# Likelihood Regions: An Axiomatic Approach* 

Fan Wang ${ }^{(r)}$, Itzhak Gilboa ${ }^{\top}$, and Stefania Minardi ${ }^{\S}$

June 28, 2023


#### 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 functions (across different databases) for which there exists a statistical model such that the choice function 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 $\mathrm{CO}_{2}$ emission. Scientists may agree that emission contributes to global warming, but have different estimates of the temperature distribution 30 years hence, given different levels of emission. ${ }^{1}$ Or, facing a new pandemic, the government has

[^0]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. 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.

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 $L(\cdot \mid D)$, for any two distributions $f, g$ in $F$, if $L(f \mid D) \geq L(g \mid D)$, 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

$$
\begin{equation*}
C_{D}(F)=\left\{f \in F \mid L(f \mid D) \geq \gamma \sup _{g \in F} L(g \mid D)\right\} \tag{1}
\end{equation*}
$$

or

$$
\begin{equation*}
C_{D}(F)=\left\{f \in F \mid L(f \mid D)>\gamma \sup _{g \in F} L(g \mid D)\right\} \tag{2}
\end{equation*}
$$

In the case of a finite set of distributions, the threshold $\gamma$ 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 $D .{ }^{2}$

The choice of the threshold $\gamma$ reflects the reasoner's degree of prudence, where a more conservative approach would be associated with a lower $\gamma$. When $\gamma$ is very low, the reasoner hardly learns anything from the data, considering as relevant many possible distributions. When $\gamma$ is close to 1 , the reasoner tends to dismiss theories quite easily. The choice of $\gamma$ 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 $\gamma$, 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).

In this paper we axiomatize choice functions 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 function $C$. 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 model in mind, and she picked likelihood regions relative to this statistical model.

More specifically, we assume that, at each period, 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

[^1]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 $p(x \mid f) \in[0,1]$, 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 $L(\cdot \mid D)$.

Monotonicity of $C_{D}(A)$ with respect to a likelihood function $L(\cdot \mid D)$ allows the threshold $\gamma$ 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, our axioms yield a quasi-representation of the choice functions 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 $f \in A$

$$
\begin{aligned}
& L(f \mid D)>\gamma \sup _{g \in A} L(g \mid D) \Rightarrow f \in C_{D}(A) \\
& L(f \mid D)<\gamma \sup _{g \in A} L(g \mid D) \Rightarrow f \notin C_{D}(A)
\end{aligned}
$$

Observe that a quasi-representation does not specify the status of theories whose likelihood value is precisely at $\gamma$ sup.

Our results can be interpreted both descriptively and normatively. A descriptive ${ }^{3}$ 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. ${ }^{4}$ Adopting a normative interpretation, our results

[^2]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 presumablyobservable 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 would apply to an economist who models the reasoner: the axiomatization might make it more reasonable to assume that reasoners might make theory selection as if they had a statistical model in mind. ${ }^{5}$

The rest of this paper is organized as follows. Section 2 is devoted to choicetheoretic foundations, where we fix the database $D$ and impose conditions on choice functions that guarantee that they be monotonic with respect to some order. Our conditions provide a new characterization of choice functions 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
specified in the formal statement of the main theorem.
${ }^{5}$ The situation is analogous to axiomatizations of subjective probability, which are not supposed to suggest that objective probabilities be ignored.
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. Section 5 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 functions that are monotonic with respect to a function $L$, that is, functions $C(A)$ for which, for every $A$ 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 $\gamma$ 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 functions that can be represented by a function $u: F \rightarrow \mathbb{R}$ and a constant $\Delta \geq 0$ such that

$$
\begin{equation*}
C(A)=\left\{f \in A \mid u(f) \geq(>) \max _{g \in A} u(g)-\Delta\right\} \tag{3}
\end{equation*}
$$

(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 functions 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 $\varepsilon$-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 $f, g$,

$$
\begin{equation*}
g P f \quad \Leftrightarrow \quad u(g)-u(f)>\Delta \tag{4}
\end{equation*}
$$

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

There is a rather rich literature on choice functions 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. ${ }^{6}$ In their setup a binary relation $P$ was assumed, and the choice function $C(A)$ was defined as the set of $P$-undominated elements from each subset $A$. Agaev 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 " $\mathcal{P O}$ " condition, which guarantees that the choice function 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 functions that can be represented as the selection of lexicographicallyundominated elements according to some sequence of semiorders (but the sequence may well be longer than 1). Frick (2016) axiomatizes choice functions that have monotone threshold representations, that is, representations in which the threshold $(\Delta)$ can depend on the set $(A)$. Importantly, she also provides a characterization for the case of a fixed $\Delta$ (Lemma 1(i), p. 765).

However, in order to relate the choice functions 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

[^3]set of axioms to be more intuitive (though this is clearly a matter of taste). We thus provide here a different characterization of choice functions that can be represented as $P$-undominated elements for a semiorder $P$. A by-product of our axiomatization is the numerical representation of these choice functions - 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 functions, 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 functions that select "top-tier elements" of each set $A$. Consider the standard axioms,

A1: Sen's $\alpha$ : If $f \in B \subset A$ and $f \in C(A)$, then $f \in C(B)$.
Sen's $\beta$ : If $f, g \in C(A), A \subset B$, and $g \in C(B)$, then $f \in C(B)$.
Sen's $\alpha$ 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 $\beta$ 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 $(B \backslash A)$, which may be better then 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 functions considered here need not satisfy Sen's $\beta$ and thus not the IIA in general. Specifically, choosing options that are close enough to the maximizers makes the choice procedure contextdependent. We note that, under Sen's $\alpha$, 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 $\beta$ 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 highlyranked 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 $\alpha$, Sen's $\beta$ implies Union.

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
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 function $C$ 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 $\left(A_{\tau}\right)_{\tau \in \Upsilon}$ is an increasing chain $\left(A_{\tau} \subset A_{\tau^{\prime}} \subset\right.$ $F$ for $\tau<\tau^{\prime}$ where $<$ is a linear order on $\left.\Upsilon\right)$. If $f \in C\left(A_{\tau}\right)$ for all $\tau \in \Upsilon$, then $f \in C\left(\cup_{\tau \in \Upsilon} A_{\tau}\right)$.

Axiom A4 states that an element $f$ that is sufficiently highly ranked to be included in the choice from any set $A_{\tau}$ 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 $\tau$ for which $f$ will already be ruled out when only $A_{\tau}$ is considered. Equipped with this axiom we can formulate the next result:

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

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

For the sake of completeness we also state a similar result for interval orders (Fishburn 1970, 1985):

Proposition 2 A choice function $C$ satisfies Sen's $\alpha$, Union, and Continuity iff there exists an interval order $P$ on $F$ such that, for all $A \subset F$,

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

Interval orders are not sufficiently structured to obtain the result we seek, because they do not induce a well-defined weak order. ${ }^{7}$ 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

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

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

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 $D, D^{\prime} \in \mathbb{D}$ and $k \geq 0, D+D^{\prime} \in \mathbb{D}$, and $k D \in \mathbb{D}$ are well-defined. Similarly, the inequality $D \geq D^{\prime}$ is read pointwise. For $D \in \mathbb{D}$ we define the support to be $\operatorname{supp} D=$ $\{x \in X \mid D(x)>0\}$.

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

[^4]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. ${ }^{8}$ 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 $D$. If $A$ is nonempty and finite, we assume that $C(A)$ is also nonempty. ${ }^{9}$

We impose the following assumptions:
P1 Choice: For every $D \in \mathbb{D}, C_{D}$ on $F$ satisfies Sen's $\alpha$, 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.

P2 Combination: For every $D, D^{\prime} \in \mathbb{D}$ and every $f, g \in F, f \succsim_{D} g$ $\left(f \succ_{D} g\right)$ and $f \succsim_{D^{\prime}} g$ imply $f \succsim_{D+D^{\prime}} g\left(f \succ_{D+D^{\prime}} g\right)$.

The Combination axiom deals with the union of databases, and it states

[^5]that, if each of two databases, $D$ and $D^{\prime}$, indicates that distribution $f$ is at least as likely as $g$, the union of the two cannot suggest the reversed ranking. Observe that, $D, D^{\prime}$ being counter vectors, the union of the databases is captured by the addition of these vectors $\left(D+D^{\prime}\right)$, corresponding to the union of disjoint sets of observations. The axiom is stated for weak orders $\left(f \succsim_{D} g\right.$ and $f \succsim_{D^{\prime}} g$ ) but it is strengthened to require that, if one of them is strict $\left(f \succ_{D} g\right)$, the same applies to the unified database $D+D^{\prime}$.

The next axiom is Archimedean in nature:
P3 Archimedeanity: For every $D, D^{\prime} \in \mathbb{D}$ and every $f, g \subset F$, if $f \succ_{D} g$, then there exists $l \in \mathbb{N}$ such that $f \succ_{l D+D^{\prime}} g$.

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 $D^{\prime}$ that favors $g$ over $f$, sufficiently many repetitions of $D$ will eventually overwhelm the evidence in $D^{\prime}$.

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.

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 functions $\left\{C_{D}\right\}_{D \in \mathbb{D}}$. Then the following two statements are equivalent:
(i) $\left\{C_{D}\right\}_{D \in \mathbb{D}}$ 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 $A$,

$$
C_{D}(A)=\left\{f \in A \mid \nexists g \in A, \quad g P_{D} f\right\}
$$

and $\succsim_{D}$, its associated weak order, is represented by $U_{D}(\cdot) \equiv \sum_{x \in X} D(x) v(\cdot, x) \cdot{ }^{10}$
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 first main result.

Proposition 3 Let there be given $X, F$, and choice functions $\left\{C_{D}\right\}_{D \in \mathbb{D}}$ that satisfy P1-P5. There exist, for each $x \in X$ and $f \in F$, a number $p(x \mid f) \in$ $(0,1]$, 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

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

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

## 4 A Fixed Threshold

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

## P6 Database Monotonicity:

$$
\begin{aligned}
& { }^{10} \text { That is, for every } f, g \in F, f \succsim_{D} g \text { iff } \\
& \qquad U_{D}(f) \geq U_{D}(g) \\
& \text { for } \\
& \qquad U_{D}(f)=\sum_{x \in \mathbb{X}} D(x) v(f, x)
\end{aligned}
$$

(i) For every $f, g \in F$, and every $D, D^{\prime} \in \mathbb{D}$, if $f P_{D} g$ and $f \succsim_{D^{\prime}} g$, then $f P_{\left(D+D^{\prime}\right)} g$.
(ii) For every $f, g \in F$, and every $D, D^{\prime} \in \mathbb{D}$, if $\neg\left(f P_{D} g\right)$ and $g \succsim_{D^{\prime}} f$, then $\neg\left(f P_{\left(D+D^{\prime}\right)} g\right)$.

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 $D^{\prime} f$ 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 $g \succ_{D^{\prime}} f$ were the case, the additional information in $D^{\prime}$ may weaken the evidence in favor of $f$ contained in $D$, and may even reverse the ordering. But the condition $f \succsim_{D^{\prime}} g$ rules out this possibility. Part (ii) of the axiom follows a very similar logic: starting with $\neg\left(f P_{D} g\right)$, 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, $D^{\prime}$, 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 $D^{\prime} \in \mathbb{D}$ there exists $l \in \mathbb{N}$ such that $f P_{l D+D^{\prime}} g$.

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 $D^{\prime}$. 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 $\Delta$.
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 $\left\{\succsim_{D}\right\}_{\{D \mid \operatorname{supp}(D)=\{x, y, z\}\}}$ 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 $\left(\mathbb{R}^{3 k}\right)$.

With these we can now state

Theorem 2 Let there be given $X, F$, and choice functions $\left\{C_{D}\right\}_{D \in \mathbb{D}}$ that satisfy P1-P8. ${ }^{11}$ There exist, for each $x \in X$ and $f \in F$, a number $p(x \mid f) \in$ $(0,1]$, 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

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

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

[^6]$f \in A$
\[

$$
\begin{aligned}
& L(f \mid D)>\gamma \sup _{g \in A} L(g \mid D) \quad \Rightarrow \quad f \in C_{D}(A) \\
& L(f \mid D)<\gamma \sup _{g \in A} L(g \mid D) \Rightarrow f \notin C_{D}(A)
\end{aligned}
$$
\]

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

## 5 Discussion

### 5.1 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 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 function $C_{D}(A)$. 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).

### 5.2 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.

### 5.3 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.

## 6 Appendix: Proofs and Related Analysis

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

### 6.1 Choice Theoretic Analysis

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

### 6.1.1 Some lemmas

Assume first that $C$ is a choice function 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 $\alpha, f \in C(A)$, 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 $f \in C(A) \cap C(B))$.

Note that the proof only uses Sen's $\alpha$.

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 $A=\left\{f, g_{1}, g_{2}, \ldots, g_{k}\right\}$. By induction on $i \leq k$, one observes that $f \in C\left(\left\{f, g_{1}, g_{2}, \ldots, g_{i}\right\}\right)$ using Union and the fact that $f \in C\left(\left\{f, g_{i+1}\right\}\right)$ 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 $(f \in B$ and) $f \in$ $C(B)$, ordered by set inclusion. Continuity guarantees that every increasing chain in this poset, $\left(B_{\tau}\right)_{\tau \in \Upsilon}\left(B_{\tau} \subset B_{\tau^{\prime}} \subset F\right.$ for $\left.\tau<\tau^{\prime}\right)$ has an upper bound, namely $\cup_{\tau \in \Upsilon} B_{\tau}$, with $f \in C\left(\cup_{\tau \in \Upsilon} B_{\tau}\right)$. The lemma implies that there exists a maximal element $B^{*}$ in the poset, so that $f \in C\left(B^{*}\right)$. If $B^{*}=A$, we are done. Otherwise, there exists $g \in A \backslash B^{*}$. We know that $f \in C(\{f, g\})$ and, using Union again, this implies that $f \in C\left(B^{*} \cup\{g\}\right)$ so that $B^{*}$ is not a maximal element. This contradiction implies that $B^{*}=A$ and $f \in C(A)$ as required.

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 $\alpha, f$ cannot be chosen out of the superset, $A$, of $\{f, g\}$.

Note that the proof above only uses Sen's $\alpha$, 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 $\alpha$, that $g \notin C(D)$. Thus $f \in C(D)$ and $f \succ g$ follows.

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 $\alpha$ ), and, similarly, $d \notin C(E \cup\{g\})$ rules out the latter, a contradiction.

Claim: $\succ$ is transitive.
Proof: Let there by 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 $\alpha$ ) $g \in C(A)$, 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 $(g \succ h)[t \notin C(E \cup\{g\})$ 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 $\alpha$ ) 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)).

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 $\alpha$. 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 $B^{\prime}=B \backslash\{g\}$ and consider $B^{\prime} \cup\{f, h\}$. If $d \notin C\left(B^{\prime} \cup\{f, h\}\right)$, then, since $d \in C\left(B^{\prime} \cup\{g, h\}\right)$ $(=C(B \cup\{h\}))$ we get $f \succ g$ (for the set $\left.B^{\prime} \cup\{h\}\right)$. If, however, $d \in$ $C\left(B^{\prime} \cup\{f, h\}\right)$ then, since $d \notin C\left(B^{\prime} \cup\{f, g\}\right)(=C(B \cup\{f\}))$ we get $g \succ h$ (for the set $B^{\prime} \cup\{f\}$ ). Thus, in both cases we obtain a contradiction.

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

Definition 2 For $f, g \in F, f P g$ 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 $f P g$ 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 $f P g$ nor $g P f$.
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 $P I P \subset P$ and $P P I \subset P .{ }^{12}$ We first observe that both $P I$ and $I P$ imply the relation $\succ$.

Lemma 8 : For all $f, g, h$ if $(f P g$ and $g I h)$ or $(f I g$ and $g P h)$, then $f \succ h$.
Proof: If $g=h, f P g$ means $f P h$ and $f \succ h$ follows. Otherwise, $f, g, h$ are distinct $(f \neq g$ because $P$ is irreflexive and $f \neq h$ because $f P g$ but $h I g)$.

Assume first that $f P g$ and $g I h$. 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 $f I g$ and $g P h$. 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 .

We now proceed to prove that $(P, I)$ satisfy Luce's axioms.
Lemma 9 : For all $f, g, h, d \in F$, if $f P g, g I h$, and $h P d$, then $f P d$.
Proof: Assume that $f P g, g I h$, and $h P d$. Because each of the relations $P, I P, P I$ 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 $d P I h$ 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.

[^7]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 $f P g, g P h$, and $h I d$, then $f P d$.
Proof: Assume that $f P g, g P h$, and $h I d$. As above, we have $f \succ g, g \succ h$, $f \succ h$. Further, as $P$ is transitive, we have $f P h$ and thus both $f P I d$ and $g P I d$ 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 $f P d$ (because $f P h$ ), 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.

Next, we wish to show that the semiorder $P$ captures all the relevant information in the choice function $C$. To this end, we remind the reader that, if $P$ is a semiorder, then $Q^{*} \equiv P I \cup I P$ is the strict part of a weak order. That is, $Q^{*}$ is irreflexive, asymmetric, and transitive, where $Q^{0} \equiv\left(Q^{*} \cup Q^{*-1}\right)^{c}$ 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, $Q P \subset P$ and $P Q \subset P$.

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

Lemma $11:\left(Q^{*}, Q^{0}\right)=(\succ, \sim)$.
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 $f Q^{*} g$ (and thus $f Q^{0} g$ 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, $h P g$. 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: $f I h$ (if $C(\{f, h\})=\{f, h\}$ ), in which case $f I P g$ and $f Q^{*} g$, or $f P h($ if $C(\{f, h\})=\{f\})$, which implies $f P g$ and $f Q^{*} g$.

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 $d P h$. In particular, we have $g I h$ or $h P g$. By the same lemma, $h \notin C(B \cup\{f\})$ implies that there exists an element $t$ in $B \cup\{f\}$ such that $t P h$, and since this element isn't in $B$, it has to be $f$. That is, $f P h$. Thus we have $f P I g$ (in case $g I h$ ) or $f P P g$, hence $f P g$ (in case $h P g$ ). In both cases, $f Q^{*} g$ follows.

### 6.1.2 Proof of Theorem 1 and of Proposition 1

## Proof of Theorem 1:

The "only if" has been proved above: we started with a choice function $C$ that satisfied A1-A4, and defined a semiorder $P$ associated with it, namely $f P g$ 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 $g P f$.

To see the converse, assume that $P$ is a semiorder and that $C(A)=$ $\{f \in A \mid \nexists g \in A, \quad g P f\}$ for every $A$. It is immediate that $C$ satisfies Sen's $\alpha$, 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 $h Q d$. We know that $\neg h P f$ and hence $f$ is not $P$-dominated by any element in $B$ (because $t P f$ for some $t \in B$ would imply $d P f$ and then $h P f$ ). 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 $d Q h$.)

We now turn to prove Monotonicity. Assume that $f, g \in A, g \in C(A)$ and
$f \succ g$. We argue that $f Q^{*} g$ (where $Q^{*}=P I \cup I P$ 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 $h P g$ (as $g$ is $P$-dominated in $D$ ) but we do not have $h P f$ (as $f$ is $P$-undominated in $D$ ). Then we have $f P h$ (in which case $f P h P g$ and then $f P g$ ) or $f I h$ (and $f I h P g$ ). In both cases, $f Q^{*} g$.

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 $g P h$ nor $d P h$ 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 $t P h$, and this $t$ can only be $f$. Thus, $f P h$. Since $\neg g P h$ we have either $h P g$ and then $f P h P g$ implying $f P g$, or $h I g$, and then $f P h I g$. In both cases, $f Q^{*} g$.

Given that $f Q^{*} g$ 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 $h P f$. But in this case we would have $h P f Q^{*} g$ and therefore $h P g$, 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 $\left(A_{\tau}\right)_{\tau \in \Upsilon}$ with $f \in C\left(A_{\tau}\right)$ for all $\tau \in \Upsilon$. If $f \in C\left(\cup_{\tau \in \Upsilon} A_{\tau}\right)$ does not hold, there exists $g \in \cup_{\tau \in \Upsilon} A_{\tau}$ that satisfies $g P f$. But then there exists $\tau \in \Upsilon$ with $g \in A_{\tau}$ contradicting $f \in C\left(A_{\tau}\right)$.

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 $f, g$,

$$
g P f \quad \Leftrightarrow \quad u(g)-u(f)>\Delta
$$

### 6.1.3 Proof of Proposition 2

We begin with the "if" part. Let there be an interval order $P$ on $F$. Define $C$ by $C(A)=\{f \in A \mid \nexists g \in A, \quad g P f\}$. Observe, first, that $C(A) \neq \varnothing$ for all $A \neq \varnothing$, because $P$ cannot have cycles. To see that it satisfies Sen's $\alpha$, let $f \in B \subset A$ and $f \in C(A)$. Clearly, there exists no $g \in B \subset A$ such that $g P f$ 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 $h P f$. As $f \in C(A)$ it has to be the case that $h \in B \backslash A$. Similarly, $g \notin C(A \cup B)$ implies that there exists $d \in A \backslash B$ such that $d P g$. Further, we know that $\neg h P g$ (by optimality of $g$ in $B$ ) and (similarly) $\neg d P f$. If $g P h$ then, by transitivity of $P, g P f$, and with $d P g$ we also get $d P f$, a contradiction. Hence $\neg h P g$ and $\neg g P h$, that is $g I h$. However, $d P g I h P f$ implies $d P f$, 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 $\alpha$, Union, and Continuity. Define $f P g$ by $g \notin C(\{f, g\})$. By Lemma 4, because $C$ satisfies Sen's $\alpha$, Union, and Continuity,

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

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

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 $P I P \subset P$. Let there be given $f, g, h, d$ such that $f P g I h P d$ and we need to show that $f P d$. Assume not. Consider $A=\{f, d\}$ and $B=\{g, h\}$. Because $\neg f P d$, we have $d \in C(A)$. We know that $g I h$ 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 $h P d$ and $g \in C(\{f, g, h, d\})$ - by $f P g$.

### 6.1.4 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 $f P g P h$. 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 $A \subset F$.

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 $A \varsubsetneqq F$; $C(\{f, g, h\})=\{f\}$. 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 $h \in C(\{h\})$.

The case of a choice function 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 $g P f$ 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 $1 \in C\left(A_{\tau}\right)$ but $1 \notin C\left(\cup_{\tau<2} A_{\tau}\right)$.

### 6.1.5 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 $(u, \Delta)$, 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.

### 6.2 Database-Related Analysis

### 6.2.1 Proof of Corollary 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

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

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

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

### 6.2.2 Proof of Theorem 2

We use the matrix $(v(f, x))_{f \in F, x \in X}$ obtained in the proof of Corollary 3, so that $\max _{f} v(f, x)=0$ for all $x$ and $\succsim_{D}$ 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 $f, g \in F$,

$$
\begin{align*}
& U_{D}(f)-U_{D}(g)>\Delta \quad \Rightarrow \quad f P_{D} g  \tag{5}\\
& U_{D}(f)-U_{D}(g)<\Delta \Rightarrow \neg f P_{D} g
\end{align*}
$$

When this is the case, we say that $\left(U_{D}, \Delta\right)$ is a pseudo-representation of $P_{D}$. Observe that, in this case, for every $D \in \mathbb{D}$, and every $A \subset F$, because

$$
C_{D}(A)=\left\{f \in A \mid \nexists g \in A, \quad g P_{D} f\right\}
$$

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

$$
\begin{aligned}
& L(f \mid D)>\gamma \sup _{g \in A} L(g \mid D) \Rightarrow f \in C_{D}(A) \\
& L(f \mid D)<\gamma \sup _{g \in A} L(g \mid D) \Rightarrow f \notin C_{D}(A)
\end{aligned}
$$

To find such a $\Delta \geq 0$ for which $\left(U_{D}, \Delta\right)$ 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 $\Delta$ is unique. Finally, we show that this $\Delta$ is the same for all pairs of theories, employing P 8 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 $\operatorname{supp} D \subset X_{0}$, (5) holds.

Proof: Consider

$$
P=\left\{D \in \mathbb{D} \mid f P_{D} g\right\}
$$

Let $v_{x}=v(f, x)-v(g, x)$ (for all $x \in X$ ) so that $f \succsim_{D} g$ iff $v D \geq 0$ (where $\left.v D=\sum_{x} v_{x} D(x)\right)$. We claim that, for all $D \in P$ and $D^{\prime} \notin P$, we have $v D \geq v D^{\prime}$. 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 $0 \leq v D^{+}<\varepsilon$. Similarly, there exists $D^{-}$ such that $D^{-}(x)>M$ for all $x \in X_{0}$ and $-\varepsilon<v D^{-} \leq 0$.

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 $\operatorname{supp} D_{x, y}^{+}=\{x, y\}$ instead of $\operatorname{supp} D_{x, y}(x)=\{x, y\}$ and $0 \leq v D_{x, y}^{+}<\frac{\varepsilon}{M\left|X_{0}\right|^{2}}$. Indeed, if $v_{x} / v_{y} \in \mathbb{Q}$, then there is such $D_{x, y}^{+}$with $v D_{x, y}^{+}=0$. If, however, $v_{x} / v_{y} \notin \mathbb{Q}$, the set $\left\{k v_{x}-l v_{y} \mid k, l \geq 0\right\}$ is dense in $\mathbb{R}$, so that we can find $D_{x, y}^{+}$with $0<v D_{x, y}^{+}<$ $\frac{\varepsilon}{M\left|X_{0}\right|^{2}}$. We can then define

$$
D^{+}=M \sum_{\left\{x, y \in X_{0} \mid v_{x}>0>v_{y}\right\}} D_{x, y}^{+}
$$

and observe that $D^{+}(x)>M$ for all $x \in X_{0}$ and $0 \leq v D^{+}<\varepsilon$. The proof for $D^{-}$is symmetric.

We prove the lemma by negation. Assume that $D \in P$ and $D^{\prime} \notin P$ satisfy $v D<v D^{\prime}$. Consider $M>D(x), D^{\prime}(x)$ for all $x \in X_{0}$ and $\varepsilon=\left(v D^{\prime}-v D\right) / 3$. Let $D^{-}$be the database provided by the claim. Consider $D^{\prime}+D^{-} \in \mathbb{D}$. Because $v D^{-} \leq 0$ we have $g \succsim_{D^{-}} f$ and, with $D^{\prime} \notin P$, by $\mathrm{P} 6, D^{\prime}+D^{-} \notin P$. Next, consider $\tilde{D}=D^{\prime}+D^{-}-D$. Notice that $D^{-}-D \geq 0$ so that $\tilde{D} \in \mathbb{D}$ (that is, $\tilde{D}(x) \geq 0$ for all $\left.x \in X_{0}\right)$. We have $v \tilde{D}=v D^{\prime}+v D^{-}-v D>v D^{-}+3 \varepsilon>$ $2 \varepsilon>0$. Hence $f \succsim_{\tilde{D}} g$. Given that $D \in P, \mathrm{P} 6$ implies $D+\tilde{D} \in P$. However, $D+\tilde{D}=D^{\prime}+D^{-}$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, P 7 implies that $P \neq \varnothing$. Thus, neither $P$ nor $P^{c}$ is empty. Define

$$
\begin{aligned}
\Delta^{+} & =\inf _{D \in P}\left[U_{D}(f)-U_{D}(g)\right] \\
\Delta^{-} & =\sup _{D \notin P}\left[U_{D}(f)-U_{D}(g)\right]
\end{aligned}
$$

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 $\Delta \in\left[\Delta^{-}, \Delta^{+}\right]$would do.

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

$$
\begin{aligned}
\Delta^{+} & =\inf _{D \in P}\left[D(x) v_{x}-D(y) v_{y}\right] \\
\Delta^{-} & =\sup _{D \notin P}\left[D(x) v_{x}-D(y) v_{y}\right]
\end{aligned}
$$

However, because $v_{x} / v_{y} \notin \mathbb{Q},\left\{D(x) v_{x}-D(y) v_{y}\right\}_{D}$ is dense in $\mathbb{R}$. Hence $\Delta=\Delta^{+}$and $\Delta$ that satisfies (5) for $X_{0}=\{x, y\}$ is unique.

Consider any other finite $X_{1} \subset X$ and apply Lemma 12 to $X^{\prime}=X_{0} \cup X_{1}$. There exists a $\Delta \geq 0$ that satisfies (5) for $X^{\prime}$ and it has to be the same $\Delta$ as found above. Finally, given an arbitrary $D \in \mathbb{D}$, we define $X_{1}$ to be its support and obtain the result.

Lemma 14 There exists $\Delta \geq 0$ such that $\left(U_{D}, \Delta\right)$ 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 $\Delta_{f g} \geq 0$ such that $\left(U_{D}, \Delta_{f g}\right)$ satisfies (5) for $(f, g)$. We will first show that for all $f, g, h$ we have $\Delta_{f g}=\Delta_{f h}$.

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

$$
\begin{aligned}
(I) U_{D}(f)-U_{D}(g) & \geq U_{D}(f)-U_{D}(h) \\
(I I) U_{D}(f)-U_{D}(h) & >\Delta_{f h} \\
(I I I) U_{D}(f)-U_{D}(g) & <\Delta_{f g}
\end{aligned}
$$

If $D$ satisfies these three inequalities, $(I)$ will imply that $U_{D}(g) \leq U_{D}(h)$, that is, $h \succsim_{D} g ;(I I)$ - 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 $(\xi, \eta)$ interpreted as $\left(U_{D}(g), U_{D}(h)\right)$. Consider the set

$$
\left\{(\xi, \eta) \in \mathbb{R}^{2} \mid \xi \leq \eta, \eta<-\Delta_{f h}, \xi>-\Delta_{f g}\right\}
$$

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 $\eta=-\Delta_{f h}$ and from the left by $\xi=-\Delta_{f g}$. However, P8 guarantees that $\left(U_{D}(g), U_{D}(h)\right)_{D \in \mathbb{D}}$ is dense in $\mathbb{R}^{2} .{ }^{13}$ Thus we can find $D \in \mathbb{D}$ that yields $\left(U_{D}(g), U_{D}(h)\right)$ in this set, contradicting the assumption $\Delta_{f g}>\Delta_{f h}$.

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

$$
\begin{aligned}
(I) U_{D}(f)-U_{D}(g) & \geq U_{D}(h)-U_{D}(g) \\
(I I) U_{D}(h)-U_{D}(g) & >\Delta_{h g} \\
(I I I) U_{D}(f)-U_{D}(g) & <\Delta_{f g}
\end{aligned}
$$

and thus $f \succsim_{D} h ; h P_{D} g$; but $\neg f P_{D} g$ - again, a contradiction.
As a result, for any $f, g, f^{\prime}, g^{\prime}$ we have $\Delta_{f g}=\Delta_{f g^{\prime}}=\Delta_{f^{\prime} g^{\prime}}$. And thus there is a single $\Delta \geq 0$ such that $\left(U_{D}, \Delta\right)$ satisfies (5) for all $(f, g)$.

## 7 Bibliography

Agaev, Rafig and Fuad Aleskerov (1993), "Interval Choice: Classic and General Cases", Mathematical Social Sciences, 26: 249-272.

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.

[^8]Beja, Avraham and Itzhak Gilboa, (1992), "Numerical Representations of Imperfectly Ordered Preferences (A Unified Geometric Exposition)", Journal of Mathematical Psychology, 36: 426-449.

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

Fechner, Gustav Theodor (1860), Elemente der Psychophisik. 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, 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

Manzini, Paola and Marco Mariotti (2012), "Choice by lexicographic semiorders." Theoretical Economics, 7: 1-23.

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
van Rooij, Robert (2008), "Revealed Preference and Satisficing Behavior", manuscript.

Weber, Ernst Heinrich (1834), De Tactu.


[^0]:    *We thank Fuad Aleskerov, Yoav Benjamini, Mira Frick, Paola Manzini, Marco Mariotti, Isaac Meilijson, Efe Ok, 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.
    ${ }^{\dagger}$ ESSEC, Singapore. b00786349@essec.edu
    ${ }^{\ddagger}$ HEC, Paris, and Tel-Aviv University. tzachigilboa@gmail.com
    ${ }^{\S}$ HEC, Paris. minardi@hec.fr
    ${ }^{1}$ See, for example, Heal and Millner (2014).

[^1]:    ${ }^{2}$ See Epstein and Schneider (2007), who use such sets to select the distributions that are to be updated in a Bayesian fashion.

[^2]:    ${ }^{3}$ We do not distinguish here between "descriptive" and "positive", nor between "normative" and "prescriptive".
    ${ }^{4}$ The statistical model we derive is unique only up to certain transformations, as will be

[^3]:    ${ }^{6}$ An interval order can be thought of as a binary relation for which a representation as above exists, where $\Delta$ is allowed to depend on $x$ or on $y$. See a formal definition and more details in the Appendix.

[^4]:    ${ }^{7}$ For 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 $g P f$ 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 $g$.

[^5]:    ${ }^{8}$ In 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.
    ${ }^{9}$ Observe that in our setup there is no loss of generality in assuming $C_{D}(A) \neq \varnothing$ even for an infinite $A$ : the likelihood function is bounded from above by 1 . Thus, the set $C(A)=$ $\left\{f \in A \mid L(f) \geq \gamma \sup _{g \in A} L(g)\right\}$ will be empty only if the $\max _{g \in A} L(g)$ is not obtained and $\gamma=1$.

[^6]:    ${ }^{11}$ As mentioned above, P7 implies P3. Hence, "P1-P8" can be replaced by "P1,P2,P4-P8".

[^7]:    ${ }^{12}$ The concatenation is defined by an existential quantifier: $x R_{1} R_{2} y$ if there exists a $z$ such that $x R_{1} z R_{2} y$. It is used inductively, so that, for example, $x P I P y$ means that there are $z, w$ such that $x P z, z I w, w P y$

[^8]:    ${ }^{13}$ To see this, observe first that, by $\mathrm{P} 8,\left(U_{D}(g), U_{D}(h)\right)_{D \in \mathbb{D}}$ includes six vectors $\left(\xi_{i}, \eta_{i}\right)_{i \leq 6}$, 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 $\left(\xi_{i}, \eta_{i}\right)_{i \leq 6}$ that enters $C$. Next consider a sequence of increasing databases $D_{t}\left(D_{t+1} \geq D_{t}\right)$ that visits $C$ infinitely many times. Since P8 precludes the sequence from visiting the same point twice, $\left(U_{D_{t}}(g), U_{D_{t}}(h)\right)_{t \geq 1}$ has an accumulation point in $C$. Take a subsequence of $\left(D_{t}\right)_{t}$ whose $\left(U_{D}(g), U_{D}(h)\right)$ values converge to a limit. Considering pairs of points in this sequence, we obtain databases $D$ (differences between such points) whose $\left(U_{D}(g), U_{D}(h)\right)$ has an arbitrarily small norm. Moreover, the sequence can be so chosen that the vectors $\left(U_{D}(g), U_{D}(h)\right)$ are not collinear. Summations of these vectors enter every open neighborhood.

