Prudent case-based prediction when experience is lacking[*]

Patrick H. O'Callaghan[†‡]

School of Economics, University of Queensland

Abstract

Before 1697, it would be difficult for a European zoologist to justify the prediction "the swans in New Holland are black". Yet the predictor might well be forward-looking enough to formulate her current forecast so as to accommodate such a prediction if, and when, a new case (observed or theoretical) arrives. Moreover, she may be prudent enough to do so in such a way that, the current forecast has the potential to extend to novel cases without revision and without violation of basic rationality conditions. In this article, we formalise these concepts and, to our knowledge, provide the first axiomatisation of the full set of additive and suitably unique similarity representations. The diverse matrix representation of Gilboa and Schmeidler (2003) is an important special case: the one where the predictor is experienced.

[*] The time stamp for this version is 14:00 (AEST), 13th April, 2018. The following link will bring you to the latest version: `https://dropbox.com/s/itchhpczp343nnc/` .

[†] `orcid.org/0000-0003-0606-5465`, `p.ocallaghan@uq.edu.au`

[‡] I thank Itzhak Gilboa for encouraging this project at an early stage.

From the past, the present acts prudently, lest it spoils future action.

Titian: Allegory of Prudence

1 Introduction

When formulating a forecast, experience is often the most important factor. Yet, outside of stylised small-world settings, even the most experienced forecasters would not claim to have access to "the full model" that includes all the relevant states of the world.

The global financial markets provide the canonical example of a large world where the set of states that are accessible to the forecaster would rarely be exhaustive. The omission (from most models) of details regarding the nature of the global financial crisis prior to 2007 serves as a strong reminder that, when trying to predict macroeconomic eventualities, it may be prudent to admit our limitations and adopt a less structured, more open framework than that which the standard Bayesian paradigm allows.

The theory of case-based decisions, and case-based prediction in particular, addresses these concerns by avoiding state spaces altogether. Instead, observations or (synonymously) cases form the basic building blocks of the model: where a given case may be empirical or theoretical. The predictor's model is then naturally bounded in size and scope by the experience of the forecaster. The present prediction problem determines the perspective (or vantage point) of this model as well as the relevance of any experience the predictor might have. Similar to unlikely states in a probabilistic framework, cases that are fanciful (such as outliers) or that have little relation to the present prediction problem receive a low weight in the present prediction problem.

In contrast with states, two cases that are identical (in terms of their information content or similarity to the present problem) for current prediction problem might provide contrasting information for another.

Although her perspective is fixed, case resampling and subsampling (as in the standard nonparametric bootstrap) allow the predictor to explore hypothetical databases and extract all the information her experience can afford. For each feasible (i.e. finite) database of resampled past cases, the predictor provides an ordinal plausibility ranking of eventualities and it is the family of such rankings (what we call a forecast) that forms the primitive data of the model.

Example 1. *Consider a search engine on a standard web browser. When a user conducts a search, the engine compiles a list of webpages, with the most plausible at the top, the second most plausible next, and so on. This list of webpages is the search engine's plausibility ranking given its current database of past cases. Through resampling and subsampling of past cases, the predictor can explore the rankings the engine would generate, for the same search, given other databases. At one extreme, the engine may yield the same plausibility ranking for every feasible database of past cases. At the other, the current data may be sufficiently rich that, for every feasible plausibility ranking of relevant webpages, there is a database that gives rise to that ranking.*

The scenarios that arise in example 1 are important for scope of the present paper. For whilst the axiomatisation of Gilboa and Schmeidler [4] (henceforth [GS]) accounts for predictors with current databases that are sufficiently rich to support every strict plausibility ranking of every subset of four eventualities (the content of their Diversity axiom), it does not accommodate predictors that are less fortunate. In other words, it does not account for predictors that are inexperienced for the purposes of the present prediction problem. In this paper, we fill this gap by dropping Diversity and replacing it with one that involves the predictor exploring a richer form of resampling. The idea is that she may mitigate her lack of experience by considering the set of feasible forecasts that arise in the presence of a novel case type. Since the predictor does not have access to this novel case, one key subtlety lies in the way we model this novel case.

Example 2. *Before the first European observation of black swans in modern day Australia in 1697, the typical European zoologist's dataset would justify the plausibility ranking "On his expedition to New Holland, it is more likely that Willem de Vlamingh will observe white swans than a black ones." It is only with a sufficiently rich knowledge of the migratory behaviour of Swans, a rich dataset documenting their absence in tropical regions, and a phylogenetic theory of Swan evolution that a zoologist could convincingly propose the reverse plausibility ranking. The fact that de Vlamingh's observation had the impact on zoology that it did reveals that few if any zoologists were capable of such a deep and convincing hypothesis. Nonetheless, a prudent zoologist might well anticipate that, since expeditions typically yield new data, it would be wise to explore the implications for her forecast of the arrival of a novel case.*

Although the inexperienced predictor can only observe past cases, she might also observe that resampling them does not yield a forecast with a diverse family of plausibility rankings. And, in turn, she may observe that, if the present prediction problem turns out to be unlike past cases, her future forecast may well be more diverse than her current one. She may then ask whether she will need to revise her current forecast in order to accommodate a forecast generated by richer databases? In the present paper, we show that this line of observation and questioning naturally leads to the following conclusion: if the inexperienced predictor restricts her resampling procedure to past cases, then she may be omitting important information that is concealed by her lack of experience. At least, that is, if she accepts the basic axioms for case-based decision theory (and prediction in particular).

The basic axioms are as follows. The first and most basic is that, conditional upon any feasible database of past cases, the plausibility ranking is both transitive and complete. The second axiom, Combination, is the requirement that whenever the plausibility rankings at two disjoint databases coincide, then the combined database (formed by taking the union) generates the same plausibility ranking. When the number of eventualities is finite, the third axiom, Archimedeanity, requires that for any pair of disjoint databases, the information associated with the first is eventually swamped by the combining it with sufficiently many copies of the second.

The formal question we shall ask is the following: for a predictor with a current forecast that satisfies the basic rationality axioms of [GS], will these axioms apply to any extension of her current forecast that is formed by considering databases involving resampled copies of a novel case? If not, then the omitted information that we refer to in the penultimate paragraph above may bias her forecast in such a way that intransitive plausibility rankings arise in every extension of her current forecast that satisfies Completeness (at each database), Combination and Archimedeanity. This is essence of the Prudence axiom that we appeal to instead of the Diversity axiom of [GS].

The fact that inexperience or novelty is so closely associated with intransitivity is well-documented in the literature on

The consequence of this bias (i.e. a violation of the Prudence axiom) is to preclude the representation of the forecast by a real-valued suitably additive weighting function of the form

$$v : \text{eventualities} \times \text{past cases} \to \mathbb{R}.$$

That is a matrix v such that for any pair of eventualities x and x' and any database of resampled past cases D,

$$x \precsim_D x' \quad \text{if, and only if,} \quad \sum_{c \in D} v(x,c) \leqslant \sum_{c \in D} v(x',c),$$

where $\precsim_D$ is the plausibility ranking that is determined by D. (For instance, in a typical empirical setting, $\precsim_D$ might be the ordinal content of the empirical likelihood conditional on D.) In the absence of this bia

Here, each case type forms a distinct dimension, where two cases are of the same type if their information content is identical for the purposes of the present prediction problem. For instance, in example 1, the (extremely) inexperienced engine has a forecast of dimension one, whereas the experienced engine has a forecast that is of dimension at least $n!$, where n is the number of plausible webpages.

The more general form of resampling that we propose includes copies of the novel case type. Since our predictor does not have access to this novel case (for then it would be a past case) we model this case as a variable (of positive arity) in contrast with past cases which are constants (of arity zero). A variable database is once again a A generalised database is then obtained by associating a plausibility ordering with that Then the higher order resampling consists of

have awe propose allows the predictor to peer into the future and contemplate the potential forecasts that she will be able to make once she has new information.

Whilst there may be settings where it is costless to update and revise Whilst it is fairly straightforward to see the value of such an exploration, there is the s

Moreover, she may wish to ensure that, should a novel case her current forecast will exte accommodates such extensions her prediction is made wish to look at potential extensions of her model She may therefore, perhaps with an eye on future prediction problems, contemplate extensions of her model thanovel cases.

Cases that, when combined with existing databases, Whilst the predictor cannot observe these cases, she can the set of feasible extensions She might well decide to ra

As in the Bayesian framework (e.g. the one-shot Kalman filter), the forecast is represented by a collection of similarity/information weights and a real-valued suitably additive function on eventualities $\times$ cases. In contrast with the Bayesian framework, no prior is assumed: the current database of cases is all the predictor needs. Via case resampling (also known as the standard nonparametric bootstrap), the predictor explores her sample by considering every possible (finite) database that can be generated from the cases we have observed. At every such database she is endowed with an ordinal plausibility ranking of eventualities. It is this family of such plausibility rankings allows us to go beyond ordinal measurement to derive the similarity weighting functions that parallels the familiar expected utility representation.

More broadly, in contrast with the probabilistic approach, one case does not exclude another and, although the set of cases is supposed to exhaust the experiences of the predictor, there is nothing to preclude new cases being added to the model. A basic contribution of the present paper is to accommodate cases that are novel and distinguish these from those that are (for the purposes of the present prediction problem) equivalent to past observations. Cases that are genuinely novel are conceptually subtle in that, should the predictor have full access to them, they would no longer be novel: they would form a part of the predictor's experience. Since the present axiomatisation holds for any prediction problem, we model past cases as constants and novel cases as variables (from the view of the predictor).

we cannot simply include them in the if they can be explicitly modelled

Indeed, as a consequence of the aforementioned case resampling the expansion of the predictor's current database to include cases that resemble past cases may change the predictor's plausibility ranking of eventualities, but it otherwise leaves the model unchanged.

Indeed, the set of cases grows with the experience of the predictor. Yet, given the ability to resample cases, a large database is not necessarily a good measure of experience, for if that database consists of identical cases, then the predictions should not change no matter how many copies we observe. In contrast, a predictor with just a handful of distinct cases would possess the measure of experience we propose is

new cases two cases are no mutually exclusive, nor is free to be expanded or These weights are directly related to the data instead of

Information about which eventualities are more plaus is most the future is most lik Those eventualities that tell us the plausibility of each object Unlike the probabilistic framework, the predictor ranks the set of available cases and present a framework derive the appropriate weight for each eventuality $\times$ case pair.

The questions we shall address in this paper are as follows: what makes for an experienced predictor? and should the predictor lack experience, what are the conditions such that her forecast is represented by an additive similarity weighting function? In particular, we extend the axiomatic framework of Gilboa and Schmeidler [4] (henceforth [GS]) to account for predictors that lack experience.

A search engine ranks one webpage over another if, and only if, it is a more plausible candidate for the search. In general, a predictor ranks one eventuality above another if, and only if, she finds it to be more plausible. The axioms we derive are necessary and sufficient for the existence of a matrix that assigns a weight to each eventuality $\times$ case pair and that represents the family of plausibility rankings that make up the forecast in the sense that, for any pair of eventualities x and x' and any database of resampled cases D:

$$x \precsim_D x' \text{ if, and only if, } \sum_{c \in D} v(x, c) \leqslant \sum_{c \in D} v(x', c),$$

where $\precsim_D$ is the plausibility ranking that is determined by D.[1]

We refer to the resulting family of (conditional) plausibility rankings as a *forecast*. Alone, each plausibility ranking is merely ordinal. Via the axioms, it is the forecast that yields the weights on cases. These weights provide a subjective measure of the *influence of* or *similarity between* the predictor's past experiences and the present prediction problem. Although our basic axioms will allow for more general representations, we will also identify minimal conditions such that the weights are *unit comparable* in the sense that the following ratio is a unique number for any eventualities x and x' and any cases c and c':

$$\big(v(x, c) - v(x', c)\big) \, / \, \big(v(x, c') - v(x', c')\big) \, .$$

When states, events and probabilities summarise the predictor's view of the world and its impact on her predictions, each state specifies

[1] In empirical settings $\precsim_D$ is simply the ordinal content of the empirical likelihood conditional on D.

a distinct dimension of the model and the probability distribution assigns a weight to each dimension. A key feature of [GS], where cases are primitive, is that the dimensions are endogenous to the problem at hand. The idea is that, for one prediction problem, two cases may may be so similar as to considered as equivalent to one another, or of the same *case type*. For another prediction problem, the same two cases may no longer be equivalent, each giving rise to different weights or indeed different plausibility rankings. It is the set of case types that determines the notion of dimension in the present model.

Our first conceptual extension of [GS] is to consider case types as one measure of experience. The scenarios we describe in example 1 support this interpretation: the search engine with one case type is, in effect, inexperienced and any resampling of past cases will not change this fact. It is clear that, the richer the data the search engine has access to, the more diverse its conditional predictions will be. Yet one forecast is more diverse than another if it contains a greater number of distinct plausibility rankings. Thus, the diversity of a forecast provides another measure the predictor's experience which is distinct from, but closely related to, case types.

The main contribution of the present work is to allow for forecasts that are nondiverse in the sense that they do not satisfy the diversity axiom of [GS]. The latter axiom requires that, for every strict ordering of four eventualities, there exists a database, conditional upon which, the predictor's ranking over all eventualities contains that ordering. In [GS], as well as in the closely related models of Gilboa and Schmeidler [3, 2], the diversity axiom is presented as a purely technical requirement for the desired representation. When diversity fails to hold, the remaining axioms do not suffice for a matrix representation that is separable across eventualities. Our answer to this fundamental difficulty (which also arises in other settings such as Azrieli [1]), is to ask that the predictor is forward looking. Our predictor is forward looking if her behaviour can be explained by adding a "novel case" to the model and studying the "potential extensions" of her forecast.

We conclude the introduction with a brief discussion these concepts and the main axiom to which they give rise. In contrast with a standard case, which determines a plausibility ranking, we define a *novel case* to be one that can be associated with any plausibility ranking. Since a novel case is intended to capture cases that lie beyond the experience of the predictor, its weight in a given a database should be indeterminate. As such, the plausibility ranking associated with any database that contains a copy of the novel case is also indeterminate.

By assigning a plausibility ranking to this case and to its affiliated databases, we obtain a *(potential) extension* of the (current) forecast. It will suffice to study the extensions that involve three or four eventualities. Two extensions will be of the same type if they feature the same degree of diversity or experience. Our main axiom requires that every extension that satisfies all the necessary axioms of [GS] (with the possible exception of transitivity), has a modification that is of the same type and that also satisfies transitivity.

The role of our main axiom, which we refer to as *prudence* is to exclude the forecasts with extensions that feature essential intransitivities. That is, our predictor fails to be prudent if she (behaves as if) she fails to foresee that, should a certain case arise, the only way she can satisfy the [GS] axioms avoid specifying an intransitive plausibility ranking at some database is to retrospectively change her current forecast or by being dogmatic and ruling out certain plausibility rankings regardless of the data. To borrow from our epigraph, by being imprudent the predictor may spoil her future action.

Through examples, we will highlight the ways in which prudence is strictly weaker than imposing the diversity axiom on the forecast. We will also show that prudence is weaker than potential diversity (the forecast admits an extension that satisfies the diversity condition).

2 Model

2.1 Framework

Where possible, we adopt the notation and interpretations of [GS]. The first primitive of our model is the nonempty set X of conceivable *eventualities* of the present prediction problem. For instance, for a search engine, an eventuality $x \in X$ might be "page such-and-such is the desired webpage". Recall that search engines present an ordered list of plausible webpages, with the most plausible appearing at the top, followed by the second most plausible, and so on. The forecaster's present prediction problem is to specify a plausibility ranking on X. Let $\mathcal{R}$ denote the set of binary (plausibility) relations on X.

Current memory The forecaster is equipped with her current memory $I^\star$. We assume that $I^\star$ is the union of a (possibly empty) finite set of past cases $D^\star$ and a *novel, variable case* $\mathfrak{n}$. The cases in $D^\star$ collectively represent the forecaster's relevant observations or experience. Formally, each $c \in D^\star$ is a constant (of arity zero). Recall that the *arity* of a variable, function, operation or relation is the number of arguments it takes. (E.g. The union operation on sets has arity two and is well-defined independently of any specific domain and range.)

Our first and most fundamental modification of the primitives of

[GS] is the inclusion of $\mathfrak{n}$ in the current memory $I^\star$. We model $\mathfrak{n}$ as a variable (of positive arity) with unspecified domain and range, to reflect the fact that the forecaster has no experience of it. We include $\mathfrak{n}$ in the current memory $I^\star$ of the forecaster in order to capture the fact that she is aware of her present prediction problem. This also allows us to model the case where the forecaster has no (relevant) data.

Plausibility given the data Like [GS], we assume the forecaster possesses a well-defined plausibility relation $\lesssim_D$ that belongs to $\mathcal{R}$, for each nonempty $D \subseteq D^\star$. In contrast, $\lesssim_{\mathfrak{n}}$ is indeterminate and not, therefore, a member of $\mathcal{R}$. It is however a well-defined free variable with values in $\mathcal{R}$. (Like $\mathfrak{n}$, the domain of $\lesssim_{\mathfrak{n}}$ is unspecified.) As such, $\lesssim_{\mathfrak{n}}$ seems to be an accurate representation of a forecaster that either has no experience or that chooses to ignore all her experience. Our purpose is to describe a framework that describes how a forecaster might exploit her experience and impose constraints on the values that the variable $\lesssim_{I^\star}$ can take.

Conceivable cases Although the set $I^\star$ of current cases is finite, the predictor is freely able to resample cases and generate hypothetical memories M in the same way that bootstrapped data sets are hypothetical samples that a forecaster might have drawn. In principle, this yields a countably infinite set of cases. In general however, resampling may take place in the ambient set (if any) from which the cases $d \in D^\star$ were drawn. In principle, this allows the set of cases to be uncountable. Thus, as in [GS], we classify resampled copies of a given case as distinct cases. Resampling $\mathfrak{n}$ is a simple matter of including additional copies of $\mathfrak{n}$, which also feature as distinct cases.

Like [GS], our framework is general enough to accommodate a forecaster that goes beyond her current memory and includes hypothetical cases that she has not experienced, but which, through reasoning or interpolation, she can clearly describe. These hypothetical cases are formally constant, like members of $D^\star$. Arguably, copies of $\mathfrak{n}$ are inconceivable: since they lie beyond the experience of the forecaster and cannot be fully described (c.f. Karni and Vierø [8] and Halpern and Rêgo [5, 6]) and do not determine a member of $\mathcal{R}$. In the present article, a copy of $\mathfrak{n}$ is well-defined and conceivable in the same way that a physical sector (the minimum storage unit) of a hard-drive is well-defined: independently of any contents that we may eventually assign to it.

Let $\mathbb{A}$ denote the resulting set of *all cases* that are relevant to the current prediction problem. The subset $[\mathfrak{n}] \subseteq \mathbb{A}$ consists of the copies of $\mathfrak{n}$ in $\mathbb{A}$. Finally, the set $\mathbb{D} \stackrel{\text{def}}{=} \mathbb{A} - [\mathfrak{n}]$ consists of the set of constant or deterministic cases.

Databases We have in mind what is known as *case resampling* in the literature on bootstrapping, where every possible resample of a given size is obtained. Indeed, like [GS], we go further and allow for every finite sized database. Let

$$\mathcal{D} \stackrel{\text{def}}{=} \{D \subseteq \mathbb{D} : 0 < \# D < \infty\}$$

denote the set of *determinate or constant databases* (these are referred to as *memories* in [GS]).

Like $D^\star$, each $D \in \mathcal{D}$ contains no copies of $\mathfrak{n}$. Perhaps through experience, in-sample reasoning, or some algorithm the forecaster possesses a well-defined plausibility ranking $\precsim_D$ for each $D \in \mathcal{D}$. Thus,

$$\precsim_{\mathcal{D}} \text{ is the point } \langle \precsim_D : D \in \mathcal{D} \rangle \text{ in } \mathcal{R}^{\mathcal{D}},$$

where $\prec_D$ and $\approx_D$ denote the asymmetric and symmetric parts of $\precsim_D$.

Let $\mathcal{I}$ denote the set of *variable* or *indeterminate* databases: those that contain at least one copy of $\mathfrak{n}$. Like $\precsim_{\mathfrak{n}}$ and $\precsim_{I^\star}$, for each $I \in \mathcal{I}$, $\precsim_I$ is a free variable with values in $\mathcal{R}$. Although, in isolation, each $\precsim_I$ is a free variable, when the axioms we introduce hold, $\precsim_I$ may well be constrained, but not determined, by the values of $\precsim_{\mathcal{D}}$.

The fact that the empty database $\varnothing$ is absent from $\mathcal{D}$ means that $\precsim_\varnothing$ is undefined (just as $\precsim_D$ is not defined when D is infinite). If $\varnothing$ were defined, then a necessary condition for the linear representation we seek is that $x \approx_\varnothing y$ for every $x, y \in X$. We leave $\precsim_\varnothing$ undefined since our axioms will accommodate the possibility that $x \prec_D y$ for every $D \in \mathcal{D}$. Indeed the situation where our forecaster is without data and without a novel case would only seem to be justified when she is entirely unaware of the prediction problem. We opt to accommodate this hypothetical scenario and simultaneously simplify our exposition by including $\varnothing$ in the set of extensions of $\mathcal{D}$, which we introduce shortly.

Let $\mathcal{A} \stackrel{\text{def}}{=} \mathcal{I} \cup \mathcal{D} \cup \{\varnothing\}$ denote the set of all databases: variable, determinate and empty. We now define the notion of a case type.

Case types As in [GS], for the present prediction problem, two cases a and a' in $\mathbb{D}$ are of the same type, written $a \sim^\star a'$ if, and only if, $\precsim_{A \cup a} = \precsim_{A \cup a'}$ for every $A \in \mathcal{A}$ such that $a, a' \notin A$ and $A \cup a, A \cup a' \in \mathcal{D}$. The proof that $\sim^\star$ is an equivalence relation on $\mathbb{D}$ follows from the corresponding observation in [GS].

Remark 1. *By virtue of the fact that $\varnothing \in \mathcal{A}$, $\precsim_a \neq \precsim_{a'}$ implies $a \nsim^\star a'$. The converse, however, is not true: the notion of case types is subtle enough to measure strength of similarity.*

We extend $\sim^\star$ to $\mathbb{A}$ by taking $[\mathfrak{n}]$ to be an equivalence class of its own. Finally, as in [GS], we extend $\sim^\star$ to databases in $\mathcal{A}$, by assuming that $A \sim^\star A'$ if, and only if, there exists an isomorphism $f : A \to A'$ such that $a \sim^\star f(a)$ for every a in A. In this way, $\sim^\star$ becomes an equivalence relation on $\mathcal{A}$.

The following assumption ensures there are enough databases to fully identify the weights on cases. It would not, of course, be expected to hold in a given application. In practice, we would settle for partial identification of the weights. It requires that case resampling is such that every finite memory is feasible.

Richness Assumption. *For every $a \in \mathbb{A}$, there are infinitely many cases $a' \in \mathbb{A}$ such that $a' \sim^\star a$.*

Potential extensions The definition that follows will allow for a precise analysis of the potential impact of novel cases. For any $Y \subseteq X$, let $\mathcal{R}_Y$ denote the set of binary plausibility relations on Y.

Definition 1. $\precsim_{\mathcal{E}} = \langle \precsim_E : E \in \mathcal{E} \rangle$ *is a* proper Y-extension *if, for some $Y \subseteq X$, there exists an* evaluation function $\mathrm{ev} : \mathcal{A} \to \mathcal{A} \times \mathcal{R}_Y$ *such that $\mathcal{E} = \mathrm{ev}(\mathcal{A})$ and,*

1. *for every $E \in \mathcal{E}$, if $E = A \times r$, then $\precsim_E \stackrel{\text{def}}{=} r$;*
2. *for every $D \in \mathcal{D}$, if $E = \mathrm{ev}(D)$, then $\precsim_E \, = \, \precsim_D \cap \, Y^2$;*
3. *for every $a \in \mathbb{A}$, $a \sim^\star \mathfrak{n}$ if, and only if, for every $A \in \mathcal{A}$ such that $a, \mathfrak{n} \notin A$, $\precsim_{\mathrm{ev}(A \cup a)} \, = \, \precsim_{\mathrm{ev}(A \cup \mathfrak{n})}$.*

Let $\mathcal{E}^*$ denote the (unique) improper extension $\precsim_{\mathcal{E}^*}$ *that satisfies* $\mathrm{ev} : \mathcal{D} \to \mathcal{D} \times \mathcal{R}_X$, *$\mathcal{E}^* = \mathrm{ev}(\mathcal{D})$, and parts 1 and 2 above.*

We often abbreviate and refer to $\precsim_{\mathcal{E}}$ or indeed $\mathcal{E}$ simply as an extension. The latter is a minor abuse of notation since, for any given

$\mathcal{E}$, the definition of its associated evaluation function $\mathrm{ev}_{\mathcal{E}}$ ensures that $\mathcal{E} = \{A \times \precsim_{\mathrm{ev}_{\mathcal{E}}(A)} : A \in \mathcal{A}'\}$ for $\mathcal{A}' = \mathcal{D}$ or $\mathcal{A}' = \mathcal{A}$. That is to say, there is an obvious isomorphism between $\mathcal{E}$ and the graph of $\precsim_{\mathcal{E}}$. We let $\mathcal{E}^\star$ denote the unique (improper) extension that is isomorphic to the graph $\{D \times \precsim_D : D \in \mathcal{D}\}$ of $\precsim_{\mathcal{D}}$. As a matter of expedience, we therefore refer to $\precsim_{\mathcal{D}}$, or $\mathcal{D}$, as an extension of itself. We will only refer to $\mathcal{E}$ as a Y*-extension* when we wish to highlight that the dependence of $\mathcal{E}$ on the subset Y of eventualities that it extends.

For any given extension $\mathrm{ev}(\mathcal{A}')$, the most natural union operation for our purposes is the following: for every $A, A' \in \mathcal{A}'$,

$$\mathrm{ev}(A) \cup \mathrm{ev}(A') \stackrel{\text{def}}{=} \mathrm{ev}(A \cup A').$$

This union operation applies only to the database dimension of $\mathrm{ev}(\mathcal{A}')$. This operation will allow us to combine potential databases independently of the orderings that ev associates with the elements of $\mathcal{A}'$.

Let $\mathcal{E}$ be arbitrary extension and let $\sim^{\mathcal{E}}$ be the binary relation on

$$\{e : e = \mathrm{ev}_{\mathcal{E}}(a) \text{ for some } a \in \mathbb{A}\}$$

such that $e \sim^{\mathcal{E}} e'$ if, and only if, for every $E \in \mathcal{E}$ such that $e, e' \notin E$, $\precsim_{E \cup e} = \precsim_{E \cup e'}$. Like $\sim^\star$, we extend $\sim^{\mathcal{E}}$ to databases in $\mathcal{E}$ by assuming that $E \sim^{\mathcal{E}} E'$ if, and only if, there exists an isomorphism $f : E \to E'$ such that $e \sim^{\mathcal{E}} f(e)$ for every $e \in E$.

When $\mathcal{E} = \mathrm{ev}(\mathcal{A}')$ is a proper extension, $\mathcal{A}' = \mathcal{A}$, so that $\mathfrak{n} \in \mathcal{A}'$, and, by part 3 of definition 2, $\mathrm{ev}_{\mathcal{E}}(\mathfrak{n}) \nsim^{\mathcal{E}} \mathrm{ev}_{\mathcal{E}}(c)$ for every $c \in \mathbb{D}$. In other words, $[\mathfrak{n}]$ forms a distinct equivalence class of $\sim^{\mathcal{E}}$ in $\mathrm{ev}_{\mathcal{E}}(\mathbb{A})$. In combination with part 2 of definition 2, this leads to

Observation 1. *For every $Y \subseteq X$, if $\mathcal{E}$ and $\mathcal{E}'$ are two proper Y-extensions, then $\sim^{\mathcal{E}} = \sim^{\mathcal{E}'}$.*

We shall see that it suffices to consider Y-extensions such that Y consists of three or four pairwise $\mathcal{D}$-distinct elements in X. For any $x, y \in X$, x and y are *$\mathcal{D}$-distinct* if $x \not\approx_D x'$ for some $D \in \mathcal{D}$. On the set of subsets Y of X, we define the counting measure $Y \mapsto |Y|_{\mathcal{D}}$ to be the number of $\mathcal{D}$-distinct elements in Y. Thus, when Y consists of $\mathcal{D}$-distinct elements alone, $|Y|_{\mathcal{D}} = \#Y$.

2.2 Axioms

In the following axioms, $\mathcal{E}$ denotes a given potential extension of $\mathcal{D}$. For reasons that will become apparent, we separate the order axiom of [GS] into the two classical axioms: transitivity and completeness.

A1 (Transitivity)**.** *For every $E \in \mathcal{E}$, $\precsim_E$ is transitive on X.*

A2 (Completeness)**.** *For every $E \in \mathcal{E}$, $\precsim_E$ is complete on X.*

A3 (Combination)**.** *For every disjoint $E, E' \in \mathcal{E}$ and every $x, y \in X$, together $x \precsim_E y$ and $x \precsim_{E'} y$ imply $x \precsim_{E \cup E'} y$, and if either premise holds strictly, then $x \prec_{E \cup E'} y$.*

A4 (Archimedeanity)**.** *For every disjoint $E, E' \in \mathcal{E}$ and every $x, y \in X$, if $x \prec_E y$, then there exist pairwise disjoint $E_1, \ldots, E_k \in \mathcal{E}$ such that $E_j \sim^{\mathcal{E}} E$ and $E_j \cap E' = \varnothing$ for each $j \leqslant k$ and $x \prec_{E_1 \cup \cdots \cup E_k \cup E'} y$.*

Recalling that $\mathcal{E}^\star$ is isomorphic to $\mathcal{D}$, we note that A1–A4 coincide with the corresponding axioms of [GS] when $\mathcal{E} = \mathcal{E}^\star$. We adopt the above, more flexible, statement of the above axioms so as to accommodate the remaining axiom of our main representation theorem which relaxes the diversity axiom of [GS].

To the same end, we introduce one final notion. The extension $\mathcal{E}'$ *features the rankings of $\mathcal{E}$* if, for every $E \in \mathcal{E}$ such that $\precsim_E$ is complete and transitive, there exists $E' \in \mathcal{E}'$ such that $\precsim_{E'} = \precsim_E$.

A5 (Prudence)**.** *For every $Y \subseteq X$, $\#Y = |Y|_{\mathcal{D}} = 3, 4$, and every proper Y-extension $\mathcal{E}$ that satisfies* A2–A4*, there exists a Y-extension that satisfies* A1–A4 *and that features the rankings of $\mathcal{E}$.*

Definition 2 ensures that A5 is a restriction on $\precsim_{\mathcal{D}}$. As such, it is meaningful to say that $\precsim_{\mathcal{D}}$ satisfies or does not satisfy A5.

3 Results

3.1 Existence

Theorem 1. *Let there be given X, $\mathbb{A}$ and $\precsim_{\mathcal{D}}$, as above, satisfying the richness condition and such that $|X|_{\mathcal{D}}$ is countable. Then $\precsim_{\mathcal{D}}$ satisfies* A1–A5 *if, and only if, some matrix $v : X \times \mathbb{D} \to \mathbb{R}$ satisfies*

(i) *for every $c, d \in \mathbb{D}$, $c \sim^{\star} d$ if, and only if, $v(\cdot, c) = v(\cdot, d)$;*

(ii) *for every $x, y \in X$ and for every $D \in \mathcal{D}$,*

$$x \precsim_D y \quad \textit{iff} \quad \sum_{c \in D} v(x, c) \leqslant \sum_{c \in D} v(y, c).$$

Proof that the axioms are sufficient for a representation. Recall that each equivalence class of $\sim^{\star}$ defines a case type. Let $\mathbb{T}$ be the set of equivalence classes of $\sim^{\star}$ in $\mathbb{A}$ and let $\mathbb{S} \overset{\text{def}}{=} \mathbb{T} - [\mathfrak{n}]$.

Until the final step of the present proof, we work with

Assumption 1. *Every pair of distinct elements in X is $\mathcal{D}$-distinct.*

In the first step of the proof we translate our model into one where databases are represented by rational vectors, the dimensions of which are the case types. This translation relies on the axioms, but also involves a translation of these

Step 1 (Translation of the model to rational vectors).

Step 1.1 (Databases as integer-valued vectors). We begin by showing that the set $\mathcal{D}_{/\sim^\star}$ of equivalence classes of $\sim^\star$ in $\mathcal{D}$ is isomorphic to the subset $\mathbb{I} \subsetneq \mathbb{Z}^{\mathbb{S}}_{\geqslant 0}$ of nonnegative integer-valued vectors

$$J : \mathbb{S} \to \mathbb{Z}_{\geqslant 0} \text{ such that } 0 < \#\{s : 0 < J(s)\} < \infty.$$

Take any $J \in \mathbb{I}$. Then for some $s_1, \ldots, s_k$, there exists $n_1, \ldots, n_k$ such that $J(s_j) = n_j$ for $j = 1, \ldots, k$. Similarly, by the richness assumption and the definition of $\mathcal{D}$, there exists $D_J \in \mathcal{D}$ such that $\#\{c \in D : c \in s_j\} = n_j$ for $j = 1, \ldots, k$. Let $i : \mathcal{D} \to \mathbb{I}$ denote the surjection $D_J \mapsto J$ for each $J \in \mathbb{I}$. We wish to take $\precsim_{i(\mathcal{D})} = \precsim_{\mathcal{D}}$. The next lemma confirms that $\precsim_J$ is well-defined member of $\mathcal{R}$ for each J. It implies that the quotient space $\mathcal{D}_{/\sim^\star}$ is isomorphic to the set $\mathbb{I}$.

Lemma 1. *For every $C, D \in \mathcal{D}$, $C \sim^\star D$ implies $\precsim_C = \precsim_D$.*

Proof of lemma 1. For the case where $C = \{c\}$ and $D = \{d\}$, $\precsim_C = \precsim_D$ follows directly from the definition of $\sim^\star$ and the fact that $\varnothing \in \mathcal{A}$. For the case where $\#D > 1$, the proof proceeds by induction. Suppose that the lemma holds for pairs of databases of cardinality k. Take C and D to be of cardinality $k+1$ and such that $C \sim^\star D$. Let $f : C \to D$ be the bijection that satisfies $c \sim^\star f(c)$ for each $c \in C$. By the induction hypothesis, $C' \sim^\star f(C')$ for some $C' \subset C$ such that $\#C' = k$. Then $C - C' = \{c'\}$ for some c' such that $c' \sim^\star f(c')$. Since C is the disjoint union of C' and c' and D is the disjoint union of $f(C')$ and $f(c')$, it follows that $\precsim_C = \precsim_D$, as required. $\square$

For each $s \in \mathbb{S}$, let $J_s \in \mathbb{I}$ denote the basis vector for dimension s, so that $J_s(s) = 1$ and $J_s(s') = 0$ for every $s' \neq s$.

Claim 1. *For every $s \in \mathbb{S}$, $\precsim_{J_s} = \precsim_c$ for every $c \in s$.*

Proof of claim 1. Let $D \in \mathcal{D}$ be such that $i(D) = J_s$. Then, by step 1.1, $\# D = 1$, so that $D = \{c\}$ and $\precsim_c = \precsim_{J_s}$ for some $c \in s$. Since $c' \sim^\star c$ for every $c' \in s$, lemma 1 implies that $\precsim_{c'} = \precsim_c$ for every such c'. $\square$

Recall that the discrete union operation is algebraically equivalent to addition (e.g. **??**). We may therefore let A1,A2 and A3 apply directly to $(\mathbb{I}, +)$ and, by lemma 1, rewrite A4 as follows. If $i(C_1) = J = (n_1, \ldots, n_k)$ and $C_1, \ldots, C_k$ are pairwise disjoint and $\sim^\star$-equivalent, then $i(C_1 \cup \cdots \cup C_k) = (k \cdot n_1, \ldots, k \cdot n_k) = kJ$. Thus

A4′ (Archimedean axiom). *For every $J, J' \in \mathbb{I}$ and every $x, x' \in Y$, if $x \prec_J x'$, then there exists $k \in \mathbb{Z}_{>0}$ such that $x \prec_{kJ+J'} x'$.*

Like [GS], we refer to the following basic result as

Claim 2. *For every $J \in \mathbb{I}$ and $k \in \mathbb{Z}_{>0}$, $\precsim_{kJ} = \precsim_J$.*

(For any given k, the proof of claim 2 follows via $k-1$ applications of A3.) Claim 2 is essential for the next substep.

Step 1.2 (The [GS] axioms for rational-valued vectors). Take $\mathbb{J} \subseteq \mathbb{Q}^{\mathbb{S}}_{\geqslant 0}$ to be the set of nonnegative rational-valued vectors J such that $0 < \# \{s : 0 < J(s)\} < \infty$. Fix $J \in \mathbb{J}$ such that $0 < J(s_j) = q_j$ for $j = 1, \ldots, k$. Since k is finite, there exists a minimal $\kappa \in \mathbb{Z}_{>0}$ such that $\kappa \cdot q_j \in \mathbb{Z}_{>0}$ for $j = 1, \ldots, k$. For each $J \in \mathbb{J}$, let $K_J \stackrel{\text{def}}{=} \kappa \cdot J$ belong to $\mathbb{I}$. Then the map $J \mapsto \precsim_J \stackrel{\text{def}}{=} \precsim_{K_J}$ is well-defined and $\precsim_{\mathbb{J}} = \langle \precsim_J \rangle_{J \in \mathbb{J}}$ belongs to $\mathcal{R}^{\mathbb{J}}$.

Claim 3. *For every $J \in \mathbb{J}$ and $q \in \mathbb{Q}_{>0}$, $\precsim_{qJ} = \precsim_J$.*

Proof of claim 3. We wish to apply claim 2. Let $J' \stackrel{\text{def}}{=} qJ$ and take $j, \kappa, \kappa' \in \mathbb{Z}_{>0}$ such that $\kappa J = K \in \mathbb{I}$ and $q = \frac{j}{\kappa'}$. Since $\kappa' J' = jJ$, we have $\kappa \cdot (\kappa' J') = \kappa \cdot (jJ) = jK$ and, by claim 2, $\precsim_{J'} = \precsim_K = \precsim_J$ $\square$

Together, A1-A3, A4$'$ and claim 3 ensure that $\precsim_{\mathbb{J}}$ satisfies

Claim 4.

A1* *For every $J \in \mathbb{J}$, $\precsim_J$ is transitive on X,*

A2* *For every $J \in \mathbb{J}$, $\precsim_J$ complete on X,*

A3* *For every $J, J' \in \mathbb{J}$, every $x, x' \in X$ and every $q, q' \in \mathbb{Q}_{>0}$, if $J'' = qJ + q'J'$, $x \precsim_J x'$ ($x \prec_J x'$) and $x \precsim_{J'} x'$, then*

$$x \precsim_{J''} x' \ (x \prec_{J''} x'),$$

A4* *For every $J, J' \in \mathbb{J}$ and every $x, x' \in X$ if $x \prec_J x'$, then there exists $q' \in \mathbb{Q} \cap (0, 1)$ such that*

$$x \prec_{(1-q)J+qJ'} x' \text{ for every } q \in \mathbb{Q} \cap (0, q').$$

Proof of claim 4. Of A1*–A4*, only A4* does not follow directly from the preceding arguments. Note that A4$'$ and the construction of $\precsim_{\mathbb{J}}$ ensure that $x \prec_{\kappa K+K'} x'$ for some $\kappa \in \mathbb{Z}_{>0}$ and $K, K' \in \mathbb{I}$ such that $jJ = K$ and $j'J' = K'$ for some $j, j' \in \mathbb{Z}_{>0}$. Let $z \stackrel{\text{def}}{=} \frac{1}{\kappa j + j'}$. Then take $q' = zj' \in \mathbb{Q} \cap (0, 1)$ and note that $1 - q' = z\kappa j$. Moreover,

$$J'' \stackrel{\text{def}}{=} (1 - q')J + q'J' = z(\kappa K + K').$$

Since $z \in \mathbb{Q}_{>0}$ and $\kappa I + I' \in \mathbb{J}$, claim 2 implies $\precsim_{J''} = \precsim_{\kappa I+I'}$. Thus, $x \prec_{J''} x'$, as required. Finally, take any $q \in \mathbb{Q} \cap (0, q')$. If it is the case that $(1-q)J + qJ'$ is rational combination of J and J'', then, via A3*, $x \prec_J x'$ and $x \prec_{J''} x'$ together imply $x \prec_{(1-q)J+qJ'} x'$, as required. From basic properties of the the real numbers, there exists $\mu < 1$ such that $q = \mu q'$ and, moreover, μ is rational. Then $(1 - q)J + qJ' = (1 - \mu)J + \mu J''$ follows from the fact that $q'(J' - J) = J'' - J$. $\square$

Step 1.3 (Y-extensions and A5 for rational-valued vectors). For a given extension $\mathcal{E}$, let $\mathbb{T}_{\mathcal{E}}$ denote the set of equivalence classes of $\sim^{\mathcal{E}}$. Observation 1 ensures that, for every $Y \subseteq X$, if $\mathcal{E}$ and $\mathcal{E}'$ are two proper Y-extensions, then the associated sets of case types are identical. As such, for every $Y \subseteq X$, the notation $\mathbb{T}_Y$ is well-defined to be the set of types generated by any given proper Y-extension $\mathcal{E}$. In turn, we let $\mathbb{S}_Y \stackrel{\text{def}}{=} \mathbb{T}_Y - [\mathfrak{n}]$. The statements of step 1.1 and step 1.2 carry over to arbitrary $Y \subseteq X$.

Let $\mathbb{L}_Y$ be the set of vectors

$$L : \mathbb{T}_Y \to \mathbb{Q}_{\geqslant 0} \text{ such that } 0 < \# \{s : 0 < L(s)\} < \infty.$$

The following is a translation of definition 2 to setting of rational vectors.

Definition 2. $\precsim_{\mathbb{L}_Y} = \langle \precsim_L : L \in \mathbb{L}_Y \rangle$ *is a* proper Y-extension *if, for some* $Y \subseteq X$*, there exists a function* $\mathrm{ev} : \mathbb{L}_Y \to \mathcal{R}_Y$*, such that*

1. *for every* $L \in \mathbb{L}_Y$*,* $\precsim_L \stackrel{\text{def}}{=} \mathrm{ev}(L)$*; and*
2. *for every* $J \in \mathbb{J}$*, if* $L = J \times 0$*, then* $\precsim_L = \precsim_J \cap Y^2$.

By step 1.1 and step 1.2, we are able to translate A5 into a statement involving rational vectors.

A5* For every $Y \subseteq X$, $\#Y = |Y|_{\mathcal{D}} = 3, 4$, and every proper Y-extension $\precsim_{\mathbb{L}_Y}$ that satisfies A2*–A4*, there exists a Y-extension that satisfies A1*–A4* and that features the rankings of $\precsim_{\mathbb{L}_Y}$.

Step 2 ($\#\mathbb{S} < \infty$ and $\#X = 2$). When $\mathbb{S}$ is finite, $\mathbb{J}$ coincides with $\mathbb{Q}^{\mathbb{S}}_{\geqslant 0} - \{0\}$. Throughout the sequel, whenever the set to which J belongs is suppressed, it is understood that $J \in \mathbb{J}$. We begin by extending lemma 1 of [GS] to allow for the fact that the Diversity axiom of [GS] may not hold.

Lemma 2. *For every $x, x' \in X$, there exists $v^{xx'}, v^{x'x} \in \mathbb{R}^{\mathbb{S}}$ such that*

1. $F_+^{xx'} \stackrel{\text{def}}{=} \{J : x \precsim_J x'\} = \left\{J : 0 \leqslant J \cdot v^{xx'}\right\}$
2. $G_+^{xx'} \stackrel{\text{def}}{=} \{J : x \prec_J x'\} = \left\{J : 0 < J \cdot v^{xx'}\right\}$
3. $v^{x'x} = -v^{xx'}$.

Proof of lemma 2. Fix an arbitrary pair $x \neq x'$ in X. If $G_+^{xx'}$ and $G_+^{x'x}$ are both nonempty, then the Diversity axiom holds for x, x' and our proof follows from that of lemma 1 of [GS]. W.l.o.g., we henceforth assume that $G_+^{x'x} = \varnothing$. By assumption 1 and the previous steps, $G_+^{xx'}$ is nonempty. (Note that the case where assumption 1 does not hold for x and x' is easily accounted for by taking $v^{xx'} = 0$).

Recall the definition of the basis vectors $\{J_s : s \in \mathbb{S}\}$ of claim 1. By A1* and the assumption that $G_+^{x'x} = \varnothing$, $\mathbb{S}$ is the disjoint union of $\mathbb{S}_{\prec} \stackrel{\text{def}}{=} \{s : x \prec_{J_s} x'\}$ and $\mathbb{S}_{\approx} \stackrel{\text{def}}{=} \{s : x \approx_{J_s} x'\}$. Take any $v^{xx'}$ such that $v^{xx'}(s) > 0$ if $s \in \mathbb{S}_{\prec}$ and $v^{xx'}(s) = 0$ otherwise. Since $J \cdot v^{xx'} \geqslant 0$ for every $J \in \mathbb{J}$, part 1 of the lemma holds.

For part 2 of the lemma we will show that $J \cdot v^{xx'} = 0$, if, and only if, $x \approx_J x'$. Fix an arbitrary $J \in \mathbb{J}$ and let $\mathbb{S}_J \stackrel{\text{def}}{=} \{s : J(s) > 0\}$ and note that $\mathbb{S}_J$ is nonempty because $J \in \mathbb{J}$. It suffices to show that $\mathbb{S}_J \subseteq \mathbb{S}_{\approx}$ if, and only if, $x \approx_J x'$. Let $1, \ldots, k$ be an enumeration of $\mathbb{S}_J$, and let $J_1, \ldots, J_k$ be the corresponding basis vectors. Then $J = q_1 J_1 + \cdots + q_k J_k$ for some $q_1, \ldots, q_k$. If $\mathbb{S}_J \subseteq \mathbb{S}_{\approx}$, then $x \approx_{J_j} x'$ for every $j = 1, \ldots, k$ and $k - 1$ applications of A3* yield $x \approx_J x'$. Now suppose that $J(s) > 0$ for some $s \in \mathbb{S}_{\prec}$, so that $\mathbb{S}_J \nsubseteq \mathbb{S}_{\approx}$. Then, for some j, $x \prec_{J_j} x'$. Since $G_+^{x'x} = \varnothing$, we have $x \precsim_{J_i} x'$ for every $i \neq j$ and, via $k - 1$ applications of A3*, we obtain $x \prec_J x'$.

For part 3, let $v^{x'x} \stackrel{\text{def}}{=} -v^{xx'}$. Then $v^{x'x}$ itself satisfies part 2 of lemma 2 since both $G_+^{x'x}$ and $\left\{J : 0 < -J \cdot v^{xx'}\right\}$ are empty. In turn, this implies that $F_+^{x'x} = \{J : x \approx_J x'\}$, and the preceding paragraph implies that $v^{x'x}$ satisfies part 1. $\square$

Lemma 2 completes the proof that the axioms are sufficient for a representation when $\#\mathbb{S} < \infty$ and $\#X = 2$. This completes step 2.

For every $x, x' \in X$, let $N_+^{x\,x'} \overset{\text{def}}{=} \{J : x \approx_J x'\}$. For each $x, x' \in X$ and each $v^{x\,x'}$ that satisfies lemma 2, let

$$G_v^{x\,x'} \overset{\text{def}}{=} \left\{J \in \mathbb{R}^{\mathbb{S}} : 0 < J \cdot v^{x\,x'}\right\} \text{ and } H_v^{x\,x'} \overset{\text{def}}{=} \left\{J \in \mathbb{R}^{\mathbb{S}} : J \cdot v^{x\,x'} = 0\right\}.$$

By assumption 1, $v^{x\,x'} \neq 0$ and $H_v^{x\,x'}$ is a well-defined hyperplane.

If $u^{x\,x'}$ is another vector that satisfies the conditions of lemma 2 and, for every $\lambda \in \mathbb{R}$, $u^{x\,x'} \neq \lambda v^{x\,x'}$, then $H_u^{x\,x'}$ and $H_v^{x\,x'}$ are distinct hyperplanes. The three mutually exclusive sufficient conditions for $H_v^{x\,x'}$ to be the unique hyperplane that contains $N_+^{x\,x'}$ are as follows:

- Both $G_+^{x\,x'}$ and $G_+^{x'x}$ are nonempty (the setting of [GS]); or
- $\mathbb{S} - \mathbb{S}_{\approx}^{x\,x'}$ is a singleton; or
- $G_+^{x\,x'} = \varnothing$ and $|\mathbb{S} - \mathbb{S}_{\approx}^{x\,x'}| > 1$, but, via the axioms, $H_v^{x\,x'}$ is determined by $\{N^{y\,z} : y \neq x \text{ or } z \neq x'\}$.

The scenarios that arise when the last of these conditions holds is central to the arguments that follow.

Step 3 ($\#\mathbb{S} < \infty$ and $\#X = 3$). Let $X = \{x, x', x''\}$. We break this step down into three substeps: $\#\mathbb{S} = 1, 2$ and $3 \leqslant \#\mathbb{S} < \infty$.

Step 3.1 ($\#\mathbb{S} = 1$). Fix $J \in \mathbb{J}$. Then, since theorem 1 requires that $|X|_{\mathcal{D}}$ is countable, A1* and A2* alone suffice for the existence of a function $v_J : X \to \mathbb{R}$ such that, for every $y, z \in X$,

$$y \precsim_J z \quad \text{if, and only if,} \quad v_J(y) \leqslant v_J(z).$$

Since $\#\mathbb{S} = 1$, the definition of $\sim^\star$ implies that $\precsim_J = \precsim_{J'}$ for every $J, J' \in \mathbb{J}$. Now take $v : X \times \mathbb{S} \to \mathbb{R}$ to satisfy $v(\cdot, s) = v_J(\cdot)$ for every $s \in \mathbb{D}$. Then, by construction, v satisfies (i) and (ii) of theorem 1.

Since $\#\mathbb{S} = 1$, A3* and A4* both follow from the observation

$$\precsim_J = \precsim_{J'} \text{ for every } J, J' \in \mathbb{J}.$$

It remains to be shown that, in the presence of A1* and A2*, the same is true of A5. For if the same is true, then the set of $\precsim_{\mathbb{J}}$ that satisfy A1* and A2* coincide with the set of $\precsim_{\mathbb{J}}$ that satisfy A1*–A5*. Then the preceding paragraph completes the proof, since A1*–A2* hold if, and only if, there exists a representation that satisfies (i) and (ii).

Step 3.2 ($\#\mathbb{S} = 2$). Recall that $0 \notin \mathbb{J}$ and suppose that $N^{xx'} \cap N^{xx''} \cap N^{x'x''}$ contains some vector J. Then, since $\#\mathbb{S} = 2$, $N^{xx'} = N^{xx''} = N^{x'x''}$ by claim 2. Let H be the hyperplane generated by 0 and J and let G and G' be the respective open half spaces generated by H. W.l.o.g., let $(x, x')(x', x'')$ be the pre-chain associated with G. Then $(x'', x')(x', x)$ is necessarily the pre-chain associated with G'. Clearly, the vectors we need should generate the open half spaces $G_v^{xx'}$, $G_v^{x'x''}$ and $G_v^{xx''}$ all equal to G. Let $v^x = 0$ and take $v^{x'} \in H_\perp$ (the linear space that is perpendicular to H) such that $J \cdot v^{x'} > 0$ if, and only if, $J \in G$. Finally, take $v^{x''} = 2v^{x'}$. Then, for every $J \in \mathbb{R}^{\mathbb{S}}$, $J \cdot (v^{x'} - v^{x''}) > 0$ if, and only if, $J \in G$, as required.

If on the other hand $N^{xx'} \cap N^{xx''} \cap N^{x'x''} = \varnothing$, then assumption 1 and the arguments of the previous paragraph together imply that any triple of vectors $v^{\cdot\cdot} \overset{\text{def}}{=} \left\{v^{ij} : i, j \in X\right\}$ that represent $\precsim_{\mathbb{J}}$ generates three distinct hyperplanes H_v^{ij}. By claim 2, the arrangement $\left\{H_v^{ij} : i, j \in X\right\}$ is linear. Since $\#\mathbb{S} = 2$, any two of the three vectors span $\mathbb{R}^{\mathbb{S}}$ (and thus the third vector). The arguments of lemma 2 [GS] then suffice to show that $v^{\cdot\cdot}$ can be chosen so as to satisfy the

Jacobi identity ******that we discuss****** next as part of the proof for the scenario $\#\mathbb{S} \geqslant 3$.

STEP 3.3 ($3 \leqslant \#\mathbb{S} < \infty$). First consider the scenario where every representation $v^{\cdot\cdot}$ of $\precsim_{\mathbb{J}}$ fails to satisfy the Jacobi identity. That is, $v^{xx'}+v^{x'x''} \neq v^{xx''}$. Then every representation $v^{\cdot\cdot}$ is such that the three vectors in $v^{\cdot\cdot}$ are linearly independent and there is no representation of $\precsim_{\mathbb{J}}$ that satisfies theorem 1.

Fix an arbitrary representation $v^{\cdot\cdot}$. By linear independence, the hyperplanes form a linear arrangement that is in general position. As such, by Zaslavsky's theorem and the fact that $\#\mathbb{S} \geqslant 3$, they separate $\mathbb{R}^{\mathbb{S}}$ into the maximal number of regions: eight. Thus, in addition to the $3! = 6$ regions regions where, together, $0 \leqslant J \cdot v^{xx'}$ and $0 \leqslant J \cdot v^{x'x''}$ imply $J \cdot v^{xx''} < 0$, there exists an open cone G such that, for every $J \in G$, $0 \leqslant J \cdot v^{xx'}$, $0 \leqslant J \cdot v^{x'x''}$ and $J \cdot v^{xx''} < 0$. The fact that $G \subset \mathbb{R}^{\mathbb{S}} - \mathbb{R}^{\mathbb{S}}_{\geqslant 0}$ follows from A1*-A4*.

Since the arrangement is linear, every neighbourhood N_0 of 0 in $\mathbb{R}^{\mathbb{S}}$ has nonempty intersection with every chamber in the arrangement. We will now derive a violation of A5.

Let $\mathbb{L} = \mathbb{Q}^{\mathbb{T}}_{\geqslant 0} - \{0\}$. We choose $\alpha^{\cdot\cdot} \stackrel{\text{def}}{=} \left\{\alpha^{ij} : i, j \in X\right\}$ such that, for each $i, j \in X$, $\langle (v \times \alpha)^{ij}, L \rangle = 0$ for some strictly positive $L \in \mathbb{L}$. This is possible by lemma 3. Endow $\mathbb{L}$ with the subspace topology, so that $G \subseteq \mathbb{L}$ is open if, and only if, $G = \mathbb{L} \cap G'$ for some open $G' \subset \mathbb{R}^{\mathbb{T}}$. Take any neighbourhood $N_L \subset \mathbb{L}$ of L.

In this paragraph, we will show that there is an isomorphism between the set of chambers that $v^{\cdot\cdot}$ generates in $\mathbb{R}^{\mathbb{S}}$ and the set of chambers that $(v \times \alpha)^{\cdot\cdot}$ generates in N_L. Pick a chamber G of $v^{\cdot\cdot}$ and an arbitrary member $J \in G$. Then $J \times 0 \in \mathbb{R}^{\mathbb{T}}$ and the ordering of X implied by the inner products $\langle v^{\cdot\cdot}, J \rangle_{\mathbb{S}}$ and the inner products

$\langle (v \times \alpha)^{\cdot\cdot}, J \times 0 \rangle_{\mathbb{T}}$ is the same. Let $K_\lambda : [0,1] \to \mathbb{L}$ be the map $\lambda \mapsto (1-\lambda)(J \times 0) + \lambda L$. Then, for every $\lambda < 1$, the ordering implied by $\langle (v \times \alpha)^{\cdot\cdot}, K_\lambda \rangle_{\mathbb{T}}$ coincides with the ordering at J. Finally note that, for some λ sufficiently close to one, $K_\lambda \in N_L$. Since G was arbitrary, this completes the argument.

Let $\precsim_{\mathcal{E}}$ be the extension that satisfies $y \precsim_L z$ if, and only if, $0 \leqslant \langle (v \times \alpha)^{yz}, L \rangle_{\mathbb{T}}$. We have shown that every pre-chain features in $\mathcal{E}$. Moreover, $\mathcal{E}$ satisfies A2–A4. Now suppose that A5 holds, so that there exists $\mathcal{E}'$ that satisfies A1–A4 and that features the rankings of $\mathcal{E}$. But then $\mathcal{E}'$ on $\mathbb{L}$ satisfies the diversity condition and hence all the axioms of [GS]. This in turn implies the existence of a representation $u : X \times \mathbb{T} \to \mathbb{R}$ such that, for every $y, z \in X$ and every $L \in \mathbb{L}$,

$$\langle u^y, L \rangle_{\mathbb{T}} \leqslant \langle u^z, L \rangle_{\mathbb{T}} \text{ if, and only if, } y \precsim_L z,$$

where $\precsim_L$ is determined by $\mathcal{E}'$. Now consider the restriction $w : X \times \mathbb{S} \to \mathbb{R}$ of u such that $w(\cdot, s) = u(\cdot, s)$ for every $s \in \mathbb{S}$. Then let $w^{xx'} = w^{x'} - w^{x}$, $w^{x'x''} = w^{x''} - w^{x'}$ and $w^{xx''} = w^{x''} - w^{x}$. It is clear that $w^{xx''} = w^{xx'} + w^{x'x''}$, so that the Jacobi identity holds. Finally, we arrive at the desired contradiction by noting the definition of an extension is such that $\mathcal{E}'$ and $\mathcal{E}$ coincide on $\mathbb{J}$, so that $w^{\cdot\cdot}$ is a representation of $\precsim_{\mathbb{J}}$ that satisfies the Jacobi identity.

****

The goal of our proof is to show that the Jacobi identity $v^{xx''} = v^{xx'} + v^{x'x''}$ holds for some $\{v^{ij} \in \mathbb{R}^{\mathbb{S}} : i, j \in X\}$. If this identity holds, then we may choose $v^{x} = 0$ and $v^{x'} = v^{xx'}$ and $v^{x''} = v^{xx''}$. Then, for $i = x', x''$, $J \cdot v^{x} \leqslant J \cdot v^{i}$ if, and only if, $x \precsim_J x'$. Moreover, since $v^{x'x''} = v^{x''} - v^{x'}$ and $0 \leqslant J \cdot v^{x'x''}$ if, and only if, $x' \precsim_J x''$, we see that the vectors $\{v^{i} \in \mathbb{R}^{\mathbb{S}} : i \in X\}$ are suitable for our purpose. In particular, to obtain a matrix $v : X \times \mathbb{D} \to \mathbb{R}$ that satisfies both (i) and (ii) of theorem 1, let $v(x,c) = 0$ for every $c \in \mathbb{D}$; and, for $i = x', x''$ and for each $s \in \mathbb{S}$, let $v(i,c) = v^i(s)$ if, and only if, $c \in s$.

We recall some facts about hyperplane arrangements (see Orlik and Terao [10]). Let $\mathcal{H} = \{H_1, \ldots, H_n\}$ denote an *arrangement* of n distinct hyperplanes in $\mathbb{R}^{\mathbb{S}}$ and let $\mathrm{ch}(\mathcal{H})$ denote the set of chambers (nonempty components) of $\mathbb{R}^{\mathbb{S}} - \bigcup_i^n H_i$ by $\mathrm{ch}(\mathcal{H})$. We will only need to consider the subarrangements $\mathcal{B} \subseteq \mathcal{H}$ such that $|\mathcal{B}| = 3, 4$. In our setting, for $|\mathcal{B}| = 3$, a typical example of a chamber will be the set $\{D : x \prec_D y \prec_D z\}$. $\mathcal{H}$ is a *linear* arrangement iff $0 \in \bigcap_i^n H_i$. When $\mathcal{H}$ is linear, each of its chambers is an open cone in $\mathbb{R}^{\mathbb{S}}$.

We also introduce some notions that are useful for the arguments that follow. $\mathcal{H}' = \{H_1', \ldots, H_n'\}$ is an extension of $\mathcal{H}$ in $\mathbb{R}^{\mathbb{T}}$ iff $\mathbb{R}^{\mathbb{S}} \cap H_i' = H_i$ for $i = 1, \ldots n$. Similarly, $G' \in \mathrm{ch}(\mathcal{H}')$ extends $G \in \mathrm{ch}(\mathcal{H})$ if $\mathbb{R}^{\mathbb{S}} \cap G' = G$. We will also find it useful to refer to the *empty chambers* $\mathrm{ech}(\mathcal{H})$ of $\mathcal{H}$. These are "chambers" that are empty because the hyperplanes intersect in a given way. For instance, when A1 holds, the set $\{D \in \mathcal{D} : x \prec_D y \prec_D z \prec_D x\}$ is empty. (Indeed, A1 ensures that $H^{xy} \cap H^{yz} \subseteq H^{xy} \cap H^{yz} \cap H^{xz}$, so that the two are in fact equal.) For $|\mathcal{B}| = 3$, when chambers are indexed by strict rankings, the largest possible number of chambers is 8: 3! transitive orderings and 2 cycles. Note that all 6 transitive orderings are only attainable if $|\mathbb{T}| \geqslant 3$, that is if $|\mathbb{S}| \geqslant 2$. For $|\mathcal{B}| = 3$, when chambers are indexed by strict rankings, the largest possible number of chambers is 8: 3 transitive orderings and 2 cycles.

The following lemma captures the key consequence of A5: that any inconsistency that is hidden in $\precsim_{\mathcal{D}}$ is revealed by some $\precsim_{\mathrm{ev}(A)}$. In this lemma, open means relative to the weak* topology on $\mathbb{R}^{\mathbb{T}}$. We recall that this is the weakest topology such that the function $\langle v, \cdot \rangle : \mathbb{R}^{\mathbb{T}} \to \mathbb{R}$

is continuous for every v in the dual space $\left(\mathbb{R}^{\mathbb{T}}\right)^*$.

We will appeal to the contrapositive of the lemma that follows. That is, we require that if every no necessary intransitive chamber is revealed by any full extension, then there is no "hidden chamber" that is intransitive. In this case, the Jacobian identity holds on $\mathbb{R}^{\mathbb{S}}$ and we may apply the results of [GS] en route to proving sufficiency of the axioms.

Lemma 3. *Let $\mathcal{H}$ be a linear arrangement in $\mathbb{R}^{\mathbb{S}}$. If $G \in \mathrm{ch}(\mathcal{H})$, then there is a linear extension $\mathcal{H}'$ of $\mathcal{H}$ in $\mathbb{R}^{\mathbb{T}}$ such that $G' \cap \mathbb{R}^{\mathbb{T}}_{>0}$ is a nonempty open cone, where $G' \in \mathrm{ch}(\mathcal{H}')$ extends G.*

Proof of lemma 3. Take n vectors $u^1, \ldots, u^k \in \left(\mathbb{R}^{\mathbb{S}}\right)^*$, each normal to a hyperplane of $\mathcal{H}$, and such that, for every $J \in G$ and every $k = 1, \ldots, n$, the inner product $\langle u^k, J\rangle$ is positive. ‘ For each $\alpha \in \mathbb{R}$ and k, let $u^k \times \alpha$ denote the corresponding vector in $\left(\mathbb{R}^{\mathbb{T}}\right)^*$. Let $J' = J \times j$ be any (strictly) positive vector in $\mathbb{R}^{\mathbb{T}}$ and fix k. Note that, for each $\alpha \in \mathbb{R}$, $\langle u^k \times \alpha, J'\rangle_{\mathbb{T}} = \langle u^k, J\rangle_{\mathbb{S}} + \alpha \cdot j$. Since $\langle u^k, J\rangle_{\mathbb{S}} \in \mathbb{R}$ and $j > 0$, there exists $\alpha^k > 0$ such that $\langle u^k \times \alpha^k, J'\rangle_{\mathbb{T}}$ is positive. By continuity of the map $J'' \mapsto \langle u^k \times \alpha^k, J''\rangle_{\mathbb{T}}$, there exists an open neighbourhood G^k of J' such that $\langle u^k \times \alpha^k, J''\rangle_{\mathbb{T}} > 0$ for every $J'' \in G^k$.

The preceding argument holds for $k = 1, \ldots, n$ and, since n is finite, $G = \bigcap_k^n G_k$ is open an open neighbourhood of J' in $\mathbb{R}^{\mathbb{T}}$. It remains to show that we can form a nonempty cone. For every positive $\rho \in \mathbb{R}$, every $J'' \in G$ and every $i = 1, \ldots, n$, the inner product $\langle u^k \times \alpha^k, \rho J''\rangle_{\mathbb{T}}$ is positive since it is equal to $\rho \cdot \langle u^k \times \alpha^k, J''\rangle_{\mathbb{T}}$. Let

$$G' = \left\{ J' \in \mathbb{R}^{\mathbb{T}}_{>0} : \langle u^k \times \alpha^k, J'\rangle_{\mathbb{T}} > 0 \text{ for } k = 1, \ldots, n \right\}.$$

We conclude that G' is the required extension of G by the following observation: for each $J \in G$, if $J' = J \times 0$, then $\langle u^k \times \alpha^k, J'\rangle_{\mathbb{T}} = \langle u^k, J\rangle_{\mathbb{S}}$ for each k. This completes the proof of lemma 3. $\square$

By lemma 3, the linear arrangements generated by $\precsim_{\mathcal{D}}$ satisfying A1-A5 are transitive. That is,

Lemma 4. *For every triple $x, y, z \in X$, the intersection of half spaces $B^{xy} \cap B^{yz} \cap B^{zx} \subseteq \mathbb{R}^{\mathbb{S}}$ generated by $\mathcal{H}_{\precsim_{\mathcal{D}}}$ has empty interior whenever it is determined.*

Proof of lemma 4. Suppose otherwise that $G \overset{\text{def}}{=} A^{xy} \cap A^{yz} \cap A^{zx} \neq \varnothing$ for some $x, y, z \in X$. Note that this is only possible if $|\mathbb{S}| \geqslant 3$. (For if $|\mathbb{S}| = 2$, then $G \neq \varnothing$ and the fact that $0 \in H^{xy} \cap H^{yz} \cap H^{zx}$ implies that, for each $i \neq j$ in $\{x, y, z\}$, $\dim H^{ij}$ is indeed a hyperplane: of dimension 1.)

Then lemma 3 implies the existence of a linear extension $\mathcal{H}'_{x,y,z}$ such that the corresponding intersection $A'^{xy} \cap A'^{yz} \cap A'^{zx} \subseteq \mathbb{R}^{\mathbb{T}}_{>0}$ is a nonempty open cone and such that w.l.o.g. $H'^{xy} \cap H'^{yz}$ is a nonempty subset of $\mathbb{R}^{\mathbb{T}}$. Then let $\mathcal{E}$ be the extension of $\mathcal{D}$ that is restricted to the complement in $\mathbb{N}^{\mathbb{T}}$ of the union of the two intransitive chambers

$$A'^{xy} \cap A'^{yz} \cap A'^{zx} \text{ and } W'_{xy} \cap W'_{yz} \cap W'_{zx} \text{ in } \mathcal{H}'_{x,y,z}.$$

and such that $x \precsim_D y$ if and only $D \in A'^{xy} \cap \mathbb{N}^{\mathbb{T}}$ and similarly for the pairs (y, z) and (x, z). Then, by construction, $\precsim_{\mathcal{E}}$ satisfies A1–A4.

It remains to be shown that there is no extension $\precsim_{\text{ev}(\mathcal{A})}$ of $\precsim_{\mathcal{E}}$ for which A1–A4 holds. We will show that there is only one full extension $\precsim_{\text{ev}(\mathcal{A})}$ of $\precsim_{\mathcal{E}}$ that satisfies A3 and A4: the one that satisfies $x \prec_D y \prec_D z \prec_D x$ for for every $D \in A'^{xy} \cap A'^{yz} \cap A'^{zx} \cap \mathbb{N}^{\mathbb{T}}$. (This will complete the proof because A1 does not hold for $\precsim_{\text{ev}(\mathcal{A})}$.)

It suffices to show that each of the hyperplanes in $\mathcal{H}'_{x,y,z}$ is determined. W.l.o.g., we show that it is true for H^{xy}. $x \prec_D y$ Since A1–A4 hold on $\mathcal{D}$ and members of $\mathcal{H}'$ are hyperplanes, in the subspace topology (obtained by considering $\mathbb{N}^{\mathbb{T}}$ as a topological subspace of $\mathbb{R}^{\mathbb{T}}$, there is an nonempty open neighbourhood G of $\mathcal{D}$ on which A1–A4 holds.

$\square$

The next lemma captures an important effect of **??**: that chambers that are hidden in

Lemma 5. *Let $\mathcal{H}$ be a transitive linear arrangement in $\mathbb{R}^{\mathbb{S}}$ and let $\mathcal{B} \subseteq \mathcal{H}$ satisfy $|\mathcal{B} = 4|$. If there exists a linear extension $\mathcal{B}'$ of $\mathcal{B}$ in $\mathbb{R}^{\mathbb{T}}$ such that $|\mathrm{ch}(\mathcal{B}')| = 24$, then $|\mathrm{ch}(\mathcal{H})| > 12$.*

Proof of lemma 5. $\square$

In case 1, case 2 and case 3 condition **??** and **??** hold vacuously.

Case 1 ($|X| = 1$). Since $X = \{x\}$, A1 and A2 together ensure that $\prec_D$ is empty for every $D \in \mathcal{D}$. Thus, there is only one type of case in $\mathcal{D}$ ($|\mathbb{T}| = 2$) and the zero matrix satisfies conditions (i) and (ii).

Conversely, for any v that satisfies the conditions, (ii) implies that $x \approx_D x$ for every $D \in \mathcal{D}$ and so once again there is only one type of case in $\mathcal{D}$ and **??** and A3 hold. (A4 holds vacuously.) For A5, we suppose that $\precsim_{\mathcal{E}}$ is an arbitrary extension of $\precsim_{\mathcal{D}}$ that satisfies **??**–A4 and prove that the conclusion of A5 holds. Clearly, if **??** holds on $\mathcal{E}$, then $x \approx_D x$ for every $D \in \mathcal{E}$. Moreover, if we take $\mathrm{ev}(\mathcal{A})$ such that $x \approx_D x$ for every $D \in \mathrm{ev}(\mathcal{A})$, then A5 holds.

Case 2 ($|X| = 2$). Let $X = \{x, y\}$. For the subcase where $x \approx_D y$ for every $D \in \mathcal{D}$, the arguments of case 1 apply. Consider the subcase where $x \prec_D y$ for every $D \in \mathcal{D}$. In this case, like in case 1 there is only one case type in $\mathcal{D}$. Let $v(x, \cdot)$ be the zero vector in $\mathbb{R}^{\mathcal{D}}$ and, similarly let $v(y, \cdot)$ be a vector of ones. Then conditions (i) and (ii) hold. Conversely, suppose that $\precsim_{\mathcal{D}}$ are represented by the latter pair of vectors. Then $\prec_D$ for every $D \in \mathcal{D}$. Thus, our only remaining concern is to show that A5 holds.

Case 3 ($|X| = 3$).

Case 4 ($|X| \geqslant 4$ and $|\mathbb{T}| \leqslant 3$). Since both the axioms and the representation fail to hold, the proof of this case is complete.

Case 5 ($|X| \geqslant 4$ and $|\mathbb{T}| \geqslant 4$).

Step 4 (Necessity).

□

3.2 Uniqueness

A forecaster with just one case type reveals that her experiences are perhaps insufficiently diverse to do case-based prediction with the above model. It is not that our representation precludes such a predictor, but rather the fact that the best we can hope for without resorting to further conditions on preferences and on the space of eventualities, is an ordinal representation. There is no hope of deriving a weighting function that is suitably unique except in the rather exceptional and uninteresting scenario where there are no $\mathcal{D}$-distinct elements in X. One could argue that our Zoologists in the black swan example have a single case type, especially if there are just two eventualities (the swans are either black or white). On the other hand, one could also argue that a more careful examination of the data would reveal that, black-and-white swans had already been discovered in South America long before 1697. This would constitute a second case type and warrant a departure from any model that predicts all swans are white for every subsample. (Though there would still be no evidence to suggest that all the swans in New Holland are black.)

A6 (Weak 2-Diversity)**.** *For every $x, y \in X$, if there exists $D \in \mathcal{D}$ such that $x \prec_D y$, then there exists $D' \in \mathcal{D}$ such that $y \prec_{D'} x$.*

When A6 holds, the scenario of a single case type only arises if there are no $\mathcal{D}$-distinct elements in X. Thus, the following corollary holds regardless of the number of case types.

Corollary 1. *If $v : X \times \mathbb{D} \to \mathbb{R}$ satisfies* (i) *and* (ii) *of theorem 1 and A6 holds, then for every $w : X \times \mathbb{D} \to \mathbb{R}$ that satisfies* (i) *and* (ii)*, there exists $\lambda > 0$ and, for each $c \in \mathbb{D}$, $\mu_c \in \mathbb{R}$ such that $w(\cdot, c) = \lambda v(\cdot, c) + \mu_c$.*

There are significantly weaker conditions than A6 that deliver uniqueness in the sense of corollary 1. In the following discussion we identify the minimal condition for identifying a unique matrix representation even though this is not a behavioural condition, it is useful since it also provides a simple way to test the model.

4 Discussion

4.1 Comparison with other approaches

The neo-Bayesian framework of Schmeidler [12], Quiggin [11] and many others since allows for ambiguity or nonuniqueness of beliefs, but it still assumes the predictor has a complete description of the states of the world. Updating of beliefs remains a futile exercise if states were absent in the first place. By allowing for nonstandard state spaces, where the set of states an agent has access to is nonexhaustive, the literature on unawareness of **FH-Unawareness**, **DLR-Nonstandard_state_spaces**, Heifetz, Meier, and Schipper [7] comes much closer to addressing the present concerns. Indeed, by allowing the agent to explicitly model the fact that the predictor acknowledges the possibility that her set of states is nonexhaustive, this literature would seem to address all the concerns that the predictor will confront in a large world setting.

However, there is also complexity to consider. Games of incomplete information where there is uncertainty about a single proposition, but where infinite hierarchies of knowledge are relevant, yield uncountably many states of the world. Even a player that is aware of this fact cannot accommodate such complexity and, from a positive (as opposed to normative) perspective, it is unreasonable to expect players to reason up through hierarchies of knowledge unless she has found it fruitful to do so in the past.

Finally, as [GS]point out,

A Proofs

References

[1] Y. Azrieli. "Axioms for Euclidean preferences with a valence dimension". In: *Journal of Mathematical Economics* 47 (2011), pages 545–553.

[2] I. Gilboa and D. Schmeidler. *A theory of case-based decisions*. Cambridge University Press, 2001.

[3] I. Gilboa and D. Schmeidler. "Act similarity in case-based decision theory". In: *Economic Theory* 9.1 (1997), pages 47–61.

[4] I. Gilboa and D. Schmeidler. "Inductive Inference: An Axiomatic Approach". In: *Econometrica* 71.1 (2003), pages 1–26.

[5] J. Y. Halpern and L. C. Rêgo. “Reasoning about knowledge of unawareness”. In: *Games and Economic Behavior* 67.2 (2009), pages 503–525.

[6] J. Y. Halpern and L. C. Rêgo. “Reasoning about knowledge of unawareness revisited”. In: *Mathematical Social Sciences* 65.2 (2013), pages 73–84.

[7] A. Heifetz, M. Meier, and B. C. Schipper. “Interactive unawareness”. In: *Journal of Economic Theory* 130.1 (2006), pages 78–94.

[8] E. Karni and M.-L. Vierø. “Awareness of unawareness: a theory of decision making in the face of ignorance”. In: *Journal of Economic Theory* 168 (2017), pages 301–328.

[9] P. H. D. O’Callaghan. “Context and Decision: three approaches to representing preferences”. In: *University of Warwick PhD Thesis* (2012).

[10] P. Orlik and H. Terao. *Arrangements of hyperplanes*. Volume 300. Springer Science & Business Media, 2013.

[11] J. Quiggin. “A theory of anticipated utility”. In: *Journal of Economic Behavior & Organization* 3.4 (1982), pages 323–343.

[12] D. Schmeidler. “Subjective probability and expected utility without additivity”. In: *Econometrica* 57.3 (1989), pages 571–587.