Ambiguity, Data and Preferences for Information — A Case-Based Approach[1]

Jürgen Eichberger[2] and Ani Guerdjikova[3]

This version: April 24, 2011

Abstract

In this paper we suggest a behavioral approach to decision making under ambiguity based on the available information. A decision situation is characterized by a set of actions, a set of outcomes, and an information context represented by a data-set containing action-outcome pairs. Decision-makers express preferences over the choice of action in a particular information context. Data-sets containing a larger number of observations, while keeping the frequency of observations constant, are considered less ambiguous. We derive a representation of preferences, which separates utility and beliefs. While the utility function is purely subjective, the beliefs of the decision maker combine objective characteristics of the data (number and frequency of observations) with the subjectively perceived relevance of observations (similarity). We identify the subjectively perceived degree of ambiguity and separate it into ambiguity due to a limited number of observations and ambiguity due to data heterogeneity. We also determine the decision maker's attitude towards ambiguity. The special case of no ambiguity represents beliefs as similarity-weighted frequencies. It provides a behavioral foundation for Billot, Gilboa, Samet and Schmeidler's (2005) representation of the mapping from data sets to probabilites over outcomes.

[1] We would like to thank to Larry Blume, Eric Danan, David Easley, Gaby Gayer, Itzhak Gilboa, John Hey, Edi Karni, David Kelsey, Nick Kiefer, Peter Klibanoff, Mark Machina, Francesca Molinari, Ben Polak, Karl Schlag, David Schmeidler, conference participants at RUD 2008 in Oxford, FUR 2008 in Barcelona and at the Workshop on Learning and Similarity 2008 in Alicante, as well as seminar participants at Cornell, UC Davis, Cergy Pontoise and George Washington University for helpful suggestions and comments.

[3] University of Heidelberg, Alfred Weber Institute, Grabengasse 14, 69117 Heidelberg, Germany, e-mail: juergen.eichberger@awi.uni-heidelberg.de

[3] Corresponding author. University of Cergy, THEMA, 33 bd. du Port, 95011 Cergy Pontoise, France, e-mail: ani.guerdjikova@u-cergy.fr

1 Introduction

The notion of induction is fundamental to human knowledge, Hume (1748). "Reasoning rests on the principle of analogy", writes Knight (1921, p. 199), "We judge the future by the past". In practice, decisions are often informed by data consisting of past observations. Randomized statistical experiments represent an ideal method of data collection, since the recorded information can be directly aggregated into a probability distribution over outcomes. In contrast, in real-life decision situations, the available data might contain a limited number of heterogenous observations with differing degrees of relevance for the decision situation at hand. Ellsberg (1961, p.657) summarizes the problem as follows: "What is at issue might be called the ambiguity of this information, a quality depending on the amount, type, reliability, and "unanimity" of information, and giving rise to one's degree of "confidence" in an estimate of relative likelihoods."

This naturally raises the question of how decision makers aggregate information in the form of data into beliefs over outcomes. In this paper, we pursue a behavioral approach to address this question. We use the case-based framework pioneered by Gilboa and Schmeidler (2001) to study agents who have to choose actions in the face of data. The natural primitives of the model are the decision maker's preferences over pairs of actions and data sets. We provide examples that illustrate the meaning of such preferences and show that they can be studied in laboratory experiments or identified in actual decision situations. We impose axioms which allow us to deduce agents' beliefs about uncertain outcomes and directly relate them to the content of the available data. We also derive a representation of preferences which shows how data influences the evaluation of actions. Furthermore, our approach allows us to determine how decision makers value information of different type.

The analysis of decisions based on data is important for several reasons: first and foremost, data represent a major source of information, but do not fit the standard classification of information into risk and uncertainty. While data carry objective information about the stochastic process of outcomes, this information might be insufficient to point-identify the probability distribution of outcomes. Limited number of observations, heterogeneity of observations and missing data are the main reasons for this ambiguity. Hence, the decision maker's beliefs will have to combine the objective characteristics of the data (such as number, frequency and type of observations) with subjective considerations (such as perception of ambiguity and similarity between observations).

The two common approaches towards beliefs cannot capture this type of considerations. In the state-based approach established by Savage (1954) a purely subjective probability distribution over states can be deduced from the choice behavior of the decision maker. Additional information is used to exclude states known not to have occurred, whereupon the subjective prior is updated according to Bayes rule. This purely behavioral approach leaves no room for using data inductively in the process of belief formation. For a Bayesian decision maker data must not be in contradiction with the description of states. Information may render the prior beliefs more precise but cannot change them.

The objective approach integrates available information directly into the decision maker's beliefs. For the special case of explicit randomizing devices, such as dice, roulette wheel, or urns containing balls of different colors, the probabilities of outcomes can be easily deduced. In the case, in which probabilistic information is not directly available, non-Bayesian statisticians use observed frequencies in the data to deduce probabilities of outcomes. Since they do not commit to a particular prior distribution, data cannot be contradictory. Yet, in this case, the number and the relevance of observations becomes an issue.

The objective and the subjective approach coincide when the amount of available data is large, and thus, the frequencies of observations approximate the true distribution of outcomes. In general, however, both approaches fail to capture important aspects of the decision situation. The former neglects the problem of ambiguity and heterogeneity of observations, while the latter completely detaches prior beliefs from relevant information.

Nevertheless, these two approaches represent limit cases of our analysis: beliefs based on a large data set resulting from a randomized statistical experiment will be (almost) objective, while purely subjective beliefs seem appropriate when the data set is small or contains cases of limited relevance to the decision at hand. The set of all possible data sets spans a universe of possible scenarios, in which both objective and subjective factors determine beliefs. Our model identifies the factors which govern the process of belief formation.

The second important insight gained from studying case-based decisions concerns the influence of data on human behavior. The institutional arrangements in an economy have an impact on the amount and type of information provided to economic agents. The Ellsberg (1961) paradox illustrates the difference in behavior in situations of risk and ambiguity. Data sets allow us to vary the ambiguity of information by changing the length of the data set, while leaving the frequencies of observations constant. The representation derived in the paper incorporates the idea that behavior depends on the perceived ambiguity of the data, as well as on the type of observations. Hence, it allows us to study the impact of different information environments on human behavior.

Finally, the distinction between subjective and objective elements of beliefs helps understand how economic agents evaluate information. If a decision maker has the possibility to collect more data before making a decision, the value of this additional information will depend both on the objective characteristics of the data, as well as on the subjective characteristics of the decision maker. Our model identifies the subjective degrees of optimism and pessimism and relates them to preferences for information precision. Preferences of this type will guide choices between different informational environments. They can serve as input for the design of institutions governing the flow of information and can be used to evaluate policies of centralized information provision.

The decision theory has so far treated the issue of data and ambiguity separately, mostly ignoring the questions above. Following Knight's (1921) work, it has concentrated on two major types of situations: risk and uncertainty. In the framework of expected utility as developed by von Neumann and Morgenstern (1944) and by Savage (1954), this distinction is inconsequential.

The theories of Knightian uncertainty, which have emerged in response to Ellsberg's (1961) experiments, highlight the importance of ambiguity and ambiguity aversion for human behavior. For a sample of different modelling approaches see Schmeidler (1989), Bewley (1986), Klibanoff, Marinacci and Mukerji (2005), Gilboa and Schmeidler (1989), Ghirardato, Maccheroni and Marinacci (2004) and Chateauneuf, Eichberger and Grant (2007). In this literature, both perceived ambiguity and ambiguity attitude are purely subjective concepts, and, hence, unrelated to any potentially available information. On the opposite side of the spectrum, several recent papers, Ahn (2008), Gajdos, Hayashi, Tallon and Vergnaud (2007), Stinchcombe (2003) assume that ambiguity can be objectively identified with a set of probability distributions. This allows them to separate the objective ambiguity from the subjective attitude towards ambiguity. Notably, this literature has neglected a third category in Knight's (1921) classification of uncertainty, "statistical probabilities", or, evidence derived from data. The case-based decision theory developed by Gilboa and Schmeidler (2001) incorporates data directly into the decision making process[4]. It allows for heterogeneity of observations and introduces the concept of similarity to capture the different relevance of these observations for the evaluation of a given action. It, however, fails to separate beliefs about the outcomes resulting from a given action from the evaluation of these outcomes. While the framework of Billot, Gilboa, Samet and Schmeidler (2005) provides a method of generating beliefs as similarity-weighted frequencies of observations, their method is not behavioral (the existence of such beliefs is postulated) and it does not take into account ambiguity of the information and the decision maker's attitude towards this ambiguity.

The relationship between ambiguity and data is largely unexplored. Some recent experimental studies by Arad and Gayer (2010) and Hau, Pleskac and Hertwig (2010) study behavior when information is provided in the form of data. Both studies report significant behavioral effects of the form in which data is provided.

Our paper combines the model of case-based decision making with the literature on ambiguity. We model a decision maker who is facing the problem of choosing among a finite set of actions knowing the set of possible outcomes. As in the case based theory of Gilboa and Schmeidler (2001), the information context of the decision situation is specified by a data set containing past observations of actions and their outcomes. Similarly to the work by Gajdos, Hayashi, Tallon and Vergnaud (2007), we assume that the decision maker can compare pairs consisting of an action and an information context. Based on behavioral axioms, we derive a representation of preferences by an α-max-min expected utility functional, as in Ghirardato, Maccheroni and Marinacci (2004) and Chateauneuf, Eichberger and Grant (2007):

$$V(a;D) = \alpha \max_{p \in H_a(D)} \sum_{r \in R} u(r)\, p(r) + (1-\alpha) \min_{p \in H_a(D)} \sum_{r \in R} u(r)\, p(r).$$

Here, each pair consisting of an action a and a data set D is evaluated by the convex combination of the maximal and minimal expected utility over outcomes $r \in R$, $\sum_{r \in R} u(r)\, p(r)$ obtained on a set of probability distributions $H_a(D)$. The beliefs $H_a(D)$ are set-valued, thus capturing the fact that information might be ambiguous. They depend on the action a, on the objective characteristics of the data D, such as the number and frequencies of cases, and on subjective factors such as similarity of cases and perceived ambiguity of the situation. The de-

[4] See also Gilboa and Schmeidler (1997, 2001) and Gilboa, Schmeidler and Wakker (2002) for alternative axiomatizations.

cision maker's degree of optimism α describes his attitude towards ambiguity and determines the weight assigned to the maximal expected utility, whereas $(1-\alpha)$, the weight assigned to the minimal expected utility, is interpreted as the degree of pessimism.

Our first contribution consists in deducing the sets of probability distributions over outcomes, $H_a(D)$ associated with the choice of a specific action a in a given information context D. Thus, we provide a behavioral foundation to the work of Billot, Gilboa, Samet and Schmeidler (henceforth, BGSS (2005)) and Eichberger and Guerdjikova (2010), in which the existence of such beliefs is exogenously assumed. Furthermore, we represent the beliefs $H_a(D)$ as a combination of objective criteria, such as the frequency and the number of observations in the data, with the subjectively perceived degree of ambiguity and relevance of observations (similarity). The obtained representation generalizes the idea of beliefs as similarity-weighted frequencies in BGSS (2005) by allowing for ambiguity.

Our second contribution consists in identifying the degree of ambiguity associated with a particular data set and behaviorally separating it from the ambiguity attitude captured by the degrees of optimism and pessimism. The perceived ambiguity can be separated into two parts: the first is due to the fact that the data set contains a limited number of observations. It can be identified by changing the number of observations in a data set while keeping the frequencies constant. As the number of observations increases, this type of ambiguity converges to 0. The second source of ambiguity is the heterogeneity of cases in the data set. Since each case contains the observation of a single action, correlation across actions cannot be learned from the data. Hence, there is ambiguity associated with predictions about the performance of a given action a using cases containing actions different from a and this ambiguity is persistent. Our model thus captures the well-known insight of identification theory in econometrics: if relevant characteristics are unobservable in the data, it might be impossible to uniquely identify the parameters of the distribution even if the data set is a complete sample of the population, see Manski (2000). The distinction between ambiguity which vanishes with a sufficiently large number of observations and ambiguity which persists for any number of observations corresponds to a similar distinction in Epstein and Schneider (2007). Our representation also extends the approach of Coignard and Jaffray (1994) and Gonzales and Jaffray (1998) to situations in which cases are heterogeneous and contain the outcome of a single action, rather than observations of the state of the world.

Our third contribution consists in using the obtained representation to derive the value of additional information. In particular, we show that the degrees of optimism and pessimism can be used as a measure of the decision maker's preferences for more precise information: the more pessimistic the decision maker is, the more he values precision of information.

The rest of the paper is organized as follows. The next section describes the framework and provides several examples illustrating the scope of our approach. Section 3 states the axioms. Section 4 derives the representation of preferences, discusses the notion of preferences for more precise information and uses the examples from Section 2 to illustrate the representation. Section 5 concludes. All proofs are collected in the Appendix.

2 Framework and Motivating Examples

We start this section by presenting the framework for our analysis.

2.1 Framework

Consider a *decision problem* $(A; R)$ consisting of a finite set of actions A and a finite set of outcomes R. The decision maker is given some information in form of data. The basic element of a *data set* is a *case* c which consists of an *action* $a \in A$ and the *outcome* $r \in R$ observed as a consequence of this action, $c = (a; r)$. The set of all possible cases is $C = A \times R$. An *information context* is identified with a data set D. A data set of length T,

$$D = (c_1 ... c_T) = ((a_1; r_1); ... (a_T; r_T)),$$

is a vector of T cases. The set of all *data sets of length* $T \in \mathbb{N}$, is denoted by $\mathbb{D}^T := C^T$. $\mathbb{D} := \underset{T \geq 1}{\cup} \mathbb{D}^T$ denotes the set of data sets of arbitrary, but finite length. The empty data set denoted $D_\varnothing$ does not contain any cases and captures a situation in which no information is available. We write $\mathbb{D}^* =: \mathbb{D} \cup \{D_\varnothing\}$ for the set of all data sets including the empty one.

We remain agnostic as to how the information context D has been generated. The presumption is that the decision maker trusts that the data is objective and reproducible and that the process determining the outcomes of the acts has not changed. Furthermore, we assume that an observation of an action per se (i.e., without reference to its outcome) does not carry additional information about the desirability of this action[5].

A *decision situation* is completely described by the triple $(A; R; \mathbb{D}^*)$, i.e., a decision problem

[5] E.g., if the observations refer to past choices, the presumption is that these choices were not made based on superior information which is not available to the decision maker.

and the set of possible information contexts arising from it.

Decision makers compare actions belonging to different information contexts. Given a decision situation $(A; R; \mathbb{D}^*)$, they are supposed to be able to rank pairs of actions and data sets, i.e. express preferences of the type $(a; D) \succsim (a'; D')$. Hence, $A \times \mathbb{D}^*$ is the domain of the decision maker's preference order $\succsim$.

2.2 Discussion of the Preference Order $\succsim$

We are not the first to include the information context of an action in the domain of preferences. Preferences over information contexts have been discussed by Gajdos, Hayashi, Tallon and Vergnaud (2007). In their framework, information is modelled by sets of probability distributions. The domain of preferences they use orders pairs, consisting of a Savage act f and a set of probability distributions over states $P \subseteq \Delta(S)$. They write:

> "*The objects $(P; f)$ are not standard (although see the discussion of related literature below). That the decision maker has preferences on such pairs means that, at least conceptually, we allow decision makers to compare the same act in different informational settings. The motivation for this formalization can be best understood going back to Ellsberg's two urns example. In urn 1 there is a known proportion of black and white balls (50–50), while in urn 2 the composition is unknown. The decision maker has the choice to bet on black in urn 1 or on black in urn 2. The action (bet on black) itself is "the same" in the two cases but the information has changed from a given probability distribution (1/2, 1/2) (urn 1) to the simplex (urn 2). We also believe that our model can be used to think of situations besides laboratory experiments. Imagine a firm in the agro business contemplating investing in various crops in different countries. Then, P and Q would capture information relative to (long term) weather forecast in different parts of the world, while f and g would capture the act of investing in a particular crop. One could also consider the example of investing in some stock in one's home country in which information is supposedly easier to acquire than in a similar stock in some exotic country. The widespread preference to invest in home country stocks (the so-called "home bias") can thus find an illustration in our model.*" (p.29)

Our approach is similar in spirit, but the information context takes the form of a data set. Since past observations provide information about the process governing the outcome realizations of different actions, the decision maker's beliefs about the likelihood of a given outcome from a given action will in general depend both on the empirical frequency and the number of observations contained in the data set. Controlling for the frequency and changing the number of observations in a data set, corresponds to varying the precision of information. As illustrated by our first example below, this suggests a natural generalization of the Ellsberg experiment. If two data sets have equal frequencies of observations, but differ in length, would the decision maker express preferences for the longer data set, regardless of the action to be chosen? To capture such preferences for information precision, the domain of preferences must allow for comparisons across information contexts.

More generally, a preference of the type $(a; D) \succ (a; D')$ for choosing action a in context D rather than in context D' means that the decision maker feels more confident choosing a in the information context of data set D rather than in the information context given by data set D'. Such preferences can reflect the fact that the frequency of good outcomes in D is higher, that D contains more precise information, that it contains more relevant data, or some combination of all of these. Such preferences can be elicited in an experimental setting by asking decision makers to make choices between urns characterized by different sets of past observations, or, more generally, between information contexts characterized by different data sets.

An alternative approach would consist in using preferences over Savage acts conditional on the information contained in the data sets. The corresponding state-space is the set of all sequences of outcome realizations. The decision maker then has to express preferences on the set of acts containing all mappings from infinite sequences of outcome realizations into outcomes. Note that these acts are quite different from the original actions a. Moreover, it might be practically impossible to implement such acts, since determining their payoff would require an infinite number of observations[6]. Furthermore, whenever Savage's P2 is violated, conditional preferences will in general depend on the payoffs ascribed to an act on unrealized events. Rather than assuming that decision makers can formulate preferences on this set of acts, we favor the case-based formulation and enrich it by introducing preferences over pairs of alternatives and information contexts.

2.3 Examples

The following examples illustrate the notion of preferences we have in mind

Example 1 *Betting on a draw from an urn*

A decision maker is offered to bet on the color of the ball drawn from one of two urns. Each of the urns contains the same total number of black and white balls. A bet on white, a_w pays 1 if the ball drawn is white and 0, otherwise. A bet on black, a_b is defined symmetrically. Hence, the set of actions is $A = \{a_w; a_b\}$ and the set of outcomes is $r \in R = \{1; 0\}$. An information context specifies the available information about urn i, $i \in \{1; 2\}$. It is given by a data set $D_i = \left(\left(a_1^i; r_1^i\right); \dots \left(a_{T^i}^i; r_{T^i}^i\right)\right)$ and contains records of bets and their outcomes based on drawings from urn i.

[6] In contrast, in our framework, the decision maker is only required to compare finite data sets.

Suppose that there are 10 observations available for urn 1 as summarized in data set D_1:

$$D_1 = ((a_w;1);(a_w;1);(a_b;0);(a_w;0);(a_w;1); \quad (a_w;0);(a_w;0);(a_b;1);(a_b;1);(a_w;1)). \tag{1}$$

Assuming that the order of the draws does not matter, D_1 can be represented by the following table:

$$\begin{array}{c|c|c|c|} & \multicolumn{3}{c}{R} \\ & & 1 & 0 \\ \cline{2-4} A & a_b & 2 & 1 \\ \cline{2-4} & a_w & 4 & 3 \\ \cline{2-4} \end{array} \tag{2}$$

The data set for urn 2, D_2, is assumed to contain 300 observations, summarized in the following table:

$$\begin{array}{c|c|c|c|} & \multicolumn{3}{c}{R} \\ & & 1 & 0 \\ \cline{2-4} A & a_b & 80 & 90 \\ \cline{2-4} & a_w & 60 & 70 \\ \cline{2-4} \end{array} \tag{3}$$

The data set associated with urn 1 implies that out of the 10 available observations, half involved a white ball being drawn and half — a black one. The data set describing urn 2 also implies equal empirical frequencies for black and white, but with a larger set of observations (300).

Preferences $\succsim$ are defined on the set $A \times \mathbb{D}^*$ with the natural interpretation that the decision maker decides which urn to bet on and which bet to place. For example, the decision maker may express preferences of the type $(a_b;D_2) \succ (a_b;D_1)$ and $(a_w;D_2) \succ (a_w;D_1)$, indicating that he prefers to bet on urn 2, regardless of the color of the ball. As in the Ellsberg two-urn paradox, such behavior could be due to the fact that data set D_2 contains more precise information than D_1.

Example 2 *Loan market*

Consider an economy, in which entrepreneurs invest in risky projects $a \in A$, such as starting an internet retail company, opening a fast food restaurant, etc. The set of possible financial returns r is given by R. Entrepreneurs do not possess capital, but can borrow from lenders. In order to start a project, an entrepreneur needs one unit of capital.

Each lender has exactly one unit of capital. A standard loan contract specifies a fixed repayment q, which is due whenever the payoff of the project exceeds q. Otherwise, the lender receives the entire return of the project.

Lenders and entrepreneurs can decide in which of two markets to be active, a well established market in a Western country (market 1), or an emerging market in an Eastern European country (market 2). The set of projects A and the set of outcomes R are identical for both markets. The information about each market $i \in \{1;2\}$ is summarized in a data set D_i containing the observed returns of projects in this market,

$$D_i = \left(\left(a_1^i;r_1^i\right) \dots \left(a_{T^i}^i;r_{T^i}^i\right)\right).$$

Preferences $\succsim$ of lenders and entrepreneurs are defined on the set $A \times \mathbb{D}^*$ expressing the idea that each agent has to choose a project and a market, in which to invest. E.g., a preference of an entrepreneur $(a_1;D_1) \succsim (a_2;D_2)$ means that the agent prefers to invest in project a_1 in market 1 described by information context D_1 to investing in project a_2 in market 2 described *by information context D_2. The two data sets can differ with respect to the number, type and frequency of observations. The informational characteristics of the two markets will determine the market participation decisions of the agents.*

Example 3 *Financial investment*

Consider an investor who can invest 1 unit of money into assets of one of several companies located either in the investor's home country (H), or in a foreign country (F). The investor considers three companies, $A = \{a_1^H; a_2^H; a_1^F\}$. All three companies are active in the same industry (e.g., internet retail), but a_1^H and a_1^F are relatively large and listed in the stock exchange of their respective countries, while a_2^H is a small company traded only across the counter in the home country. R represents the set of possible returns. The investor has access to past returns of the three companies summarized in a data set D:

$$D = \left(\left(a_1^H; r_1\right) ... \left(a_1^H; r_{T_1^H}\right); \left(a_2^H; \tilde{r}_1\right) ... \left(a_2^H; \tilde{r}_{T_2^H}\right); \left(\left(a_1^F; \hat{r}_{T_2^H}\right) ... \left(a_1^F; \hat{r}_{T_1^F}\right) \right) \right) \tag{4}$$

Given the information contained in D, the investor can compare the prospects of investing in a listed internet retail company a_1^H in his home market to investing in the corresponding company in a foreign market a_1^F, expressing preferences of the type $\left(a_1^H; D\right) \succsim \left(a_1^F; D\right)$. Notice that the information of the investor contains the performance of a_2^H, which does not provide direct evidence about a_1^H or a_1^F. However, since a_2^H is a company in the same industry, the investor could consider cases involving this company to be relevant for the evaluation of a_1^H and a_1^F. In particular, the information about the returns of the non-listed company a_2^H might be considered to be more relevant for the prediction about the listed home company a_1^H than for the prediction about the listed foreign company a_1^F. If the information about the returns of a_2^H is favorable for the evaluation of a_1^H, preferences may exhibit the well-known home-bias phenomenon[7].

Example 4 *Medical treatment*

Consider a doctor who has to choose a treatment for a patient with a particular disease. The possible treatment options are given by $A = \{a_1; a_2; a_3\}$, where a_1 stands for administering a new drug, a_2 for using the traditional treatment, and a_3 for applying a placebo. The potential outcomes are $R = \{r_1; r_2; r_3\}$, with r_1 complete recovery, r_2 several weeks of illness, and r_3 long-term chronic disease.

The information context can capture the doctor's personal experience, results of clinical studies or records from hospitals. It can be represented by a data set D consisting of cases $(a_t; r_t)$ which describe observed treatments and their outcomes. The following table summarizes a particular data set D by listing the number of occurrences for each case:

	R		
A	r_1	r_2	r_3
a_1	15	20	0
a_2	35	80	10
a_3	5	70	15

The data in this table reflect a limited experience with the new drug a_1 as compared to the traditional treatment a_2 and the placebo a_3. The doctor's choice of a treatment will reflect her

7 The empirical fact that cross-country asset returns are uncorrelated is the main reason for expecting a risk-averse decision-maker to choose an internationally diversified portfolio. Quality of information is usually neglected in this argument.

preferences over actions in the light of the information in D, e.g.,

$$(a_1; D) \succ (a_2; D) \succsim (a_3; D) .$$

The doctor can also express preferences for additional information. E.g., given the small sample of cases containing observations of the new drug a_1, she might decide to conduct an additional study or buy another data set. If the so obtained data set, D', contains more observations of a_1, she might feel more confident in the positive impact of the new treatment:

$$(a_1; D') \succsim (a_1; D) .$$

Note that such preferences for data sets related to a given action a_1 do not imply the availability of both data sets. While the doctor can specify the type and quantity of observations, she cannot control the outcomes in the new data set. The evaluation of the benefits from a study of 100 additional cases may require, however, comparing the benefit from potential data sets generated by such a study. Such preferences for information are thus hypothetical, but experimentally testable.

2.4 Notation

We conclude this section by introducing some notation which will be used throughout the paper. We use $\mathbb{N}$ to denote the set of natural numbers, which does not include 0. For a case c, we denote by a_c the action observed in case c. For a data set $D = (c_1 ... c_T)$, and a natural number k, we denote by D^k the k-fold of D, i.e.:

$$D^k = \left(\underbrace{c_1 ... c_T ; c_1 ... c_T ; ... ; c_1 ... c_T}_{k\text{-times}} \right) .$$

c^k denotes the data set which contains k observations of case c. The frequency of a data set $D \in \mathbb{D}^T$ for $T \in \mathbb{N}$ is given by

$$f_D = (f_D(c))_{c \in C} =: \left(\frac{|\{t \mid c_t = c\}|}{T} \right)_{c \in C} .$$

Note that the length of the empty data set $D_\varnothing$ is $0 \notin \mathbb{N}$ and its frequency is not well defined. δ_c denotes the Dirac measure on case c. It represents the frequency of a data set containing only observations of case c. The set of frequencies of the data sets of length T is given by:

$$F^T =: \left\{ f \in \Delta^{|C|-1} \text{ with } f(c) \in \left\{ 0; \frac{1}{T} ... \frac{k}{T} ... \frac{T-1}{T}; 1 \right\} \text{ for all } c \in C \right\} .$$

For a given $\mu \in [0; 1]$, a convex combination of two frequencies f and f' is defined as:

$$\mu f + (1 - \mu) f' = (\mu f(c) + (1 - \mu) f'(c))_{c \in C} .$$

Note that the convex combination of two frequencies need not itself be a frequency of a data set.

$\mathbb{D}_a$ is the set of data sets containing only observations of action a,

$$\mathbb{D}_a = \{D \in \mathbb{D} \,|\, f_D (a'; r) = 0 \text{ for all } a' \neq a \text{ and all } r \in R\}.$$

$\mathbb{D}_a^T =: \mathbb{D}_a \cap \mathbb{D}^T$ stands for the set of data sets of length T containing only observations of action a and F_a^T is the corresponding set of frequencies. Finally, δ_r stands for the Dirac measure on outcome r.

3 Axioms

We now suggest a set of axioms on preferences which characterize an α-MEU representation, in which the set of probabilities over outcomes depends on the data set associated with the action. The following ten axioms can be roughly divided into three categories: Axiom 1 (Complete order), Axiom 2 (Invariance), Axiom 6 (Most preferred and least preferred outcome) and Axiom 7 (Continuity) are standard in the literature. In particular, Invariance is often used in case-based decision theory to ensure that the data-generating process satisfies an exchangeability condition and, hence that learning from data is possible. Axioms 3, 4, 5 and 10 all imply some sort of separability of preferences. Axiom 3 (Betweenness) ensures that the frequency of observations can be evaluated separately from the length of the data set. Axiom 4 (Independence) guarantees that cases containing observations of the same action, will be considered equally relevant in the evaluation of a given action, regardless of the outcomes observed. According to Axiom 5 (Action-independent evaluation of outcomes), the evaluation of an outcome does not depend on the action from which it has resulted. Axiom 10 (Length independence) guarantees that the utility over outcomes, the similarity across actions and the subjectively perceived correlation between actions are independent of the length of the data set. The last group of axioms, Axioms 8 and 9, deal with the issues of ambiguity and the value of information. Axiom 8 (Neutral outcome) is the key axiom which allows us to calibrate the decision maker's attitude towards ambiguity by determining the evaluation of an action in absence of information. Axiom 9 (Decreasing ambiguity) ensures that perceived ambiguity decreases as the number of observations grows.

Axiom 1 Complete order

The preference relation $\succsim$ on $A \times \mathbb{D}^*$ is complete and transitive.

Axiom 1 is standard and without it a real-valued representation is impossible. While transitivity seems to be an innocuous assumption, completeness might be too demanding in this setting. In particular, it requires the decision maker to be able to imagine any two hypothetical data sets D and D' and to be able to compare the prospects of any two actions a and a' with respect to these two data sets[8]. In a given choice situation, however, only subsets of action-data set pairs may be feasible. The examples discussed in Section 4.1 illustrate this point and demonstrate the applicability of this approach.

Axiom 2 Invariance

For a given $T \in \mathbb{N}$, let π be a one-to-one mapping $\pi : \{1...T\} \to \{1...T\}$. Then, for any action $a \in A$ and any data set $D = \left((c_t)_{t=1}^T\right) \in \mathbb{D}^T$,

$$\left(a; (c_t)_{t=1}^T\right) \sim \left(a; \left(c_{\pi(t)}\right)_{t=1}^T\right).$$

This is an exchangeability condition, which implies that the order in which data arrive is irrelevant. It is a standard assumption in case-based theory, see e.g., BGSS (2005), and ensures that learning from data is possible. Axiom 2 implies that for any $T \in \mathbb{N}$, an information context $D \in \mathbb{D}^T$ is fully characterized by its length T and the frequency of observations f_D. We can thus write

$$D = (f_D; T).$$

We will use the notation D and $(f; T)$ interchangeably for data sets in $\mathbb{D}$. Note, however, that a combination $(f; T)$ defines a data set only if $f \in F^T$.

Axiom 3 Betweenness for sets of equal length

For any $a \in A$, $T, T' \in \mathbb{N}$, $f \in F^T \cap F^{T'}$, $f' \in F^T \cap F^{T'}$, if $(a; (f; T)) \underset{(\sim)}{\succ} (a; (f'; T))$, then $(a; (f; T')) \underset{(\sim)}{\succ} (a; (f'; T'))$, and for any $\mu \in (0; 1)$ such that $\mu f + (1-\mu) f' \in F^{T'}$

$$(a; (f; T')) \underset{(\sim)}{\succ} (a; (\mu f + (1-\mu) f'; T')) \underset{(\sim)}{\succ} (a; (f'; T')).$$

If a decision maker prefers to choose a when the frequency in the data is f rather than f', then his prediction about the outcome of a given f is more favorable than his prediction for a given f'. Axiom 3 suggests that the length of the data set can be separated from the frequency of observations when evaluating the information context of action a. Intuitively, the frequency determines how much support the information context provides for the choice of a, while the length determines the precision of the information. Keeping the precision constant across two

[8] Similarly, in the Savage framework, the decision maker has to express preferences over acts he might never be able to afford, or acts which appear utterly unreasonable, in that their consequences contradict the state, Savage (1954). In order to avoid such absurd examples, the sets of states and consequences, respectively, the set of cases have to be chosen carefully.

information contexts, they can be ranked based solely on their content. Hence, the preference between $(a;(f;T))$ and $(a;(f';T))$ should not change, if the length of the two sets is changed to T', while keeping the frequencies constant. Moreover, since the linear combination of the two frequencies $\mu f+(1-\mu)f'$ contains the preferred evidence from f and the less preferred one from f' in proportions μ and $(1-\mu)$, $\mu f+(1-\mu)f'$ should be ranked between f and f' as long as the length of all three data sets is equal.

Axiom 4 Independence

For all T, $T'\in\mathbb{N}$, all a and $a'\in A$, all f_1, $f_2\in F_a^T$, f_1', $f_2'\in F_{a'}^{T'}$ and all $\mu\in(0;1]$ such that $\mu f_1+(1-\mu)f_2\in F_a^T$ and $\mu f_1'+(1-\mu)f_2'\in F_{a'}^{T'}$

$$(a;(f_1;T))\underset{(\prec)}{\succ}(a;(f_1';T')) \tag{5}$$

$$(a;(f_2;T))\underset{(\precsim)}{\succsim}(a;(f_2';T'))$$

implies:

$$(a;(\mu f_1+(1-\mu)f_2;T))\underset{(\prec)}{\succ}(a;(\mu f_1'+(1-\mu)f_2';T')) \tag{6}$$

and if $(a;(f_2;T))\sim(a;(f_2';T'))$, then the two statements (5) and (6) are equivalent.

If the evidence $(f_1;T)$ provides more support for the choice of a than $(f_1';T')$, and if $(f_2;T)$ provides more support for the choice of a than $(f_2';T')$, then the evidence from the convex combination of the frequencies f_1 and f_2 should give stronger support for the choice of a than the convex combination (with the same coefficient μ) of the frequencies f_1' and f_2'. This is justified since the data sets $(f_1;T)$ and $(f_2;T)$ are of the same length and contain only observations of the same action a. Hence, if the decision maker considers cases containing the observation of the same action to be equally relevant regardless of the outcomes observed, in the evaluation of $(\mu f_1+(1-\mu)f_2;T)$, he should put a weight of μ on the evidence from $(f_1;T)$ and a weight of $(1-\mu)$ on the evidence from $(f_2;T)$. A symmetric argument applies to $(f_1';T')$ and $(f_2';T')$. Hence, the independence assumption appears reasonable if the decision maker does not consistently overweigh or underweigh the evidence from cases based on the observed outcomes.

Axiom 5 Action-independent evaluation of outcomes

For all a, $a'\in A$, $(a;D_\varnothing)\sim(a';D_\varnothing)$ and for all $T\in\mathbb{N}$, $f\in F_a^T$, $f'\in F_{a'}^T$ such that for all $r\in R$, $f(a;r)=f'(a';r)$,

$$(a;(f;T))\sim(a';(f';T)).$$

Axiom 5 requires the decision maker to choose between actions based entirely on the available observations. Whenever two actions have performed identically for the same number of periods, their evaluation is the same. Similarly, if no information is available, i.e., $D = D_{\varnothing}$, the decision maker should be indifferent among all available actions. This axiom allows us to separate the evaluation of payoffs from the specific action they have resulted from.

Axiom 6 Most preferred and least preferred outcome

There exist $\bar{r}$ and $\underline{r} \in R$ such that for all $a \in A$, $(a;(a;\bar{r})) \succsim (a;D_{\varnothing}) \succsim (a;(a;\underline{r}))$ and for all $c \in C$,

$$(a;(a;\bar{r})) \succsim (a;c) \succsim (a;(a;\underline{r})) \,. \tag{7}$$

Furthermore, for each $a \in A$, there is at least one case $c \in C$ for which both inequalities in (7) are strict.

This axiom postulates that there is an outcome $\bar{r}$ such that the observation of the case $(a;\bar{r})$, provides the most preferred evidence in favor of choosing a among all possible cases in C. It is also weakly preferred to obtaining no information at all. Similarly, there is an outcome $\underline{r}$, such that the observation of $(a;\underline{r})$ is at most as good as obtaining no information and provides the least preferred evidence for a among all cases. Intuitively, there is a best and a worst outcome which define the best and the worst scenario for any action. By Axiom 5, $\bar{r}$ and $\underline{r}$ will coincide for all actions $a \in A$, and we disregard the potential dependence on a in the statement of the axiom.

By requiring that at least one case c satisfies the condition with strict rather than weak inequalities, we also obtain a richness condition on preferences, which is not necessary for the representation, but which implies that the utility function over outcomes is non-constant and allows us to uniquely identify the similarity across actions.

Remark 3.1 *Note that in combination with Axiom 3, Axiom 6 implies that for all $a \in A$, and all $T \in \mathbb{N}$,*

$$\left(a;(a;\bar{r})^T\right) \succ \left(a;(a;\underline{r})^T\right)$$

and, furthermore, for all $D \in \mathbb{D}^T$,

$$\left(a;(a;\bar{r})^T\right) \succsim (a;D) \succsim \left(a;(a;\underline{r})^T\right).$$

Axiom 7 Continuity

For any a, $a' \in A$, any T and $T' \in \mathbb{N}$ and any $D \in \mathbb{D}^T \cup \{D_{\varnothing}\}$, $D' \in \mathbb{D}^{T'} \cup \{D_{\varnothing}\}$, if $(a;D) \succ (a';D')$, there is a $k \in \mathbb{N}$, $k \geq \max\{T;T'\}$ and $\mu \in \left\{\frac{1}{k!};\frac{2}{k!}...\frac{k!-1}{k!}\right\}$ such that

$$(a;D) \succ \left(a;\left(\mu\delta_{(a;\bar{r})} + (1-\mu)\,\delta_{(a;\underline{r})};k!\right)\right) \succ (a';D')\,.$$

According to the axiom, whenever the decision maker has strict preferences over two pairs $(a;D)$ and $(a';D')$, there exists a sufficiently long data set containing only observations of the best and the worst outcome together with action a, which can be nested between the two pairs.

Axiom 8 Neutral outcome

There exists an $\hat{r} \in R$ such that for all $a \in A$ and all $k \in \mathbb{N}$,

$$(a; D_{\varnothing}) \sim \left(a; (a; \hat{r})^k\right).$$

Axiom 8 plays a key role for calibrating the decision maker's attitude towards ambiguity. It allows us to identify the degrees of optimism and pessimism by determining the decision maker's evaluation of an action in absence of information. It postulates the existence of an outcome $\hat{r}$ such that the observation of $(a; \hat{r})$ is identical to receiving no information about the performance of a. Hence, additional observations of the case $(a; \hat{r})$ do not change the evaluation of action a. For instance, $\hat{r}$ could indicate the missing record of an outcome of an action, a problem often encountered in empirical analysis, see Manski (2000).

Axiom 9 Decreasing ambiguity

For all $a \in A$, $k \in \mathbb{N}$ and all $D \in \mathbb{D}$, $(a; D) \succ (a; (a; \hat{r}))$ implies $\left(a; D^{k+1}\right) \succ \left(a; D^k\right)$ and $(a; (a; \hat{r})) \succ (a; D)$ implies $\left(a; D^k\right) \succ \left(a; D^{k+1}\right)$.

Axiom 9 captures the idea that ambiguity decreases with increasing number of observations. It establishes the connection between preferences for information precision and the content of a data set. If the choice of a given some data set D is preferred to choosing a given an observation of the neutral outcome, then the content of D provides positive evidence for choosing a, i.e., evidence which is preferred to receiving no additional information. As the number of observations increases, while the frequency remains constant, the information in the data set becomes more precise, thus providing stronger evidence in favor of a, $\left(a; D^{k+1}\right) \succ \left(a; D^k\right)$. In contrast, if the data set under consideration represents evidence which is worse than the neutral outcome w.r.t. a, then as the precision of the data set increases, the bad evidence will be confirmed and the action a will appear less desirable.

The last axiom will guarantee that the utility over outcomes, the similarity of cases and the perceived correlation between actions do not depend on the length of the data sets[9]. Consider an arbitrary action a and a data set D. We wish to find a data set which contains only observations of $(a; \bar{r})$ and $(a; \underline{r})$ such that the decision maker is indifferent between the action-data set pair

[9] The first two properties are standard in case-based theory, see Gilboa and Schmeidler (2001). The last property of case-based beliefs was introduced in BGSS (2005). One may question these properties, compare the discussion in Gilboa and Schmeidler (2001) and Eichberger and Guerdjikova (2010), but it makes sense to impose them on our representation in order to make it comparable with most of the case-based literature.

$(a;D)$ and the action a in combination with the data set $\left(\mu_D^a \delta_{(a;\bar{r})} + (1-\mu_D^a)\,\delta_{(a;\underline{r})};\hat{T}\right)$ for some $\hat{T}\in\mathbb{N}$, i.e., $\left(a;\left(\mu_D^a \delta_{(a;\bar{r})} + (1-\mu_D^a)\,\delta_{(a;\underline{r})};\hat{T}\right)\right)\sim(a;D)$. Since the mixture coefficient μ_D^a of the data sets $\left(\delta_{(a;\bar{r})};\hat{T}\right)$ and $\left(\delta_{(a;\underline{r})};\hat{T}\right)$ has to be rational-valued, such a data set need not exist even for large values of $\hat{T}$. We can show, however that an arbitrarily precise approximation of $(a;D)$ is possible for large $\hat{T}$.

In order to state our last Axiom, we introduce some notation. For some $T\in\mathbb{N}$, let $D\in\mathbb{D}^T\cup\{D_\varnothing\}$. For every $k\in\mathbb{N}$, $k\geq T$, define $\mu_{k!}^a(D)$ and $\nu_{k!}^a(D)$ as the frequencies of occurrence of $(a;\bar{r})$ in the data sets of length $k!$ which best approximate D from above and from below. Formally,

$$\mu_{k!}^a(D)=\left(\mu\in\left\{0;\frac{1}{k!}\ldots\frac{k!-1}{k!};1\right\}\mid\begin{array}{l}\left(a;\left(\mu\delta_{(a;\bar{r})}+(1-\mu)\,\delta_{(a;\underline{r})};k!\right)\right)\succsim(a;D)\\ \text{and there is no }\mu'\in\left\{0;\frac{1}{k!}\ldots\frac{k!-1}{k!};1\right\}\text{ such that}\\ \left(a;\left(\mu\delta_{(a;\bar{r})}+(1-\mu)\,\delta_{(a;\underline{r})};k!\right)\right)\succ\\ \succ\left(a;\left(\mu'\delta_{(a;\bar{r})}+(1-\mu')\,\delta_{(a;\underline{r})};k!\right)\right)\succsim(a;D)\end{array}\right)\tag{8}$$

and, similarly,

$$\nu_{k!}^a(D)=\left(\nu\in\left\{0;\frac{1}{k!}\ldots\frac{k!-1}{k!};1\right\}\mid\begin{array}{l}(a;D)\succsim\left(a;\left(\nu\delta_{(a;\bar{r})}+(1-\nu)\,\delta_{(a;\underline{r})};k!\right)\right)\\ \text{and there is no }\nu'\in\left\{0;\frac{1}{k!}\ldots\frac{k!-1}{k!};1\right\}\text{ such that}\\ (a;D)\succsim\left(a;\left(\nu'\delta_{(a;\bar{r})}+(1-\nu')\,\delta_{(a;\underline{r})};k!\right)\right)\succ\\ \succ\left(a;\left(\nu\delta_{(a;\bar{r})}+(1-\nu)\,\delta_{(a;\underline{r})};k!\right)\right)\end{array}\right).\tag{9}$$

We denote the common limit of $\mu_{k!}^a(D)$ and $\nu_{k!}^a(D)$ by μ_D^a and call it *the unambiguous equivalent of data set D with respect to a*, or *the unambiguous equivalent* of $(a;D)$.

Definition 3.1 *The unambiguous equivalent of $(a;D)$, μ_D^a, is the common limit of $\mu_{k!}^a(D)$ and $\nu_{k!}^a(D)$ as $k\to\infty$.*

The repeated observation of cases $(a;\bar{r})$ and $(a;\underline{r})$ represents an outcome of a statistical experiment with respect to a. As the number of observations becomes arbitrarily large, the ambiguity caused by a limited number of observations vanishes. As k goes to ∞, the decision maker's prediction about a associated with the data set $\mu_{k!}^a(D)$ should converge to the probability distribution which assigns a probability μ_D^a to $\bar{r}$ and $(1-\mu_D^a)$ to $\underline{r}$. Since this limit distribution is considered to provide the same information for the choice of a as D, the two are regarded as equivalent.

In the Appendix, we prove the following lemma, which shows that under Axioms 1-9 the unambiguous equivalent μ_D^a exists for any $a\in A$ and any $D\in\mathbb{D}^*$.

Lemma 3.1 *Under Axioms 1-9, for any $D \in \mathbb{D}^*$ and any $a \in A$, the sequences $\mu^a_{k!}(D)$ and $\nu^a_{k!}(D)$ converge to a common limit μ^a_D. Hence, the unambiguous equivalent of $(a; D)$ exists.*

Lemma 6.1 in the Appendix shows that under Axioms 1-9, the function $V(a; D) =: \mu^a_D$ can be used to represent $\succsim$ on $A \times \mathbb{D}^*$. Hence, consider three data sets with the same number of observations T, but different frequencies, f, f' and $f'' \in F^T$ such that:

$$(a; (f; T)) \succ (a; (f''; T)) \succ (a; (f'; T)).$$

The corresponding unambiguous equivalents then satisfy $\mu^a_{(f;T)} > \mu^a_{(f'';T)} > \mu^a_{(f';T)}$. This allows us to state the following definition:

Definition 3.2 *For any $a \in A$, any $T \in \mathbb{N}$ and any three frequencies f, f' and $f'' \in F^T$ such that*

$$(a; (f; T)) \succ (a; (f''; T)) \succ (a; (f'; T))$$

the coefficient $\lambda(f; f'; f''; T) \in (0; 1)$ is defined by:

$$\lambda(f; f'; f''; T)\, \mu^a_{(f;T)} + (1 - \lambda(f; f'; f''; T))\, \mu^a_{(f';T)} = \mu^a_{(f'';T)}.$$

The so-defined coefficient $\lambda(\cdot)$ has different meanings depending on the specific frequencies of the three data sets. For data sets with frequencies which are averages of other frequencies, $f'' = \eta f + (1 - \eta) f'$ for some $\eta \in (0; 1)$, the weight $\lambda(f; f'; f''; T)$ reflects the relative similarity between the action under consideration a and the different actions observed in the two data sets $(f; T)$ and $(f'; T)$. In particular, if $f = \delta_{(a;r)}$, $f' = \delta_{(a';r)}$ and $f'' = \frac{1}{2}\delta_{(a;r)} + \frac{1}{2}\delta_{(a';r)}$, the relative weight put on the evidence from the observation of a' for the evaluation of a (similarity between a and a') is given by the ratio $\frac{1-\lambda(f;f';f'';T)}{\lambda(f;f';f'';T)}$.

For three data sets, each of which contains only observations of a with a single outcome, e.g., $f = \delta_{(a;\bar{r})}$, $f' = \delta_{(a;\underline{r})}$ and $f'' = \delta_{(a;r)}$, $\lambda\left(\delta_{(a;\bar{r})}; \delta_{(a;\underline{r})}; \delta_{(a;r)}; T\right)$ represents the evaluation of outcome r relative to the best and the worst outcome. In particular, if the utility of the best outcome is normalized to 1 and the utility of the worst outcome is set to 0, the utility of r is given by $\lambda\left(\delta_{(a;\bar{r})}; \delta_{(a;\underline{r})}; \delta_{(a;r)}; T\right)$.

Finally, for three data sets with frequencies $f = \delta_{(a;\bar{r})}$, $f' = \delta_{(a;\underline{r})}$ and $f'' = \delta_{(a';r')}$ with $a' \neq a$, $\lambda\left(\delta_{(a;\bar{r})}; \delta_{(a;\underline{r})}; \delta_{(a';r')}; T\right)$ represents the probability assigned to action a resulting in outcome $\bar{r}$ given the observation $(a'; r')$. Hence, it reflects the perceived correlation between the outcomes of a and a'.

Axiom 10 Length independence

Let $a \in A$. Suppose that for some f, f' and $f'' \in F^T$, $(a; (f; T)) \succ (a; (f''; T)) \succ (a; (f'; T))$ and

(i) either $\eta f + (1 - \eta) f' = f''$ for some $\eta \in (0; 1)$,

(ii) or $f = \delta_{(a;\bar{r})}$, $f' = \delta_{(a;\underline{r})}$ and $f'' = \delta_c$ for some $c \in C$.

Then, for any T' such that f, f' and $f'' \in F^{T'}$, $\lambda (f; f'; f''; T) = \lambda (f; f'; f''; T')$.

Axiom 10 requires that the coefficient $\lambda (f; f'; f''; T)$ is independent of T. Consider first case (i). Intuitively, the relevance of a case for the evaluation of an action is based on some a priori information, which is encoded in the structure of the action set A and which cannot be learned from the data. Hence, $\lambda (f; f'; f''; T)$ should not depend on the number of observations. While a change in the number of observations from T to T' such that f, f' and $f'' \in F^{T'}$ will certainly influence the evaluation of the data sets, i.e., the unambiguous equivalents of $(a; (f; T'))$ and $(a; (f'; T'))$ will be different from those of $(a; (f; T))$ and $(a; (f'; T))$, once $\mu^a_{(f;T')}$ and $\mu^a_{(f';T')}$ have been fixed, the unambiguous equivalent of $\mu^a_{(f'';T')}$ can be determined as their weighted average with the fixed coefficient $\lambda (f; f'; f''; T)$. In case (ii), if $f = \delta_{(a;\bar{r})}$, $f' = \delta_{(a;\underline{r})}$ and $f'' = \delta_{(a';r')}$, the weight $\lambda \left(\delta_{(a;\bar{r})}; \delta_{(a;\underline{r})}; \delta_{(a';r')}; T\right)$ reflects the perceived utility of an outcome relative to the best and the worst outcome (if $a = a'$) or a perceived correlation between the realizations of a and a' (if $a \neq a'$), both of which cannot be learned from the data. Hence, in this case $\lambda \left(\delta_{(a;\bar{r})}; \delta_{(a;\underline{r})}; \delta_{(a';r')}; T\right)$ is also a subjective characteristic of the decision maker which should not depend on the number of observations in the data set.

4 The Representation

In this section, we derive an α-max-min representation of preferences over action-data-set pairs. We identify the utility function over outcomes and the decision maker's beliefs and show how beliefs can be represented as a combination of the objective characteristics of the data set and the subjective characteristics of the decision maker such as similarity perception and perception of ambiguity. To state the main theorem, we assume that there are more than three outcomes[10], $|R| > 3$.

[10] While this condition is not necessary for the representation we wish to derive and, hence, does not restrict the application of our model, combined with the second part of Axiom 6, it ensures that the similarity function over actions can be uniquely determined. We can also prove the statement of the main Theorem for $|R| = 3$, and a somewhat more restrictive assumption on the class of preference orders. Details are available from the authors upon request.

Theorem 4.1 *Let $|R| > 3$. A preference relation $\succsim$ on $A \times \mathbb{D}^*$ satisfies Axioms 1–10 if and only if there exist a utility function over outcomes $u : R \to \mathbb{R}$, a prediction function $\rho : A \times C \to R$, a family of similarity functions $s_a : A \to \mathbb{R}_{++}$, $a \in A$, degrees of optimism, α, and pessimism, $(1-\alpha)$, a sequence of perceived degrees of ambiguity $(\gamma_T)_{T\in\mathbb{N}}$, and minimal coefficients of perceived ambiguity depending on the cases and the actions, $\gamma_a^c : A \times C \to [0;1]$ such that $\succsim$ can be represented by the function:*

$$V(a;D) = \alpha \max_{p\in H_a(D)} u \cdot p + (1-\alpha) \min_{p\in H_a(D)} u \cdot p, \tag{10}$$

where for all $a \in A$, $H_a(D_\varnothing) = \Delta^{|R|-1}$ and for a given action a and a data set $D \in \mathbb{D}$ with frequency f_D and length T, the set of probability distributions $H_a(D)$ is defined as:

$$H_a(D) = \left[\gamma_T + (1-\gamma_T)\frac{\sum_{c\in C}\gamma_a^c f_D(c)\, s_a(a_c)}{\sum_{c'\in C} f_D(c')\, s_a(a_{c'})}\right]\Delta^{|R|-1} + (1-\gamma_T)\left[\frac{\sum_{c\in C}(1-\gamma_a^c) f_D(c)\, s_a(a_c)\, \delta_{\rho_a^c}}{\sum_{c'\in C} f_D(c')\, s_a(a_{c'})}\right]. \tag{11}$$

The elements of the representation satisfy the following conditions:

(i) *u is unique up to affine-linear transformations;*

(ii) *ρ is unique up to indifference*[11] *and $\rho_a^{(a;r)} = r$ for all $a \in A$ and all $r \in R$;*

(iii) *each of the functions s_a is unique up to a multiplication by a positive number;*

(iv) *$\alpha \in [0;1]$ is unique and for all $a \in A$, $V(a;D_\varnothing) = u(\hat{r}) = \alpha u(\bar{r}) + (1-\alpha) u(\underline{r})$, where $\bar{r}$ is the best, $\underline{r}$ is the worst and $\hat{r}$ is the neutral outcome;*

(v) *the sequence $(\gamma_T)_{T\in\mathbb{N}}$ is unique, strictly decreasing with $\gamma_T \in (0;1)$ and $\lim_{T\to\infty}\gamma_T = 0$*

(vi) *the (minimal) coefficients γ_a^c are unique and satisfy $\gamma_a^{(a;r)} = 0$ for all $a \in A$ and all $r \in R$.*

The α-MEU representation in (10) says that when evaluating the choice of a for a given data set D, the decision maker considers a set of probability distributions $H_a(D)$. He assigns a weight of α (his degree of optimism) to the expected utility derived using the best probability distribution in this set and a weight $(1-\alpha)$ (his degree of pessimism) to the expected utility derived using the worst probability distribution in $H_a(D)$. The representation shows that these weights reflect the evaluation of an action in absence of information. They are thus naturally related by property (iv) to the empty data set $D_\varnothing$ and to the neutral outcome, $\hat{r}$, the observation of which (by Axiom 8) is identical to obtaining no additional information about a.

The axiomatizations of the α-MEU model proposed so far in the literature[12] differ from ours in two respects: first, they make use of the Savage framework and, thus, derive a single set of prior

[11] I.e., if ρ and $\tilde{\rho}$ are two functions which can be used in the representation of $\succsim$, $(a;(a;\rho_a^c)) \sim (a;(a;\tilde{\rho}_a^c))$ holds for all $a \in A$ and all $c \in C$, see Lemma 6.5 and its proof for details.

[12] There does not seem to exist an axiomatisaton for the general α-MEU representation in the Savage framework. Ghirardato, Maccheroni and Marinacci (2004) and Eichberger, Grant, Kelsey and Koshevoy (2011) are relevant references. A special case has been axiomatised in Chateauneuf, Eichberger and Grant (2007).

distributions over the states of the world; second, the set of priors is purely subjective. In the context of the case-based approach proposed in this paper, the set of probability distributions over outcomes $H_a(D)$ depends on the evaluated action and on the information context and incorporates objective features of the data set.

It is worth having a closer look at how the set of priors depends on the data in $D \in \mathbb{D}$. Expression (11) states that beliefs $H_a(D)$ can be represented as a convex combination of the simplex, $\Delta^{|R|-1}$ with the single probability distribution:

$$\sum_{c \in C} \frac{(1-\gamma_a^c) f_D(c) s_a(a_c)}{\sum_{c' \in C} (1-\gamma_a^{c'}) f_D(c') s_a(a_{c'})} \delta_{\rho_a^c}. \tag{12}$$

This probability distribution aggregates the objective information from the data set by assigning to each case c a deterministic prediction about the outcome of a given the observation of case c, $\rho_a^c \in R$. The probability distribution concentrated on such a prediction ρ_a^c, $\delta_{\rho_a^c}$, is weighted by the frequency of case c in the data, $f_D(c)$, by its similarity to the action a, $s_a(a_c)$, as well as by the degree of confidence, $(1-\gamma_a^c)$ assigned by the decision maker to this prediction. The so obtained probability distribution represents *the unambiguous prediction* based on the frequency of the data set D, f_D.

In a second step, the decision maker forms a convex combination of this "unambiguous" prediction with the simplex $\Delta^{|R|-1}$. The weight assigned to $\Delta^{|R|-1}$

$$\gamma_T + (1-\gamma_T) \sum_{c \in C} \gamma_a^c \frac{f_D(c) s_a(a_c)}{\sum_{c' \in C} f_D(c') s_a(a_{c'})}$$

captures the subjectively perceived ambiguity about a given data set D. It is composed of the perception of ambiguity due to a limited number of observations in the data, γ_T and of the perception of ambiguity due to the heterogeneity of cases in the data, captured by the weighted average of the coefficients γ_a^c. For the empty data set, the number of observations is 0 and the corresponding degree of ambiguity can be set to 1, implying that the $H_a(D_\varnothing)$ coincides with the simplex $\Delta^{|R|-1}$. In general, when the number of observations T is small, γ_T is close to 1 and the impact of ambiguity due to heterogeneity is relatively small. As T increases, the ambiguity due to limited number of observations converges to 0 and the entire perceived ambiguity can be attributed to the heterogeneity in the data.

For the special case of statistical experiments, the data set contains only observations of action a, and since $\gamma_a^{(a;r)} = 0$, ambiguity completely vanishes as $T \to \infty$. Furthermore, $\rho_a^{(a;r)} = r$ implies that the unambiguous prediction coincides with the empirical frequency of observed payoffs and reflects the objective character of the data. In contrast, when the data set consists of heterogeneous cases, the lack of objective information about the correlation between the two actions means that the ambiguity γ_a^c about the prediction ρ_a^c does not disappear as the number of observations becomes large[13]. The unambiguous prediction for this case is subjective and reflects the decision maker's beliefs about the unobserved correlation. By choosing γ_a^c to be minimal for each $a \in A$ and each $c \in C$, we implicitly assume that the decision maker adopts the prediction he is most confident in.

Remark 4.1 *An alternative and equivalent way to write the representation in equation (*10*) is given by:*

$$\begin{aligned} V(a;D) = \left(\gamma_T + (1-\gamma_T)\frac{\sum_{c\in C}\gamma_a^c f_D(c)s_a(a_c)}{\sum_{c'\in C} f_D(c')s_a(a_{c'})}\right)\left[\alpha u(\bar{r}) + (1-\alpha)u(\underline{r})\right] \\ +(1-\gamma_T)\frac{\sum_{c\in C}(1-\gamma_a^c) f_D(c)s_a(a_c)}{\sum_{c'\in C} f_D(c')s_a(a_{c'})} u(\rho_a^c) \end{aligned} \tag{13}$$

if $D \in \mathbb{D}$*, whereas* $V(a;D_\varnothing) = \left[\alpha u(\bar{r}) + (1-\alpha)u(\underline{r})\right]$.

In contrast to most of the literature on ambiguity, which, following Gilboa and Schmeidler (1989), represents ambiguity as a purely subjective phenomenon, here the perceived ambiguity is directly related to the precision of information and in particular to the number and type of cases in a data set. A decision maker who does not perceive ambiguity related to the number of observations will express preferences which are independent of the length of the data set. We next present a modification of Axiom 3, which allows us to capture this type of behavior.

Axiom 3A Betweenness for sets of arbitrary length

For any $a \in A$, $T, T' \in \mathbb{N}$, $f \in F^T$, $f' \in F^{T'}$, if $(a;(f;T)) \underset{(\sim)}{\succ} (a;(f';T'))$, then

$$(a;(f;T)) \underset{(\sim)}{\succ} \left(a;\left(\frac{T}{T+T'}f + \frac{T'}{T+T'}f';T+T'\right)\right) \underset{(\sim)}{\succ} (a;(f';T')).$$

Axiom 3A is the behavioral counterpart of the Concatenation axiom in BGSS (2005). It states that the concatenation of two data sets $(f;T)$ and $(f';T')$, which has frequency $\frac{T}{T+T'}f + \frac{T'}{T+T'}f'$ and length $T+T'$ is always evaluated in between the two original data sets. Axiom 3A has the following two implications: first, for all $T \in \mathbb{N}$, $f \in F^T$ and all $k \in \mathbb{N}$

$$(a;(f;T)) \sim (a;(f;kT)). \tag{14}$$

Second, if $(a;(f;T)) \succ (a;(f';T'))$ for some T and T' satisfying the conditions of the axiom,

13 Eichberger and Guerdjikova (2010) provide an example.

then

$$\left(a;\left(f;\hat{T}\right)\right) \succ \left(a;\mu f+(1-\mu)\,f';\hat{T}''\right) \succ \left(a;\left(f';\hat{T}'\right)\right)$$

for all $\hat{T}$, $\hat{T}'$ and $\hat{T}'' \in \mathbb{N}$ such that $f \in F^{\hat{T}}$, $f' \in F^{\hat{T}'}$ and $\mu f + (1-\mu) f' \in F^{\hat{T}''}$. Hence, Axiom 3A strengthens Axiom 3 by requiring the comparison between two frequencies f and f' and their mixtures not to depend on the lengths of the three data sets.

Note that if Axiom 3A holds, we have $(a;(a;r)) \sim \left(a;(a;r)^k\right)$ for all $a \in A$, $r \in R$ and $k \in \mathbb{N}$. However, under Axiom 6, the observation of $(a;r)$ is not always considered equivalent to the empty data set $D_{\varnothing}$. Hence, Axioms 3A and 8 are inconsistent whenever $\succsim$ is non-trivial. Similarly, under Axiom 3A, Axiom 9 is violated — the decision maker perceives the ambiguity of a data set of any given frequency as constant rather than as strictly decreasing in the number of observations. Axiom 3A also implies that Axiom 10 is trivially satisfied.

This allows us to state the following theorem:

Theorem 4.2 *Let $|R| > 3$. A preference relation $\succsim$ on $A \times \mathbb{D}^*$ satisfies Axioms 1, 2 3A and 4-7 if and only if there exist a utility function over outcomes $u : R \to \mathbb{R}$, a prediction function $\rho : A \times C \to R$, a family of similarity functions $s_a : A \to \mathbb{R}_{++}$, $a \in A$, degrees of optimism, α, and pessimism, $(1-\alpha)$ and minimal coefficients of perceived ambiguity depending on the cases and the actions, $\gamma_a^c : A \times C \to [0;1]$ such that $\succsim$ can be represented by the function:*

$$V(a;D) = \alpha \max_{p \in H_a(D)} u \cdot p + (1-\alpha) \min_{p \in H_a(D)} u \cdot p,$$

where for all $a \in A$, $H_a(D_{\varnothing}) = \Delta^{|R|-1}$ and for a given action a and a data set $D \in \mathbb{D}$ with frequency f_D and length T, the set of probability distributions $H_a(D)$ is defined as:

$$H_a(D) = \frac{\sum_{c \in C} \gamma_a^c f_D(c)\, s_a(a_c)}{\sum_{c' \in C} f_D(c')\, s_a(a_{c'})} \Delta^{|R|-1} + \frac{\sum_{c \in C} (1-\gamma_a^c) f_D(c)\, s_a(a_c)}{\sum_{c' \in C} f_D(c')\, s_a(a_{c'})} \delta_{\rho_a^c}.$$

The elements of the representation satisfy the following conditions:

(i) *u is unique up to affine-linear transformations;*

(ii) *ρ is unique up to indifference and $\rho_a^{(a;r)} = r$ for all $a \in A$ and all $r \in R$.*

(iii) *each of the functions s_a is unique up to a multiplication by a positive number;*

(iv) *$\alpha \in [0;1]$ is unique and satisfies $V(a;D_{\varnothing}) = \alpha u(\bar{r}) + (1-\alpha) u(\underline{r})$, where $\bar{r}$ is the best and $\underline{r}$ is the worst outcome;*

(v) *the (minimal) coefficients γ_a^c are unique and satisfy $\gamma_a^{(a;r)} = 0$ for all $a \in A$ and all $r \in R$.*

This representation is a special case of the representation derived in Theorem 4.1, in that $\gamma_T = 0$ for all $T \in \mathbb{N}$. I.e., the decision maker does not perceive ambiguity related to the limited number of observations in the data. Hence, he associates identical probability distributions over outcomes with data sets of *different length*, but *identical frequencies*. Ambiguity is entirely due to the unobservable correlation between actions and is captured by the coefficients γ_a^c. The evaluation of an action in absence of information gives insight about the decision maker's attitude towards ambiguity. Hence, the degrees of optimism and pessimism are naturally related to the evaluation of the empty data set $D_{\varnothing}$, as in property (iv).

A special case of the representation in Theorem 4.2 is derived if the decision maker perceives no ambiguity at all, i.e., if $\gamma_T = 0$ for all $T \in \mathbb{N}$ and $\gamma_a^c = 0$ for all $a \in A$ and all $c \in C$. The representation on $A \times \mathbb{D}$ then takes the form:

$$V(a; D) = u \cdot \sum_{c \in C} \frac{f_D(c)\, s_a(a_c)}{\sum_{c' \in C} f_D(c')\, s_a(a_{c'})} \delta_{\rho_a^c}.$$

This representation provides a behavioral foundation for the result of BGSS (2005), in that beliefs are represented as similarity-weighted frequencies of observations in the data. This special case obtains if preferences satisfy Axioms 1, 2, 3A, 4-7 and in addition the following condition: for any $a \in A$ and any $c \in C$, there is an $\rho_a^c \in R$ such that $(a; c) \sim (a; (a; \rho_a^c))$.

4.1 Examples Resumed

In this section we reconsider the examples introduced in Section 2 and show how the representation derived in Theorem 4.1 can be applied to these decision problems.

Example 1 (resumed) *Betting on a draw from an urn*

Our first example shows how the Ellsberg paradox can be generalized to account for different degrees of information precision. Instead of stating that "there are 50 white and 50 black balls in urn 2" and providing no information about urn 1, we provide data for both urns and ask the decision maker to choose which urn and which color he would like to bet on.

Suppose that urn 1 is characterized by data set D_1 in (2), whereas urn 2 is described by D_2 in (3). An important characteristic of this example is that the observation of the outcome of a given bet, say a_w, uniquely identifies the color of the ball drawn from the urn, and with this the (counterfactual) outcome of the other bet, a_b. Hence, the frequency of white balls drawn from an urn can be obtained by adding the frequency of bets on white won, $(a_w; 1)$ and bets on black lost, $(a_b; 0)$. The empirical frequency with which white has been drawn, thus equals $\frac{1}{2}$ for both urns. The two urns differ, however, in the number of observations — with only 10 observations, the ambiguity about the composition of urn 1 is much larger than for urn 2, for which 300 observations are available.

In order to evaluate bets on different urns, we need to specify the similarity function s and the predictions ρ_a^c. Since the outcome of a bet on white (black) identifies the outcome of the bet on black (white), we assume that all observations are equally relevant for the prediction to be made:

$$s_{a_b}(a_b) = s_{a_b}(a_w) = s_{a_w}(a_b) = s_{a_w}(a_w) = 1.$$

Furthermore, since winning with a_w and losing with a_b provide the same information of a white *ball drawn from the urn, we assume:*

$$\rho_{a_b}^c = \begin{cases} 1 & \text{for } c = (a_b; 1) \text{ or } \quad c = (a_w; 0) \\ 0 & \text{for } c = (a_b; 0) \text{ or } \quad c = (a_w; 1) \end{cases}$$

and, similarly,

$$\rho_{a_w}^c = \begin{cases} 0 & \text{for } c = (a_b; 1) \text{ or } \quad c = (a_w; 0) \\ 1 & \text{for } c = (a_b; 0) \text{ or } \quad c = (a_w; 1) \end{cases} .$$

Since in this example the correlation between actions is known, we assume that there is no ambiguity due to heterogeneity of cases, $\gamma_a^c = 0$ for all $a \in A$ and $c \in C$. Assuming a strictly decreasing sequence γ_T, which describes the perceived degree of ambiguity related to limited number of observations, we obtain that as T goes to ∞, the beliefs associated with a given data set D of length T converge to the empirical frequencies of observations of black and white balls drawn from the urn. The following expressions describe the sets of probabilities assigned to $r = 1$, $H_a(D)(1)$ for a given data set D of length T:

$$H_{a_b}(D)(1) = [(1-\gamma_T)(f_D(a_b;1) + f_D(a_w;0)); (1-\gamma_T)(f_D(a_b;1) + f_D(a_w;0)) + \gamma_T]$$
$$H_{a_w}(D)(1) = [(1-\gamma_T)(f_D(a_w;1) + f_D(a_b;0)); (1-\gamma_T)(f_D(a_w;1) + f_D(a_b;0)) + \gamma_T].$$

Further assuming a degree of optimism α and a degree of pessimism $(1-\alpha)$, we derive:

$$\begin{aligned} V(a_b; D_1) &= V(a_w; D_1) = \gamma_{10}[\alpha u(1) + (1-\alpha)u(0)] + (1-\gamma_{10})\frac{u(1)+u(0)}{2}, \quad (15) \\ V(a_b; D_2) &= V(a_w; D_2) = \gamma_{300}[\alpha u(1) + (1-\alpha)u(0)] + (1-\gamma_{300})\frac{u(1)+u(0)}{2}. \end{aligned}$$

Hence, the comparison between the bets on urn 1 and urn 2 is completely determined by the degrees of optimism and pessimism. Any of the two bets on urn 2 is preferred to any of the two bets on urn 1 if and only if the decision maker's degree of pessimism exceeds his degree of optimism, $\alpha < (1-\alpha)$. In particular, a purely pessimistic decision maker with $\alpha = 0$, will exhibit the usual Ellsberg preferences, choosing to bet on the urn, for which more precise information is available, regardless of the bet, a_w or a_b:

$$V(a_w; D_2) = V(a_b; D_2) > V(a_w; D_1) = V(a_b; D_1).$$

The reason for this result is that a pessimistic decision maker overweighs the probability of the worst outcome, 0, relative to its frequency in the data. As the number of observations increases, the weight assigned to the worst outcome diminishes. Hence, controlling for the frequency of observations, a pessimistic decision maker prefers longer data sets.

If the decision maker is a pure "frequentist", i.e., if he satisfies Axiom 3A, he will not perceive any ambiguity, regardless of the length of the data set. We can thus substitute $\gamma_{10} = \gamma_{300} = 0$ in (15) and obtain:

$$V(a_b; D_1) = V(a_b; D_2) = V(a_w; D_1) = V(a_w; D_2) = \frac{1}{2}[u(1) + u(0)].$$

Hence, a frequentist will be indifferent between any of the bets on urn 1 and urn 2.

This example shows how the Ellsberg paradox can be extended to deal with various degrees of information precision. Information differences regarding the urns are a characteristic feature of the Ellsberg paradox. The notion of a data set allows us to capture the *"amount, quality,* *and unanimity of information"*, Ellsberg (1961, p.657) as objective characteristics of a decision problem and incorporate this objectivity in the beliefs of the decision maker. The parameters of our model, the degrees of optimism and pessimism as well as the perception of ambiguity, can then be used to characterize the decision maker's behavior in face of imprecise information.

Example 2 (resumed) *Loan market*

Reconsider the loan market, in which entrepreneurs (E) and lenders (L) choose projects from the set A to obtain returns in R. Agents can invest either in a well established market 1 or in an emerging Eastern European market 2. All agents have identical utility functions over outcomes, $u(\cdot)$ and, for a data set of a given length T, the same degree of perceived ambiguity γ_T. The degree of optimism for an agent of type $i \in \{E; L\}$ is given by α^i, and the degree of pessimism is $1-\alpha^i$. Hence, $\alpha^i = 1$ describes a pure optimist, whereas $\alpha^i = 0$ characterizes a pure pessimist.

To simplify the model, we assume that there is only one type of project, $A = \{a\}$, e.g., opening a fast food chain, which can result in a high return, $\bar{r}$, or a low return, $\underline{r}$. Hence, $R = \{\bar{r}; \underline{r}\}$. The information about the well established market 1 is contained in the data set D_1, whereas the information about the emerging Eastern European market is given by D_2. In order to focus on the effect of information precision on market outcomes, we assume that both D_1 and D_2 contain the same frequency of high payoff realizations, $f := f(\bar{r}) \in (0;1)$, but differ in length, $D_1 \in \mathbb{D}^{T_1}$, $D_2 \in \mathbb{D}^{T_2}$ with $T_1 > T_2$.

For an agreed upon repayment amount $q \in (\underline{r}; \bar{r})$, the payoff of a lender is given by the minimum of q and $\underline{r}$, $\min\{q; \underline{r}\}$, and the payoff of an entrepreneur is $\max\{\bar{r} - q; 0\}$. Assuming that the repayment q is identical for both markets, and using equation (13) in Remark 4.1, the evaluation of project a in a given information context D_j, $j \in \{1; 2\}$, is:

$$V_L(a; D_j) = \gamma_{T_j}\left[\alpha^L u(q) + \left(1-\alpha^L\right) u(\underline{r})\right] + \left(1-\gamma_{T_j}\right)\left[f u(q) + (1-f) u(\underline{r})\right]$$

for the lender, and

$$V_E(a; D_j) = \gamma_{T_j}\left[\alpha^E u(\bar{r}-q) + \left(1-\alpha^E\right) u(0)\right] + \left(1-\gamma_{T_j}\right)\left[f u(\bar{r}-q) + (1-f) u(0)\right]$$

for the entrepreneur.

Both types of agents will prefer to invest in the established market 1, i.e., $V_i(a; D_1) > V_i(a; D_2)$ for $i \in \{E; L\}$, if and only if the following two conditions are satisfied:

$$\begin{aligned} \left[f u(q) + (1-f) u(\underline{r})\right] &> \left[\alpha^L u(q) + \left(1-\alpha^L\right) u(\underline{r})\right] \\ \left[f u(\bar{r}-q) + (1-f) u(0)\right] &> \left[\alpha^E u(\bar{r}-q) + \left(1-\alpha^E\right) u(0)\right]. \end{aligned}$$

Both conditions will be satisfied for sufficiently low degrees of optimism α^L and α^E. In particular, if $\alpha^L = \alpha^E = 0$, both types will prefer the market with more precise information D_1, regardless of the frequency f. In this case, there will be trade only in the well established market and no transactions in the emerging market[14].

More interesting is the case, in which the two types differ significantly in their degrees of

[14] The assumption that repayment is equal across markets is inconsequential for this result. Increasing the price of credit in market 1 relative to market 2 would only increase the incentives of lenders to participate in market 1. Decreasing the relative price of credit in market 1 would give more incentives to entrepreneurs to choose this market. Hence, the only equilibrium involves both lenders and entrepreneurs participating in market 1.

pessimism. We assume that lenders are more conservative than entrepreneurs and consider the extreme case where the entrepreneurs are pure optimists and the lenders pure pessimists, $1 = \alpha^E > \alpha^L = 0$*. Lenders will provide funding for entrepreneurs investing in both economies only if the agreed upon repayment* q_2 *in the market with more ambiguous information* D_2 *is sufficiently higher than the repayment* q_1 *for* D_1*. Assuming that there are no other investment opportunities, an equilibrium system of repayments* $(q_1^*; q_2^*)$ *must be such that both types of agents are indifferent between investing in the two markets:*

$$V_L\left(a; D_1|q_1^*\right) = V_L\left(a; D_2|q_2^*\right) \quad \text{and} \quad V_E\left(a; D_1|q_1^*\right) = V_E\left(a; D_2|q_2^*\right).$$

Hence, one obtains the equilibrium conditions:

$$\begin{aligned} & \gamma_{T_1} u\left(\underline{r}\right) + \left(1 - \gamma_{T_1}\right)\left[f u\left(q_1^*\right) + (1-f)\, u\left(\underline{r}\right)\right] \\ = \;& \gamma_{T_2} u\left(\underline{r}\right) + \left(1 - \gamma_{T_2}\right)\left[f u\left(q_2^*\right) + (1-f)\, u\left(\underline{r}\right)\right], \end{aligned}$$

$$\begin{aligned} & \gamma_{T_1} u\left(\bar{r} - q_1^*\right) + \left(1 - \gamma_{T_1}\right)\left[f u\left(\bar{r} - q_1^*\right) + (1-f)\, u(0)\right] \\ = \;& \gamma_{T_2} u\left(\bar{r} - q_2^*\right) + \left(1 - \gamma_{T_2}\right)\left[f u\left(\bar{r} - q_2^*\right) + (1-f)\, u(0)\right]. \end{aligned}$$

For the case of a linear utility function, $u(r) = r$ for $r \in R$*, straightforward computations yield the explicit solution for the equilibrium prices*

$$q_1^* = \frac{\left(\gamma_{T_1} - \gamma_{T_2}\right)\left[\left(1 - \gamma_{T2}\right)(1-f)\,\bar{r} + \left(f + \gamma_{T2}\,(1-f)\right) f \underline{r}\right]}{\left(1 - \gamma_{T_2}\right)\left(\gamma_{T_1} + (1-\gamma_{T_1}) f\right) - \left(1 - \gamma_{T_1}\right)\left(\gamma_{T_2} + (1-\gamma_{T_2}) f\right) f}$$

and

$$q_2^* = \frac{q_1^*\left(1 - \gamma_{T_1}\right) + \underline{r}\left(\gamma_{T_1} - \gamma_{T_2}\right)}{1 - \gamma_{T_2}} f.$$

It is easy to check that $q_2^* \in (\underline{r}; \bar{r})$*, while* $q_1^* \in (\underline{r}; q_2^*)$*.*

It follows that the cost of credit is lower in the more informative market. If one would compare the empirical distribution of returns without taking into account the informativeness of data, one would expect that both markets would be served at the same price, i.e., $\gamma_{T_1} = \gamma_{T_2}$ *implies* $q_1^* = q_2^*$*.*

Example 2 illustrates the potential of our approach to model the choice of market participation based on informational differences across markets. It allows us to generate new and testable hypothesis about market outcomes which could not be obtained with expected utility theory. Given the same frequencies of outcomes, the Bayesian model predicts that a decision maker with a given prior for the probability of high returns prefers to trade in the market with more observations, D_1, if the realized frequency f of the better outcome exceeds her prior. Otherwise, she would want to trade in the market D_2. In contrast, a purely pessimistic decision maker will always choose to trade in the more informative market, regardless of the realized frequency f, provided that the price of credit in both markets is equal.

Our next example will focus on the impact of similarity on the decision maker's evaluation of actions.

Example 3 (resumed) *Financial Investment*

Suppose that the investor considers investing in the listed company in his home market, a_1^H, or the listed company in the foreign market, a_1^F, given the information in D, equation (4). To examine the effect of information about the non-listed home company a_2^H on this decision, we assume that the two listed assets a_1^H and a_1^F are essentially identical, except for their similarity to a_2^H. In particular, we consider an observation of a_2^H to be more relevant for the evaluation of a_1^H than for the evaluation of a_1^F. We thus assume:

(i) *The number and frequency of observations of a_1^H, a_1^F and a_2^H satisfy: $T_1^H = T_1^F =: T_1$, $T_2^H =: T_2$, $f_D\left(a_1^H; r\right) = f_D\left(a_1^F; r\right) =: f_D\left(a_1; r\right)$ and $f_D\left(a_2^H; r\right) =: f_D\left(a_2; r\right)$ for all $r \in R$.*

(ii) *The unambiguous predictions satisfy: $\rho_a^{(a';\tilde{r})} = \tilde{r}$, whenever a, $a' \in \{a_H; a_F\}$ and $\rho_{a_1^F}^{(a_2^H;\tilde{r})} = \rho_{a_1^H}^{(a_2^H;\tilde{r})} =: \rho^{(a_2^H;\tilde{r})}$ for all $\tilde{r} \in R$,*

(iii) *The similarity function satisfies:*

$$s_{a_1^H}(a_1^H) = s_{a_1^H}(a_1^F) = s_{a_1^F}(a_1^H) = s_{a_1^F}(a_1^F) = 1$$

and $s_{a_1^H}(a_2^H) = s_H > s_{a_1^F}(a_2^H) = s_F$,

(iv) *The coefficients of perceived ambiguity do not depend on the observed outcomes $\tilde{r}$ and satisfy $\gamma_{a_1^H}^{(a';\tilde{r})} = \gamma_{a_1^H}^{a'} = \gamma_{a_1^F}^{a'} = \gamma_{a_1^F}^{(a';\tilde{r})} =: \gamma^{a'}$ for all $a' \in A$, $\tilde{r} \in R$.*

Note that under these assumptions, given the data set D of length $T =: 2T_1 + T_2$, the sets of beliefs associated with action a_1^i, $i \in \{H; F\}$ satisfy:

$$\begin{aligned} H_{a_1^i}(D) \;=\;& \left[\gamma_T + (1-\gamma_T)\sum_{r\in R}\frac{\left(1-\gamma^{a_2^H}\right)s_i f_D(a_2;r)}{\sum_{r'\in R}\left[2f_D(a_1;r') + s_i f_D(a_2;r')\right]}\right]\Delta^{|R|-1} \\ &+ (1-\gamma_T)\sum_{r\in R}\frac{\left[2f_D(a_1;r)\,\delta_r + \left(1-\gamma^{a_2^H}\right)s_i f_D(a_2;r)\,\delta_{\rho\left(a_2^H;r\right)}\right]}{\sum_{r'\in R}\left[2f_D(a_1;r') + s_i f_D(a_2;r')\right]}, \end{aligned}$$

where $j \in \{H; F\}$, $j \neq i$. Since the correlation between the assets is not observed in the data set, beliefs do not necessarily converge to a single probability distribution even as the length of the data set T goes to ∞. We show in the Appendix that under the assumptions made above, the comparison between a_1^H and a_1^F depends on the sign of the expression:

$$V\left(a_1^H; D\right) - V\left(a_1^F; D\right) = \tag{16}$$

$$\frac{2T_1T_2\left(s_H - s_F\right)}{\left[2T_1 + s_HT_2\right]\left[2T_1 + s_FT_2\right]}\left[\left(1-\gamma^{a_2^H}\right)u\cdot\frac{\sum_{r\in R} f_D(a_2;r)\,\delta_{\rho\left(a_2^H;r\right)}}{\sum_{r'\in R} f_D(a_2;r')} - u\cdot\frac{\sum_{r\in R} f_D(a_1;r)\,\delta_r}{\sum_{r'\in R} f_D(a_1;r')}\right].$$

Since by assumption, $s_H > s_F$, a_1^H will be preferred to a_1^F if the term in the square brackets is positive. This would be true, if the expected utility of a_1^H predicted based on the information about a_2^H and discounted by the degree of confidence in this prediction $\left(1-\gamma^{a_2^H}\right)$ is higher than the expected utility based on the prediction from the directly relevant data about a_1^H and a_1^F. Hence, information about another non-listed asset in the home country may induce a strict preference for the listed company in the home country if this information is considered to be more relevant for the home asset a_1^H than for the foreign asset a_1^F and if the predictions based on *this information are not too ambiguous. Note that if the similarity values were equal, $s_H = s_F$, the two assets would have the same evaluation. Hence, similarity perceptions alone can explain the strict preference for home assets.*

Example 3 highlights the role of similarity among actions. More information which the decision maker considers relevant, i.e., similar to the case under consideration, may influence the choice between otherwise identical assets in different countries. This may help explain biases in portfolio choice between home and foreign countries, without appealing to differences in perceived ambiguity, which in turn may further re-enforce the effect.

4.2 Optimism, Pessimism and Preferences for More Precise Information

The novel aspect of our approach is the domain of preferences: the decision maker can order information contexts in which a given action is chosen. Such preferences can reflect different criteria for evaluation of information. Examples 1 and 2 illustrate how the amount of data, together with the decision maker's attitude towards ambiguity can bias the choice of all actions in favor of the larger (or smaller) data set. In Example 3, the composition of the data set and similarity perceptions affect the decision maker's evaluation of an action, which, in turn, can lead to preferences for data sets containing specific types of cases.

In our framework, increasing the length of the data set, while keeping frequencies unchanged increases the precision of information. More precise information and, hence, less ambiguity, must not necessarily be desirable. Grant, Kaji and Polak (1998, p. 234) quote the New York Times:

> "There are basically two types of people. There are "want-to-knowers" and there are "avoiders." There are some people who, even in the absence of being able to alter outcomes, find information of this sort beneficial. The more they know, the more their anxiety level goes down. But there are others who cope by avoiding, who would rather stay hopeful and optimistic and not have the unanswered questions answered."

We now demonstrate how preferences for more precise information can be directly related to the decision maker's degrees of optimism and pessimism in the spirit of the quotation above. Consider two decision makers, i and j whose preferences $\succsim_i$ and $\succsim_j$ on $A \times \mathbb{D}^*$ satisfy Axioms 1-10 and can, therefore, be represented as in Theorem 4.1. To compare i and j with respect to their preferences for information precision, we have to control for the other parameters of the representation, which are unrelated to information precision — the utility functions over outcomes, u, the similarity functions, s, the predictions ρ and the coefficients of perceived ambiguity γ_a^c. The following Lemma provides conditions under which these elements of the representation can be taken to be identical for two preference relations $\succsim_i$ and $\succsim_j$.

Lemma 4.3 *Let $\succsim_i$ and $\succsim_j$ be preference relations on $A \times \mathbb{D}^*$ satisfying Axioms 1-10. Suppose that for any $a \in A$ and any D and $D' \in \mathbb{D}^T$ for some $T \in \mathbb{N}$, we have that $(a; D) \succsim_i (a; D')$ if and only if $(a; D) \succsim_j (a; D')$ and let $\lambda^i (f; f'; f'') = \lambda^j (f; f'; f'')$ for any three frequencies satisfying the conditions of Axiom 10. Then u^i is an affine-linear transformation of u^j, for all $a \in A$, $s_a^i = K_a s^j$ for some positive constants K_a, the prediction functions ρ^i and ρ^j are identical up to indifference and the coefficients of perceived ambiguity $\gamma_a^{i;c}$ and $\gamma_a^{j;c}$ satisfy $\gamma_a^{i;c} = \gamma_a^{j;c}$ for all $a \in A$ and all $c \in C$.*

We will say that i values information precision more than j if, whenever j prefers to obtain a longer data set to a shorter one with the same frequency of observations, so does i:

Definition 4.1 *For two preference relations $\succsim_i$ and $\succsim_j$ on $A \times \mathbb{D}^*$ satisfying the conditions of Lemma 4.3, we say that $\succsim_i$ values information precision more than $\succsim_j$ if $\left(a; D^k\right) \succ_j (a; D)$ for some $k \in \{2, 3, ...\}$ implies $\left(a; D^k\right) \succ_i (a; D)$.*

Our next Proposition shows that this definition is equivalent to i having a smaller optimism parameter than j, i.e. $\alpha_i \leq \alpha_j$:

Proposition 4.4 *Let $\succsim_i$ and $\succsim_j$ be preference relations on $A \times \mathbb{D}^*$ satisfying the conditions of Lemma 4.3. i values information precision more than j if and only if $\alpha_i \leq \alpha_j$.*

The following example which resumes Example 4 provides an opportunity to illustrate the potential of our approach for dealing with the value of information precision. In our model, given an action a, receiving additional information is considered valuable if the choice of a in the new information context is preferred to the choice of a in the old information context. More generally, additional information is considered valuable if the choice of the best available action given the new information context is preferred to the choice of the best available action given the old information context.

Example 4 (resumed) *Medical treatment*

Consider again the medical doctor who has to choose a treatment from a set of actions A with outcomes in R given her information $D \in \mathbb{D}^T$. Suppose that there is a new study which offers a data set $D' \in \mathbb{D}^{T'}$. For a data set D, let $a^* (D)$ denote the optimal choice of an action given the information context D. The new study provides valuable information if the best possible choice given the concatenation of D and D', i.e., the data set with length $T + T'$ and frequency

$\left(\frac{T}{T+T'}f_D + \frac{T'}{T+T'}f_{D'}\right)$, *denoted* $D \circ D'$ *is evaluated higher than the optimal choice in the old information context* D, *i.e., if* $V\left(a^*\left(D \circ D'\right); D \circ D'\right) > V\left(a^*\left(D\right); D\right)$.

Usually, when deciding whether to acquire new information, the doctor will have control over the type of cases and the number of observations contained in the data-set. However, she will not be able to influence the observed outcomes. Hence, to determine whether or not the additional information would be valuable, she has to form a prediction about the resulting data set D', *which in turn will determine her beliefs about the performance of actions in* A *upon acquiring the additional information. In general, the evaluation of the additional information will depend both on the type and frequency cases, in particular the doctor will consider whether the new data set contains cases which are more or less similar to the cases in the original data set. If similarity considerations do not play a role, it seems natural to assume that the doctor will not expect the new information to alter her initial prediction.*

In the special case, in which the data sets D *and* $D \circ D'$ *differ only with respect to their length, but have identical frequencies,* $f_D = f_{D \circ D'}$, *the unambiguous predictions associated with any action* $a \in A$ *for the two data sets coincide:*

$$\frac{\sum_{c\in C}\left(1-\gamma_a^c\right) f_D\left(c\right) s_a\left(a_c\right)}{\sum_{c'\in C}\left(1-\gamma_a^{c'}\right) f_D\left(c'\right) s_a\left(a_{c'}\right)}\delta_{\rho_a^c} = \frac{\sum_{c\in C}\left(1-\gamma_a^c\right) s_a\left(a_c\right) f_{D\circ D'}\left(c\right)}{\sum_{c'\in C}\left(1-\gamma_a^{c'}\right) s_a\left(a_{c'}\right) f_{D\circ D'}\left(c'\right)}\delta_{\rho_a^c} \tag{17}$$

Hence, the optimal choices in the two information contexts are identical:

$$a^*\left(D \circ D'\right) = a^*\left(D\right) =: a^*.$$

Furthermore, it is straightforward to check that $V\left(a^*\left(D \circ D'\right); D \circ D'\right) - V\left(a^*\left(D\right); D\right) > 0$ *if and only if*

$$\left(\gamma_T - \gamma_{T+T'}\right)\left[u \cdot \frac{\sum_{c\in C}\left(1-\gamma_{a^*}^c\right) f_D\left(c\right) s_{a^*}\left(a_c\right)\delta_{\rho_{a^*}^c}}{\sum_{c'\in C}\left(1-\gamma_{a^*}^{c'}\right) f_D\left(c'\right) s_{a^*}\left(a_{c'}\right)} - u \cdot \left(\alpha\delta_{\bar{r}} + \left(1-\alpha\right)\delta_{\underline{r}}\right)\right] > 0.$$

Hence, the decision maker will prefer to obtain the additional information in D' *if:*

- *the additional information reduces the perceived ambiguity,* $\gamma_T - \gamma_{T+T'} > 0$ *and*
- *the expected utility of action* a^* *given the information in the data,* $u \cdot \frac{\sum_{c\in C}\left(1-\gamma_{a^*}^c\right) f_D(c) s_{a^*}(a_c)\delta_{\rho_{a^*}^c}}{\sum_{c'\in C}\left(1-\gamma_{a^*}^{c'}\right) f_D(c') s_{a^*}(a_{c'})}$ *exceeds the evaluation of* a^* *in the absence of any information* $\left(\alpha u\left(\bar{r}\right) + \left(1-\alpha\right) u\left(\underline{r}\right)\right)$.

In particular, a pure pessimist with $\alpha = 0$ *will prefer an increase in the precision of information, regardless of its content, i.e., regardless of the unambiguous prediction associated with it. In contrast, a pure optimist (*$\alpha = 1$*) will always prefer to avoid receiving additional information.*■

Example 4.2 illustrates minimal conditions under which additional information in form of data is of value for the doctor. Our example illustrates how different degrees of optimism and pessimism will lead to different preferences for information precision.

For the example of a medical study, it might be realistic to assume that the doctor is a pure pessimist, who associates "not knowing" with the occurrence of the worst possible outcome and who, therefore, prefers to receive all possible evidence, both favorable and unfavorable, in order to best evaluate the different treatment options. In contrast, an optimistic patient, who is about to undergo a certain treatment might try to avoid additional negative information about the treatment, since it would make her "feel worse". As we demonstrated in example 3, our model also allows us to consider preferences for "particular types of information". One could, e.g., analyze whether the doctor will prefer to include observations of similar (but not identical) treatments to the new treatment a_1. This is a further field which can be modelled and explored with the approach suggested in this paper.

5 Conclusion

In this paper, we analyze decisions informed by data. Introducing preferences on the set of action-data-set pairs allows us to derive an α-MEU representation. In particular, we are able to separate the utility over outcomes from beliefs represented by sets of probability distributions over outcomes. We identify the subjectively perceived degree of ambiguity and separate it from the decision maker's attitude towards ambiguity as represented by his degrees of optimism and pessimism. We distinguish between two types of ambiguity: ambiguity due to a limited number of observations and ambiguity due to heterogeneity in the data. While the first type of ambiguity decreases as the number of observations grows, the second persists even for large data sets. We show how beliefs depend on the perceived ambiguity of information and represent them as a function of the frequency of cases in the data set, the relevance of each of the cases for the prediction to be made and the unambiguous prediction associated with each case. For the case of controlled statistical experiments, we demonstrate that the beliefs of the decision maker converge to the frequency of observed outcomes as the number of observations becomes large. Finally, we define preferences for information precision and relate them to the decision maker's degrees of optimism and pessimism.

Assuming that the decision maker can compare pairs of actions and data sets is a novel feature of our model. We show that the Ellsberg paradox can be easily generalized to capture such preferences. This example shows that preferences of this type are in principle observable. Assumptions about the preference relation can thus be tested in a controlled experiment. Field data for such preferences can be obtained, e.g., from market participation decisions for markets characterized by different information, from observed choices of technologies, for which different quality and amount of data is available, or from observed decisions about data acquisition.

The value added by a new model of decision making depends foremost on its applicability to real-life phenomena and on its ability to generate novel predictions. Our examples illustrate the wide scope of economic situations which can be described using our approach. We demonstrate how market participation decisions will be influenced by the information structure of the markets and show that information differences across markets can have a significant impact on prices and allocations. We also illustrate the role of information heterogeneity and of similarity perception on investment decisions. Extending these examples can provide additional insights into the evolution of the information structure of markets and explain differences in price dynamics across markets. An important application of our approach is to learning under ambiguity. Eichberger and Guerdjikova (2011) use this framework to model technology adoption triggered by climate change.

Preferences for information are central to our approach and allow us to derive the value of additional information depending on its content and on the subjective characteristics of the decision maker. Hence, our framework can be also used to evaluate the welfare effects of different policies of information provision and to design efficient institutions governing the flow of information.

6 Appendix

Proof of Lemma 3.1:

We proceed in three steps. Step 1 proves an intermediate result, which is then used to show in step 2 that $\lim_{k\to\infty} \left(\mu^a_{k!}(D) - \nu^a_{k!}(D)\right) = 0$ and, in step 3, that $\mu^a_{k!}(D)$ converges. These two statements imply the result of the Lemma.

Step 1: For any $a \in A$, $T \in \mathbb{N}$ and μ, $\mu' \in [0;1]$ such that $\mu\delta_{(a;\bar{r})} + (1-\mu)\,\delta_{(a;\underline{r})}$ and $\mu'\delta_{(a;\bar{r})} + (1-\mu')\,\delta_{(a;\underline{r})} \in F^T$, $\mu' < \mu$ iff

$$\left(a; \left(\mu\delta_{(a;\bar{r})} + (1-\mu)\,\delta_{(a;\underline{r})}; T\right)\right) \succ \left(a; \left(\mu'\delta_{(a;\bar{r})} + \left(1-\mu'\right)\delta_{(a;\underline{r})}; T\right)\right). \tag{18}$$

Proof of Step 1

Axioms 3 and 6 imply that for all $T \in \mathbb{N}$,

$$\left(a; (a;\bar{r})^T\right) \succ \left(a; (a;\underline{r})^T\right). \tag{19}$$

Axiom 4 then implies that for any $\mu \in (0;1]$,

$$\left(a; \left(\mu\delta_{(a;\bar{r})} + (1-\mu)\,\delta_{(a;\underline{r})}; T\right)\right) \succ \left(a; \left(\delta_{(a;\underline{r})}; T\right)\right). \tag{20}$$

Let first $\mu' < \mu$. For $\mu' = 0$, (20) is equivalent to (18). If $0 < \mu' < \mu$, a repeated application of Axiom

4 gives:

$$\left(a;\left(\mu\delta_{(a;\bar{r})}+(1-\mu)\,\delta_{(a;\underline{r})};T\right)\right)\succ\left(a;\left(\frac{\mu'}{\mu}\left(\mu\delta_{(a;\bar{r})}+(1-\mu)\,\delta_{(a;\underline{r})}\right)+\left(1-\frac{\mu'}{\mu}\right)\delta_{(a;\underline{r})};T\right)\right),$$

which is equivalent to (18). Now suppose that (18) holds and note that by the argument above $\mu'\geq\mu$ implies

$$\left(a;\left(\mu'\delta_{(a;\bar{r})}+\left(1-\mu'\right)\delta_{(a;\underline{r})};T\right)\right)\succsim\left(a;\left(\mu\delta_{(a;\bar{r})}+(1-\mu)\,\delta_{(a;\underline{r})};T\right)\right),$$

a contradiction. Hence, $\mu'<\mu$.■

<u>**Step 2**</u>: For any $a\in A$, any $T\in\mathbb{N}$ and any $D\in\mathbb{D}^T\cup\{D_\varnothing\}$, $\lim_{k\to\infty}\left(\mu^a_{k!}(D)-\nu^a_{k!}(D)\right)=0$.

<u>**Proof of Step 2**</u>

By Axioms 6 and 9, sequences $\left(\mu^a_{k!}(D)\right)_{\substack{k\in\mathbb{N}\\k\geq T}}$ and $\left(\nu^a_{k!}(D)\right)_{\substack{k\in\mathbb{N}\\k\geq T}}$ satisfying (8) and (9) exist. Step 1 allows us to rewrite (8) and (9) as:

$$\mu^a_{k!}(D)=\min_{\mu\in\left\{0;\frac{1}{k!}\dots\frac{k!-1}{k!};1\right\}}\left\{\mu\mid\left(a;\left(\mu\delta_{(a;\bar{r})}+(1-\mu)\,\delta_{(a;\underline{r})};k!\right)\right)\succsim(a;D)\right\}$$
$$\nu^a_{k!}(D)=\max_{\nu\in\left\{0;\frac{1}{k!}\dots\frac{k!-1}{k!};1\right\}}\left\{\nu\mid(a;D)\succsim\left(a;\left(\nu\delta_{(a;\bar{r})}+(1-\nu)\,\delta_{(a;\underline{r})};k!\right)\right)\right\}.$$

By step 1, unless $\mu^a_{k!}(D)=\nu^a_{k!}(D)$, it must be that $\mu^a_{k!}(D)-\nu^a_{k!}(D)=\frac{1}{k!}$. Hence, $\mu^a_{k!}(D)-\nu^a_{k!}(D)\leq\frac{1}{k!}$, implying $\lim_{k\to\infty}\left(\mu^a_{k!}(D)-\nu^a_{k!}(D)\right)=0$.■

<u>**Step 3**</u>: For any $a\in A$, any $T\in\mathbb{N}$ and any $D\in\mathbb{D}^T\cup\{D_\varnothing\}$, $\left(\mu^a_{k!}(D)\right)_{\substack{k\in\mathbb{N}\\k\geq T}}$ converges.

<u>**Proof of Step 3**</u>

By Axiom 8, there is an outcome $\hat{r}$ such that $(a;(a;\hat{r}))\sim(a;(a;\hat{r})^n)$ for any $n\in\mathbb{N}$. Suppose that $(a;D)\succsim(a;(a;\hat{r}))$. By (8),

$$\left(a;\left(\mu^a_{k!}(D)\,\delta_{(a;\bar{r})}+\left(1-\mu^a_{k!}(D)\right)\delta_{(a;\underline{r})};k!\right)\right)\succsim(a;D)\succsim(a;(a;\hat{r})).$$

By Axiom 9, this implies

$$\left(a;\left(\mu^a_{k!}(D)\,\delta_{(a;\bar{r})}+\left(1-\mu^a_{k!}(D)\right)\delta_{(a;\underline{r})};(k+1)!\right)\right)\succsim\left(a;\left(\mu_{k!}(D)\,\delta_{(a;\bar{r})}+\left(1-\mu_{k!}(D)\right)\delta_{(a;\underline{r})};k!\right)\right).$$

Since $\mu^a_{(k+1)!}(D)\in\left\{0;\frac{1}{(k+1)!}\dots\frac{(k+1)!-1}{(k+1)!};1\right\}$, by step 1 we obtain $\mu^a_{(k+1)!}(D)\leq\mu^a_{k!}(D)$. Hence, the sequence $\left(\mu^a_{k!}(D)\right)_{\substack{k\in\mathbb{N}\\k\geq T}}$ is bounded and decreasing and, thus, it converges. For the case $(a;(a;\hat{r}))\succ(a;D)$, a symmetric argument shows that $\nu^a_{k!}(D)$ converges. Step 2 then implies convergence for $\mu^a_{k!}(D)$.■

<u>**Proof of Theorem 4.1**</u>:

It is straightforward to verify the necessity of the axioms for the representation. The sufficiency part of the Theorem is proved in four consecutive Lemmas. Lemma 6.1 shows that the Axioms imply a utility representation

$$V(a;D)=u\cdot\hat{h}_a(D).$$

Here, $u:R\to\mathbb{R}$ is a utility function over outcomes. $\hat{h}:A\times\mathbb{D}^*\to\Delta^{|R|-1}$ is any selection of a maximal with respect to set inclusion correspondence $\hat{H} : A \times \mathbb{D}^* \rightrightarrows \Delta^{|R|-1}$, with the property that $u \cdot \hat{h}_a(D) = u \cdot \hat{h}'_a(D)$ for all $\hat{h}_a(D)$ and $\hat{h}'_a(D) \in \hat{H}_a(D)$. In Lemma 6.2, we use the result proven in Eichberger and Guerdjikova (2010) to show that $V(a; D)$ on $A \times \mathbb{D}$ can be expressed as:

$$V(a; D) = u \cdot \sum_{c \in C} \frac{s_a(a_c) f_D(c) \hat{h}_a(c^T)}{\sum\limits_{c' \in C} s_a(a_{c'}) f_D(c')},$$

where for $a \in A$, $s_a : A \to \mathbb{R}_{++}$ is a family of similarity functions across actions, each of which is unique up to a multiplication by a positive number and $\hat{h}_a(c^T)$ is any element of $\hat{H}_a(c^T)$. In Lemma 6.4, we identify the coefficients α and $(\gamma_T)_{T \in \mathbb{N}}$ and show that $\hat{H}_a(c^T)$ can be written as:

$$\hat{H}_a(c^T) = \left\{ h \in \Delta^{|R|-1} \mid u \cdot h = u \cdot \left[\gamma_T \left(\alpha \delta_{\bar{r}} + (1-\alpha) \delta_{\underline{r}} \right) + (1-\gamma_T) \hat{h}_a^c \right] \right\},$$

where $\hat{h}_a^c \in \lim_{T \to \infty} \hat{H}_a(c^T)$ and $\hat{h}_a^{(a;r)} = \delta_r$ for all $c \in C$, $a \in A$ and $r \in R$. In Lemma 6.5, we identify the prediction function ρ and the coefficients of ambiguity γ_a^c and show that $\hat{h}_a^c$ satisfies:

$$u \cdot \hat{h}_a^c = u \cdot \left[\gamma_a^c \left(\alpha \delta_{\bar{r}} + (1-\alpha) \delta_{\underline{r}} \right) + (1-\gamma_a^c) \delta_{\rho_a^c} \right].$$

We then combine all four steps to show that $H_a(D)$ can be chosen so as to have the desired structure in (11) and derive the desired representation in (10).

Lemma 6.1 *The preference relation $\succsim$ on $A \times \mathbb{D}^*$ can be represented by a utility function*

$$V(a; D) = u \cdot \hat{h}_a(D)$$

where $u : R \to \mathbb{R}$ is a utility function over outcomes and $\hat{h} : A \times \mathbb{D}^* \to \Delta^{|R|-1}$ is any selection of a maximal with respect to set inclusion correspondence $\hat{H} : A \times \mathbb{D}^* \rightrightarrows \Delta^{|R|-1}$ with the property that $u \cdot \hat{h}_a(D) = u \cdot \hat{h}'_a(D)$ for all $\hat{h}_a(D)$ and $\hat{h}'_a(D) \in \hat{H}_a(D)$ and $\delta_{\hat{r}} \in \hat{H}(a; D_\varnothing)$ for all $a \in A$.

Proof of Lemma 6.1:

We proceed to prove the Lemma in 4 steps. In step 1, we define the function V using the unambiguous equivalents μ_D^a. In step 2, we demonstrate that the so defined V represents $\succsim$. In step 3, we elicit the von-Neumann-Morgenstern utility function over outcomes u. In step 4, we construct the correspondence $\hat{H}$.

Step 1: Define the function $V : A \times \mathbb{D}^* \to [0; 1]$ as:

$$V(a; D) =: \mu_D^a.$$

By Lemma 3.1 the unambiguous equivalents $\mu_D^a \in [0; 1]$ are well defined and so is the function V.

Remark 6.1 *Note that by Definition 3.1, $\lim_{T \to \infty} \mu^a_{(a;\bar{r})^T} = 1$ and $\lim_{T \to \infty} \mu^a_{(a;\underline{r})^T} = 0$. Hence,* *the definition of* $V(a;D)$ *implies*

$$\begin{aligned}\lim_{T\to\infty} V\left(a;(a;\bar{r})^T\right) &= 1\\ \lim_{T\to\infty} V\left(a;(a;\bar{r})^T\right) &= 0.\end{aligned}$$

<u>**Step 2**</u>: The function V defined in Step 1 represents $\succsim$.

<u>**Proof of Step 2**</u>:

To see that the function V represents $\succsim$, consider two actions a and $a' \in A$, and for T and $T' \in \mathbb{N}$, two data sets $D \in \mathbb{D}^T \cup \{D_\varnothing\}$ and $D' \in \mathbb{D}^{T'} \cup \{D_\varnothing\}$. Let $\hat{T} = \max\{T;T'\}$. By Axiom 5, we have that for all $k \geq \hat{T}$:

$$\left(a';\left(\mu_{k!}^{a'}(D')\,\delta_{(a';\bar{r})} + \left(1-\mu_{k!}^{a'}(D')\right)\delta_{(a';\underline{r})};k!\right)\right) \sim \left(a;\left(\mu_{k!}^{a'}(D')\,\delta_{(a;\bar{r})} + \left(1-\mu_{k!}^{a'}(D')\right)\delta_{(a;\underline{r})};k!\right)\right) \tag{21}$$

Suppose (w.l.o.g.) that $(a;D) \succsim (a';D')$. Then, the construction of the unambiguous equivalents in (8) together with (21) imply[15]

$$\left(a;\left(\mu_{k!}^{a}(D)\,\delta_{(a;\bar{r})} + (1-\mu_{k!}^{a}(D))\,\delta_{(a;\underline{r})};k!\right)\right) \succsim \left(a;\left(\mu_{k!}^{a'}(D')\,\delta_{(a;\bar{r})} + \left(1-\mu_{k!}^{a'}(D')\right)\delta_{(a;\underline{r})};k!\right)\right).$$

By step 1 of Lemma 3.1, this is equivalent to $\mu_{k!}^{a}(D) \geq \mu_{k!}^{a'}(D')$ for all $k \geq \hat{T}$. Hence, by Lemma 3.1 and by the definition of V,

$$V(a;D) = \mu_D^a = \lim_{k\to\infty}\mu_{k!}^{a}(D) \geq \lim_{k\to\infty}\mu_{k!}^{a'}\left(D'\right) = \mu_{D'}^{a'} = V\left(a';D'\right).$$

Now suppose that $V(a;D) \geq V(a';D')$, or $\mu_D^a \geq \mu_{D'}^{a'}$. Note that the sequences $\mu_{k!}^{a}(D)$ and $\mu_{k!}^{a'}(D')$ satisfy $\mu_{k!}^{a}(D) \geq \mu_{k!}^{a'}(D')$ for all $k \geq \hat{T}$, whenever $(a;D) \succsim (a';D')$ and $\mu_{k!}^{a}(D) \leq \mu_{k!}^{a'}(D')$ for all $k \geq \hat{T}$ whenever $(a';D') \succsim (a;D)$. Furthermore, $\mu_D^a > \mu_{D'}^{a'}$, if and only if there is a $k \geq \hat{T}$ such that $\mu_{k!}^{a}(D) > \mu_{k!}^{a'}(D')$. Since $\mu_D^a = \lim_{k\to\infty}\mu_{k!}^{a}(D) \geq \lim_{k\to\infty}\mu_{k!}^{a'}(D') = \mu_{D'}^{a'}$, it follows that $\mu_{k!}^{a}(D) \geq \mu_{k!}^{a'}(D')$ for all $k \geq \hat{T}$. Let first $\mu_D^a > \mu_{D'}^{a'}$. Then, there is a k such that:

$$\begin{aligned}\left(a;\left(\mu_{k!}^{a}(D)\,\delta_{(a;\bar{r})} + (1-\mu_{k!}^{a}(D))\,\delta_{(a;\underline{r})};k!\right)\right) &\succ \left(a;\left(\mu_{k!}^{a'}(D')\,\delta_{(a;\bar{r})} + \left(1-\mu_{k!}^{a'}(D')\right)\delta_{(a;\underline{r})};k!\right)\right)\\ &\sim \left(a';\left(\mu_{k!}^{a'}(D')\,\delta_{(a';\bar{r})} + \left(1-\mu_{k!}^{a'}(D')\right)\delta_{(a';\underline{r})};k!\right)\right)\end{aligned}$$

Assuming that $(a';D') \succsim (a;D)$ then contradicts the fact that $\mu_{k!}^{a}(D)$ is the minimum of the set in (8) and we conclude that $(a;D) \succ (a';D')$. Assume now that $\mu_D^a = \mu_{D'}^{a}$, and hence, $\mu_{k!}^{a}(D) = \mu_{k!}^{a'}(D')$

[15] Note that if

$$\left(a;\left(\mu_{k!}^{a'}(D')\,\delta_{(a;\bar{r})} + \left(1-\mu_{k!}^{a'}(D')\right)\delta_{(a;\underline{r})};k!\right)\right) \succ \left(a;\left(\mu_{k!}^{a}(D)\,\delta_{(a;\bar{r})} + (1-\mu_{k!}^{a}(D))\,\delta_{(a;\underline{r})};k!\right)\right),$$

we would have

$$\begin{aligned}\left(a';\left(\mu_{k!}^{a'}(D')\,\delta_{(a';\bar{r})} + \left(1-\mu_{k!}^{a'}(D')\right)\delta_{(a';\underline{r})};k!\right)\right) &\succ \left(a';\left(\mu_{k!}^{a}(D)\,\delta_{(a';\bar{r})} + (1-\mu_{k!}^{a}(D))\,\delta_{(a';\underline{r})};k!\right)\right)\\ &\succsim (a;D) \succsim (a';D'),\end{aligned}$$

thus contradicting the fact that $\mu_{k!}^{a'}(D')$ satisfies the condition in (8) for $(a';D')$.

for all $k \geq \hat{T}$. If $(a'; D') \succ (a; D)$, then we must have:

$$\begin{aligned}\left(a'; \left(\mu^{a'}_{k!}\left(D'\right)\delta_{(a';\bar{r})} + \left(1 - \mu^{a'}_{k!}\left(D'\right)\right)\delta_{(a';\underline{r})}; k!\right)\right) &\sim \left(a; \left(\mu^{a}_{k!}\left(D\right)\delta_{(a;\bar{r})} + \left(1 - \mu^{a}_{k!}\left(D\right)\right)\delta_{(a;\underline{r})}; k!\right)\right) \\ &\succsim \left(a'; D'\right) \succ (a; D)\end{aligned}$$

for all $k \geq \hat{T}$. Axiom 7, however, states that there exists a $\bar{k} \in \mathbb{N}$, $\bar{k} \geq \hat{T}$ and a $\mu \in \left\{\frac{1}{\bar{k}!} \dots \frac{\bar{k}!-1}{\bar{k}!}\right\}$ such that:

$$\left(a'; D'\right) \succ \left(a'; \left(\mu\delta_{(a';\bar{r})} + (1-\mu)\,\delta_{(a';\underline{r})}; \bar{k}!\right)\right) \succ (a; D)\,.$$

Hence, $\mu^{a}_{\bar{k}!}(D)$ does not satisfy the definition (8) for $(a; D)$ and we obtain a contradiction to the assumption $(a'; D') \succ (a; D)$ for the case $\mu^a_D = \mu^a_{D'}$. It follows that $(a; D) \succsim (a'; D')$ whenever $V(a; D) \geq V(a'; D')$.■

Step 3: Eliciting the function $u : R \to \mathbb{R}$

For given $a \in A$, $T \in \mathbb{N}$ and $r \in R$, let $\lambda_r =: \lambda\left(\delta_{(a;\bar{r})}; \delta_{(a;\underline{r})}; \delta_{(a;r)}; T\right)$, see Definition 3.2, and thus:

$$V\left(a; (a; r)^T\right) = \lambda_r V\left(a; (a; \bar{r})^T\right) + (1 - \lambda_r)\, V\left(a; (a; \underline{r})^T\right). \tag{22}$$

Such a $\lambda_r \in [0; 1]$ exists by Axiom 6. By Axioms 5 and 10, λ_r is independent of a and T and is thus well-defined. We define the function $u : R \to \mathbb{R}$ by:

$$u(r) =: \lambda_r$$

for all $r \in R$. Obviously, $u(\bar{r}) = 1$ and $u(\underline{r}) = 0$. Furthermore, by Remark 6.1, taking $T \to \infty$ on both sides of (22), we obtain for all $r \in R$:

$$\lim_{T\to\infty} V\left(a; (a; r)^T\right) = u(r)\,. \tag{23}$$

Note that by Axiom 8, $V(a; D_\varnothing) = u(\hat{r})$ for all $a \in A$.

Step 4: Identifying the correspondence $\hat{H} : A \times \mathbb{D}^* \rightrightarrows \Delta^{|R|-1}$

For any $a \in A$ and $D \in \mathbb{D}^*$, let $\hat{h}_a(D) \in \Delta^{|R|-1}$ be such that

$$u \cdot \hat{h}_a(D) = \mu^a_D, \tag{24}$$

where u is the function defined in step 3. Note that such an $\hat{h}_a(D)$ exists for all a and D, e.g.,

$$\hat{h}_a(D)(r) = \begin{cases} \mu^a_D & \text{if } r = \bar{r} \\ 1 - \mu^a_D & \text{if } r = \underline{r} \\ 0 & \text{else} \end{cases}$$

satisfies (24). Denote the set of all $\hat{h}_a(D)$ satisfying (24) by:

$$\hat{H}_a(D) =: \left\{\hat{h} \in \Delta^{|R|-1} \mid u \cdot \hat{h} = \mu^a_D\right\}. \tag{25}$$

Note that $\hat{H}_a(D)$ defines a correspondence:

$$\hat{H} : A \times \mathbb{D}^* \rightrightarrows \Delta^{|R|-1}$$

with the properties specified in the statement of the Lemma. In particular, for all $a \in A$, $\delta_{\hat{r}} \in \hat{H}_a(D_\varnothing)$

and for any selection $\hat{h} : A \times \mathrm{D}^* \rightarrow \Delta^{|R|-1}$ from $\hat{H}$, the preference relation $\succsim$ can be represented by $V(a; D) = u \cdot \hat{h}_a(D)$.■

Lemma 6.2 *The preference relation $\succsim$ on $A \times \mathbb{D}$ can be represented by:*

$$V(a; D) = u \cdot \sum_{c \in C} \frac{s_a(a_c) f_D(c) \hat{h}_a(c^T)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')}, \tag{26}$$

for some and thus all $\hat{h}_a(c^T) \in \hat{H}_a(c^T)$, where $\hat{H}_a(c^T)$ are defined as in (25) and where for $a \in A$, $s_a : A \rightarrow \mathbb{R}_{++}$ is a family of similarity functions, each of which is unique up to a multiplication by a positive number.

Proof of Lemma 6.2:

We proceed in 4 steps. In step 1, we construct a correspondence $P : A \times \mathbb{D} \rightrightarrows \Delta^{|R|-1}$ such that for all $D \in \mathbb{D}$, $u \cdot p = u \cdot p'$ for all p and $p' \in P_a(D)$ and such that $\tilde{V}(a; D) =: u \cdot p$ for some and, thus, all $p \in P_a(D)$ represents $\succsim$. In step 2, we show that the so constructed correspondence P satisfies a list of properties $(B1)$ — $(B5)$ stated below. In step 3, we restate a Theorem which appears in Eichberger and Guerdjikova (2010) and which implies that a correspondence $P : A \times \mathbb{D} \rightrightarrows \Delta^{|R|-1}$ satisfying properties $(B1)$ — $(B5)$ can be represented as:

$$P_a(D) = \sum_{c \in C} \frac{s_a(c) f_D(c) P_a(c^T)}{\sum_{c' \in C} s_a(c') f_D(c')}, \tag{27}$$

where $s_a : A \times R \rightarrow \mathbb{R}_{++}$ is a family of similarity functions, each of which is unique up to a multiplication by a positive number. In step 4, we show that all s_a are independent of r. Furthermore, for each $a \in A$, $(s_a(a_c))_{c \in C}$ are the unique up to a multiplication by a positive number $K_a > 0$ coefficients such that for each $D \in \mathbb{D}^T$, μ_D^a can be represented as:

$$\mu_D^a = \frac{\sum_{c \in C} \mu_{(c)^T}^a \tilde{s}_a(a_c) f_D(c)}{\sum_{c' \in C} \tilde{s}_a(a_{c'}) f_D(c')}.$$

Combining this result with the definitions of V and the sets $\hat{H}_a(c^T)$ in (25) gives the result of the Lemma.

Step 1: Note that by Axiom 9, for any $a \in A$, $c \in C$, the sequence $\left(\mu_{c^T}^a\right)_{T \in \mathbb{N}}$ is either constant or strictly monotonic. Since $\mu_{c^T}^a \in [0; 1]$, it follows that the limit $\lim_{T \rightarrow \infty} \mu_{c^T}^a$ exists.

For a given $a \in A$, let $C_a^* =: \{c \in C \mid (a; c) \not\sim (a; (a; \bar{r}))$ and $(a; c) \not\sim (a; (a; \underline{r}))\}$. By Axiom 6, C_a^* is non-empty.

Choose an $\epsilon \in (0; 1)$ such that

$$\frac{\epsilon}{2} < \min_{a \in A,\, c \in C_a^*} \min \left\{ \lim_{T \rightarrow \infty} \mu_{c^T}^a ; \left(1 - \lim_{T \rightarrow \infty} \mu_{c^T}^a\right) \right\}.$$

By Axiom 6 and Remark 6.1, such an $\epsilon \in (0;1)$ exists. Let the set P^* be defined as:

$$P^* =: \left\{ p \in \Delta^{|R|-1} \mid u \cdot p = \frac{1}{2} \right\}$$

and for each $D \in \mathbb{D}$, define

$$P_a(D) =: \epsilon P^* + (1-\epsilon) \left[\mu_D^a \delta_{\bar{r}} + (1-\mu_D^a)\, \delta_{\underline{r}} \right]. \tag{28}$$

The sets $P_a(D) \subset \Delta^{|R|-1}$ are non-empty, convex and compact and each $P_a(D)$ is a translation of the set $\epsilon P^* + (1-\epsilon)\,\delta_{\underline{r}}$ with the property that $u \cdot p = \left(1 - \frac{\epsilon}{2}\right) \mu_D^a + \frac{\epsilon}{2}$ for all $p \in P_a(D)$. Hence, the function $\tilde{V}$ defined by $\tilde{V}(a;D) =: u \cdot p$ for some and thus, all $p \in P_a(D)$, satisfies $\tilde{V}(a;D) = \left(1 - \frac{\epsilon}{2}\right) V(a;D) + \frac{\epsilon}{2}$ for all $a \in A$ and $D \in \mathbb{D}$, and, thus, represents $\succsim$ on $A \times \mathbb{D}$.

Step 2: Fix $a \in A$ and consider the projection of the correspondence P on $\mathbb{D}$ for this a, $P_a(D) : \mathbb{D} \rightrightarrows \Delta^{|R|-1}$. $P_a(D)$ satisfies the following properties:

$(B1)$ $P_a(D)$ depends only on the frequency and the length of D, but not on the order of cases in D.

$(B2)$ Let $(f_i)_{i=1}^n$ and f be in F^T. Whenever $\sum_{i=1}^n \eta_i f_i = f$, for some coefficients $\eta_i \in (0;1)$ such that $\sum_{i=1}^n \eta_i = 1$, there are coefficients $(\lambda_i)_{i=1}^n \in (0;1)$ with $\sum_{i=1}^n \lambda_i = 1$ such that $P_a(f;T) = \sum_{i=1}^n \lambda_i P_a(f_i;T)$.

$(B3)$ Under the conditions listed in $(B2)$, $P_a(f;T) = \sum_{i=1}^n \lambda_i P_a(f_i;T)$ if and only if $P_a\left(f;\hat{T}\right) = \sum_{i=1}^n \lambda_i P_a\left(f_i;\hat{T}\right)$ holds for any $\hat{T} \in \mathbb{N}$, such that $(f_i)_{i=1}^n$ and $f \in F^{\hat{T}}$, i.e., the vector $(\lambda_i)_{i=1}^n$ does not depend on T.

$(B4)$ For all $c \in C$, the sequences $\left(P_a\left(c^T\right)\right)_{T \in \mathbb{N}}$ have a limit, $\lim_{T\to\infty} P_a\left(c^T\right)$.

$(B5)$ No three of the sets $\lim_{T\to\infty} P_a\left(c^T\right)$ of dimension 0 or 1 are collinear.

Proof of Step 2

$(B1)$: follows directly from Axiom 2.

$(B2)$ and $(B3)$: Take three data sets D, D' and $D'' \in \mathbb{D}^T$ with corresponding frequencies f, f' and $f'' \in F^T$ such that $\eta f + (1-\eta) f'' = f'$ for some $\eta \in (0;1)$. If $(f;T) \succ (f'';T)$. Axiom 3 implies

$$\left(f;\hat{T}\right) \succ \left(f';\hat{T}\right) \succ \left(f'';\hat{T}\right)$$

for all $\hat{T} \in \mathbb{N}$ such that f, f' and $f'' \in F^{\hat{T}}$. Hence, by step 2 of Lemma 6.1, $\mu^a_{(f;T)} > \mu^a_{(f';T)} > \mu^a_{(f'';T)}$ and therefore, by Axiom 10, there is a $\lambda(f;f'';f') \in (0;1)$ independent of T such that

$$\mu^a_{(f';T)} = \lambda\left(f;f'';f'\right) \mu^a_{(f;T)} + \left(1 - \lambda\left(f;f'';f'\right)\right) \mu^a_{(f'';T)}.$$

If $(f;T) \sim (f'';T)$, Axiom 3 implies

$$\left(f;\hat{T}\right) \sim \left(f';\hat{T}\right) \sim \left(f'';\hat{T}\right)$$

for all $\hat{T} \in \mathbb{N}$. Hence, $\mu^a_{(f;\hat{T})} = \mu^a_{(f';\hat{T})} = \mu^a_{(f'';\hat{T})}$ for all $\hat{T} \in \mathbb{N}$ and the coefficient $\lambda(f;f'';f')$ can be chosen arbitrarily.

Applying the same reasoning inductively, we obtain that for any frequencies $(f_i)_{i=1}^n$ and f in F^T such that $\sum_{i=1}^{n}\eta_i f_i = f$, for some coefficients $\eta_i \in (0;1)$ with $\sum_{i=1}^{n}\eta_i = 1$, there are coefficients $(\lambda_i)_{i=1}^{n} \in (0;1)$ with $\sum_{i=1}^{n}\lambda_i = 1$ such that

$$\mu^a_{(f;\hat{T})} = \sum_{i=1}^{n}\lambda_i \mu^a_{(f_i;\hat{T})}$$

for all $\hat{T} \in \mathbb{N}$ such that $(f_i)_{i=1}^{n}$ and $f \in F^{\hat{T}}$. It follows that for all $\hat{T} \in \mathbb{N}$ such that $(f_i)_{i=1}^{n}$ and $f \in F^{\hat{T}}$,

$$\begin{aligned} P_a\left(f;\hat{T}\right) &= \epsilon P^* + (1-\epsilon)\left[\mu^a_{(f;\hat{T})}\delta_{\bar{r}} + \left(1-\mu^a_{(f;\hat{T})}\right)\delta_{\underline{r}}\right] = \\ &= \sum_{i=1}^{n}\lambda_i\left[\epsilon P^* + (1-\epsilon)\left[\mu^a_{(f_i;\hat{T})}\delta_{\bar{r}} + \left(1-\mu^a_{(f_i;\hat{T})}\right)\delta_{\underline{r}}\right]\right] = \sum_{i=1}^{n}\lambda_i P_a\left(f_i;\hat{T}\right). \blacksquare \end{aligned}$$

$(B4)$: By step 1, $\lim_{T\to\infty}\mu^a_{c^T}$ exists. Hence, by the definition of $P_a\left(c^T\right)$,

$$\lim_{T\to\infty} P_a\left(c^T\right) = \epsilon P^* + (1-\epsilon)\left[\lim_{T\to\infty}\mu^a_{(c;T)}\delta_{\bar{r}} + \left(1-\lim_{T\to\infty}\mu^a_{(c;T)}\right)\delta_{\underline{r}}\right]. \tag{29}$$

$(B5)$: Note that since $|R| > 3$ and $\epsilon \in (0;1)$, the set P^* is a subset of a hyperplane in $\Delta^{|R|-1}$ and has a dimension 2 or higher and so do the sets in (29) for all $c \in C$. Hence, there are no three sets of dimension 0 or 1 and $(B5)$ is trivially satisfied.$\blacksquare$

Step 3: In Eichberger and Guerdjikova (2010), we prove the following Theorem:

Theorem 6.3 *Let P_a be a correspondence $P_a : \mathbb{D} \rightrightarrows \Delta^{|R|-1}$ the images of which are non-empty convex, and compact sets and which satisfies the Axioms $(B4)$ Learning and $(B5)$ Non-collinearity. Then the following two statements are equivalent:*

P satisfies the Axioms $(B1)$ Invariance, $(B2)$ Concatenation restricted to data sets of equal length, and $(B3)$ Constant Similarity.

There exists a unique, up to multiplication by a positive number, function

$$s_a : C \to \mathbb{R}_{++}$$

such that for all $T \in \mathbb{N}$ and any $D \in \mathbb{D}^T$,

$$P_a(D) = \frac{\sum_{c\in C} s_a(c)\, f_D(c)\, P_a\left(c^T\right)}{\sum_{c'\in C} s_a(c')\, f_D(c')}. \tag{30}$$

Step 4: The similarity functions $s_a : C \to \mathbb{R}_{++}$ derived in Theorem 6.3 do not depend on the observed outcomes and can be written as $s_a : A \to \mathbb{R}_{++}$. Furthermore, for each $a \in A$, $(s_a(a_c))_{c\in C}$ are the unique up to a multiplication by a positive number $K_a > 0$ coefficients such that for each $D \in \mathbb{D}^T$, μ^a_D can be represented as:

$$\mu^a_D = \frac{\sum_{c\in C}\mu^a_{(c)^T}\tilde{s}_a(a_c)\, f_D(c)}{\sum_{c'\in C}\tilde{s}_a(a_{c'})\, f_D(c')}.$$

Proof of Step 4:

Take a, $a' \in A$ and for some $T \in \mathbb{N}$, consider a data set $D \in \mathbb{D}^T_{a'}$. We will show that $\mu^a_D =$

$\sum_{r\in R} f_D(a';r)\,\mu^a_{(a';r)^T}$, which combined with (28), and (30) implies:

$$\begin{aligned}
P_a(D) &= \sum_{r\in R} f_D\left(a';r\right) P_a\left(a';r\right)^T \\
\frac{\sum_{r\in R} s_a(a';r)\, f_D(a';r)\, P_a(a';r)^T}{\sum_{r'\in R} s_a(a';r')\, f_D(a';r')} &= \sum_{r\in R} f_D\left(a';r\right) P_a\left(a';r\right)^T \text{, or} \\
s_a\left(a';r\right) &= s_a\left(a';r'\right) \text{ for all } r,\, r' \in R.
\end{aligned}$$

To see that $\mu^a_D = \sum_{r\in R} f_D(a';r)\,\mu^a_{(a';r)^T}$, construct for each $r \in R$, the sequences $\left(\mu^{*a}_{k!}(a';r)^T\right)_{k\geq T}$ and $\left(\nu^{*a}_{k!}(a';r)^T\right)_{k\geq T}$ as in (8) and (9), however restricting them to be in $\left\{0; \frac{T}{k!}; \frac{2T}{k!} ... \frac{k!-T}{k!}; 1\right\}$ rather than $\left\{0; \frac{1}{k!}; \frac{2}{k!} ... \frac{k!-1}{k!}; 1\right\}$. Note that since for all $k \geq T$, $\mu^{*a}_{k!}(a';r)^T - \nu^{*a}_{k!}(a';r)^T \leq \frac{T}{k!}$, the convergence results obtained in Lemma 3.1 apply. Furthermore, since $\mu^{*a}_{k!}(a';r)^T - \mu^{a}_{k!}(a';r)^T \leq \frac{T}{k!}$, $\lim_{k\to\infty} \mu^{*a}_{k!}(a';r)^T = \lim_{k\to\infty} \mu^{a}_{k!}(a';r)^T = \mu^a_{(a';r)^T}$ for all $r \in R$. Note that for each $k \geq T$,

$$\sum_{r\in R} f_D\left(a';r\right) \mu^{*a}_{k!}\left(a';r\right)^T \in \left\{0; \frac{1}{k!}; \frac{2}{k!} ... \frac{k!-1}{k!}; 1\right\}.$$

Applying iteratively Axiom 4, we conclude that

$$\begin{aligned}
&\left(a; \left(\sum_{r\in R} \left(\mu^{*a}_{k!}\left(a';r\right)^T \delta_{(a;\bar{r})} + \left(1 - \mu^{*a}_{k!}\left(a';r\right)^T\right) \delta_{(a;\underline{r})}\right) f_D\left(a';r\right); k!\right)\right) \succsim (a; D) \\
&\succsim \left(a; \left(\sum_{r\in R} \left(\nu^{*a}_{k!}\left(a';r\right)^T \delta_{(a;\bar{r})} + \left(1 - \nu^{*a}_{k!}\left(a';r\right)^T\right) \delta_{(a;\underline{r})}\right) f_D\left(a';r\right); k!\right)\right).
\end{aligned}$$

Note that $\lim_{k\to\infty} \sum_{r\in R} \mu^{*a}_{k!}(a';r)^T f_D(a';r) = \lim_{k\to\infty} \sum_{r\in R} \nu^{*a}_{k!}(a';r)^T f_D(a';r) = \sum\limits_{r\in R} \mu^a_{(a';r)^T} f_D(a';r)$. Assume that $\mu^a_D > \sum\limits_{r\in R} \mu^a_{(a';r)^T} f_D(a';r)$. Then there is a $k \in \mathbb{N}$ such that

$$\mu^a_{k!}(D) \geq \nu^a_{k!}(D) > \sum_{r\in R} \mu^{*a}_{k!}\left(a';r\right)^T f_D\left(a';r\right),$$

a contradiction. A symmetric argument applies to the case $\mu^a_D < \sum\limits_{r\in R} \mu^a_{(a';r)^T} f_D(a';r)$ and we conclude that $\mu^a_D = \sum\limits_{r\in R} \mu^a_{(a';r)^T} f_D(a';r)$.

By Axiom 6, for any a, a', there are outcomes r and r' such that $(a;(a;r)) \not\sim (a;(a';r'))$ and, hence, $\mu^a_{(a;r)^T} \neq \mu^a_{(a';r')^T}$ for all $T \in \mathbb{N}$. Since s_a does not depend on r, (30) implies:

$$P_a\left(\left(\frac{1}{2}\delta_{(a;r)} + \frac{1}{2}\delta_{(a';r')}; T\right)\right) = \frac{s_a(a)\, P_a(a;r)^T + s_a(a')\, P_a(a';r')^T}{s_a(a) + s_a(a')}.$$

whereas by the definition of $P_a(D)$ in (28), we obtain that for all a, $a' \in A$, $s_a(a)$ and $s_a(a')$ are the unique up to a multiplication by a positive number coefficients that satisfy:

$$\mu^a_{\left(\frac{1}{2}\delta_{(a;r)} + \frac{1}{2}\delta_{(a';r')}; T\right)} = \frac{s_a(a)\, \mu^a_{(a;r)^T} + s_a(a')\, \mu^a_{(a';r')^T}}{s_a(a) + s_a(a')}.$$

It follows that the $s_a(a_c)$ are the unique up to a multiplication by a positive number $K_a > 0$ coefficients such that for each $D \in \mathbb{D}^T$, μ_D^a can be represented as:

$$\mu_D^a = \frac{\sum_{c \in C} \mu_{(c)^T}^a \tilde{s}_a(a_c) f_D(c)}{\sum_{c' \in C} \tilde{s}_a(a_{c'}) f_D(c')}.$$

We thus obtain that V on $A \times \mathbb{D}$ satisfies:

$$V(a; D) = u \cdot \frac{\sum_{c \in C} s_a(a_c) f_D(c) \hat{h}_a(c^T)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')} \text{ for each } \hat{h} \in \hat{H}_a(c^T). \blacksquare$$

Lemma 6.4 *There exist a coefficient $\alpha \in [0;1]$ satisfying $\alpha u(\bar{r}) + (1-\alpha) u(\underline{r}) = u(\hat{r})$, a strictly decreasing sequence $(\gamma_T)_{T \in \mathbb{N}}$ satisfying $\gamma_T \in (0;1)$ and $\lim_{T \to \infty} \gamma_T = 0$ and a function $\hat{h} : A \times C \to \Delta^{|R|-1}$ with $\hat{h}_a^c \in \lim_{T \to \infty} \hat{H}_a(c^T)$ and $\hat{h}_a^{(a;r)} = \delta_r$ for all $a \in A$, $c \in C$ and $r \in R$ such that*

$$\hat{H}_a(c^T) = \left\{ h \in \Delta^{|R|-1} \mid u \cdot h = u \cdot \left[\gamma_T (\alpha \delta_{\bar{r}} + (1-\alpha)\delta_{\underline{r}}) + (1-\gamma_T) \hat{h}_a^c \right] \right\}.$$

The coefficient α and the sequence $(\gamma_T)_{T \in \mathbb{N}}$ are unique.

Proof of Lemma 6.4:

Step 1: Define $\gamma_T =: \alpha_T + \beta_T$, with $\alpha_T =: \mu_{(a;\underline{r})^T}^a$ and $\beta_T =: 1 - \mu_{(a;\bar{r})^T}^a$. Then, $\gamma_T > 0$, the sequence $(\gamma_T)_{T \in \mathbb{N}}$ is decreasing and converges to 0.

Proof of Step 1

Observe that by Axioms 6 and 9, $(\alpha_T)_{T \in \mathbb{N}}$, $(\beta_T)_{T \in \mathbb{N}}$ and $(\gamma_T)_{T \in \mathbb{N}}$ are decreasing. If $(a;(a;\hat{r})) \succ (a;(a;\underline{r}))$, α_T, is strictly decreasing and converges to 0, whereas if $(a;(a;\hat{r})) \sim (a;(a;\underline{r}))$, $\alpha_T = 0$ for all $T \in \mathbb{N}$. If $(a;(a;\bar{r})) \succ (a;(a;\hat{r}))$, then β_T is strictly decreasing and converges to 0, whereas if $(a;(\alpha;\hat{r})) \sim (a;(a;\bar{r}))$, $\beta_T = 0$ for all $T \in \mathbb{N}$. Since by Axiom 6, $(a;(a;\bar{r})) \succ (a;(a;\underline{r}))$, $(\gamma_T)_{T \in \mathbb{N}}$ is always strictly decreasing, converges to 0 and $\gamma_T > 0$.

Step 2: There is a function $\hat{h} : A \times C \to \mathbb{R}$ such that for all $a \in A$, $r \in R$, $c \in C$ and $T \in \mathbb{N}$, $\hat{h}_a^c \in \lim_{T \to \infty} \hat{H}_a(c^T)$, $\hat{h}_a^{(a;r)} = \delta_r$ and

$$\hat{H}_a(c^T) = \left\{ \hat{h} \in \Delta^{|R|-1} \mid u \cdot h = u \cdot \left[\alpha_T \delta_{\bar{r}} + \beta_T \delta_{\underline{r}} + (1-\gamma_T) \hat{h}_a^c \right] \right\}.$$

Proof of Step 2

Consider a case $c \in C$. By Axioms 6 and 10, there is a $\lambda_c =: \lambda\left(\delta_{(a;\bar{r})}; \delta_{(a;\underline{r})}; \delta_c\right) \in [0;1]$ independent of T such that

$$\begin{aligned} V(a; c^T) &= \mu_{c^T}^a = \lambda_c \mu_{(a;\bar{r})^T}^a + (1-\lambda_c) \mu_{(a;\underline{r})^T}^a \\ &= \alpha_T + (1-\gamma_T) \lambda_c \end{aligned}$$

By step 1 of Lemma 6.2, we can take $\lim_{T \to \infty}$ on both sides of the equation and obtain $\lambda_c = \lim_{T \to \infty} \mu_{c^T}^a$. Thus, for any

$$\hat{h}_a^c \in \left\{ \hat{h} \mid u \cdot \hat{h} = \lim_{T \to \infty} \mu_{c^T}^a \right\} = \lim_{T \to \infty} \hat{H}_a(c^T),$$

$$V\left(a;c^{T}\right)=u\cdot\left[\alpha_{T}\delta_{\bar{r}}+\beta_{T}\delta_{\underline{r}}+\left(1-\gamma_{T}\right)\hat{h}_{a}^{c}\right], \tag{31}$$

and, thus

$$\hat{H}_{a}\left(c^{T}\right)=\left\{\hat{h}\in\Delta^{|R|-1}\mid u\cdot\hat{h}=u\cdot\left[\alpha_{T}\delta_{\bar{r}}+\beta_{T}\delta_{\underline{r}}+\left(1-\gamma_{T}\right)\hat{h}_{a}^{c}\right]\right\}.$$

In particular, since by (23), for every $a\in A$, $r\in R$, $\delta_{r}\in\lim_{T\to\infty}\hat{H}_{a}\left((a;r)^{T}\right)$, we can set $\hat{h}_{a}^{(a;r)}=\delta_{r}$.■

<u>Step 3</u>: The ratio $\frac{\alpha_{T}}{\alpha_{T}+\beta_{T}}$ does not depend on T and equals $u\left(\hat{r}\right)$. Hence, we can define $\alpha:=\frac{\alpha_{T}}{\alpha_{T}+\beta_{T}}$ and obtain:

$$\hat{H}_{a}\left(c^{T}\right)=\left\{h\in\Delta^{|R|-1}\mid u\cdot h=u\cdot\left[\gamma_{T}\left(\alpha\delta_{\bar{r}}+\left(1-\alpha\right)\delta_{\underline{r}}\right)+\left(1-\gamma_{T}\right)\hat{h}_{c}^{a}\right]\right\}.$$

<u>Proof of Step 3</u>

By Axiom 8, the sequence $\left(V\left(a;(a;\hat{r})^{T}\right)\right)_{T\in\mathbb{N}}$ is constant and by (31), it can be written as:

$$\begin{aligned}V\left(a;(a;\hat{r})^{T}\right) &= \alpha_{T}+\left(1-\alpha_{T}-\beta_{T}\right)u\left(\hat{r}\right)\\ &= u\left(\hat{r}\right)\mu_{(a;\bar{r})^{T}}^{a}+\left(1-u\left(\hat{r}\right)\right)\mu_{(a;\underline{r})^{T}}^{a}\end{aligned}$$

By step 1 of Lemma 6.2, we can take limits on both sides and obtain $\lim_{T\to\infty}V\left(a;(a;\hat{r})^{T}\right)=V\left(a;(a;\hat{r})^{T}\right)=u\left(\hat{r}\right)$ for all $T\in\mathbb{N}$. It follows that:

$$\begin{aligned}\alpha_{T}+\left(1-\alpha_{T}-\beta_{T}\right)u\left(\hat{r}\right) &= u\left(\hat{r}\right) \qquad (32)\\ \frac{\alpha_{T}}{\left(\alpha_{T}+\beta_{T}\right)} &= u\left(\hat{r}\right)\in[0;1].\blacksquare\end{aligned}$$

<u>Step 4</u>: $\gamma_{T}\in(0;1)$ for all $T\in\mathbb{N}$. The coefficients α and $\left(\gamma_{T}\right)_{T\in\mathbb{N}}$ are unique.

<u>Proof of Step 4</u>

Using (32), we obtain for any $a\in A$, $r\in R$, $T\in\mathbb{N}$,

$$V\left(a;(a;r)^{T}\right)=\gamma_{T}\alpha+\left(1-\gamma_{T}\right)u\left(r\right)=\gamma_{T}u\left(\hat{r}\right)+\left(1-\gamma_{T}\right)u\left(r\right). \tag{33}$$

If $(a;(a;r))\succ(a;(a;\hat{r}))$, Axiom 9 implies that for all $T\in\mathbb{N}$,

$$V\left(a;(a;r)^{T+1}\right)>V\left(a;(a;r)^{T}\right)>u\left(\hat{r}\right)$$

Since by (23) $\lim_{T\to\infty}V\left(a;(a;r)^{T}\right)=u\left(r\right)$, we obtain that for all $T\in\mathbb{N}$, $V\left(a;(a;r)^{T}\right)\in\left(u\left(\hat{r}\right);u\left(r\right)\right)$. It follows that $\gamma_{T}\in(0;1)$ for all $T\in\mathbb{N}$. The argument for $(a;(a;\hat{r}))\succ(a;(a;r))$ is symmetric. Note that the case in which $(a;(a;r))\sim(a;(a;\hat{r}))$ holds for all $r\in R$ is excluded by Axiom 6.

The uniqueness of α follows immediately from the requirement $\alpha u\left(\bar{r}\right)+\left(1-\alpha\right)u\left(\underline{r}\right)=u\left(\hat{r}\right)$ together with the fact that $u\left(\bar{r}\right)>u\left(\underline{r}\right)$. Once α is determined in this way, the uniqueness of the sequence γ_{T} follows from (33) and the fact that there is at least one $r\in R$ such that $u\left(r\right)\neq u\left(\hat{r}\right)$.■

Lemma 6.5 *There exist minimal coefficients $\gamma_a^c \in [0;1]$ and a prediction function $\rho : A \times C \to R$ such that for every $c \in C$ and every $a \in A$,*

$$u \cdot \hat{h}_a^c = u \cdot \left[\gamma_a^c \left(\alpha \delta_{\bar{r}} + (1-\alpha)\,\delta_{\underline{r}}\right) + (1-\gamma_a^c)\,\delta_{\rho_a^c}\right], \tag{34}$$

where $\hat{h}_a^c$ and α are those identified in Lemma 6.4. The minimal coefficients γ_a^c are unique and satisfy $\gamma_a^{(a;r)} = 0$ for all $a \in A$ and all $r \in R$. The prediction function is unique up to indifference: if ρ and $\tilde{\rho}$ are two functions which satisfy (34), $(a;(a;\rho_a^c)) \sim (a;(a;\tilde{\rho}_a^c))$ holds for all $a \in A$ and all $c \in C$. Furthermore, $\rho_a^{(a;r)} = r$ for all $a \in A$ and all $r \in R$.

Proof of Lemma 6.5:

For each a, $a' \in A$ define the function $\rho_a^{(a';r)}$ as follows: for r such that $(a;(a;\hat{r})) \succ (a;(a';r))$, let

$$\rho_a^{(a';r)} =: \left\{ \begin{array}{c} \tilde{r} \in R \mid (a;(a';r)) \succsim (a;(a;\tilde{r})) \text{ and there is no } r' \in R \text{ such that} \\ (a;(a';r)) \succsim (a;(a;r')) \succ (a;(a;\tilde{r})) \end{array} \right\}, \tag{35}$$

for $(a;(a';r)) \succ (a;(a;\hat{r}))$, let

$$\rho_a^{(a';r)} =: \left\{ \begin{array}{c} \tilde{r} \in R \mid (a;(a;\tilde{r})) \succsim (a;(a';r)) \text{ and there is no } r' \in R \text{ such that} \\ (a;(a;\tilde{r})) \succ (a;(a;r')) \succsim (a;(a';r)) \end{array} \right\} \tag{36}$$

and for r such that $(a;(a';r)) \sim (a;(a;\hat{r}))$, let $\rho_a^{(a';r)} = \hat{r}$. If for a given case $(a';r)$ several outcomes $\tilde{r}$ satisfy condition (35) or (36), fix one of these outcomes arbitrarily and set $\rho_a^{(a';r)}$ equal to the so selected outcome.

Now define the coefficients $\gamma_a^{(a';r)}$ so that they satisfy:

$$\gamma_a^{(a';r)} \lim_{T\to\infty} \mu^a_{(a';\hat{r})^T} + \left(1 - \gamma_a^{(a';r)}\right) \lim_{T\to\infty} \mu^a_{\left(a;\rho_a^{(a';r)}\right)^T} = \lim_{T\to\infty} \mu^a_{(a';r)^T}. \tag{37}$$

By the definition of $\rho_a^{(a';r)}$, such coefficients $\gamma_a^{(a';r)} \in [0;1]$ always exist. They are unique, except for the case of $(a;(a';r)) \sim (a;(a;\hat{r}))$ and we set $\gamma_a^{(a';r)} = 0$ for this case.

Since $\hat{h}_a^{(a';r)} \in \lim_{T\to\infty} H_a(a';r)^T$, (23) and (25) imply:

$$u \cdot \hat{h}_a^{(a';r)} = \gamma_a^{(a';r)} u(\hat{r}) + \left(1-\gamma_a^{(a';r)}\right) u\left(\rho_a^{(a';r)}\right) = u \cdot \left[\gamma_a^{(a';r)} \left(\alpha\delta_{\bar{r}} + (1-\alpha)\,\delta_{\underline{r}}\right) + \left(1-\gamma_a^{(a';r)}\right) \delta_{\rho_a^{(a';r)}}\right]. \tag{38}$$

To see that the so defined coefficients γ_a^c are minimal, suppose that there exists a $\tilde{\rho} \neq \rho$ and a corresponding set of coefficients $\tilde{\gamma}_a^c$ which satisfy (34). Expression (38) implies that $\tilde{\rho}$ and $\tilde{\gamma}_a^c$ have to satisfy (37) for all $a \in A$ and $c \in C$. Since $\tilde{\rho} \neq \rho$, there exists an $a \in A$ and a $c \in C$ such that $(a;(a;\tilde{\rho}_a^c)) \nsim (a;(a;\rho_a^c))$. The definition of ρ_a^c together with (37) then implies that $\tilde{\gamma}_a^c > \gamma_a^c$. Hence, the definition of ρ ensures that for each $a \in A$, $c \in C$, the coefficient γ_a^c is minimal. In particular, for $a_c = a$, we have $\rho_a^{(a;r)} = r$ and, hence, $\gamma_a^{(a;r)} = 0$ for all $r \in R$. Finally, by (38), once the minimal coefficients γ_a^c have been determined, $\rho_a^{(a';r)}$ is unique up to indifference, i.e., $\tilde{r}$ and $\tilde{r}'$ both satisfy the definition of $\rho_a^{(a';r)}$ if and only if $u(\tilde{r}) = u(\tilde{r}')$, or $(a;(a;\tilde{r})) \sim (a;(a;\tilde{r}'))$. ■

For $a \in A$ and $D \in \mathbb{D}$, define $H_a(D)$ as

$$H_a(D) = \left[\gamma_T + (1-\gamma_T)\frac{\sum_{c\in C}\gamma_a^c f_D(c)\, s_a(a_c)}{\sum_{c'\in C} f_D(c')\, s_a(a_{c'})}\right]\Delta^{|R|-1} + (1-\gamma_T)\frac{\sum_{c\in C}(1-\gamma_a^c) f_D(c)\, s_a(a_c)\,\delta_{\rho_a^c}}{\sum_{c'\in C} f_D(c')\, s_a(a_{c'})}, \tag{39}$$

where $(s_a)_{a\in A}$ is the family of similarity functions derived in Lemma 6.2, $(\gamma_T)_{T\in\mathbb{N}}$ is the sequence of perceived degrees of ambiguity derived in Lemma 6.4, ρ is the prediction function and γ_a^c are the coefficients of perceived ambiguity derived in Lemma 6.5. Let $H_a(D_\varnothing) = \Delta^{|R|-1}$.

By Lemma 6.1, the function V represents $\succsim$. By Lemmas 6.2, 6.4 and 6.5, V on $A \times \mathbb{D}$ can be written as:

$$V(a;D) = \tag{40}$$

$$= u\cdot\sum_{c\in C}\frac{s_a(a_c)\, f_D(c)}{\sum_{c'\in C} s_a(a_{c'})\, f_D(c')}\left[\gamma_T\left(\alpha\delta_{\bar{r}} + (1-\alpha)\,\delta_{\underline{r}}\right) + (1-\gamma_T)\left[\gamma_a^c\left(\alpha\delta_{\bar{r}} + (1-\alpha)\,\delta_{\underline{r}}\right) + (1-\gamma_a^c)\,\delta_{\rho_a^c}\right]\right]$$

$$= \alpha u\cdot\left(\left[\gamma_T + (1-\gamma_T)\sum_{c\in C}\gamma_a^c\frac{s_a(a_c)\, f_D(c)}{\sum_{c'\in C} s_a(a_{c'})\, f_D(c')}\right]\delta_{\bar{r}} + (1-\gamma_T)\frac{\sum_{c\in C}(1-\gamma_a^c)\, s_a(a_c)\, f_D(c)\,\delta_{\rho_a^c}}{\sum_{c'\in C} s_a(a_{c'})\, f_D(c')}\right)$$

$$+ (1-\alpha)\, u\cdot\left(\left[\gamma_T + (1-\gamma_T)\sum_{c\in C}\gamma_a^c\frac{s_a(a_c)\, f_D(c)}{\sum_{c'\in C} s_a(a_{c'})\, f_D(c')}\right]\delta_{\underline{r}} + (1-\gamma_T)\frac{\sum_{c\in C}(1-\gamma_a^c)\, s_a(a_c)\, f_D(c)\,\delta_{\rho_a^c}}{\sum_{c'\in C} s_a(a_{c'})\, f_D(c')}\right),$$

whereas by Lemma 6.1, $V(a;D_\varnothing) = u(\hat{r}) = \alpha u\cdot\delta_{\bar{r}} + (1-\alpha)\, u\cdot\delta_{\underline{r}}$.

Combining (39) and (40), we obtain

$$V(a;D) = \alpha \max_{p\in H_a(D)} u\cdot p + (1-\alpha)\min_{p\in H_a(D)} u\cdot p$$

which completes the proof of the existence part of the Theorem. It remains to verify the uniqueness of the utility function u, which follows immediately from the fact that $\rho_a^{(a;r)} = r$ and, hence, $\delta_r \in \lim_{T\to\infty}\hat{H}_a(a;r)^T$ for all $a \in A$ and all $r \in R$.

<u>Sketch of the proof of Theorem 4.2</u>:

Since the arguments of the proof follow very closely those in the proof of Theorem 4.1, we just provide a sketch. Start, as in Lemma 3.1 by showing that the unambiguous equivalents for every data set, μ_D^a exist. Note that in the proof of Step 3 of Lemma 3.1, we can use Axiom 3A instead of Axiom 9 to conclude that for all k,

$$\left(a;\left(\mu_{k!}^a(D)\,\delta_{(a;\bar{r})} + (1-\mu_{k!}^a(D))\,\delta_{(a;\underline{r})};(k+1)!\right)\right) \sim \left(a;\left(\mu_{k!}(D)\,\delta_{(a;\bar{r})} + (1-\mu_{k!}(D))\,\delta_{(a;\underline{r})};k!\right)\right)$$

and obtain the convergence result. In analogy to Lemma 6.1, show that $V(a;D) =: \mu_D^a$ represents $\succsim$. Note that by Axiom 3A, $\mu_{c^T}^a = \mu_c^a$ for all $a \in A$, all $c \in C$ and all $T \in \mathbb{N}$ and hence, λ_r can be defined as in (22). This identifies $u(r)$ and consequently, $\hat{H}_a(D)$. Use the same arguments as in Lemma 6.2 to identify the similarity function. In Lemma 6.4, define $\alpha =: \mu_{D_\varnothing}^a$. This α is obviously unique and satisfies property (iv) of the Theorem. Note that since $\mu_{(a;r)^T}^a = const$ for all T, defining γ_T as in the proof of Lemma 6.4 implies $\gamma_T = 0$ for all $T \in \mathbb{N}$. The so identified α and $\gamma_T = 0$ can be used to represent $\hat{H}_a\left(c^T\right)$ as in the statement of Lemma 6.4. Finally, replicate Lemma 6.5 to identify γ_a^c and the function ρ.

Proof of (16) for Example 3:

Using assumptions (i) — (iv), one easily derives the following evaluation of a_1^H:

$$V\left(a_1^H;D\right)=$$

$$=(1-\gamma_T)\left[\frac{\sum_{r\in R}\left[2f_D\left(a_1;r\right)\delta_r+\left(1-\gamma^{a_2^H}\right)s_Hf_D\left(a_2;r\right)\right]u\left(\rho^{\left(a_2^H;r\right)}\right)}{\sum_{r'\in R}\left[2f_D\left(a_1;r'\right)+s_Hf_D\left(a_2;r'\right)\right]}\right]$$
$$+\left[\gamma_T+(1-\gamma_T)\sum_{r\in R}\frac{\left(1-\gamma^{a_2^H}\right)s_Hf_D\left(a_2;r\right)}{\sum_{r'\in R}\left[2f_D\left(a_1;r'\right)+s_Hf_D\left(a_2;r'\right)\right]}\right]\left(\alpha u\left(\bar{r}\right)+(1-\alpha)u\left(\underline{r}\right)\right)$$

Similarly, for a_1^F, we obtain:

$$V\left(a_1^F;D\right)=$$

$$=\quad(1-\gamma_T)\left[\frac{\sum_{r\in R}2f_D\left(a_1;r\right)\delta_r+\left(1-\gamma^{a_2^H}\right)s_Ff_D\left(a_2;r\right)u\left(\rho^{\left(a_2^H;r\right)}\right)}{2\sum_{r'\in R}f_D\left(a_1;r'\right)+s_F\sum_{r'\in R}f_D\left(a_2;r'\right)}\right]$$
$$+\left[\gamma_T+(1-\gamma_T)\sum_{r\in R}\frac{\left(1-\gamma^{a_2^H}\right)s_Ff_D\left(a_2;r\right)}{\sum_{r'\in R}\left[2f_D\left(a_1;r'\right)+s_Ff_D\left(a_2;r'\right)\right]}\right]\left(\alpha u\left(\bar{r}\right)+(1-\alpha)u\left(\underline{r}\right)\right)$$

Observing that $\sum_{r\in R}f_D\left(\left(a_1;r\right)\right)=\frac{T_1}{T}$ and $\sum_{r\in R}f_D\left(\left(a_2;r\right)\right)=\frac{T_2}{T}$, implies:

$$\begin{aligned}&V\left(a_1^H;D\right)-V\left(a_1^F;D\right)\\=\;&\frac{\sum_{r\in R}\left[2f_D\left(a_1;r\right)\delta_r+\left(1-\gamma^{a_2^H}\right)s_Hf_D\left(\left(a_2;r\right)\right)u\left(\rho^{\left(a_2^H;r\right)}\right)\right]}{\sum_{r'\in R}\left[2f_D\left(a_1;r'\right)+s_Hf_D\left(a_2;r'\right)\right]}\\&-\frac{\sum_{r\in R}\left[2f_D\left(a_1;r\right)\delta_r+\left(1-\gamma^{a_2^H}\right)s_Ff_D\left(a_2;r\right)u\left(\rho^{\left(a_2^H;r\right)}\right)\right]}{\sum_{r'\in R}\left[2f_D\left(a_1;r'\right)+s_Ff_D\left(a_2;r'\right)\right]}\\=\;&\frac{2T_1T_2\left(s_H-s_F\right)}{\left[2T_1+s_HT_2\right]\left[2T_1+s_FT_2\right]}\left[\left(1-\gamma^{a_2^H}\right)\sum_{r\in R}\frac{f_D\left(a_2;r\right)u\left(\rho^{\left(a_2^H;r\right)}\right)}{\sum_{r'\in R}f_D\left(a_2;r'\right)}-\left[\sum_{r\in R}\frac{f_D\left(\left(a_1;r\right)\right)u\left(r\right)}{\sum_{r'\in R}f_D\left(a_1;r'\right)}\right]\right].\blacksquare\end{aligned}$$

Proof of Lemma 4.3:

Examination of the proof of Theorem 4.1 shows that the utility function satisfies $u\left(r\right)=\lambda\left(\delta_{(a;\bar{r})};\delta_{(a;\underline{r})};\delta_{(a;r)}\right)$, the similarity function is given by $s_a\left(a_c\right)=\frac{1-\lambda\left(\delta_{(a;r)};\delta_c;\frac{1}{2}\delta_{(a;r)}+\frac{1}{2}\delta_c\right)}{\lambda\left(\delta_{(a;r)};\delta_c;\frac{1}{2}\delta_{(a;r)}+\frac{1}{2}\delta_c\right)}$ for any $r\in R$, ρ_a^c is identified by preferences on $A\times C$ and $\gamma_a^c=\lambda\left(\delta_{(a';\hat{r})};\delta_{(a;\rho_a^c)};\delta_c\right)$. It follows that if two individuals have the same preferences on the set of data sets of equal length and identical $\lambda\left(\cdot\right)$-functions, then their utility functions are identical up to an affine linear transformation, their similarity functions are identical up to a multiplication by a positive constant, their prediction functions ρ are identical up to indifference and their minimal coefficients of perceived ambiguity γ_a^c are identical.

Proof of Proposition 4.4:

Suppose that $\alpha^i \leq \alpha^j$. According to Axiom 9, $\left(a; D^k\right) \succ_i (a; D)$ will hold whenever $\left(a; D^k\right) \succ_i (a; \hat{r}_i)$, or whenever,

$$\left(\gamma_T^i + \left(1-\gamma_T^i\right) \frac{\sum_{c\in C} \gamma_a^c f_D(c) s_a(a_c)}{\sum_{c'\in C} f_D(c') s_a(a_{c'})}\right) \left[\alpha^i u(\bar{r}) + \left(1-\alpha^i\right) u(\underline{r})\right] +$$
$$+\left(1-\gamma_T^i\right) \frac{\sum_{c\in C} (1-\gamma_a^c) f_D(c) s_a(a_c)}{\sum_{c'\in C} f_D(c') s_a(a_{c'})} u(\rho_a^c) \geq \left[\alpha^i u(\bar{r}) + \left(1-\alpha^i\right) u(\underline{r})\right],$$

or

$$\left(1-\gamma_T^i\right)\left[\sum_{c\in C} (1-\gamma_a^c) \frac{f_D(c) s_a(a_c)}{\sum_{c'\in C} f_D(c') s_a(a_{c'})}\right]\left[\sum_{c\in C} \frac{(1-\gamma_a^c) f_D(c) s_a(a_c) u(\rho_a^c)}{\sum_{c'\in C}\left(1-\gamma_a^{c'}\right) f_D(c') s_a(a_{c'})} - \left[\alpha^i u(\bar{r}) + \left(1-\alpha^i\right)^i u(\underline{r})\right]\right] \tag{41}$$

Since $\alpha^i \leq \alpha^j$, and therefore,

$$\left[\alpha^i u(\bar{r}) + \left(1-\alpha^i\right) u(\underline{r})\right] \leq \left[\alpha^j u(\bar{r}) + \left(1-\alpha^j\right) u(\underline{r})\right],$$

(41) implies that $\left(a; D^k\right) \succ_j (a; D)$.

Conversely, if for all $(a; D)$, $\left(a; D^k\right) \succ_j (a; D)$ implies $\left(a; D^k\right) \succ_i (a; D)$, it follows by Axiom 9 that whenever

$$\begin{aligned} (a; D) &\succ_j (a; (a; \hat{r}_j)), \\ (a; D) &\succ_i (a; (a; \hat{r}_i)). \end{aligned}$$

Since the utility functions of i and j are identical, this implies that $u(\hat{r}_j) \geq u(\hat{r}_i)$. Normalizing $u(\bar{r}) = 1$, $u(\underline{r}) = 0$ and noting that $u(\hat{r}_j) = \alpha^j \geq u(\hat{r}_i) = \alpha^i$, implies the result of the proposition.■

7 References

AHN, D. (2008). "Ambiguity Without a State Space", *Review of Economic Studies 71*, 3-28.

ARAD, A., AND GAYER, G. (2010)."Imprecise Datasets as a Source for Ambiguity: A Model and Experimental Evidence". *mimeo*, Tel Aviv University.

BILLOT, A., GILBOA, I., SAMET, D. AND SCHMEIDLER, D. (2005). "Probabilities as Similarity-Weighted Frequencies". *Econometrica 73*, 1125-1136.

BEWLEY, T. F. (1986). "Knightian Decision Theory: Part I", Discussion Paper, *Cowles Foundation*.

CHATEAUNEUF, A., EICHBERGER , J., AND GRANT, S. (2007). "Choice Under Uncertainty with the Best and the Worst in Mind: Neo-Additive Capacities", *Journal of Economic Theory 137*, 538-567.

COIGNARD, Y., AND JAFFRAY, J.-Y. (1994). "Direct Decision Making" in: Decision Theory and Decision Analysis: Trends and Challenges", Rios, S. (ed.). Boston: Kluwer Academic Publishers.

EICHBERGER, J., GRANT, S., KELSEY, D. AND KOSHEVOY, G. (2011). "Differentiating

Ambiguity: A Comment", *Journal of Economic Theory,* forthcoming

EICHBERGER, J., AND GUERDJIKOVA, A. (2010). "Case-Based Belief Formation under Ambiguity", *Mathematical and Social Sciences 60,* 161-177.

EICHBERGER, J., AND GUERDJIKOVA, A. (2011). "Technology Adoption and Adaptation to Climate Change — A Case-Based Approach", *mimeo.*

ELLSBERG, D. (1961). "Risk, Ambiguity and the Savage Axioms", *Quarterly Journal of Economics 75*, 643-669.

EPSTEIN, L. AND SCHNEIDER, M. (2007). "Learning Under Ambiguity", *Review of Economic Studies 74,* 1275-1303

FISHBURN, P. (1970). *Utility Theory for Decision Making* , New York: John Wiley and Sons.

GAJDOS, TH., HAYASHI, T., TALLON, AND J.-M., VERGNAUD, J.-C. (2007). "Attitude Towards Imprecise Information", *Journal of Economic Theory, 141*, 68-99.

GHIRARDATO, P., MACCHERONI, F., AND MARINACCI, M. (2004). "Differentiating Ambiguity and Ambiguity Attitude", *Journal of Economic Theory 118*, 133-173.

GILBOA, I., LIEBERMAN, O. AND SCHMEIDLER, D. (2004). "Empirical Similarity", *Review of Economics and Statistics,* forthc.

GILBOA, I., AND SCHMEIDLER, D. (2001). *A Theory of Case-Based Decisions.* Cambridge, UK: Cambridge University Press.

GILBOA, I., AND SCHMEIDLER, D. (1997). ”Act Similarity in Case-Based Decision Theory”, *Economic Theory 9*, 47-61.

GILBOA, I., AND SCHMEIDLER, D. (1989). ”Maxmin Expected Utility with a Non-Unique Prior”, *Journal of Mathematical Economics 18,* 141-153.

GILBOA, I., SCHMEIDLER, D., WAKKER, P. (2002). ”Utility in Case-Based Decision Theory”, *Journal of Economic Theory 105,* 483-502.

GONZALES, CH. AND JAFFRAY, J.-Y. (1998). "Imprecise Sampling and Direct Decision Making", *Annals of Operations Research 80*, 207-235.

GRANT, S., KAJI, A., POLAK, B. (1998). "Instrinsic Preference for Information", *Journal of Economic Theory, 83,* 233-259.

HAU, R., PLESKAC, T., AND HERTWIG, R. (2010). "Decisions from Experience and Statistical Probabilities: Why They Trigger Different Choices than A Priory Probabilities". *Journal of Behavioral Decision Making 23,* 48-68.

HUME, D. (1748). An Enquiry Concerning Human Understanding, Oxford: Clarendon Press.

KEYNES, J. M. (1921). *A Treatise on Probability*. London:Macmillan.

KNIGHT, F.H. (1921). *Risk, Uncertainty and Profit.* New York: Dover Publications 2006.

KLIBANOFF, P., MARINACCI, M., AND MUKERJI, S. (2005). "A Smooth Model of Decision Making Under Uncertainty", *Econometrica 73,* 1849-1892.

MANSKI, C. (2000). “Identi cation Problems and Decisions under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice”, *Journal of Econometrics 95,* 415–432.

SAVAGE, L. J. (1954). *The Foundations of Statistics.* New York: John Wiley and Sons.

SCHMEIDLER, D. (1989). “Subjective Probability and Expected Utility Without Additivity,” *Econometrica, 57,* 571—587.

STINCHCOMBE, M. (2003). "Choice with Ambiguity as Set of Probabilities", *mimeo*, University of Texas, Austin.

VON NEUMANN, J., MORGENSTERN, O. (1944). *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ.