

Ambiguity, Data and Preferences for Information — A Case-Based Approach¹

Jürgen Eichberger² and Ani Guerdjikova³

This version: April 3, 2012

Abstract

In this paper we suggest a behavioral approach to decision making under ambiguity based on available information. A decision situation is characterized by a set of actions, a set of outcomes, and data consisting of action-outcome pairs. Decision-makers express preferences over actions and data sets. 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 subjective features of the decision maker (similarity of observations and perceived ambiguity). 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 and provides a behavioral foundation for Billot, Gilboa, Samet and Schmeidler's (2005) representation.

¹ This research was supported by grant PCIG-GA-2011-293529 of the European Research Council as well as by grant ANR 2011 CHEX 006 01 of the Agence Nationale de Recherche. We would like to thank 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, George Washington University, Ecole Polytechnique and University of Paris 2 for helpful suggestions and comments.

² 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

Since Hume's (1748) "Enquiry Concerning Human Understanding", it has been recognized that the notion of induction is fundamental to human knowledge. "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 heterogeneous 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 from data into beliefs over outcomes. In this paper, we pursue a behavioral approach to this issue. We use the case-based framework pioneered by Gilboa and Schmeidler (2001) to study agents who 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 impose axioms which allow us to deduce agents' beliefs about uncertain outcomes and directly relate them to the content of the available data. In particular, we identify the perceived ambiguity of information and decompose it into two parts: ambiguity due to limited number of observations and ambiguity due to data heterogeneity. We also derive a representation of preferences which shows how data influence 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, since 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, see Manski (2000). Hence, the agent's beliefs will have to combine 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. The state-based approach established by Savage (1954) derives a purely subjective probability distribution over states. It thus completely detaches prior beliefs from objectively given information and leaves no room for using data inductively in the process of belief formation.

The objective approach integrates available probabilistic information directly into the decision maker's beliefs. When such information is not directly available, non-Bayesian statisticians use observed frequencies in the data to infer probabilities of outcomes. In this case, the number and the relevance of observations become an issue.

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. By considering the set of all possible data sets, our model spans a universe of possible scenarios, in which both objective and subjective factors determine beliefs.

Decision theory has so far treated the issues of data and ambiguity separately. Following Knight's (1921) work, the literature deals with two types of situations: risk and uncertainty. In the expected utility framework developed by von Neumann and Morgenstern (1944) and by Savage (1954), this distinction is inconsequential. The models of Knightian uncertainty, which have emerged in response to Ellsberg's (1961) experiments, highlight the importance of ambiguity and ambiguity aversion for human behavior. Different models of ambiguity and ambiguity attitude have been provided by Bewley (1986), Schmeidler (1989), Gilboa and Schmeidler (1989), Ghirardato, Maccheroni and Marinacci (2004), Klibanoff, Marinacci and Mukerji (2005) 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. At the opposite end of the spectrum, several recent papers, Stinchcombe (2003), Gajdos, Hayashi, Tallon and Vergnaud (2007) (henceforth, GHTV (2007)), Ahn (2008), assume that ambiguity can be related to an objectively given set of probability distributions. This allows them to separate the "objective ambiguity" from the "subjective attitude" towards ambiguity.

Case-based decision theory developed by Gilboa and Schmeidler (2001) incorporates data di-

rectly into the decision making process⁴. 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. The different representations fail, however, to separate beliefs over outcomes from the evaluation of these outcomes. Billot, Gilboa, Samet and Schmeidler (2005) (henceforth, BGSS (2005)) represent beliefs as similarity-weighted frequencies of observations. However, their method is not behavioral, i.e., the existence of such beliefs is postulated. Moreover, it does not take into account ambiguity of data and attitude towards such ambiguity.

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

In this paper, we combine the case-based approach to information processing with the literature on ambiguity. We model a decision maker who chooses from 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. Similar to GHTV (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 -MEU functional, as in Ghirardato, Maccheroni and Marinacci (2004) and Chateauneuf, Eichberger and Grant (2007):

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

A pair $(a; D)$, 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 over outcomes $H_a(D)$ are set-valued, thus capturing the fact that information might be ambiguous. They depend on the action a and on the characteristics of the data D . The decision maker's attitude towards ambiguity is described by a degree of optimism α , the weight assigned to the maximal expected utility, and the degree of pessimism $(1 - \alpha)$, the weight assigned to the minimal expected utility.

Our first contribution is to deduce the sets of probability distributions over outcomes, $H_a(D)$, associated with a specific action a in a given information context D . Thus, we provide a behav-

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

ioral foundation for the work of BGSS (2005) and Eichberger and Guerdjikova (2010). The set $H_a(D)$ combines objective criteria, such as the frequency and the number of observations in the data, with subjective features of the decision maker, such as the perceived degree of ambiguity and the similarity of observations. The representation obtained generalizes the idea of beliefs as similarity-weighted frequencies in BGSS (2005) by allowing for ambiguity.

Our second contribution is to identify the degree of ambiguity associated with a particular data set and behaviorally separate it from ambiguity attitude captured by α . Perceived ambiguity can be separated into two parts: vanishing ambiguity due to a limited number of observations and persistent ambiguity due to the heterogeneity of cases in the data set. Since cases contain 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. We thus capture the well-known insight of identification theory: if relevant characteristics are unobservable, the model might not be identifiable, Manski (2000). The distinction between vanishing and persistent ambiguity corresponds to a similar distinction in Epstein and Schneider (2007). Our model also extends the approach of Coignard and Jaffray (1994) and Gonzales and Jaffray (1998) to incorporate data heterogeneity.

Our third contribution is to demonstrate that the representation obtained in this paper can serve to model preferences for additional information. In particular, we show that the degrees of optimism and pessimism determine the decision maker's preferences for more precise information: the more pessimistic the decision maker, 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 which illustrate the scope of our approach. Section 3 states the axioms. Section 4 contains the main representation theorem, relates it to other representations, and considers important special cases. The notion of preferences for more precise information is discussed in Section 5. 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's information is given in form of data. A *data set* consists of *cases* c which are tuples 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 cases is $C = A \times R$. An *information context* is identified with a data set D . A data set of length T ,

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

is a vector of T cases. The set of all *data sets of length* $T \in \mathbb{N}$, is denoted by $\mathbb{D}^T := C^T$. $\mathbb{D} := \bigcup_{T \geq 1} \mathbb{D}^T$ denotes the set of data sets of finite length. The empty data set denoted D_\emptyset does not contain any cases and captures a situation in which no information is available. $\mathbb{D}^* :=: \mathbb{D} \cup \{D_\emptyset\}$ is 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 actions 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 action⁵.

A *decision situation* is completely described by the triple $(A; R; \mathbb{D}^*)$, i.e., a decision problem 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 rank pairs of actions and data sets, expressing 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. GHTV (2007) model preferences over pairs of Savage acts and sets of probability distributions over states. The interpretation of the two preference relations is similar. As GHTV (2007, p.29) note, it can be best understood using Ellsberg's two urn example. In Urn 1, there are 50 black and 50 white balls. The composition of urn 2 is unknown. The decision maker chooses the color and the urn to bet on. We can describe Urn II by the empty data set and Urn 1 by a very large data set, in which the frequency of success is $\frac{1}{2}$ for both bets. While the bets are the same, the information context of the bet varies across the two urns, rendering natural the idea of preferences on action-data-set pairs.

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.

Describing urns by data sets suggests a generalization of the Ellsberg experiment. Given two urns characterized by equal frequencies, but different number of observations, would the decision maker prefer the longer data set, regardless of the bet? By incorporating comparisons across information contexts, we can model such preferences for information precision.

More generally, information contexts can describe distinct economic environments (markets, locations, see Section 2.3 for examples), which offer the same set of choices, but differ with respect to available information. Since data provide information about the process governing the outcome realizations of actions, they affect the decision maker's beliefs about the outcomes of a given action in a given information context. The preference $(a; D) \succsim (a; D')$ means that the decision maker feels more confident choosing a based on the information in D than in D' . This might reflect a higher frequency of good outcomes, better information precision, or higher relevance of observations in the environment described by D , or some combination of all of these. \succsim 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 sets⁶.

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 (black), a_w (a_b) pays 1 if the ball drawn is white (black) and 0, otherwise. Hence, the set of actions is $A = \{a_w; a_b\}$ and the set of outcomes is $R = \{0; 1\}$. An information context specifies the available information about urn i , $i \in \{1; 2\}$. It is given by a data set $D_i = ((a_1^i; r_1^i); \dots (a_{T^i}^i; r_{T^i}^i))$ and contains records of bets and their outcomes based on drawings from urn i .

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

$$D_1 = ((a_w; 1); (a_w; 1); (a_b; 0); (a_w; 0); (a_w; 1); (a_w; 0); (a_w; 0); (a_b; 1); (a_b; 1); (a_w; 1)).$$

⁶ An alternative approach is to use preferences over Savage acts conditional on a data set. The corresponding state-space is the set of all sequences of outcome realizations. The set of acts contains all mappings from infinite sequences of outcome realizations into outcomes and is quite different from the original set of actions A . 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 actions and data sets.

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

		R		
			1	
A	a_b	2	1	(1)
	a_w	4	3	

The data set for urn 2, D_2 , contains 300 observations, summarized in the following table:

		R		
			1	
A	a_b	80	90	(2)
	a_w	60	70	

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 for 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. E.g., 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, $D_i = ((a_1^i; r_1^i) \dots (a_{T^i}^i; r_{T^i}^i))$.

Agents' preferences \succsim are defined on the set $A \times \mathbb{D}^*$ expressing the idea that each of them has to choose a project and a market, in which to invest. E.g., a preference of an entrepreneur $(a_1; D_1) \succ (a_2; D_2)$ means that he prefers to invest in project a_1 in market 1 described by information context D_1 to investing in a_2 in market 2 described by 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 out of several companies located either in his 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 of which are active in the same industry (e.g., internet retail). a_1^H and a_1^F are relatively large and listed in the stock exchange of their respective coun-

tries, while a_2^H is a small company traded only across the counter in the home country. The set of possible returns is R . The investor has access to past returns summarized in a data set D :

$$D = \left((a_1^H; r_1) \dots (a_1^H; r_{T_1^H}) ; (a_2^H; \tilde{r}_1) \dots (a_2^H; \tilde{r}_{T_2^H}) ; \left((a_1^F; \hat{r}_{T_2^H}) \dots (a_1^F; \hat{r}_{T_1^F}) \right) \right) \quad (3)$$

Given the information in D , the investor can compare an investment in a listed company a_1^H in his home market to an investment in the foreign company a_1^F , expressing preferences of the type $(a_1^H; D) \succsim (a_1^F; D)$. Note that D 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, observations of its returns might be relevant for the evaluation of a_1^H and a_1^F . In particular, the information about a_2^H might be considered 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 phenomenon⁷.

Example 4 Medical treatment

A doctor has to choose a treatment for a patient with a particular disease. The possible treatment options, $A = \{a_1; a_2; a_3\}$, are a_1 – administering a new drug, a_2 – using the traditional treatment, and a_3 – 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 $(a; r)$ of observed treatments and their outcomes. The following table summarizes a particular data set D by listing the number of occurrences for each case:

		R		
		r_1	r_2	r_3
A	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 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. Hence, the decision to collect additional observations will depend on the perceived benefits from potential data sets generated by such a study. Such preferences for information are thus hypothetical, but experimentally testable.

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.

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 , a_c is the action observed in case c . For a data set $D = (c_1 \dots c_T)$, and $k \in \mathbb{N}$, D^k denotes the k -fold of D :

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

c^k is the data set with k observations of case c . The frequency of $D \in \mathbb{D}^T$ 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_\emptyset 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 all data sets of length T is given by:

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

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 be a frequency of a data set.

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

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

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

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. Axioms 3, 4, 5 and 10 all imply some sort of separability of preferences. 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 eval-

uation 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 sets⁸. 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 $\pi : \{1 \dots T\} \rightarrow \{1 \dots T\}$. Then, for any action $a \in A$ and any data set $D = ((c_t)_{t=1}^T) \in \mathbb{D}^T$,

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

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 $(a; (f; T)) \succsim_{(\sim)} (a; (f'; T))$, then $(a; (f; T')) \succsim_{(\sim)} (a; (f'; T'))$, and for any $\mu \in (0; 1)$ such that $\mu f + (1 - \mu) f' \in F^{T'}$

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

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 data provide for the choice of a , while the length determines the preci-

⁸ Similarly, in the Savage framework, the decision maker has to express preferences over acts he might never be able to afford, or acts whose consequences contradict the state, Savage (1954). To avoid such absurd examples, the sets of states and consequences, respectively, the set of cases have to be carefully chosen.

sion of the information. Keeping the precision constant across two data sets, 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' . 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 $\mu f'_1 + (1 - \mu) f'_2 \in F_{a'}^{T'}$

$$\begin{aligned} (a; (f_1; T)) &\succ_{(\prec)} (a; (f'_1; T')) & (4) \\ (a; (f_2; T)) &\tilde{\succ}_{(\tilde{\prec})} (a; (f'_2; T')) \end{aligned}$$

implies:

$$(a; (\mu f_1 + (1 - \mu) f_2; T)) \succ_{(\prec)} (a; (\mu f'_1 + (1 - \mu) f'_2; T')) \quad (5)$$

and if $(a; (f_2; T)) \sim (a; (f'_2; T'))$, then the two statements (4) and (5) 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 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 f'_1 and f'_2 . This is justified since $(f_1; T)$ and $(f_2; T)$ have the same length and contain only observations of the same action a . If cases containing the observation of the same action are considered equally relevant, regardless of the outcomes observed, in the evaluation of $(\mu f_1 + (1 - \mu) f_2; T)$, the weight on the evidence from $(f_1; T)$ should be μ and that of $(f_2; T)$ should be $(1 - \mu)$. The same argument applies to $(f'_1; T')$ and $(f'_2; T')$. This axiom seems 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 $a, a' \in A, (a; D_\emptyset) \sim (a'; D_\emptyset)$ 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 that preferences on actions depend only on the available observations. If two actions have performed identically for the same number of periods, their evaluation is the same. In absence of information, $D = D_\emptyset$, the decision maker should be indifferent among all actions.

Hence, the evaluation of payoffs can be separated from the 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 $a \in A$, $(a; (a; \bar{r})) \succsim (a; D_\emptyset) \succsim (a; (a; \underline{r}))$ and for all $c \in C$,

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

Furthermore, for all $a, a' \in A$, there are r' and $r'' \in R$ such that

$$(a; (a; \bar{r})) \succsim (a; (a'; r')) \succ (a; (a'; r'')) \succsim (a; (a; \underline{r}))$$

and at least one of the two weak inequalities is strict.

This axiom postulates that there is a best outcome \bar{r} such that the observation of the case $(a; \bar{r})$, provides the most preferred evidence in favor of choosing a compared to any other case in C or to obtaining no information. The least preferred outcome \underline{r} is defined symmetrically. 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. The second part of the axiom is a richness condition, which is not necessary for the representation, but guarantees that the utility function over outcomes is non-constant and the similarity function is unique.

Remark 3.1 *Axioms 3 and 6 imply that for all $a \in A$, and all $T \in \mathbb{N}$, $(a; (a; \bar{r})^T) \succ (a; (a; \underline{r})^T)$ and for all $D \in \mathbb{D}^T$, $(a; (a; \bar{r})^T) \succsim (a; D) \succsim (a; (a; \underline{r})^T)$.*

Axiom 7 Continuity

For all $a, a' \in A$, $D, D' \in \mathbb{D}^*$, if $(a; D) \succ (a'; D')$, there is a $\xi > 0$ and for each $k \in \mathbb{N}$, there exist a $\bar{k} > k$ and $\mu_1^{\bar{k}}, \mu_2^{\bar{k}} \in \left\{0; \frac{1}{\bar{k}!}; \dots; \frac{\bar{k}!-1}{\bar{k}!}; 1\right\}$ with $\left|\mu_1^{\bar{k}} - \mu_2^{\bar{k}}\right| \geq \xi$ such that

$$(a; D) \succsim \left(a; \left(\mu_1^{\bar{k}} \delta_{(a; \bar{r})} + \left(1 - \mu_1^{\bar{k}}\right) \delta_{(a; \underline{r}); \bar{k}!}\right)\right) \succ \left(a; \left(\mu_2^{\bar{k}} \delta_{(a; \bar{r})} + \left(1 - \mu_2^{\bar{k}}\right) \delta_{(a; \underline{r}); \bar{k}!}\right)\right) \succsim (a'; D').$$

Here, $\mu_1^{\bar{k}}$ and $\mu_2^{\bar{k}}$ identify two data sets of length $\bar{k}!$, which contain only observations of $(a; \bar{r})$ and $(a; \underline{r})$ and which can be nested between $(a; D)$ and $(a'; D')$. For a given \bar{k} , choosing $\mu_1^{\bar{k}}$ and $\mu_2^{\bar{k}}$ so that their difference is maximized amounts to choosing the best approximations to $(a; D)$ and $(a'; D')$ on the set of data sets of length $\bar{k}!$ containing only $(a; \bar{r})$ and $(a; \underline{r})$. The axiom requires that if preferences between $(a; D)$ and $(a'; D')$ are strict, the difference between their best approximations is bounded away from 0.

Axiom 8 Neutral outcome

There exists an $\hat{r} \in R$ such that for all $a \in A$ and all $k \in \mathbb{N}$, $(a; D_\emptyset) \sim (a; (a; \hat{r})^k)$.

Axiom 8 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 $(a; \hat{r})$ do not change the evaluation of a . E.g., \hat{r} could indicate the missing record of an outcome of an action, a problem often encountered in empirical analysis, see Manski (2000). This axiom 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.

Axiom 9 Decreasing ambiguity

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

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 D is preferred to choosing a absent any information, then the content of D provides positive evidence for choosing a . 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 , $(a; D^{k+1}) \succ (a; D^k)$. The reverse holds for evidence considered worse than no information at all.

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 sets⁹. 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 $(a; \underline{r})$, i.e, a data set $(\mu_D^a \delta_{(a; \bar{r})} + (1 - \mu_D^a) \delta_{(a; \underline{r})}; \hat{T})$ for some $\hat{T} \in \mathbb{N}$ such that $(a; (\mu_D^a \delta_{(a; \bar{r})} + (1 - \mu_D^a) \delta_{(a; \underline{r})}; \hat{T})) \sim (a; D)$. Since the mixture coefficient μ_D^a has to be rational-valued, such a data set may not exist for any \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 $D \in \mathbb{D}^T \cup \{D_\emptyset\}$. 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

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 to most of the case-based literature.

from below. Formally,

$$\mu_{k!}^a(D) = \left(\mu \in \left\{ 0; \frac{1}{k!} \dots \frac{k!-1}{k!}; 1 \right\} \mid \begin{array}{l} (a; (\mu\delta_{(a;\bar{r})} + (1-\mu)\delta_{(a;\underline{r})}; k!)) \succsim (a; D) \\ \text{and there is no } \mu' \in \left\{ 0; \frac{1}{k!} \dots \frac{k!-1}{k!}; 1 \right\} \text{ such that} \\ (a; (\mu\delta_{(a;\bar{r})} + (1-\mu)\delta_{(a;\underline{r})}; k!)) \succ \\ \succ (a; (\mu'\delta_{(a;\bar{r})} + (1-\mu')\delta_{(a;\underline{r})}; k!)) \succsim (a; D) \end{array} \right) \quad (6)$$

and, similarly,

$$\nu_{k!}^a(D) = \left(\nu \in \left\{ 0; \frac{1}{k!} \dots \frac{k!-1}{k!}; 1 \right\} \mid \begin{array}{l} (a; D) \succsim (a; (\nu\delta_{(a;\bar{r})} + (1-\nu)\delta_{(a;\underline{r})}; k!)) \\ \text{and there is no } \nu' \in \left\{ 0; \frac{1}{k!} \dots \frac{k!-1}{k!}; 1 \right\} \text{ such that} \\ (a; D) \succsim (a; (\nu'\delta_{(a;\bar{r})} + (1-\nu')\delta_{(a;\underline{r})}; k!)) \succ \\ \succ (a; (\nu\delta_{(a;\bar{r})} + (1-\nu)\delta_{(a;\underline{r})}; k!)) \end{array} \right). \quad (7)$$

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 $k \rightarrow \infty$.*

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 $k!$ grows, the ambiguity caused by a limited number of observations vanishes. The limit prediction about a associated with the sequence of data sets $\mu_{k!}^a(D)$ should assign 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 $\mu_{k!}^a(D)$ and $\nu_{k!}^a(D)$ converge to a common limit μ_D^a . Hence, the unambiguous equivalent of $(a; D)$ exists.*

Lemma A.1 in the Appendix shows that under Axioms 1-9, the function $V(a; D) =: \mu_D^a$ can be used to represent \succsim on $A \times \mathbb{D}^*$. Consider three data sets with equal number of observations T , and 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 satisfy $\mu_{(f;T)}^a > \mu_{(f'';T)}^a > \mu_{(f';T)}^a$. This allows us to state:

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:

$$\lambda(f; f'; f''; T) \mu_{(f; T)}^a + (1 - \lambda(f; f'; f''; T)) \mu_{(f'; T)}^a = \mu_{(f''; T)}^a.$$

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 similarity between a and a' is the weight put on the evidence from the observation of a' in the evaluation of a and is given by $s_a(a') = \frac{1 - \lambda(f; f'; f''; T)}{\lambda(f; f'; f''; T)}$.

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 $f'' = \delta_{(a; r)}$, $\lambda(\delta_{(a; \bar{r})}; \delta_{(a; \underline{r})}; \delta_{(a; r)}; T)$ represents the evaluation of outcome r relative to the best and the worst outcome. Normalizing the utility of \bar{r} to 1 and that of \underline{r} to 0, the utility of r is given by $\lambda(\delta_{(a; \bar{r})}; \delta_{(a; \underline{r})}; \delta_{(a; r)}; T)$.

Finally, for three data sets with frequencies $f = \delta_{(a; \bar{r})}$, $f' = \delta_{(a; \underline{r})}$ and $f'' = \delta_{(a'; r')}$ with $a' \neq a$, $\lambda(\delta_{(a; \bar{r})}; \delta_{(a; \underline{r})}; \delta_{(a'; r')}; T)$ represents the evaluation of action a 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. In case (ii), if $f = \delta_{(a; \bar{r})}$, $f' = \delta_{(a; \underline{r})}$ and $f'' = \delta_{(a'; r')}$, the weight $\lambda(\delta_{(a; \bar{r})}; \delta_{(a; \underline{r})}; \delta_{(a'; r')}; T)$ 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 $a \neq a'$), neither of which can be learned from the data. Hence, in this case $\lambda(\delta_{(a; \bar{r})}; \delta_{(a; \underline{r})}; \delta_{(a'; r')}; T)$ is also a subjective

characteristic of the decision maker which should not depend on the length of the data set T .

In Appendix B, we state a behavioral version of Length Independence, which does not make use of Definition 3.2 and is shown to be equivalent to Axiom 10.

Showing independence is straightforward¹⁰ for axioms A1, A3–A9 and (the behavioral version of) A10. A2 (Invariance) implies that only the frequency and the length of the data set determine preferences. It allows us to use the notation $(f; T)$ to write the rest of the axioms in a more concise form. In this sense, the rest of the axioms are not independent of A2.

4 The Representation

In this section, we derive an α -MEU 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 and perception of ambiguity. To state the main theorem, we assume that there are more than three outcomes¹¹, $|R| > 3$.

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 $u : R \rightarrow \mathbb{R}$, a prediction function $\rho : A \times C \rightarrow R$, a family of similarity functions $s_a : A \rightarrow \mathbb{R}_{++}$, $a \in A$, 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 cases and actions, $\gamma_a^c : A \times C \rightarrow [0; 1)$ such that \succsim can be represented by the function:*

$$V(a; D) = \alpha \max_{p \in H_a(D)} u \cdot p + (1 - \alpha) \min_{p \in H_a(D)} u \cdot p, \quad (8)$$

where for all $a \in A$, $H_a(D_\emptyset) = \Delta^{|R|-1}$ 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:

$$H_a(D) = \left[\gamma_T + \frac{(1-\gamma_T) \sum_{c \in C} \gamma_a^c f_D(c) s_a(a_c)}{\sum_{c' \in C} f_D(c') s_a(a_{c'})} \right] \Delta^{|R|-1} + \frac{(1-\gamma_T) \sum_{c \in C} (1-\gamma_a^c) s_a(a_c) f_D(c)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')} \left\{ \frac{\sum_{c \in C} (1-\gamma_a^c) f_D(c) s_a(a_c) \delta_{\rho_a^c}}{\sum_{c \in C} (1-\gamma_a^c) s_a(a_c) f_D(c)} \right\} \quad (9)$$

The elements of the representation satisfy the following conditions:

- (i) u is unique up to affine-linear transformations;
- (ii) ρ is unique up to indifference¹² and satisfies $\rho_a^{(a;r)} = r$ for all $a \in A$ and all $r \in R$;

¹⁰ Proofs are available from the authors upon request.

¹¹ This condition is not necessary for the representation and, hence, does not restrict the application of our model. Combined with Axiom 6, it ensures that the similarity function is unique. We can also prove the statement of the main Theorem for $|R| = 3$, and a somewhat more restrictive assumption on the preference order. Details are available from the authors upon request.

¹² Lemma A.5 and its proof provide more details.

- (iii) the functions s_a are unique up to a multiplication by a positive number;
- (iv) $\alpha \in [0; 1]$ is unique and for all $a \in A$, $V(a; D_\emptyset) = u(\hat{r}) = \alpha u(\bar{r}) + (1 - \alpha) u(\underline{r})$, 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 $\lim_{T \rightarrow \infty} \gamma_T = 0$;
- (vi) the minimal coefficients γ_a^c are unique and satisfy $\gamma_a^{(a;r)} = 0$ for all $a \in A$ and $r \in R$;
- (vii) for all $a, a' \in A$, there are r' and $r'' \in R$ such that $V(a; (a; \bar{r})) \geq V(a; (a'; r')) > V(a; (a'; r'')) \geq V(a; (a; \underline{r}))$ and at least one of the weak inequalities is strict.

4.1 Interpretation of the Representation

The α -MEU representation in (8) says that when evaluating the choice of a for a given data set D , the decision maker considers a set of probability distributions over outcomes $H_a(D)$. 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 $H_a(D)$. These weights reflect the evaluation of an action in absence of information, i.e., given the empty data set D_\emptyset .

The α -MEU approaches in the literature¹³ differ from ours in two respects. First, they use the Savage framework and, thus, derive a single set of priors over the states of the world. The case-based framework does not specify states of the world, nor are these observable in the data. Hence, uncertainty cannot be dissociated from the actions. $H_a(D)$ is thus a *set of probability distributions over payoffs for a given action a* . It is comparable to a marginal probability distribution obtained in the Savage framework by conditioning the prior on the choice of a given action. Without additional assumptions specifying a state space, there is no sense in which such beliefs should be consistent across actions. While a general analysis falls outside the scope of this paper, in Section 4.5 below, we use the example of a randomized statistical experiment to show how a notion of states can be introduced in this framework.

The second important distinction between the standard models of ambiguity and our approach lies in the fact that 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 set¹⁴. In particular, we take the stand that if the number of observations is large, frequencies

¹³ There does not seem to exist an axiomatisation 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).

¹⁴ In this sense, our representation resembles Carnap's (1980) inductive approach to probabilities.

alone determine beliefs. The frequencies of observations in the data can then be taken as a benchmark of unambiguous beliefs¹⁵. Deviations from this benchmark are interpreted to mean that the agent perceives ambiguity. In absence of any information, this ambiguity is maximal and the derived representation coincides with the Hurwicz criterion, Hurwicz (1951).

It is worth having a closer look at how beliefs depend on data. We restate (9) giving interpretation to its various elements:

$$H_a(D) = \underbrace{\left[\underbrace{\underbrace{\gamma_T}_{\text{vanishing ambiguity}} + \frac{(1 - \gamma_T) \sum_{c \in C} \gamma_a^c f_D(c) s_a(a_c)}{\sum_{c' \in C} f_D(c') s_a(a_{c'})}}_{\text{perceived ambiguity}} \right]}_{\text{perceived ambiguity}} \Delta^{|R|-1} + \underbrace{\frac{(1 - \gamma_T) \sum_{c \in C} (1 - \gamma_a^c) s_a(a_c) f_D(c)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')}}_{\text{perceived confidence}} \underbrace{\left\{ \frac{\sum_{c \in C} (1 - \gamma_a^c) f_D(c) s_a(a_c) \delta_{\rho_a^c}}{\sum_{c \in C} (1 - \gamma_a^c) s_a(a_c) f_D(c)} \right\}}_{\text{prediction}}$$

Observe first that $H_a(D)$ is a convex combination of the simplex, $\Delta^{|R|-1}$ with a probability distribution referred to as *prediction*, which is computed based on the frequency of D , f_D . It aggregates the observations in the data set by assigning to each case c an *unambiguous prediction* about the outcome of a given the observation of this case, $\rho_a^c \in R$. The probability distribution concentrated on such an unambiguous 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.

ρ_a^c is interpreted as the decision maker's unambiguous prediction about the outcome of a based on a very large data set containing only observations of case c . If c contains the action under consideration, i.e., $c = (a; r)$, the agent should be persuaded that the payoff of a is r upon observing $D = (c^T)$ for large values of T . His confidence in this prediction should be 1. Hence, $\rho_a^{(a;r)} = r$ and $\gamma_a^{(a;r)} = 0$, reflecting the objective character of the data. In contrast, when c contains an action different from a , the lack of objective information about the correlation between the actions means that the unambiguous prediction ρ_a^c is subjective and the confidence in it may be less than 1¹⁶, i.e., $\gamma_a^c > 0$. By choosing γ_a^c to be minimal for each $a \in A$ and $c \in C$, we implicitly assume that the decision maker adopts the prediction he is most confident in.

The weight assigned to the simplex captures the *subjectively perceived ambiguity* about the outcome of a given data set D . It is composed of *vanishing ambiguity* due to a limited number of

¹⁴ For a family of conditional preferences over acts, Wakker (2002) provides an axiomatic foundation for Carnap's formula. We are grateful to Peter Wakker for referring us to this strand of literature.

¹⁵ This is similar, e.g., to GHTV (2007), who choose the Steiner point of the set of priors as a benchmark for ambiguity-neutrality.

¹⁶ Eichberger and Guerdjikova (2010) provide an example.

observations in the data, γ_T and of *persistent ambiguity* due to unobservable correlation, i.e., 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 $H_a(D_\emptyset)$ 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 unobservables is relatively small. As T increases, the ambiguity due to limited number of observations goes to 0 and the entire ambiguity can be attributed to data heterogeneity.

4.2 Special cases

The representation in Theorem 4.1 is very general. It combines elements from the case-based decision theory and the α -max-min approach. Hence, we expect special cases to be of particular interest for applications in economics and other fields. Without claim to completeness, we discuss some cases which correspond to well-known representations.

1. *Objective probabilities as limit frequencies:* Assume that D contains only observations of a given action a , $f_D \in F_a$ and the number of observations T is large. According to Theorem 4.1, $\rho_a^{(a;r)} = r$ and $\gamma_a^{(a;r)} = 0$. As T grows, $\lim_{T \rightarrow \infty} \gamma_T = 0$ and ambiguity completely vanishes. We obtain:

$$\lim_{T \rightarrow \infty} V(a; (f_D; T)) = \sum_{r \in R} u(r) f_D(a; r)$$

Hence, for statistical experiments with a large number of observations, beliefs coincide with the limit frequencies of outcomes. Preferences are represented by EU with respect to these limit frequencies.

2. *Similarity-weighted frequencies:* Assume that the decision maker perceives no ambiguity about unobservables, i.e. $\gamma_a^c = 0$ for all $a \in A$, $c \in C$. As the number of observations in a data set D with frequency f_D grows, we obtain

$$\lim_{T \rightarrow \infty} V(a; (f_D; T)) = \sum_{r \in R} u(r) \sum_{c \in C} \frac{f_D(c) s_a(a_c) \delta_{\rho_a^c}(r)}{\sum_{c' \in C} f_D(c') s_a(a_{c'})}.$$

Since D contains heterogeneous cases ($a_c \neq a_{c'}$ for some $c, c' \in D$), limit beliefs have to take into account the similarity between different observations¹⁷. The representation then corresponds to EU with respect to similarity-weighted frequencies. For an arbitrary T , beliefs are centered around the similarity-weighted frequencies and given by:

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

This is an extension of BGSS (2005) to the case in which the decision maker perceives ambiguity due to the limited number of observations in the data.

¹⁷ BGSS (2005) discuss why decisions may have to be based on heterogeneous data sets.

3. *Hurwicz criterion*: If the decision maker has no information, i.e., $D = D_\emptyset$, we obtain:

$$V(a; D_\emptyset) = \alpha \max_{p \in \Delta^{|R|-1}} u \cdot p + (1 - \alpha) \min_{p \in \Delta^{|R|-1}} u \cdot p.$$

In this case, the representation coincides with the Hurwicz criterion, Hurwicz (1951), applicable to the case of complete ignorance about the true probability distribution.

4. *NEO-additive capacity*: Assume that all observations are equally relevant, $s_a(a') = 1$ for all $a, a' \in A$ and that there is no ambiguity about unobservables, $\gamma_a^c = 0$ for all $a \in A, c \in C$. We can then define a state space, which coincides with the set of cases, $S = C$. The outcome of action a in state $s_c = c$ is given by the unambiguous prediction ρ_a^c . For a data set $D = (f_D; T)$, the representation can be written as:

$$V(a; D) = \alpha \max_{p \in \Pi(D)} \sum_{c \in C} u(\rho_a^c) p(s_c) + (1 - \alpha) \min_{p \in \Pi(D)} \sum_{c \in C} u(\rho_a^c) p(s_c),$$

$$\Pi(D) := \gamma_T \Delta^{|C|-1} + (1 - \gamma_T) \{f_D\}.$$

This representation is equivalent to Choquet expected utility with respect to a NEO-additive capacity ν^D on $S = C$, $\nu^D(E) := \alpha \gamma_T + (1 - \gamma_T) \sum_{\{c | s_c \in E\}} f_D(c)$ for $\emptyset \neq E \subsetneq S$, $\nu(S) =$

1, $\nu(\emptyset) = 0$, with a frequency-based additive belief on S given by $f_D(s_c) = f_D(c)$, see Chateauneuf, Eichberger and Grant (2007).

We conclude by stating an alternative way of writing the representation in (8), which will be useful in the applied examples in Section 4.4.

Remark 4.1 Since R is finite, for $D \in \mathbb{D}$, the representation in (8) can be rewritten as¹⁸

$$V(a; D) = \left(\gamma_T + \frac{(1 - \gamma_T) \sum_{c \in C} \gamma_a^c f_D(c) s_a(a_c)}{\sum_{c' \in C} f_D(c') s_a(a_{c'})} \right) [\alpha u(\bar{r}) + (1 - \alpha) u(\underline{r})] + \frac{(1 - \gamma_T) \sum_{c \in C} (1 - \gamma_a^c) f_D(c) s_a(a_c) u(\rho_a^c)}{\sum_{c' \in C} f_D(c') s_a(a_{c'})} \quad (10)$$

4.3 Comparative Statics and Parameter Independence

Our representation has multiple parameters. In this section, we show that these parameters are independent, i.e., that different parameters affect the evaluations of distinct action-data-set-pairs. W.l.o.g., we normalize $u(\bar{r}) = 1$ and $u(\underline{r}) = 0$. For concreteness, we fix two actions $a, a' \in A, T \in \mathbb{N}$ and $r \in R$ and consider the following data sets and their evaluations:

¹⁸ See, e.g., Chateauneuf, Eichberger, and Grant (2007, p.543, Remark 3.2).

$D_1 = D_\emptyset$	$V(a; D_1) = \alpha$
$D_2 = (a; \bar{r})^T$	$V(a; D_2) = \alpha\gamma_T + (1 - \gamma_T)$
$D_3 = (a; \underline{r})^T$	$V(a; D_3) = \alpha\gamma_T$
$D_4 = (a'; r)^T$	$V(a; D_4) = \alpha\gamma_T + (1 - \gamma_T) \left[\alpha\gamma_a^{(a'; r)} + \left(1 - \gamma_a^{(a'; r)}\right) u\left(\rho_a^{(a'; r)}\right) \right]$
$D_5 = \left((a; \hat{r})^{T-k}; (a'; r)^k \right)$	$V(a; D_5) = \alpha\gamma_T + (1 - \gamma_T) \left[\frac{(T-k)\alpha + s_a(a')k}{T-k+s_a(a')k} \left[\alpha\gamma_a^{(a'; r)} + \left(1 - \gamma_a^{(a'; r)}\right) u\left(\rho_a^{(a'; r)}\right) \right] \right]$

We summarize the comparative statics in the following table:

	$V(a; D_1)$	$V(a; D_2)$	$V(a; D_3)$	$V(a; D_4)$	$V(a; D_5)$
α	+	+	+	+	+
γ_T	0	0 if $\alpha=1$ - if $\alpha<1$	+ if $\alpha>0$ 0 if $\alpha=0$	+ if $u\left(\rho_a^{(a'; r)}\right) < \alpha$ - if $u\left(\rho_a^{(a'; r)}\right) > \alpha$	+ if $u\left(\rho_a^{(a'; r)}\right) < \alpha$ - if $u\left(\rho_a^{(a'; r)}\right) > \alpha$
$\gamma_a^{(a'; r)}$	0	0	0	+ if $u\left(\rho_a^{(a'; r)}\right) < \alpha$ - if $u\left(\rho_a^{(a'; r)}\right) > \alpha$	+ if $u\left(\rho_a^{(a'; r)}\right) < \alpha$ - if $u\left(\rho_a^{(a'; r)}\right) > \alpha$
$s_a(a')$	0	0	0	0	+ if $u\left(\rho_a^{(a'; r)}\right) > \alpha$ - if $u\left(\rho_a^{(a'; r)}\right) < \alpha$

Each of the parameters α , γ_T , $s_a(a')$ and $\gamma_a^{(a'; r)}$ affects the evaluations of a different subset of $\{(a; D_n)_{n=1}^5\}$. Hence, we can distinguish between changes caused by these parameters by observing which of the values $V(a; D_n)$ have changed as a result of an exogenous shock.

4.4 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 agent about his preferred color-urn combination.

An important characteristic of this example is that counterfactuals are observable, i.e., the observation of the outcome of a given bet, say $(a_w; 1)$, uniquely identifies the color of the ball drawn from the urn (white), and with this the (counterfactual) outcome of the other bet, a_b , 0. We thus make the following assumptions about the parameters of the representation:

(i) *all observations are equally relevant for the prediction to be made:*

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

(ii) *for $i, j \in \{b; w\}$, $i \neq j$, the unambiguous predictions $\rho_{a_i}^c$ satisfy:*

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

(iii) *there is no ambiguity due to heterogeneity of cases, $\gamma_a^c = 0$ for all $a \in A$ and $c \in C$.*

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 :

$$H_{a_b}(D)(1) = [(1 - \gamma_T)(f_D(a_b; 1) + f_D(a_w; 0)); (1 - \gamma_T)(f_D(a_b; 1) + f_D(a_w; 0)) + \gamma_T] \quad (11)$$

$$H_{a_w}(D)(1) = [(1 - \gamma_T)(f_D(a_w; 1) + f_D(a_b; 0)); (1 - \gamma_T)(f_D(a_w; 1) + f_D(a_b; 0)) + \gamma_T]. \quad (12)$$

Note that beliefs are centered around the empirical frequencies and converge to them as the number of observations becomes large, i.e., $\lim_{T \rightarrow \infty} \gamma_T = 0$.

Let urn 1 be characterized by data set D_1 in (1), and let urn 2 be described by D_2 in (2). For a given degree of optimism α , we obtain:

$$V(a_b; D_1) = V(a_w; D_1) = \gamma_{10} [\alpha u(1) + (1 - \alpha) u(0)] + (1 - \gamma_{10}) \frac{u(1) + u(0)}{2},$$

$$V(a_b; D_2) = V(a_w; D_2) = \gamma_{300} [\alpha u(1) + (1 - \alpha) u(0)] + (1 - \gamma_{300}) \frac{u(1) + u(0)}{2}.$$

Since the empirical frequencies of D_1 and D_2 are equal and $\gamma_{10} > \gamma_{300}$, the comparison between the bets on urn 1 and urn 2 is completely determined by α . 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, $\alpha < \frac{1}{2}$. A purely pessimistic decision maker with $\alpha = 0$, exhibits the usual Ellsberg preferences, choosing to bet on the less ambiguous urn, regardless of the bet:

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

Intuitively, a pessimist 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.

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 , where $\alpha^i = 1$ stands for a pure optimist, and $\alpha^i = 0$ for a pure pessimist.

To simplify the model, we assume that there is only one type of project, $A = \{a\}$, and two outcomes, high, \bar{r} , and low, \underline{r} . Hence, $R = \{\bar{r}; \underline{r}\}$. Data set D_1 describes the established market 1, whereas D_2 contains the information about the emerging Eastern European market. 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 in market $j \in \{1; 2\}$, $q_j \in (\underline{r}; \bar{r})$, the payoff of a lender is given by $\min\{q_j; \underline{r}\}$, and the payoff of an entrepreneur is $\max\{\bar{r} - q_j; 0\}$. Using equation (10) in Remark 4.1, the evaluation of project a in a given information context D_j , $j \in \{1; 2\}$, is:

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

for the lender, and

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

for the entrepreneur.

If the repayment is identical for both markets, $q_1 = q_2 = q$, both types of agents will prefer to invest in the established market 1, if and only if the following two conditions are satisfied:

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

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 market¹⁹.

More interesting is the case, in which the two types differ significantly in their degrees of optimism α . 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:

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

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

$$q_1^* = (1 - \gamma_{T_2}) [(1 - f) \bar{r} + f \underline{r}] + \gamma_{T_2} \underline{r}$$

and

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

¹⁹ 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.

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 were to 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 shows how our approach can be used to model market participation choices 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 identical frequencies of outcomes and equal price of credit in the two markets, a Bayesian expected utility maximizer prefers to trade in the market with more observations, D_1 , if and only if the realized frequency f of the better outcome exceeds her prior. In contrast, a purely pessimistic decision maker always trades in the more informative market, regardless of f .

Our next example will focus on the impact of similarity.

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 (3). 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: $T_1^H = T_1^F =: T_1$, $T_2^H =: T_2$, $f_D(a_1^H; r) = f_D(a_1^F; r) =: f_D(a_1; r)$ and $f_D(a_2^H; r) =: f_D(a_2; r)$ 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 $a' \in A$, $\tilde{r} \in R$. Hence, by property (vi) of Theorem 4.1, we obtain $\gamma_{a_1^H}^{a_1^H} = \gamma_{a_1^F}^{a_1^F} = 0$ and $\gamma^{a_2^H} \in [0; 1)$.

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:

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

It is easy to show that under the assumptions made above, the comparison between a_1^H and a_1^F

depends on the sign of the expression:

$$V(a_1^H; D) - V(a_1^F; D) = \frac{2T_1 T_2 (s_H - s_F)}{[2T_1 + s_H T_2][2T_1 + s_F T_2]} \left[\left(1 - \gamma^{a_2^H}\right) u \cdot \frac{\sum_{r \in R} f_D(a_2; r) \delta_{\rho(a_2^H; r)}}{\sum_{r' \in R} f_D(a_2; r')} - u \cdot \frac{\sum_{r \in R} f_D(a_1; r) \delta_r}{\sum_{r' \in R} f_D(a_1; r')} \right]$$

Since by assumption, $s_H > s_F$, a_1^H is preferred to a_1^F , 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 $(1 - \gamma^{a_2^H})$ exceeds the expected utility based on the prediction from the directly relevant data about a_1^H and a_1^F . Hence, favorable information about a non-listed asset in the home country may induce a strict preference for the listed company in the home country if this information is not too ambiguous. For $s_H = s_F$, the two assets 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 showing that it can help explain the home bias in portfolio choice without appealing to differences in perceived ambiguity, which in turn may further re-enforce the effect.

4.5 Relating the Representation to the State-Based Model of Ambiguity

Most of the definitions of ambiguity in the literature are related to the violation of the Savage P2 axiom for non-comonotonic acts, and, thus, rely on the notion of a state space. In the case-based model, we define perceived ambiguity as deviation of beliefs from observed frequencies. In this section, the connection between the two approaches is illustrated for statistical experiments, for which the definition of a state space is implied by the description of the problem.

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

The state-space in this example is given by the colors of the ball in the urn: $S = \{B; W\}$. For a given $D \in \mathbb{D}^T$, define π_D as the probability distribution which assigns to state $s \in \{B; W\}$ the frequency with which this state has been observed in the data:

$$\pi_D = (\pi_D(W); \pi_D(B)) = (f_D(a_b; 0) + f_D(a_w; 1); f_D(a_b; 1) + f_D(a_w; 0)).$$

Define $\Pi(D)$ as

$$\Pi(D) = (1 - \gamma_T) \{\pi_D\} + \gamma_T \Delta^1$$

and observe that for $a \in A$, $H_a(D)$ defined by (11) and (12) gives the corresponding set of marginal distributions over outcomes for action $a \in A$. Hence, a case-based agent satisfying A1-A10 behaves identically to an α -MEU maximizer with subjective set of priors $\Pi(D)$. Since $\Pi(D)$ is centered around the empirical frequency of states, π_D , γ_T can be interpreted as the perceived ambiguity due to an insufficient number of observations.

To see how the coefficient α is related to attitude towards ambiguity, consider a data set $\bar{D} \in \mathbb{D}^T$, in which both states have occurred with equal frequency, $\pi_{\bar{D}}(W) = \pi_{\bar{D}}(B) = \frac{1}{2}$. Normalizing w.l.o.g. $u(1) = 1$ and $u(0) = 0$, we obtain:

$$V(a_b; \bar{D}) = V(a_w; \bar{D}) = \gamma_T \alpha + \frac{1}{2} (1 - \gamma_T).$$

If $\alpha = \frac{1}{2}$, these evaluations are consistent with EU with additive beliefs $\pi(W) = \pi(B) = \frac{1}{2}$.

Such a decision maker perceives ambiguity described by γ_T , but is neutral towards it. For $\alpha \neq \frac{1}{2}$, no additive belief is consistent with these evaluations, but beliefs can be represented by the capacity $\nu(W) = \nu(B) = \alpha\gamma_T + \frac{1}{2}\gamma_T$. This capacity is convex for $\alpha < \frac{1}{2}$, which corresponds to ambiguity aversion in the Ellsberg experiment. It is concave for $\alpha > \frac{1}{2}$, expressing ambiguity loving.

Finally, to illustrate the impact of additional information, we replicate the data set \bar{D} k -times.

$$V(a_b; \bar{D}^k) = V(a_w; \bar{D}^k) \begin{matrix} > \\ < \end{matrix} V(a_b; \bar{D}) = V(a_w; \bar{D})$$

holds iff $\alpha \begin{matrix} < \\ > \end{matrix} \frac{1}{2}$. Thus, additional information can force the evaluations of two perfectly negatively correlated actions, a_w and a_b , to move in the same direction. The reason for this is that incoming information decreases ambiguity for both actions (Axiom 9) and this has a positive (negative) effect on their evaluation if the decision maker is ambiguity-averse (-loving). In contrast, for a Bayesian expected utility maximizer, additional information can never strictly increase (decrease) the evaluations of both actions simultaneously. Since the two actions are perfectly negatively correlated, a Bayesian, who interprets incoming information as a positive signal for one of the actions, will necessarily consider it a negative signal for the other one.

4.6 No Ambiguity Due to the Size of Data Sets

In this section, we consider a decision maker who perceives no ambiguity regarding the number of observations. Preferences are thus, independent of the length of the data set. We model such preferences using a modification of Axiom 3 and provide a behavioral foundation for BGSS (2005) similarity-weighted frequencies.

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 $(a; (f; T)) \begin{matrix} \succ \\ \sim \end{matrix} (a; (f'; T'))$, then

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

Axiom 3A is the behavioral counterpart of the Concatenation axiom in BGSS (2005). It requires a concatenation of two data sets $(f; T)$ and $(f'; T')$ with frequency $\frac{T}{T+T'}f + \frac{T'}{T+T'}f'$ and length $T + T'$ to be evaluated 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 $k \in \mathbb{N}$, $(a; (f; T)) \sim (a; (f; kT))$. Second, if $(a; (f; T)) \succ (a; (f'; T'))$ for some T and T' satisfying the conditions of the axiom, 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. 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 $u : R \rightarrow \mathbb{R}$, a prediction function $\rho : A \times C \rightarrow R$, a family of similarity functions $s_a : A \rightarrow \mathbb{R}_{++}$, $a \in A$, degrees of optimism, α , and pessimism, $(1 - \alpha)$ and minimal coefficients of perceived ambiguity depending on the cases and the actions, $\gamma_a^c : A \times C \rightarrow [0; 1)$ 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 $a \in A$, $H_a(D_\emptyset) = \Delta^{|R|-1}$ 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:

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

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 $V(a; D_\emptyset) = \alpha u(\bar{r}) + (1 - \alpha) u(\underline{r})$, 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 $r \in R$;
- (vi) for all $a, a' \in A$, there are r' and $r'' \in R$ such that $V(a; (a; \bar{r})) \geq V(a; (a'; r')) > V(a; (a'; r'')) \geq V(a; (a; \underline{r}))$ and at least one of the weak inequalities is strict.

This representation can be obtained from the representation derived in Theorem 4.1 by setting $\gamma_T = 0$ for all $T \in \mathbb{N}$. A decision maker whose preferences satisfy Axiom 3A does not perceive ambiguity related to the number of observations in the data. Hence, data sets of *different length*, but *identical frequencies* will be associated with identical probability distributions over outcomes. 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_\emptyset , as in (iv).

If, in addition, the decision maker perceives no persistent ambiguity, i.e., if $\gamma_a^c = 0$ for all $a \in A$ and all $c \in C$, he behaves like an EU-maximizer with beliefs given by similarity-weighted frequencies as in BGSS (2005):

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

This special case obtains if, in addition to Axioms 1, 2, 3A, 4-7, the following condition holds:

for any $a \in A$ and any $c \in C$, there is a $\rho_a^c \in R$ such that $(a; c) \sim (a; (a; \rho_a^c))$.

4.7 Intuition for the Proof: Identifying the Components of the Representation

In this section, we provide some intuition for the proof of Theorem 4.1. The detailed proof is contained in Appendix A.

In Lemma A.1, we show that μ_D^a represents \succsim on $A \times \mathbb{D}$. Define $V(a; D) =: \mu_D^a$. The utility function over outcomes $u : R \rightarrow \mathbb{R}$ is determined by

$$u(r) =: \lim_{T \rightarrow \infty} V(a; (a; r)^T) = \lim_{T \rightarrow \infty} \mu_{(a; r)^T}^a.$$

Intuitively, for a very large data set containing only observations of the same case $(a; r)$ the decision maker believes that a pays r with certainty. Hence, $V(a; (a; r)^T)$ converges to $u(r)$.

For any a and $a' \in A$, the similarity coefficient $s_a(a')$ can be identified from the equation:

$$\frac{V(a; (a; r)^{2T}) + s_a(a') V(a; (a'; r')^{2T})}{1 + s_a(a')} = V(a; ((a; r)^T; (a'; r')^T)).$$

Axioms 4 and 10 imply that $s_a(a')$ is independent of r, r' and T .

Axiom 8 yields the equation $u(\hat{r}) = \alpha u(\bar{r}) + (1 - \alpha) u(\underline{r})$ which uniquely determines α .

Identifying γ_T relies on the fact that ambiguity based on a limited number of observations disappears as $T \rightarrow \infty$. Hence, the difference between $V(a; (a; r)^T)$ and $u(r)$ determines γ_T ,

$$V(a; (a; r)^T) = \gamma_T u(\hat{r}) + (1 - \gamma_T) u(r).$$

Axioms 5 and 10 imply that γ_T is independent of r and on a and, by Axiom 9, γ_T is strictly decreasing. For $D \in \mathbb{D}_a^T$, Axiom 4 thus implies:

$$V(a; D) = \gamma_T [\alpha u(\bar{r}) + (1 - \alpha) u(\underline{r})] + (1 - \gamma_T) \sum_{r \in R} u(r) f_D(a; r).$$

We define the predicted outcome of a based on data consisting only of observations of the case $c, \rho_a^c \in R$, by

$$\rho_a^c =: \arg \min_{r \in R} \left| \frac{\lim_{T \rightarrow \infty} V(a; c^T) - u(r)}{u(\hat{r}) - u(r)} \right|.$$

Obviously, this ρ_a^c is unique up to indifference. The perceived ambiguity with respect to this prediction, γ_a^c , is now determined from the equation

$$\gamma_a^c u(\hat{r}) + (1 - \gamma_a^c) u(\rho_a^c) = \lim_{T \rightarrow \infty} V(a; c^T).$$

By construction, the perceived degree of ambiguity γ_a^c is minimal for each a and c , i.e., ρ_a^c is the best guess of the decision maker as to the outcome to be expected from a given the observation

of case c . Note that imposing the condition of minimal ambiguity uniquely identifies γ_a^c .

This allows us to write $V(a; D)$ as in (4.1). Noting that

$$[\alpha u(\bar{r}) + (1 - \alpha) u(\underline{r})] = \alpha \max_{p \in \Delta^{|R|-1}} u \cdot p + (1 - \alpha) \min_{p \in \Delta^{|R|-1}} u \cdot p,$$

and defining $H_a(D)$ as in (9) gives us the representation in Equation (8).

5 Optimism, Pessimism and Preferences for More Precise Information

The novel aspect of our approach is the domain of preferences: the decision maker can rank information contexts in which a given action is chosen. Such preferences can reflect different criteria for evaluation of information, such as number of observations, as in Examples 1 and 2, or type and relevance of observations, as in Example 3.

Increasing the length of the data set, while keeping frequencies unchanged increases the precision of information in an objective sense. More precise information, i.e., 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 show 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 prediction functions 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 5.1 *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 $a \in A$, $s_a^i = K_a s_a^j$ 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 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 5.1 For two preference relations \succsim_i and \succsim_j on $A \times \mathbb{D}^*$ satisfying the conditions of Lemma 5.1, we say that \succsim_i values information precision more than \succsim_j if for every $a \in A$, $D \in \mathbb{D}$, $k \in \mathbb{N}$, $(a; D^k) \succ_j (a; D)$ implies $(a; D^k) \succ_i (a; D)$.

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 5.2 Let \succsim_i and \succsim_j be preference relations on $A \times \mathbb{D}^*$ satisfying the conditions of Lemma 5.1. i values information precision more than j if and only if $\alpha_i \leq \alpha_j$.

We use Example 4 to illustrate these ideas:

Example 4 (resumed) Medical treatment

Consider again the medical doctor choosing a treatment in A given her information $D \in \mathbb{D}^T$. Suppose that there is a new study offering a data set $D' \in \mathbb{D}^{T'}$. For a data set D , let $a^*(D)$ denote the optimal action given D . Acquiring the new study results in a new data set with length $T + T'$ and frequency $(\frac{T}{T+T'} f_D + \frac{T'}{T+T'} f_{D'})$, denoted $D \circ D'$. The new study is valuable if $V(a^*(D \circ D'); D \circ D') > V(a^*(D); D)$, i.e., if it increases the value of the optimal choice.

When deciding whether to acquire new information, the doctor might have control over the type and number of additional observations. However, she cannot influence the observed outcomes. To determine whether the additional information is valuable, she has to form a prediction about the resulting data set D' , which in turn will determine her beliefs about the performance of the actions upon acquiring the information. In general, these beliefs will depend both on the frequency and on the type of cases, i.e., on the similarity of the new cases to those 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.

If 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 predictions associated with any action $a \in A$ for the two data sets coincide:

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

Hence, the optimal choices in the two information contexts are identical, $a^*(D \circ D') = a^*(D) =: a^*$.

It is straightforward to check that $V(a^*(D \circ D'); D \circ D') - V(a^*(D); D) > 0$ iff

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

Since $\gamma_T - \gamma_{T+T'} > 0$, i.e., additional information reduces perceived ambiguity, the doctor prefers obtaining the additional information in D' if the expected utility of a^* with respect to the prediction based on f_D exceeds the evaluation of a^* under complete ignorance. In particular, a pure pessimist with $\alpha = 0$ prefers an increase in the precision of information, regardless of its content. A pure optimist ($\alpha = 1$) always prefers to avoid additional information. ■

Example 4.2 illustrates minimal conditions under which additional information in form of data is of value for the doctor. In this example, it might be realistic to assume that the doctor is a pure pessimist, who associates ignorance with the worst outcome and thus, 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, since it would make her "feel worse". As in example 3, we could also consider preferences for particular *type* of information, and model preferences for including observations of treatments similar (but not identical) to the new treatment a_1 . This is a further field which can be explored using our approach.

6 Conclusion

In this paper, we analyze decisions informed by data. Introducing preferences on 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 of cases. 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 randomized statistical experiments, beliefs converge to the frequency of observed outcomes as the number of observations grows. 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. Our examples show 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.

Appendix A. Proofs

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 $\lim_{k \rightarrow \infty} (\mu_{k!}^a(D) - \nu_{k!}^a(D)) = 0$ and, in step 3, that $\mu_{k!}^a(D)$ converges. These two statements imply the result of the Lemma.

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

$(1 - \mu') \delta_{(a; \underline{x})} \in F^T$, $\mu' < \mu$ iff

$$(a; (\mu \delta_{(a; \bar{r})} + (1 - \mu) \delta_{(a; \underline{x})}; T)) \succ (a; (\mu' \delta_{(a; \bar{r})} + (1 - \mu') \delta_{(a; \underline{x})}; T)). \quad (13)$$

Proof of Step 1

Axioms 3 and 6 imply that for all $T \in \mathbb{N}$, $(a; (a; \bar{r})^T) \succ (a; (a; \underline{x})^T)$. Axiom 4 then implies that for any $\mu \in (0; 1]$,

$$(a; (\mu \delta_{(a; \bar{r})} + (1 - \mu) \delta_{(a; \underline{x})}; T)) \succ (a; (\delta_{(a; \underline{x})}; T)). \quad (14)$$

Let first $\mu' < \mu$. For $\mu' = 0$, (14) is equivalent to (13). If $\mu' \in (0; \mu)$, Axiom 4 gives:

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

which is equivalent to (13). Now assume (13) and note that by the argument above $\mu' \geq \mu$ implies

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

a contradiction. Hence, $\mu' < \mu$. ■

Step 2: For any $a \in A$, any $T \in \mathbb{N}$ and any $D \in \mathbb{D}^T \cup \{D_\emptyset\}$, $\lim_{k \rightarrow \infty} (\mu_{k!}^a(D) - \nu_{k!}^a(D)) = 0$.

Proof of Step 2

By Axioms 6 and 9, sequences $(\mu_{k!}^a(D))_{\substack{k \in \mathbb{N} \\ k \geq T}}$ and $(\nu_{k!}^a(D))_{\substack{k \in \mathbb{N} \\ k \geq T}}$ satisfying (6) and (7) exist. Step 1 allows us to rewrite (6) and (7) as:

$$\begin{aligned} \mu_{k!}^a(D) &= \min_{\mu \in \{0; \frac{1}{k!}; \dots; \frac{k!-1}{k!}; 1\}} \left\{ \mu \mid (a; (\mu \delta_{(a; \bar{r})} + (1 - \mu) \delta_{(a; \underline{x})}; k!)) \succ (a; D) \right\} \\ \nu_{k!}^a(D) &= \max_{\nu \in \{0; \frac{1}{k!}; \dots; \frac{k!-1}{k!}; 1\}} \left\{ \nu \mid (a; D) \succ (a; (\nu \delta_{(a; \bar{r})} + (1 - \nu) \delta_{(a; \underline{x})}; k!)) \right\}. \end{aligned}$$

By step 1, unless $\mu_{k!}^a(D) = \nu_{k!}^a(D)$, it must be that $\mu_{k!}^a(D) - \nu_{k!}^a(D) = \frac{1}{k!}$. Hence, $\mu_{k!}^a(D) - \nu_{k!}^a(D) \leq \frac{1}{k!}$, implying $\lim_{k \rightarrow \infty} (\mu_{k!}^a(D) - \nu_{k!}^a(D)) = 0$. ■

Step 3: For any $a \in A$, any $T \in \mathbb{N}$ and any $D \in \mathbb{D}^T \cup \{D_\emptyset\}$, $(\mu_{k!}^a(D))_{\substack{k \in \mathbb{N} \\ k \geq T}}$ converges.

Proof of Step 3

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 $(a; D) \succ (a; (a; \hat{r}))$. By (6),

$$(a; (\mu_{k!}^a(D) \delta_{(a; \bar{r})} + (1 - \mu_{k!}^a(D)) \delta_{(a; \underline{x})}; k!)) \succ (a; D) \succ (a; (a; \hat{r})).$$

By Axiom 9, this implies

$$(a; (\mu_{k!}^a(D) \delta_{(a; \bar{r})} + (1 - \mu_{k!}^a(D)) \delta_{(a; \underline{x})}; (k+1)!)) \succ (a; (\mu_{k!}^a(D) \delta_{(a; \bar{r})} + (1 - \mu_{k!}^a(D)) \delta_{(a; \underline{x})}; k!)).$$

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

Proof of Theorem 4.1:

It is straightforward to verify necessity of the axioms. Sufficiency is proved in four consecutive Lemmas.

Lemma A.1 shows that the axioms imply a utility representation

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

Here, $u : R \rightarrow \mathbb{R}$ is a utility function over outcomes. $\hat{h} : A \times \mathbb{D}^* \rightarrow \Delta^{|R|-1}$ 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 A.2, we use the result proved in Eichberger and Guerdjikova (2010) to show that $V(a; D)$ on $A \times \mathbb{D}$ can be expressed as:

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

where T is the length of D and where for $a \in A$, $s_a : A \rightarrow \mathbb{R}_{++}$ 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 A.4, we identify the coefficients α and $(\gamma_T)_{T \in \mathbb{N}}$ and show that

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

where $\hat{h}_a^c \in \lim_{T \rightarrow \infty} \hat{H}_a(c^T)$ and $\hat{h}_a^{(a;r)} = \delta_r$ for all $c \in C$, $a \in A$ and $r \in R$. In Lemma A.5, we identify the prediction function ρ and the minimal coefficients of ambiguity γ_a^c and show that

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

We then combine all four steps to show that $H_a(D)$ can be chosen so as to have the desired structure in (9) and derive the representation in (8).

Lemma A.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 $u : R \rightarrow \mathbb{R}$ is a utility function over outcomes and $\hat{h} : A \times \mathbb{D}^* \rightarrow \Delta^{|R|-1}$ 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 $\delta_{\bar{r}} \in \hat{H}(a; D_\emptyset)$ for all $a \in A$.

Proof of Lemma A.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 utility function over outcomes u . In step 4, we construct the correspondence \hat{H} .

Step 1: Define the function $V : A \times \mathbb{D}^* \rightarrow [0; 1]$ 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 A.1 *Note that by Definition 3.1, $\lim_{T \rightarrow \infty} \mu_{(a; \bar{r})^T}^a = 1$ and $\lim_{T \rightarrow \infty} \mu_{(a; r)^T}^a = 0$. Hence, the definition of $V(a; D)$ implies $\lim_{T \rightarrow \infty} V(a; (a; \bar{r})^T) = 1$ and $\lim_{T \rightarrow \infty} V(a; (a; r)^T) = 0$.*

Step 2: The function V defined in Step 1 represents \succsim .

Proof of Step 2:

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 $D \in \mathbb{D}^T \cup \{D_\emptyset\}$ and $D' \in \mathbb{D}^{T'} \cup \{D_\emptyset\}$. Let $\hat{T} = \max\{T; T'\}$. By Axiom 5, for all $k \geq \hat{T}$:

$$\left(a'; \left(\mu_{k!}^{a'}(D') \delta_{(a'; \bar{r})} + \left(1 - \mu_{k!}^{a'}(D')\right) \delta_{(a'; \underline{r}); k!}\right)\right) \sim \left(a; \left(\mu_{k!}^{a'}(D') \delta_{(a; \bar{r})} + \left(1 - \mu_{k!}^{a'}(D')\right) \delta_{(a; \underline{r}); k!}\right)\right) \quad (15)$$

Suppose (w.l.o.g.) that $(a; D) \succsim (a'; D')$. Then, (6) together with (15) implies

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

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 ,

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

Now suppose that $V(a; D) < V(a'; D')$, or $\mu_D^a < \mu_{D'}^{a'}$. If $\mu_D^a > \mu_{D'}^{a'}$, there is a k such that

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

Assuming that $(a'; D') \succsim (a; D)$ together with (15) contradicts (6) and we conclude $(a; D) \succ (a'; D')$.

Let now $\mu_D^a = \mu_{D'}^{a'}$, and suppose that $(a'; D') \succ (a; D)$. By Axiom 7, there exists an $\xi > 0$ such that for each k , there is a $\bar{k} > k$ and $\mu_1^{\bar{k}}, \mu_2^{\bar{k}} \in \left\{0; \frac{1}{k!} \dots \frac{\bar{k}! - 1}{\bar{k}!}; 1\right\}$ with $\left|\mu_1^{\bar{k}} - \mu_2^{\bar{k}}\right| \geq \xi$ such that:

$$\left(a; D\right) \succ \left(a; \left(\mu_1^{\bar{k}} \delta_{(a; \bar{r})} + \left(1 - \mu_1^{\bar{k}}\right) \delta_{(a; \underline{r}); \bar{k}!}\right)\right) \succ \left(a; \left(\mu_2^{\bar{k}} \delta_{(a; \bar{r})} + \left(1 - \mu_2^{\bar{k}}\right) \delta_{(a; \underline{r}); \bar{k}!}\right)\right) \succ \left(a'; D'\right).$$

Combining this with (13), (6) and (7), we obtain $\nu_{\bar{k}!}^a(D) \geq \mu_1^{\bar{k}} \geq \mu_2^{\bar{k}} + \xi \geq \mu_{\bar{k}!}^{a'}(D') + \xi$. Hence, for each k , there is a $\bar{k} > k$ such that $\nu_{\bar{k}!}^a(D) - \mu_{\bar{k}!}^{a'}(D') \geq \xi$, in contradiction to the assumption that

$$\lim_{k \rightarrow \infty} \nu_{k!}^a(D) = \lim_{k \rightarrow \infty} \mu_{k!}^{a'}(D') = \mu_D^a = \mu_{D'}^{a'}.$$

We conclude that $(a; D) \succsim (a'; D')$ if and only if $V(a; D) \geq V(a'; D')$. ■

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

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

$$V\left(a; (a; r)^T\right) = \lambda_r V\left(a; (a; \bar{r})^T\right) + (1 - \lambda_r) V\left(a; (a; \underline{r})^T\right). \quad (16)$$

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. Define $u : R \rightarrow \mathbb{R}$ by $u(r) =: \lambda_r$ for all $r \in R$. Obviously, $u(\bar{r}) = 1$ and $u(\underline{r}) = 0$. By Remark A.1, taking $T \rightarrow \infty$ on both sides of (16), we obtain for all $r \in R$:

$$\lim_{T \rightarrow \infty} V\left(a; (a; r)^T\right) = u(r). \quad (17)$$

Note that by Axiom 8, $V(a; D_\emptyset) = u(\hat{r})$ 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

$$u \cdot \hat{h}_a(D) = \mu_D^a, \quad (18)$$

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., $\hat{h}_a(D)$ with $\hat{h}_a(D)(\bar{r}) = \mu_D^a$, $\hat{h}_a(D)(\underline{r}) = (1 - \mu_D^a)$ satisfies (18). The set of all $\hat{h}_a(D)$ satisfying (18) is:

$$\hat{H}_a(D) =: \left\{ \hat{h} \in \Delta^{|R|-1} \mid u \cdot \hat{h} = \mu_D^a \right\} \quad (19)$$

The correspondence $\hat{H} : A \times \mathbb{D}^* \rightrightarrows \Delta^{|R|-1}$ satisfies the properties in the statement of the Lemma. ■

Lemma A.2 *The preference relation \succsim on $A \times \mathbb{D}$ can be represented by:*

$$V(a; D) = u \cdot \sum_{c \in C} \frac{s_a(a_c) f_D(c) \hat{h}_a(c^T)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')}, \quad (20)$$

for some and thus all $\hat{h}_a(c^T) \in \hat{H}_a(c^T)$, where T is the length of D , $\hat{H}_a(c^T)$ are defined as in (19) and $(s_a : A \rightarrow \mathbb{R}_{++})_{a \in A}$ is a family of similarity functions, each of which is unique up to a multiplication by a positive number.

Proof of Lemma A.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 $D \in \mathbb{D}$, $u \cdot p = u \cdot p' =: \tilde{V}(a; D)$ for all p and $p' \in P_a(D)$ and $\tilde{V}(a; D)$ 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:

$$P_a(D) = \sum_{c \in C} \frac{s_a(c) f_D(c) P_a(c^T)}{\sum_{c' \in C} s_a(c') f_D(c')}, \quad (21)$$

where T is the length of D and $s_a : A \times R \rightarrow \mathbb{R}_{++}$ 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 $\mu_D^a = \frac{\sum_{c \in C} \mu_{c^T}^a s_a(a_c) f_D(c)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')}$. This result combined with the definitions of V and the sets $\hat{H}_a(c^T)$ in (19) proves the Lemma.

Step 1: Note that by Axiom 9, for any $a \in A$, $c \in C$, the sequence $(\mu_{c^T}^a)_{T \in \mathbb{N}}$ 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 $C_a^* =: \{c \in C \mid (a; c) \not\sim (a; (a; \bar{r})) \text{ and } (a; c) \not\sim (a; (a; \underline{r}))\}$. By Axiom 6, C_a^* is

non-empty. By Axiom 6 and Remark A.1, we can choose an $\epsilon \in (0; 1)$ such that

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

Define the set P^* as $P^* =: \{p \in \Delta^{|R|-1} \mid u \cdot p = \frac{1}{2}\}$ and for each $D \in \mathbb{D}$, define

$$P_a(D) =: \epsilon P^* + (1 - \epsilon) [\mu_D^a \{\delta_{\bar{r}}\} + (1 - \mu_D^a) \{\delta_{\underline{r}}\}]. \quad (22)$$

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 $u \cdot p = (1 - \epsilon) \mu_D^a + \frac{\epsilon}{2}$ 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 $\tilde{V}(a; D) = (1 - \epsilon) V(a; D) + \frac{\epsilon}{2}$ 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 $\eta_i \in (0; 1)$, $\sum_{i=1}^n \eta_i = 1$, 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 $P_a(f; \hat{T}) = \sum_{i=1}^n \lambda_i P_a(f_i; \hat{T})$ holds for any $\hat{T} \in \mathbb{N}$, such that $(f_i)_{i=1}^n$ and $f \in F^{\hat{T}}$.
- (B4) For all $c \in C$, the sequences $(P_a(c^T))_{T \in \mathbb{N}}$ have a limit, $\lim_{T \rightarrow \infty} P_a(c^T)$.
- (B5) No three of the sets $\lim_{T \rightarrow \infty} P_a(c^T)$ 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 $(f; \hat{T}) \succ (f'; \hat{T}) \succ (f''; \hat{T})$ for all $\hat{T} \in \mathbb{N}$ such that f, f' and $f'' \in F^{\hat{T}}$. Hence, by step 2 of Lemma A.1, $\mu_{(f; T)}^a > \mu_{(f'; T)}^a > \mu_{(f''; T)}^a$ and therefore, by Axiom 10, there is a $\lambda(f; f''; f') \in (0; 1)$ independent of T such that $\mu_{(f; T)}^a = \lambda(f; f''; f') \mu_{(f'; T)}^a + (1 - \lambda(f; f''; f')) \mu_{(f''; T)}^a$. If $(f; T) \sim (f''; T)$, Axiom 3 implies $(f; \hat{T}) \sim (f'; \hat{T}) \sim (f''; \hat{T})$ for all $\hat{T} \in \mathbb{N}$. Hence, $\mu_{(f; \hat{T})}^a = \mu_{(f'; \hat{T})}^a = \mu_{(f''; \hat{T})}^a$ 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 $\mu_{(f; \hat{T})}^a = \sum_{i=1}^n \lambda_i \mu_{(f_i; \hat{T})}^a$ 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 $f \in F^{\hat{T}}$,

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

(B4): By step 1, $\lim_{T \rightarrow \infty} \mu_{c^T}^a$ exists. Hence, by the definition of $P_a(c^T)$,

$$\lim_{T \rightarrow \infty} P_a(c^T) = \epsilon P^* + (1 - \epsilon) \left\{ \lim_{T \rightarrow \infty} \mu_{c^T}^a \delta_{\bar{r}} + \left(1 - \lim_{T \rightarrow \infty} \mu_{c^T}^a \right) \delta_{\underline{r}} \right\}. \quad (23)$$

(B5): Note that since $|R| > 3$, the set P^* is a subset of a hyperplane in $\Delta^{|R|-1}$ and has a dimension 2 or higher and since $\epsilon \in (0, 1)$, so do the sets in (23) for all $c \in C$. Hence, there are no three sets of dimension 0 or 1 and (B5) is trivially satisfied. ■

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

Theorem A.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 (B4) and (B5). Then P_a satisfies (B1), (B2) and (B3) if and only if there exists a unique, up to multiplication by a positive number, function $s_a : C \rightarrow \mathbb{R}_{++}$ such that for all $T \in \mathbb{N}$ and any $D \in \mathbb{D}^T$,*

$$P_a(D) = \frac{\sum_{c \in C} s_a(c) f_D(c) P_a(c^T)}{\sum_{c' \in C} s_a(c') f_D(c')}. \quad (24)$$

Step 4: The similarity functions $s_a : C \rightarrow \mathbb{R}_{++}$ derived in Theorem A.3 do not depend on the observed outcomes and can be written as $s_a : A \rightarrow \mathbb{R}_{++}$. 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_D^a can be represented as $\mu_D^a = \frac{\sum_{c \in C} \mu_{c^T}^a s_a(a_c) f_D(c)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')}$.

Proof of Step 4:

Take any $a, a' \in A$. Note that by Axiom 6, for any $r \in R$, there is an $r' \in R$ such that $(a; (a'; r)) \not\sim (a; (a'; r'))$. Hence, $\mu_{(a'; r)}^a \neq \mu_{(a'; r')}^a$ and $P_a(a'; r)^T \neq P_a(a'; r')^T$. For some even number $T \in \mathbb{N}$, consider a data set $D \in \mathbb{D}_{a'}^T$ such that $f_D(a'; r) = f_D(a'; r') = \frac{1}{2}$. We will show that $\mu_D^a = \sum_{\tilde{r} \in \{r; r'\}} \frac{1}{2} \mu_{(a'; \tilde{r})}^a$, which combined with (22) and (24) implies:

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

or, $s_a(a'; r) = s_a(a'; r')$. Since for all $r'' \in R \setminus \{r; r'\}$, we have either $(a; (a'; r'')) \not\sim (a; (a'; r'))$ or $(a; (a'; r'')) \not\sim (a; (a'; r))$, an analogous argument implies that $s_a(a'; r'') = s_a(a'; r) = s_a(a'; r')$.

To see that $\mu_D^a = \sum_{\tilde{r} \in \{r; r'\}} \frac{1}{2} \mu_{(a'; \tilde{r})}^a$, construct for $\tilde{r} \in \{r; r'\}$ the sequences $\left(\mu_{k!}^{*a}(a'; \tilde{r}) \right)_{\substack{k \in \mathbb{N} \\ k \geq T}}$ and $\left(\nu_{k!}^{*a}(a'; \tilde{r}) \right)_{\substack{k \in \mathbb{N} \\ k \geq T}}$ as in (6) and (7), however restricting them to be in $\{0; \frac{2}{k!}; \frac{4}{k!} \dots \frac{k!-2}{k!}; 1\}$ rather than $\{0; \frac{1}{k!}; \frac{2}{k!} \dots \frac{k!-1}{k!}; 1\}$. Note that since for all $k \geq T$, $\mu_{k!}^{*a}(a'; \tilde{r})^T - \nu_{k!}^{*a}(a'; \tilde{r})^T \leq \frac{2}{k!}$, the convergence

results obtained in Lemma 3.1 apply. Since $\mu_{k!}^{*a}(a'; \tilde{r})^T - \mu_{k!}^a(a'; \tilde{r})^T \leq \frac{1}{k!}$, $\lim_{k \rightarrow \infty} \mu_{k!}^{*a}(a'; \tilde{r})^T = \lim_{k \rightarrow \infty} \mu_{k!}^a(a'; \tilde{r})^T = \mu_{(a'; \tilde{r})}^a$. Furthermore, for each $k \geq T$, $\sum_{\tilde{r} \in \{r; r'\}} \frac{1}{2} \mu_{k!}^{*a}(a'; \tilde{r})^T \in \{0; \frac{1}{k!}; \frac{2}{k!} \dots \frac{k!-1}{k!}; 1\}$.

Applying Axiom 4, we conclude that

$$\left(a; \left(\sum_{\tilde{r} \in \{r; r'\}} \frac{1}{2} \left(\mu_{k!}^{*a}(a'; \tilde{r})^T \delta_{(a; \tilde{r})} + \left(1 - \mu_{k!}^{*a}(a'; \tilde{r})^T \right) \delta_{(a; \underline{r})} \right); k! \right) \right) \succsim (a; D) \quad (25)$$

$$\succsim \left(a; \left(\sum_{\tilde{r} \in \{r; r'\}} \frac{1}{2} \left(\nu_{k!}^{*a}(a'; \tilde{r})^T \delta_{(a; \tilde{r})} + \left(1 - \nu_{k!}^{*a}(a'; \tilde{r})^T \right) \delta_{(a; \underline{r})} \right); k! \right) \right).$$

Note that $\lim_{k \rightarrow \infty} \sum_{\tilde{r} \in \{r; r'\}} \frac{1}{2} \mu_{k!}^{*a}(a'; \tilde{r})^T = \lim_{k \rightarrow \infty} \sum_{\tilde{r} \in \{r; r'\}} \frac{1}{2} \nu_{k!}^{*a}(a'; \tilde{r})^T = \sum_{\tilde{r} \in \{r; r'\}} \frac{1}{2} \mu_{(a'; \tilde{r})}^a$. Assume that $\mu_D^a > \sum_{\tilde{r} \in \{r; r'\}} \frac{1}{2} \mu_{(a'; \tilde{r})}^a$. Then there is a $k \in \mathbb{N}$ such that $\mu_{k!}^a(D) \geq \nu_{k!}^a(D) > \sum_{\tilde{r} \in \{r; r'\}} \frac{1}{2} \mu_{k!}^{*a}(a'; \tilde{r})^T$, in contradiction to (6), (7) and (25). A symmetric argument applies to the case $\mu_D^a < \sum_{\tilde{r} \in \{r; r'\}} \frac{1}{2} \mu_{(a'; \tilde{r})}^a$ and

we conclude that $\mu_D^a = \sum_{\tilde{r} \in \{r; r'\}} \frac{1}{2} \mu_{(a'; \tilde{r})}^a$.

Since s_a does not depend on r , (24) implies $P_a(D) = \sum_{c \in C} \frac{s_a(a_c) f_D(c) P_a(c^T)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')}$, which by the definition of $P_a(D)$ in (22) is equivalent to $\mu_D^a = \sum_{c \in C} \frac{s_a(a_c) f_D(c) \mu_{c^T}^a}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')}$. The uniqueness of $s_a(a_c)$ is established by Theorem A.3. Hence, V on $A \times \mathbb{D}$ satisfies $V(a; D) = u \cdot \frac{\sum_{c \in C} s_a(a_c) f_D(c) \hat{h}_a(c^T)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')}$ for each $\hat{h} \in \hat{H}_a(c^T)$. ■

Lemma A.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 $\lim_{T \rightarrow \infty} \gamma_T = 0$ and a function $\hat{h} : A \times C \rightarrow \Delta^{|R|-1}$ with $\hat{h}_a^c \in \lim_{T \rightarrow \infty} \hat{H}_a(c^T)$ 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 A.4:

Step 1: Define $\gamma_T =: \alpha_T + \beta_T$, with $\alpha_T =: \mu_{(a; \underline{r})}^a$ and $\beta_T =: 1 - \mu_{(a; \bar{r})}^a$. Then, $\gamma_T > 0$, the sequence $(\gamma_T)_{T \in \mathbb{N}}$ is strictly 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 $(a; (a; \bar{r})) \sim (a; (a; \hat{r}))$, $\beta_T = 0$ 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 $\hat{h} : A \times C \rightarrow \Delta^{|R|-1}$ such that for all $a \in A, r \in R, c \in C$ and $T \in \mathbb{N}$, $\hat{h}_a^c \in$

$\lim_{T \rightarrow \infty} \hat{H}_a(c^T), \hat{h}_a^{(a;r)} = \delta_r$ 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 $\lambda_c =: \lambda(\delta_{(a;\bar{r})}; \delta_{(a;\underline{r})}; \delta_c) \in [0; 1]$ independent of T such that $V(a; c^T) = \mu_{c^T}^a = \lambda_c \mu_{(a;\bar{r})^T}^a + (1 - \lambda_c) \mu_{(a;\underline{r})^T}^a = \alpha_T + (1 - \gamma_T) \lambda_c$. By step 1 of Lemma A.2, we can take $\lim_{T \rightarrow \infty}$ on both sides of the equation and obtain $\lambda_c = \lim_{T \rightarrow \infty} \mu_{c^T}^a$. Thus, for any $\hat{h}_a^c \in \left\{ \hat{h} \mid u \cdot \hat{h} = \lim_{T \rightarrow \infty} \mu_{c^T}^a \right\} = \lim_{T \rightarrow \infty} \hat{H}_a(c^T)$,

$$V(a; c^T) = u \cdot \left[\alpha_T \delta_{\bar{r}} + \beta_T \delta_{\underline{r}} + (1 - \gamma_T) \hat{h}_a^c \right], \quad (26)$$

and, thus $\hat{H}_a(c^T) = \left\{ \hat{h} \in \Delta^{|R|-1} \mid u \cdot \hat{h} = u \cdot \left[\alpha_T \delta_{\bar{r}} + \beta_T \delta_{\underline{r}} + (1 - \gamma_T) \hat{h}_a^c \right] \right\}$. Since by (17), for every $a \in A, r \in R, \delta_r \in \lim_{T \rightarrow \infty} \hat{H}_a((a; r)^T)$, we can set $\hat{h}_a^{(a;r)} = \delta_r$. ■

Step 3: The ratio $\frac{\alpha_T}{\alpha_T + \beta_T}$ does not depend on T and equals $u(\hat{r})$. Defining $\alpha := \frac{\alpha_T}{\alpha_T + \beta_T}$, we obtain:

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

Proof of Step 3

By Axiom 8, the sequence $\left(V(a; (a; \hat{r})^T) \right)_{T \in \mathbb{N}}$ is constant and by (26), it can be written as:

$$V(a; (a; \hat{r})^T) = \alpha_T + (1 - \alpha_T - \beta_T) u(\hat{r}) = u(\hat{r}) \mu_{(a;\bar{r})^T}^a + (1 - u(\hat{r})) \mu_{(a;\underline{r})^T}^a$$

By step 1 of Lemma A.2, we can take limits on both sides and obtain $\lim_{T \rightarrow \infty} V(a; (a; \hat{r})^T) = V(a; (a; \hat{r})^T) = u(\hat{r})$ for all $T \in \mathbb{N}$. It follows that: $\alpha_T + (1 - \alpha_T - \beta_T) u(\hat{r}) = u(\hat{r})$, or

$$\alpha := \frac{\alpha_T}{\alpha_T + \beta_T} = u(\hat{r}) \in [0; 1]. \quad \blacksquare \quad (27)$$

Step 4: $\gamma_T \in (0; 1)$ for all $T \in \mathbb{N}$. The coefficients α and $(\gamma_T)_{T \in \mathbb{N}}$ are unique.

Proof of Step 4

Using (27), we obtain for any $a \in A, r \in R, T \in \mathbb{N}$,

$$V(a; (a; r)^T) = \gamma_T \alpha + (1 - \gamma_T) u(r) = \gamma_T u(\hat{r}) + (1 - \gamma_T) u(r). \quad (28)$$

If $(a; (a; r)) \succ (a; (a; \hat{r}))$, Axiom 9 implies that for all $T \in \mathbb{N}, V(a; (a; r)^{T+1}) > V(a; (a; r)^T) > u(\hat{r})$. Since by (17) $\lim_{T \rightarrow \infty} V(a; (a; r)^T) = u(r)$, we obtain that for all $T \in \mathbb{N}, V(a; (a; r)^T) \in (u(\hat{r}); u(r))$. 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. The case in which $(a; (a; r)) \sim (a; (a; \hat{r}))$ for all $r \in R$ is excluded by Axiom 6.

The uniqueness of α follows immediately from the requirement that for all $T \in \mathbb{N}$,

$$V(a; (a; \hat{r})^T) = \gamma_T [\alpha u(\bar{r}) + (1 - \alpha) u(\underline{r})] + (1 - \gamma_T) u(\hat{r}) = u(\hat{r})$$

together with the fact that $u(\bar{r}) > u(\underline{r})$. Once α is determined, the uniqueness of the sequence γ_T follows from (28) and the fact that there is at least one $r \in R$ such that $u(r) \neq u(\hat{r})$. ■

Lemma A.5 *There exist minimal coefficients $\gamma_a^c \in [0; 1)$ and a prediction function $\rho : A \times C \rightarrow R$*

such that for every $c \in C$ and every $a \in A$,

$$u \cdot \hat{h}_a^c = u \cdot [\gamma_a^c (\alpha \delta_{\tilde{r}} + (1 - \alpha) \delta_r) + (1 - \gamma_a^c) \delta_{\rho_a^c}], \quad (29)$$

where \hat{h}_a^c and α are those identified in Lemma A.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 (29), $(a; (a; \rho_a^c)) \sim (a; (a; \tilde{\rho}_a^c))$ 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 A.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

$$\rho_a^{(a';r)} =: \left\{ \tilde{r} \in R \mid \begin{array}{l} (a; (a'; r)) \succsim (a; (a; \tilde{r})) \text{ and there is no } r' \in R \text{ such that} \\ (a; (a'; r)) \succsim (a; (a; r')) \succ (a; (a; \tilde{r})) \end{array} \right\}, \quad (30)$$

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

$$\rho_a^{(a';r)} =: \left\{ \tilde{r} \in R \mid \begin{array}{l} (a; (a; \tilde{r})) \succsim (a; (a'; r)) \text{ and there is no } r' \in R \text{ such that} \\ (a; (a; \tilde{r})) \succ (a; (a; r')) \succsim (a; (a'; r)) \end{array} \right\} \quad (31)$$

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 (30) or (31), set $\rho_a^{(a';r)}$ equal to one of these outcomes.

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

$$\gamma_a^{(a';r)} \lim_{T \rightarrow \infty} \mu_{(a'; \hat{r})}^a T + (1 - \gamma_a^{(a';r)}) \lim_{T \rightarrow \infty} \mu_{(a; \rho_a^{(a';r)})}^a T = \lim_{T \rightarrow \infty} \mu_{(a'; r)}^a T. \quad (32)$$

By the definition of $\rho_a^{(a';r)}$ and Axiom 3, such $\gamma_a^{(a';r)} \in [0; 1]$ always exists. If $(a; (a'; r)) \not\sim (a; (a; \hat{r}))$, $\gamma_a^{(a';r)}$ is unique and satisfies $\gamma_a^{(a';r)} < 1$. If $(a; (a'; r)) \sim (a; (a; \hat{r}))$, set $\gamma_a^{(a';r)} = 0$.

Since $\hat{h}_a^{(a';r)} \in \lim_{T \rightarrow \infty} H_a(a'; r)^T$, (17) and (19) imply:

$$u \cdot \hat{h}_a^{(a';r)} = \gamma_a^{(a';r)} u(\hat{r}) + (1 - \gamma_a^{(a';r)}) u(\rho_a^{(a';r)}) = u \cdot \left[\gamma_a^{(a';r)} (\alpha \delta_{\tilde{r}} + (1 - \alpha) \delta_r) + (1 - \gamma_a^{(a';r)}) \delta_{\rho_a^{(a';r)}} \right]. \quad (33)$$

To see that the so defined coefficients γ_a^c are minimal, suppose that there exists a $\tilde{\rho} \neq \rho$ such that $(a; (a; \tilde{\rho}_a^c)) \not\sim (a; (a; \rho_a^c))$ for some $a \in A$ and $c \in C$. Let $\tilde{\gamma}_a^c$ be the corresponding set of coefficients which satisfy (29). Expression (33) implies that $\tilde{\rho}$ and $\tilde{\gamma}_a^c$ have to satisfy (32) for all $a \in A$ and $c \in C$. Since $(a; (a; \tilde{\rho}_a^c)) \not\sim (a; (a; \rho_a^c))$, the definition of ρ_a^c together with (32) implies that $\tilde{\gamma}_a^c > \gamma_a^c$. 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 (33), if the coefficients γ_a^c have been chosen to be minimal, $\rho_a^{(a';r)}$ 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

$$H_a(D) = \left[\gamma_T + \frac{(1 - \gamma_T) \sum_{c \in C} \gamma_a^c f_D(c) s_a(a_c)}{\sum_{c' \in C} f_D(c') s_a(a_{c'})} \right] \Delta^{|R|-1} + \frac{(1 - \gamma_T) \sum_{c \in C} (1 - \gamma_a^c) s_a(a_c) f_D(c)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')} \left\{ \frac{\sum_{c \in C} (1 - \gamma_a^c) f_D(c) s_a(a_c) \delta_{\rho_a^c}}{\sum_{c \in C} (1 - \gamma_a^c) s_a(a_c) f_D(c)} \right\} \quad (34)$$

where $(s_a)_{a \in A}$ is the family of similarity functions derived in Lemma A.2, $(\gamma_T)_{T \in \mathbb{N}}$ is the sequence

of perceived degrees of ambiguity derived in Lemma A.4, ρ is the prediction function and γ_a^c are the coefficients of perceived ambiguity derived in Lemma A.5. Let $H_a(D_\emptyset) = \Delta^{|R|-1}$.

By Lemma A.1, V represents \succsim . By Lemmas A.2, A.4 and A.5, V on $A \times \mathbb{D}$ can be written as:

$$\begin{aligned} V(a; D) &= \tag{35} \\ &= u \cdot \sum_{c \in C} \frac{s_a(a_c) f_D(c)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')} [\gamma_T (\alpha \delta_{\bar{r}} + (1 - \alpha) \delta_{\underline{r}}) + (1 - \gamma_T) [\gamma_a^c (\alpha \delta_{\bar{r}} + (1 - \alpha) \delta_{\underline{r}}) + (1 - \gamma_a^c) \delta_{\rho_a^c}]] \\ &= \alpha u \cdot \left(\left[\gamma_T + (1 - \gamma_T) \sum_{c \in C} \gamma_a^c \frac{s_a(a_c) f_D(c)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')} \right] \delta_{\bar{r}} + (1 - \gamma_T) \frac{\sum_{c \in C} (1 - \gamma_a^c) s_a(a_c) f_D(c) \delta_{\rho_a^c}}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')} \right) \\ &\quad + (1 - \alpha) u \cdot \left(\left[\gamma_T + (1 - \gamma_T) \sum_{c \in C} \gamma_a^c \frac{s_a(a_c) f_D(c)}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')} \right] \delta_{\underline{r}} + (1 - \gamma_T) \frac{\sum_{c \in C} (1 - \gamma_a^c) s_a(a_c) f_D(c) \delta_{\rho_a^c}}{\sum_{c' \in C} s_a(a_{c'}) f_D(c')} \right), \end{aligned}$$

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

Combining (34) and (35), 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, $\delta_r \in \lim_{T \rightarrow \infty} \hat{H}_a(a; r)^T$ for all $a \in A$ and all $r \in R$. ■

Sketch of the proof of Theorem 4.2:

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 ,

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

and obtain the convergence result. In analogy to Lemma A.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 (16). This identifies $u(r)$ and consequently, $\hat{H}_a(D)$. Use the same arguments as in Lemma A.2 to identify the similarity function. In Lemma A.4, define $\alpha =: \mu_{D_\emptyset}^a$. This α is obviously unique and satisfies property (iv) of the Theorem. Note that since $\mu_{(a;r)^T}^a = \text{const}$ for all T , defining γ_T as in the proof of Lemma A.4 implies $\gamma_T = 0$ for all $T \in \mathbb{N}$. The so identified α and $\gamma_T = 0$ can be used to represent $\hat{H}_a(c^T)$ as in the statement of Lemma A.4. Finally, replicate Lemma A.5 to identify the coefficients γ_a^c and the function ρ . ■

Proof of Lemma 5.1:

Examination of the proof of Theorem 4.1 shows that the utility function satisfies $u(r) = \lambda(\delta_{(a;\bar{r})}; \delta_{(a;\underline{r})}; \delta_{(a;r)})$, the similarity function is given by $s_a(a_c) = \frac{1 - \lambda(\delta_{(a;r)}; \delta_{c; \frac{1}{2}\delta_{(a;r)} + \frac{1}{2}\delta_c})}{\lambda(\delta_{(a;r)}; \delta_{c; \frac{1}{2}\delta_{(a;r)} + \frac{1}{2}\delta_c})}$ for any $r \in R$, ρ_a^c is identified by preferences on $A \times \mathbb{D}^1$ and $\gamma_a^c = \lambda(\delta_{(a';\bar{r})}; \delta_{(a';\rho_a^c)}; \delta_c)$. It follows that if two individuals have the same preferences on the set of data sets of equal length and identical $\lambda(\cdot)$ -functions, then the elements of their

respective representations satisfy the conditions of the Lemma. ■

Proof of Proposition 5.2:

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

$$\sum_{c \in C} \frac{(1 - \gamma_a^c) f_D(c) s_a(a_c) u(\rho_a^c)}{\sum_{c' \in C} (1 - \gamma_a^{c'}) f_D(c') s_a(a_{c'})} - [\alpha^j u(\bar{r}) + (1 - \alpha^j) u(\underline{r})] > 0. \quad (36)$$

Since $\alpha^i \leq \alpha^j$, and therefore, $[\alpha^i u(\bar{r}) + (1 - \alpha^i) u(\underline{r})] \leq [\alpha^j u(\bar{r}) + (1 - \alpha^j) u(\underline{r})]$, (36) implies that $(a; D^k) \succ_i (a; D)$.

Conversely, if for all $(a; D)$, $(a; D^k) \succ_j (a; D)$ implies $(a; D^k) \succ_i (a; D)$, it follows by Axiom 9 that whenever $(a; D) \succ_j (a; (a; \hat{r}_j))$, $(a; D) \succ_i (a; (a; \hat{r}_i))$. 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. ■

Appendix B. Behavioral Formulation of Axiom 10

In this part of the Appendix, we provide an equivalent behavioral formulation of Axiom 10. We make use of the definitions of $\mu_{k!}^a(D)$ and $\nu_{k!}^a(D)$ in (6) and (7).

Axiom 10A For some $a \in A$ and $T \in \mathbb{N}$, let f, f' and $f'' \in F^T$ be such that $(a; (f; T)) \succ (a; (f''; T)) \succ (a; (f'; T))$ and

- (i) either $\eta f + (1 - \eta) f' = f''$ for some $\eta \in (0; 1)$,
- (ii) or $f = \delta_{(a; \bar{r})}$, $f' = \delta_{(a; \underline{r})}$, $f'' = \delta_c$ for some $c \in C$.

If there exists a $\lambda \in (0; 1)$ such that for all $k \in \mathbb{N}$, $k \geq T$ and all²⁰ $\bar{\lambda}, \underline{\lambda} \in [0; 1]$ such that $\bar{\lambda} \geq \lambda \geq \underline{\lambda}$,

$$(a; ([\bar{\lambda} \mu_{k!}^a(f; T) + (1 - \bar{\lambda}) \mu_{k!}^a(f'; T)] \delta_{(a; \bar{r})} + [\bar{\lambda} (1 - \mu_{k!}^a(f; T)) + (1 - \bar{\lambda}) (1 - \mu_{k!}^a(f'; T))]) \delta_{(a; \underline{r}); k!}) \quad (37)$$

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

$$(a; ([\underline{\lambda} \nu_{k!}^a(f; T) + (1 - \underline{\lambda}) \nu_{k!}^a(f'; T)] \delta_{(a; \bar{r})} + [\underline{\lambda} (1 - \nu_{k!}^a(f; T)) + (1 - \underline{\lambda}) (1 - \nu_{k!}^a(f'; T))]) \delta_{(a; \underline{r}); k!})$$

then for at least one λ satisfying (37), all $T' \in \mathbb{N}$ such that f, f' and $f'' \in F^{T'}$, all $k \in \mathbb{N}$, $k \geq T'$ and all $\bar{\lambda}, \underline{\lambda} \in [0; 1]$ such that $\bar{\lambda} \geq \lambda \geq \underline{\lambda}$,

$$(a; ([\bar{\lambda} \mu_{k!}^a(f; T') + (1 - \bar{\lambda}) \mu_{k!}^a(f'; T')] \delta_{(a; \bar{r})} + [\bar{\lambda} (1 - \mu_{k!}^a(f; T')) + (1 - \bar{\lambda}) (1 - \mu_{k!}^a(f'; T'))]) \delta_{(a; \underline{r}); k!}) \quad (38)$$

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

$$(a; ([\underline{\lambda} \nu_{k!}^a(f; T') + (1 - \underline{\lambda}) \nu_{k!}^a(f'; T')] \delta_{(a; \bar{r})} + [\underline{\lambda} (1 - \nu_{k!}^a(f; T')) + (1 - \underline{\lambda}) (1 - \nu_{k!}^a(f'; T'))]) \delta_{(a; \underline{r}); k!}).$$

²⁰ In this and the following statement, "all $\bar{\lambda}, \underline{\lambda} \in [0; 1]$ " is short for "all $\bar{\lambda}, \underline{\lambda} \in [0; 1]$ such that the frequencies of the respective data sets on the first and the third line of equations (37), (38) are in $F^{k!}$ ".

Note that differently from Axiom 10, Axiom 10A does not postulate the existence or the uniqueness of a coefficient $\lambda(f; f'; f''; T)$. Rather, it says that if for three data sets $(f; T)$, $(f'; T)$ and $(f''; T)$ satisfying the conditions of Axiom 10, there exists a coefficient λ , such that:

- (a) the linear combination with any coefficient larger than λ of the data sets of length $k!$ which best approximate $(a; (f; T))$ and $(a; (f'; T))$ from above is weakly preferred to $(a; (f''; T))$ and such that
- (b) the linear combination with any coefficient smaller than λ of the data sets of length $k!$ which best approximate $(a; (f; T))$ and $(a; (f'; T))$ from below is weakly worse than $(a; (f''; T))$,

then (a) and (b) should continue to hold if the length of the three data sets is changed to T' , for at least one such λ . It is easy to show that under Axioms 1-9, a unique coefficient λ exists for any three data sets satisfying the conditions of Axiom 10A and that it equals $\lambda(f; f'; f''; T)$ from Definition 3.2. Hence, just as Axiom 10, Axiom 10A postulates that the coefficients $\lambda(f; f'; f''; T)$ are independent of T . The interpretation of Axiom 10A is thus, exactly the same as in the discussion following Axiom 10.

Lemma B.1 *Under Axioms 1-9, Axiom 10A is equivalent to Axiom 10.*

Proof of Lemma B.1:

To show that Axiom 10A implies Axiom 10, note that for any $a \in A$, $T \in \mathbb{N}$ and f_1 and $f_2 \in F_a^T$, and any $\lambda_1, \lambda_2 \in (0; 1)$ such that $\lambda_1 + \lambda_2 = 1$ and $\lambda_1 f_1 + \lambda_2 f_2 \in F_a^T$, we have

$$\mu_{(\lambda_1 f_1 + \lambda_2 f_2; T)}^a = \lambda_1 \mu_{(f_1; T)}^a + \lambda_2 \mu_{(f_2; T)}^a. \quad (39)$$

The proof uses Axiom 4 and is identical to the one in step 4 of Lemma A.2, except that the sequences $(\mu_{k!}^{*a}(f_i; T))_{\substack{k \in \mathbb{N} \\ k \geq T}}$ and $(\nu_{k!}^{*a}(f_i; T))_{\substack{k \in \mathbb{N} \\ k \geq T}}$ for $i \in \{1; 2\}$ have to be restricted to be in $\left\{0; \frac{1}{\lambda_i k!}; \frac{2}{\lambda_i k!} \dots \frac{k! - \frac{1}{\lambda_i}}{k!}; 1\right\}$ rather than $\left\{0; \frac{2}{k!}; \frac{4}{k!} \dots \frac{k! - 2}{k!}; 1\right\}$.

Observe next that for every $\epsilon \in (0; 1)$, there is a \bar{k} such that for all $k \geq \bar{k}$,

$$\min_{\kappa \in \left\{0; \frac{1}{\mu_{k!}^a(f; T) - \mu_{k!}^a(f'; T)}; \dots; \frac{\mu_{k!}^a(f; T) - \mu_{k!}^a(f'; T) - 1}{\mu_{k!}^a(f; T) - \mu_{k!}^a(f'; T)}; 1\right\}} |\lambda - \kappa| \leq \epsilon. \quad (40)$$

This follows from $(a; (f; T)) \succ (a; (f'; T))$ and Axiom 7, which implies that the cardinality of the set $\left\{0; \frac{1}{\mu_{k!}^a(f; T) - \mu_{k!}^a(f'; T)}; \dots; \frac{\mu_{k!}^a(f; T) - \mu_{k!}^a(f'; T) - 1}{\mu_{k!}^a(f; T) - \mu_{k!}^a(f'; T)}; 1\right\}$ exceeds $\xi k!$ for some $\xi > 0$, and, hence, the minimum of (40) is smaller than $\frac{1}{\xi k!}$ for every k .

Let $\mu_{\mu_{k!}^a(f; T)}^a =: \mu_{(\mu_{k!}^a(f; T)\delta_{(a; \bar{r})} + (1 - \mu_{k!}^a(f; T))\delta_{(a; \bar{z})}; k!)}^a$ and note that the data set on the first line of (37) has an unambiguous equivalent given by $\bar{\lambda} \mu_{\mu_{k!}^a(f; T)}^a + (1 - \bar{\lambda}) \mu_{\mu_{k!}^a(f'; T)}^a$ such that for every k ,

$$\bar{\lambda} \mu_{\mu_{k!}^a(f; T)}^a + (1 - \bar{\lambda}) \mu_{\mu_{k!}^a(f'; T)}^a \geq \mu_{(f''; T)}^a$$

holds for all $\bar{\lambda} \geq \lambda + \frac{1}{\xi k!}$. Since $\lim_{k \rightarrow \infty} \mu_{\mu_{k!}^a(f; T)}^a = \mu_{(f; T)}^a$ and $\lim_{k \rightarrow \infty} \mu_{\mu_{k!}^a(f'; T)}^a = \mu_{(f'; T)}^a$, it follows that $\bar{\lambda} \mu_{(f; T)}^a + (1 - \bar{\lambda}) \mu_{(f'; T)}^a \geq \mu_{(f''; T)}^a$ for all $\bar{\lambda} \geq \lambda$. By a symmetric argument on $\nu_{k!}^a$ using (37), we obtain $\underline{\lambda} \mu_{(f; T)}^a + (1 - \underline{\lambda}) \mu_{(f'; T)}^a \leq \mu_{(f''; T)}^a$ for all $\underline{\lambda} \leq \lambda$. Hence, condition (37)

implies $\lambda\mu_{(f;T)}^a + (1 - \lambda)\mu_{(f';T)}^a = \mu_{(f'';T)}^a$, and hence, λ is unique. Condition (38) then implies $\lambda\mu_{(f;T')}^a + (1 - \lambda)\mu_{(f';T')}^a = \mu_{(f'';T')}^a$, and, hence, $\lambda(f; f'; f''; T) = \lambda(f; f'; f''; T')$ for all T' such that f, f' and $f'' \in F^{T'}$.

To show that Axiom 10 implies Axiom 10A, suppose that $\lambda\mu_{(f;T)}^a + (1 - \lambda)\mu_{(f';T)}^a = \mu_{(f'';T)}^a$. Since $(a; (f; T)) \succ (a; (f'; T))$, this implies that:

$$\bar{\lambda}\mu_{(f;T)}^a + (1 - \bar{\lambda})\mu_{(f';T)}^a \geq \mu_{(f'';T)}^a$$

for all $\bar{\lambda} \geq \lambda$. By the definition of $\mu_{k_1}^a(f; T)$, we have, $\mu_{\mu_{k_1}^a(f;T)}^a \geq \mu_{(f;T)}^a$ and $\mu_{\mu_{k_1}^a(f';T)}^a \geq \mu_{(f';T)}^a$ hence, $\bar{\lambda}\mu_{\mu_{k_1}^a(f;T)}^a + (1 - \bar{\lambda})\mu_{\mu_{k_1}^a(f';T)}^a \geq \mu_{(f'';T)}^a$ for all $\bar{\lambda} \geq \lambda$ which combined with a symmetric argument w.r.t. $\nu_{k_1}^a$ and (39) implies (37). The fact that $\lambda(f; f'; f''; T) = \lambda(f; f'; f''; T')$ then implies (38). ■

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*.
- CARNAP, R. (1980). "A Basic System of Inductive Logic" in: JEFFREY, R. (ed.), *Studies in Inductive Logic and Probability, Vol. II*, UCP, Berkeley, 7-155.
- 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). "Intrinsic 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 Priori Probabilities". *Journal of Behavioral Decision Making* 23, 48-68.
- HUME, D. (1748). *An Enquiry Concerning Human Understanding*, Oxford: Clarendon Press.
- HURWICZ, L. (1951). "Some Specification Problems and Applications to Econometric Models". *Econometrica*, 19, 343-344.
- 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). "Identification 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.
- WAKKER, P. (2002). "Decision-Principles to Justify Carnap's Updating Method and to Suggest Corrections of Probability Judgements" in: DARWICHE, A. AND FRIEDMAN, N. (eds.), *UAI '02, Proceedings of the 18th Conference in Uncertainty in Artificial Intelligence*, Morgan Kaufmann, Alberta, 544-551.