ruled out by hPd and $g \in C(\{f, g, h, d\})$ – by fPg . \square

7.4 Independence of the Axioms

We wish to verify that axioms A1-A4 are independent. We first note that the first two are not implied by the others.

Remark 1 A1 does not follow from A2, A3, and A4.

Proof: Let $F = \{f, g, h\}$ and define C by $C(\{f, g\}) = C(\{f, h\}) = \{f\}$; $C(\{g, h\}) = \{g\}$; $C(\{f, g, h\}) = \{f, g, h\}$. To see that C satisfies A2, observe that, apart from $C(F)$, C picks the maximizer of a strict order P with $fPgPh$. As for A3, C defines the relation $f \succ g, g \succ h, f \succ h$ (and no other pairs are in \succ). The function C clearly satisfies Monotonicity with respect to this \succ . However, A1 doesn’t hold as $C(F)$ consists of all elements, including those that are excluded from $C(A)$ for subsets $A \subset F$. \square

Remark 2 A2 does not follow from A1, A3, and A4.

Proof: Let $F = \{f, g, h\}$ and define C by $C(A) = C(A)$ for all $A \subsetneq F$; $C(\{f, g, h\}) = \{f\}$. A1 is satisfied, because, for all $B \subset F$ with $|B| \leq 2$, we have $C(B) = B$. As for A3, we first have to identify the \succ relation. We have $f \succ g$ and $f \succ h$ because $C(\{f, g, h\}) = \{f\}$ (and clause (i) of the definition of \succ). There are no other \succ rankings (again, because no element is excluded from $C(B)$ for $|B| \leq 2$), and Monotonicity is satisfied. Yet, Union isn't satisfied as can be verified by considering $g \in C(\{f, g\})$ and $h \in C(\{h\})$. \square

The case of a choice correspondence that satisfies A1, A2, and A4 but not A3 is characterized in Proposition 2.

Finally, we note that

Remark 3 A4 does not follow from A1, A2, and A3.

Proof: Let $F = [0, 2)$, and define C by

$$C(A) = \left\{ f \in A \mid f > \sup_{g \in A} g - 1 \right\}$$

Thus, for finite sets A , $C(A)$ is the P -undominated elements according to the semiorder gPf iff $g - f \geq 1$. It can be easily seen that C satisfies A1-A3. However, A4 fails: define $A_\tau = [0, \tau)$ for $\tau \in [0, 2)$. Note that $1 \in C(A_\tau)$ but $1 \notin C(\cup_{\tau < 2} A_\tau)$. \square

7.5 The Definition of \succ

The definition we use is

Definition: For $f, g \in F$, $f \succ g$ if one of the following holds:

- (i) there exists a set A such that $f, g \in A$, $f \in C(A)$ and $g \notin C(A)$;
- (ii) there exists a set B such that $f, g \notin B$, and $h \in B$ such that $h \notin C(B \cup \{f\})$ and $h \in C(B \cup \{g\})$.

It is worthwhile to note that both clauses of the definition are needed to capture the intended meaning of the relation. Specifically, if we have a representation of C by (u, Δ) , and we wish to have $f \succ g$ whenever $u(f) > u(g)$, both clauses are needed.

To see that (i) is needed, assume that there is a strict transitive order P such that C selects only its (unique) maximizer. Specifically, we may have $F = \{f, g, h\}$ with $C(\{f, g, h\}) = C(\{f, g\}) = C(\{f, h\}) = \{f\}$; and $C(\{g, h\}) = \{g\}$. This would correspond to the representation $u(f) = 4; u(g) = 2; u(h) = 0$ and $\Delta = 0$. If we use only clause (ii) of the definition, we have only $f \succ h$. More generally, if there is a strict order on F and C picks its unique maximizer (in each set A), clause (ii) would fail to capture the ranking between any two consecutive alternatives.

Conversely, if we only use clause (i), we may have $F = \{f, g, h\}$ with $C(\{f, g, h\}) = C(\{f, g\}) = \{f, g\}$, $C(\{g, h\}) = \{g, h\}$ and $C(\{f, h\}) = \{f\}$. Clause (i) would capture $f \succ h$ and $g \succ h$ but not $f \succ g$. This would correspond to the representation $u(f) = 4; u(g) = 2; u(h) = 0$ and $\Delta = 3$. Thus, the presence of h reveals that f is more highly ranked than g , but clause (i) of the definition fails to capture that.

8 Database-Related Analysis

8.1 Proof of Proposition 3

Given Lemma 1, we have a representation of \succsim_D by $U_D(\cdot) \equiv \sum_{x \in X} D(x)v(\cdot, x)$. The representation (Gilboa and Schmeidler, 2003) allows shifting every ‘‘column’’, $(v(\cdot, x))$, by an additive constant. Given axiom P5, we know that, for each x , there exists $g \in F$ such that $v(g, x) = \max_{f \in F} v(f, x)$. This means that we can add $-v(g, x)$ to all the values $v(\cdot, x)$ and obtain a representation in which $v(f, x) \leq 0$ for all x, f . We then define

$$p(x|f) = \exp(v(f, x)) \in (0, 1]$$

and the ranking given by $\sum_{x \in X} D(x)v(\cdot, x)$ is equivalent to that given by

$$L(f|D) = \prod_{\{x \in X \mid D(x) > 0\}} [p(x|f)]^{D(x)}.$$

□

8.2 Proof of Theorem 2

We use the matrix $(v(f, x))_{f \in F, x \in X}$ obtained in the proof of Proposition 3, so that $\max_f v(f, x) = 0$ for all x and \lesssim_D is represented by $U_D(\cdot) \equiv \sum_{x \in X} D(x)v(\cdot, x)$ for all D . We wish to show that, under the additional axioms P6-P8, there exists $\Delta \geq 0$ such that, for every $D \in \mathbb{D}$, and every $f, g \in F$,

$$\begin{aligned} U_D(f) - U_D(g) > \Delta &\Rightarrow f P_D g \\ U_D(f) - U_D(g) < \Delta &\Rightarrow \neg f P_D g \end{aligned} \tag{5}$$

When this is the case, we say that (U_D, Δ) is a *pseudo-representation* of P_D . Observe that, in this case, for every $D \in \mathbb{D}$, and every $A \subset F$, because

$$C_D(A) = \{f \in A \mid \nexists g \in A, \quad g P_D f\}$$

we will have

$$\begin{aligned} U_D(f) > \sup_{g \in A} U_D(g) - \Delta &\Rightarrow \\ \nexists g \in A, \quad g P_D f &\Rightarrow f \in C_D(A) \\ U_D(f) < \sup_{g \in A} U_D(g) - \Delta &\Rightarrow \\ \exists g \in A, \quad g P_D f &\Rightarrow f \notin C_D(A) \end{aligned}$$

or, equivalently, for $\gamma = \exp(-\Delta) \leq 1$,

$$\begin{aligned} L(f|D) > \gamma \sup_{g \in A} L(g|D) &\Rightarrow f \in C_D(A) \\ L(f|D) < \gamma \sup_{g \in A} L(g|D) &\Rightarrow f \notin C_D(A) \end{aligned}$$

To find such a $\Delta \geq 0$ for which (U_D, Δ) is a pseudo-representation of P_D (for all D), we proceed in three steps: first, we fix two theories, f, g , and a finite support $X_0 \subset X$, and show that we can find $\Delta \geq 0$ such that (5) holds for all D with support in X_0 . The main assumption that yields this result is P6, which guarantees that we can linearly separate the databases D for which $f P_D g$ from the other databases (for which $\neg f P_D g$). We then show that the same $\Delta \geq 0$ can serve as the threshold for all finite supports. Axiom P8 will guarantee that this Δ is unique. Finally, we show that this Δ is the same for all pairs of theories, employing P8 again.

Lemma 12 Fix $f, g \in F$ and a finite subset $X_0 \subset X$ such that, for some $x, y \in X_0$, $f \succ_{D_x} g$ and $g \succ_{D_y} f$. There exists $\Delta \geq 0$ such that, for all $D \in \mathbb{D}$ with $\text{supp}D \subset X_0$, (5) holds.

Proof: Consider

$$P = \{ D \in \mathbb{D} \mid f P_D g \}$$

Let $v_x = v(f, x) - v(g, x)$ (for all $x \in X$) so that $f \succsim_D g$ iff $vD \geq 0$ (where $vD = \sum_x v_x D(x)$). We claim that, for all $D \in P$ and $D' \notin P$, we have $vD \geq vD'$. To see this, first prove:

Claim 1 For every natural $M (> 0)$ and every $\varepsilon > 0$ there exists D^+ such that $D^+(x) > M$ for all $x \in X_0$ and $0 \leq vD^+ < \varepsilon$. Similarly, there exists D^- such that $D^-(x) > M$ for all $x \in X_0$ and $-\varepsilon < vD^- \leq 0$.

Proof: By our assumption on X_0 , there exist x, y such that $v_x > 0 > v_y$. We argue that, for each such pair x, y there is $D_{x,y}^+$ with $\text{supp}D_{x,y}^+ = \{x, y\}$ instead of $\text{supp}D_{x,y}(x) = \{x, y\}$ and $0 \leq vD_{x,y}^+ < \frac{\varepsilon}{M|X_0|^2}$. Indeed, if $v_x/v_y \in \mathbb{Q}$, then there is such $D_{x,y}^+$ with $vD_{x,y}^+ = 0$. If, however, $v_x/v_y \notin \mathbb{Q}$, the set $\{kv_x - lv_y \mid k, l \geq 0\}$ is dense in \mathbb{R} , so that we can find $D_{x,y}^+$ with $0 < vD_{x,y}^+ < \frac{\varepsilon}{M|X_0|^2}$. We can then define

$$D^+ = M \sum_{\{x,y \in X_0 \mid v_x > 0 > v_y\}} D_{x,y}^+$$

and observe that $D^+(x) > M$ for all $x \in X_0$ and $0 \leq vD^+ < \varepsilon$. The proof for D^- is symmetric. \square

We prove the lemma by negation. Assume that $D \in P$ and $D' \notin P$ satisfy $vD < vD'$. Consider $M > D(x), D'(x)$ for all $x \in X_0$ and $\varepsilon = (vD' - vD)/3$. Let D^- be the database provided by the claim. Consider $D' + D^- \in \mathbb{D}$. Because $vD^- \leq 0$ we have $g \succsim_{D^-} f$ and, with $D' \notin P$, by P6, $D' + D^- \notin P$. Next, consider $\tilde{D} = D' + D^- - D$. Notice that $D^- - D \geq 0$ so that $\tilde{D} \in \mathbb{D}$ (that is, $\tilde{D}(x) \geq 0$ for all $x \in X_0$). We have $v\tilde{D} = vD' + vD^- - vD > vD^- + 3\varepsilon > 2\varepsilon > 0$. Hence $f \succsim_{\tilde{D}} g$. Given that $D \in P$, P6 implies $D + \tilde{D} \in P$. However, $D + \tilde{D} = D' + D^-$ which was shown not to be in P – a contradiction.

To complete the proof of the lemma, observe that $0 \in P^c$. On the other hand, P7 implies that $P \neq \emptyset$. Thus, neither P nor P^c is empty. Define

$$\begin{aligned}\Delta^+ &= \inf_{D \in P} [U_D(f) - U_D(g)] \\ \Delta^- &= \sup_{D \notin P} [U_D(f) - U_D(g)]\end{aligned}$$

we have shown that $\Delta^+ \geq \Delta^-$ where $\Delta^- \geq 0$ follows from $0 \in P^c$. In case $\Delta^+ = \Delta^-$, set $\Delta = \Delta^+ = \Delta^-$, and it is the unique value satisfying (5). Otherwise, any value $\Delta \in [\Delta^-, \Delta^+]$ would do. \square

Lemma 13 Fix $f, g \in F$. There exists a unique $\Delta \geq 0$ such that, for all $D \in \mathbb{D}$ (5) holds.

Proof: Given P8, there exists $X_0 = \{x, y\}$ such that, $v_x \equiv v(f, x) - v(g, x) > 0 > v(f, y) - v(g, y) \equiv v_y$ and $v_x/v_y \notin \mathbb{Q}$. Apply Lemma 12 to X_0 . Observe that

$$\begin{aligned}\Delta^+ &= \inf_{D \in P} [D(x)v_x - D(y)v_y] \\ \Delta^- &= \sup_{D \notin P} [D(x)v_x - D(y)v_y]\end{aligned}$$

However, because $v_x/v_y \notin \mathbb{Q}$, $\{D(x)v_x - D(y)v_y\}_D$ is dense in \mathbb{R} . Hence $\Delta = \Delta^+$ and Δ that satisfies (5) for $X_0 = \{x, y\}$ is unique.

Consider any other finite $X_1 \subset X$ and apply Lemma 12 to $X' = X_0 \cup X_1$. There exists a $\Delta \geq 0$ that satisfies (5) for X' and it has to be the same Δ as found above. Finally, given an arbitrary $D \in \mathbb{D}$, we define X_1 to be its support and obtain the result. \square

Lemma 14 There exists $\Delta \geq 0$ such that (U_D, Δ) is a pseudo-representation of P_D for all $D \in \mathbb{D}$.

Proof: Lemma 13 shows that for every f, g there exists a unique $\Delta_{fg} \geq 0$ such that (U_D, Δ_{fg}) satisfies (5) for (f, g) . We will first show that for all f, g, h we have $\Delta_{fg} = \Delta_{fh}$.

Let there be given distinct f, g, h . Assume that $\Delta_{fg} = \Delta_{fh}$ does not hold. Without loss of generality, suppose that $\Delta_{fg} > \Delta_{fh}$. We wish to find a

database D such that

$$\begin{aligned} (I) \quad U_D(f) - U_D(g) &\geq U_D(f) - U_D(h) \\ (II) \quad U_D(f) - U_D(h) &> \Delta_{fh} \\ (III) \quad U_D(f) - U_D(g) &< \Delta_{fg} \end{aligned}$$

If D satisfies these three inequalities, (I) will imply that $U_D(g) \leq U_D(h)$, that is, $h \succsim_D g$; (II) – that $f P_D h$; but (III) will imply that $\neg f P_D g$, and this will be a contradiction to the monotonicity of P_D with respect to \succsim_D .

To see that such D indeed exists, assume, w.l.o.g., that $U_D(f) = 0$, and consider \mathbb{R}^2 with a generic point (ξ, η) interpreted as $(U_D(g), U_D(h))$. Consider the set

$$\{ (\xi, \eta) \in \mathbb{R}^2 \mid \xi \leq \eta, \eta < -\Delta_{fh}, \xi > -\Delta_{fg} \}$$

i.e., the half-open triangle that is bounded (from below and to the right) by the main diagonal (including the edge) $\xi \leq \eta$, from above by $\eta = -\Delta_{fh}$ and from the left by $\xi = -\Delta_{fg}$. However, P8 guarantees that $(U_D(g), U_D(h))_{D \in \mathbb{D}}$ is dense in \mathbb{R}^2 .¹⁷ Thus we can find $D \in \mathbb{D}$ that yields $(U_D(g), U_D(h))$ in this set, contradicting the assumption $\Delta_{fg} > \Delta_{fh}$.

Next we wish to show by similar reasoning that, for all f, g, h , $\Delta_{fg} = \Delta_{hg}$. If not, say $\Delta_{fg} > \Delta_{hg}$, we can find a D with

$$\begin{aligned} (I) \quad U_D(f) - U_D(g) &\geq U_D(h) - U_D(g) \\ (II) \quad U_D(h) - U_D(g) &> \Delta_{hg} \\ (III) \quad U_D(f) - U_D(g) &< \Delta_{fg} \end{aligned}$$

¹⁷To see this, observe first that, by P8, $(U_D(g), U_D(h))_{D \in \mathbb{D}}$ includes six vectors $(\xi_i, \eta_i)_{i \leq 6}$, one in each of the cones defined by $\xi = 0, \eta = 0, \xi = \eta$ (that is, the second and fourth orthants, and the four half-orthants obtained by intersection of the first and the third with the main diagonal). Consider a circle C around the origin whose radius is large enough that, starting from any point in the plane, one can find a sequence of "steps" chosen from $(\xi_i, \eta_i)_{i \leq 6}$ that enters C . Next consider a sequence of increasing databases D_t ($D_{t+1} \geq D_t$) that visits C infinitely many times. Since P8 precludes the sequence from visiting the same point twice, $(U_{D_t}(g), U_{D_t}(h))_{t \geq 1}$ has an accumulation point in C . Take a subsequence of $(D_t)_t$ whose $(U_D(g), U_D(h))$ values converge to a limit. Considering pairs of points in this sequence, we obtain databases D (differences between such points) whose $(U_D(g), U_D(h))$ has an arbitrarily small norm. Moreover, the sequence can be so chosen that the vectors $(U_D(g), U_D(h))$ are not collinear. Summations of these vectors enter every open neighborhood.

and thus $f \succsim_D h$; $h P_D g$; but $\neg f P_D g$ – again, a contradiction.

As a result, for any f, g, f', g' we have $\Delta_{fg} = \Delta_{fg'} = \Delta_{f'g'}$. And thus there is a single $\Delta \geq 0$ such that (U_D, Δ) satisfies (5) for all (f, g) . \square

8.3 Proof of Theorem 3

In light of the analysis above, it suffices to show that, under P10, P11, the functions $v(\cdot, x) : F \rightarrow \mathbb{R}$ of Lemma 1 are continuous for every x . Assume, then, that, for each D , \succsim_D is continuous. We wish to show that so is $v(\cdot, x)$ for each $x \in X$. By Debreu’s (1954) result, we know that, for every x , there is a continuous function $u_x : F \rightarrow \mathbb{R}$ that represents $\succsim_{\mathbf{1}_{\{x\}}}$. We know that $U_D(\cdot) = v(\cdot, x)$ also represents $\succsim_{\mathbf{1}_{\{x\}}}$, so that $v(\cdot, x)$ has to be ordinally equivalent to u_x . Specifically, there exists a strictly increasing $\varphi_x : \mathbb{R} \rightarrow \mathbb{R}$ such that $v(\cdot, x) = \varphi_x(u_x(\cdot))$. If φ_x is continuous, $v(\cdot, x)$ is clearly also continuous. We wish to show that this is indeed the case for every x . If not, for a given $x \in X$, φ_x has a “jump”: for some $a_0 \in \mathbb{R}$,

$$\varphi^- \equiv \limsup_{a < a_0} \varphi_x(a) < \liminf_{a > a_0} \varphi_x(a) \equiv \varphi^+$$

Choose f^0 with $u_x(f^0) = a_0$. Let y be an observation for which the Richness Assumption holds. Consider $E_{\mathbf{1}_{\{x\}}}(f^0)$. Choose a connected path in $E_{\mathbf{1}_{\{x\}}}(f^0)$ connecting g' and g with $g \succ_{\mathbf{1}_{\{y\}}} g'$. Thus, both $u_y(\cdot)$ and $v(\cdot, y) = \varphi_y(u_y(\cdot))$ are increasing on the path. The function φ_y may have only countably many discontinuities between $u_y(g')$ and $u_y(g)$. Hence, there exists f^* such that $u_x(f^*) = a_0$ is a discontinuity point of φ_x but a continuity point of φ_y . We wish to show that this leads to a contradiction.

Choose two distributions $f_1^- = h, f_1^+ = g$ where g, h are as in part (ii) of the Richness Assumption. Thus, $v(f_1^-, x) < a_0 < v(f_1^+, x)$ and $v(f_1^-, y) > v(f^*, y) > v(f_1^+, y)$. Consider a path from f_1^- to f_1^+ . By continuity of $\succsim_{\mathbf{1}_{\{x\}}}$, we can find two sequences of distributions, $(f_n^-)_{n \geq 1}, (f_m^+)_{m \geq 1}$, and a (limit) distribution f^* such that

$$\begin{aligned} f_n^- &\rightarrow_{n \rightarrow \infty} f^*, & f_m^+ &\rightarrow_{m \rightarrow \infty} f^* \\ v(f_n^-, x) &\nearrow_{n \rightarrow \infty} \varphi^-, & v(f_m^+, x) &\searrow_{m \rightarrow \infty} \varphi^+ \end{aligned}$$

In other words, we pick a point f^* such that $u_x(f^*) = a_0$ is a discontinuity point of φ_x , and pick two sequences of distributions that converge to f^* in the space F , and whose $v(\cdot, x)$ values converge monotonically, from below and from above respectively, to the supremum below and infimum above, φ^- and φ^+ . Note that this also implies that $f_{n+1}^- \succ_{\mathbf{1}_{\{x\}}} f_n^-$ for all n , and $f_m^+ \succ_{\mathbf{1}_{\{x\}}} f_{m+1}^+$ for all m .

By similar arguments, because $\succ_{\mathbf{1}_{\{y\}}}$ is also continuous, we can select subsequences of $(f_n^-)_{n \geq 1}$ and of $(f_m^+)_{m \geq 1}$, $(n_\nu)_{\nu \geq 1}$, $(m_\mu)_{\mu \geq 1}$ so that $f_{n_\nu}^- \succ_{\mathbf{1}_{\{y\}}} f_{n_{\nu+1}}^-$ and $f_{m_\mu}^+ \succ_{\mathbf{1}_{\{y\}}} f_{m_{\mu+1}}^+$ for all ν, μ . Wlog we drop the sub-indices and assume that

$$\begin{aligned} f_n^- &\succ_{\mathbf{1}_{\{y\}}} f_{n+1}^- \quad \forall n \\ f_{m+1}^+ &\succ_{\mathbf{1}_{\{y\}}} f_m^+ \quad \forall m \\ f_n^- &\succ_{\mathbf{1}_{\{y\}}} f_m^+ \quad \forall n, m \end{aligned}$$

Consider $v(\cdot, y)$. We know that it represents $\succ_{\mathbf{1}_{\{y\}}}$ so that $v(f_n^-, y) > v(f_{n+1}^-, y) > v(f_{m+1}^+, y) > v(f_m^+, y)$ for all n, m . Furthermore, as we picked a continuity point of φ_y , we have

$$\sup v(f_m^+, y) = \psi = \inf v(f_n^-, y)$$

Find a large enough $l \geq 1$ so that $f_1^- \succ_{\mathbf{1}_{\{x\}} + l\mathbf{1}_{\{y\}}} f_1^+$ (by the Archimedean axiom, such an l exists). Using the fact that $\succ_{\mathbf{1}_{\{x\}} + l\mathbf{1}_{\{y\}}}$ is also continuous, we can select subsequences that are decreasing according to $\succ_{\mathbf{1}_{\{x\}} + l\mathbf{1}_{\{y\}}}$ as well. Wlog we therefore assume that

$$\begin{aligned} f_n^- &\succ_{\mathbf{1}_{\{x\}} + l\mathbf{1}_{\{y\}}} f_{n+1}^- \quad \forall n \\ f_{m+1}^+ &\succ_{\mathbf{1}_{\{x\}} + l\mathbf{1}_{\{y\}}} f_m^+ \quad \forall m \\ f_n^- &\succ_{\mathbf{1}_{\{x\}} + l\mathbf{1}_{\{y\}}} f_m^+ \quad \forall n, m \end{aligned}$$

For large enough n, m , we have

$$0 < v(f_n^-, y) - v(f_m^+, y) < \frac{\varphi^+ - \varphi^-}{l}$$

and then

$$v(f_n^-, x) + lv(f_n^-, y) < v(f_m^+, x) + lv(f_m^+, y)$$

so that $v(\cdot, x) + lv(\cdot, y)$ fails to represent $\succ_{\mathbf{1}_{\{x\}} + l\mathbf{1}_{\{y\}}}$. \square

Bibliography

Agaev, Rafiq and Fuad Aleskerov (1993), “Interval Choice: Classic and General Cases”, *Mathematical Social Sciences*, **26**: 249-272.

Aguilera, Juan P. and Levent Ulku (2017), “On the Maximization of Menu-Dependent Interval Orders”, *Social Choice and Welfare*, **48**: 357-366.

Akaike, Hirotugu (1974), “A New Look at the Statistical Model Identification”, *IEEE Transactions on Automatic Control*, **19**: 716-723.

Aleskerov, Fuad, Denis Bouyssou, and Bernard Monjardet (2007), *Utility Maximization, Choice and Preference*. Second edition. Springer, Berlin.

Beja, Avraham and Itzhak Gilboa (1992), “Numerical Representations of Imperfectly Ordered Preferences (A Unified Geometric Exposition)”, *Journal of Mathematical Psychology*, **36**: 426-449.

Berger, Loïc, Nicolas Berger, Valentina Bosetti, Itzhak Gilboa, Lars Peter Hansen, Christopher Jarvis, and Richard D. Smith (2021), “Rational Policy-making during a Pandemic”, *Proceedings of the National Academy of Sciences*, **118**: 1-7.

Berger, Loïc, Johannes Emmerling, and Massimo Tavoni (2016), “Managing Catastrophic Climate Risks Under Model Uncertainty Aversion”, *Management Science*, **63**: 749-765.

Berger, Loïc and Massimo Marinacci (2020), “Model Uncertainty in Climate Change Economics: A Review and Proposed Framework for Future Research”, *Environmental and Resource Economics*, **77**: 475-501.

Debreu, Gerard (1954). “Representation of a Preference Ordering by a Numerical Function”. Reprinted in *Mathematical economics: Twenty papers of Gerard Debreu*; Introduction by Werner Hildenbrand (1st pbk. ed.). Cambridge Cambridge University Press.

Epstein, Larry G. and Martin Schneider (2007), “Learning Under Ambiguity”, *Review of Economic Studies*, **74**: 1275-1303.

- Fechner, Gustav Theodor (1860), *Elemente der Psychophysik*. Breitkopf und Hartel, Leipzig.
- Fishburn, Peter C. (1970), “Intransitive Indifference with Unequal Indifference Intervals”, *Journal of Mathematical Psychology*, **7**: 144-149.
- Fishburn, Peter C. (1975), “Semiorders and Choice Functions”, *Econometrica*, **43**: 975-977.
- Fishburn, Peter C. (1985), *Interval Orders and Interval Graphs, a Study of Partially Ordered Sets*. Wiley, New York.
- Frick, Mira (2016), “Monotone Threshold Representations”, *Theoretical Economics*, **11**: 757-772. doi: 10.3982/TE1547
- Gilboa, Itzhak and David Schmeidler (2003), “Inductive Inference: An Axiomatic Approach”, *Econometrica*, **71**: 1-26.
- Gilboa, Itzhak and David Schmeidler (2010), “Likelihood and Simplicity: An Axiomatic Approach”, *Journal of Economic Theory*, **145**: 1757-1775.
- Heal, Geoffrey and Antony Millner (2014), “Uncertainty and Decision Making in Climate Change Economics”, *Review of Environmental Economics and Policy*, **8**: 120-137. doi: 10.1093/reep/ret023
- Jamison, Dean T. and Lawrence J. Lau (1973), “Semiorder and Theory of Choice”, *Econometrica*, **41**: 901-912.
- Luce, R. Duncan (1956), “Semiorders and a Theory of Utility Discrimination”, *Econometrica*, **24**: 178-191. <https://doi.org/10.2307/1905751>
- Manski, Charles F. (2018), “Communicating uncertainty in policy analysis”, *Proceedings of the National Academy of Sciences*, **116**: 7634-7641.
- Manski, Charles F. (2019), *Patient Care under Uncertainty*. Princeton University Press.
- Manzini, Paola and Marco Mariotti (2012), “Choice by lexicographic semiorders.” *Theoretical Economics*, **7**: 1-23.
- Millner, Antony, Simon Dietz, and Geoffrey Heal (2013), “Scientific Ambiguity and Climate Policy” *Environmental and Resource Economics* , **55**: 21-46.

- Pindyck, Robert (2007), “Uncertainty in environmental economics”, *Review of Environmental Economics and Policy*, **1**: 45-65.
- Savage, Leonard J. (1954), *The Foundations of Statistics*. New York: John Wiley and Sons. (Second addition in 1972, Dover)
- Schwarz, Gideon E. (1978), “Estimating the Dimension of a Model”, *Annals of Statistics*, **6**: 461-464. doi: 10.1214/aos/1176344136
- Schwarz, Thomas (1976), “Choice Functions, ‘Rationality’ Conditions, and Variations on the Weak Axiom of Revealed Preferences”, *Journal of Economic Theory*, **13**: 414-427. [https://doi.org/10.1016/0022-0531\(76\)90050-8](https://doi.org/10.1016/0022-0531(76)90050-8)
- van Rooij, Robert (2008), “Revealed Preference and Satisficing Behavior”, manuscript.
- Weber, Ernst Heinrich (1834), *De Tactu*.