SONDERFORSCHUNGSBEREICH 504

Rationalitätskonzepte,
Entscheidungsverhalten und
ökonomische Modellierung

No. 08-32

Case-Based Expected Utility: Preferences over Actions and Data

Jürgen Eichberger*
and Ani Guerdjikova**

December 2008

Financial support from the Deutsche Forschungsgemeinschaft, SFB 504, at the University of Mannheim, is gratefully acknowledged.

*Sonderforschungsbereich 504, email: juergen.eichberger@awi.uni-heidelberg.de

**Cornell University, email: ag334@cornell.edu

Universität Mannheim
L 13,15
68131 Mannheim

Case-Based Expected Utility: Preferences over Actions and Data[1]

Jürgen Eichberger[2] and Ani Guerdjikova[3]

3 September 2008

We consider a decision-situation in which the available information is given by a data-set. The decision-maker can express preferences over data-set-action pairs. In particular, he can compare different actions given the available information contained in a data-set and also different data-sets w.r.t. to the evidence they give in support of the choice of a given action. Three characteristics of a data-set can be used to evaluate the evidence it provides with respect to the payoff of a given action: the frequency of observations, the number of observations and the relevance of observations to the action under consideration. We state axioms, which ensure that the decision maker is indifferent among data-sets with identical frequencies, but different lengths. We then provide an expected utility representation of preferences, in which the beliefs of the decision maker about the outcome of a given action can be expressed as similarity-weighted frequencies of observed cases, as in BGSS (2005). The similarity weights reflect the subjectively perceived relevance of observations for the specific action.

[1] We would like to thank to Larry Blume, David Easley, Gaby Gayer, Itzhak Gilboa, John Hey, Edi Karni, 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 for helpful suggestions and comments.
[3] University of Heidelberg, Alfred Weber Institute, Grabengasse 14, 69117 Heidelberg, Germany, e-mail: juergen.eichberger@awi.uni-heidelberg.de
[3] Corresponding author. Cornell University, Department of Economics, 462 Uris Hall, e-mail: ag334@cornell.edu.

1 Introduction

Most of modern decision theory under uncertainty has been conducted in the framework of Savage (1954), where acts, which associate states with outcomes, are primitives and decision makers can order these acts. Given his well-known axioms these preferences can be represented by a subjective expected utility functional, where the subjective probability distribution over states as well as the utility evaluation of outcomes are endogenously derived.

More recently, Gilboa and Schmeidler (2001) developed an alternative framework, case-based decision theory, in which actions and data-sets of cases are the primitive concepts. Preferences are assumed to be about actions conditional on information given in the data-set. Gilboa and Schmeidler (1997, 2001) as well as Gilboa, Schmeidler and Wakker (2002) provide axioms which allow us to evaluate an action by the frequency-weighted sum of its case-by-case evaluations.

Both approaches are behavioral, taking the observable preferences of the decision maker as a primitive concept. They differ however in the objects which the decision maker is assumed to rank. For Savage (1954) all the information necessary for the evaluation of an action is encoded in the states. In the case-based approach, information comes in the form of a set of observations (cases).

Gilboa and Schmeidler (2001) do not require the decision maker to condense the information from a data-set of cases into states of the world. This is an attractive feature of their approach. Their functional representation of preferences in the case-based context lacks however the intuitive appeal of the subjective expected utility approach of Savage (1954). In particular, it does not provide a separation of utility and beliefs.

Most recently, Billot, Gilboa, Samet and Schmeidler (2005), henceforth BGSS (2005), showed that one can derive probability distributions over outcomes as similarity-weighted frequencies of the cases. Eichberger and Guerdjikova (2007), henceforth EG (2007), extend this idea to a context of multiple-priors, hence allowing beliefs to also reflect the confidence of the decision maker in the informational content of the data. These results assume, however, a mapping from cases to probabilities over outcomes as primitive concepts and, therefore, lack a behavioral foundation.

In the light of this literature, we suggest in this paper a behavioral approach which allows us to derive a representation of the subjective expected utility type in the case-based model. We deduce the mapping from data-sets to probability distributions over outcomes, which were exogenous concepts in BGSS (2005) and EG (2007), endogenously and apply it to an expected utility evaluation of actions. The key difference between the approach suggested in this paper and Gilboa and Schmeidler (2001) concerns the domain of preferences. Gilboa and Schmeidler (2001) assume a family of preference orders over actions. In our approach the decision maker's preferences order *data-sets and actions*. We derive an expected utility representation from preferences over the product space of actions and data-sets, where expected utility is computed with respect to frequency-weighted subjective probabilities as in BGSS (2005).

Our approach resembles the model of Gajdos, Hayashi, Tallon and Vergnaud (2007), henceforth GHTV (2007), who consider a preference order defined on pairs of acts and sets of probability distributions. Hence, in their model, a decision maker can compare an act across several different situations characterized by different information about the probability of the states of the world. Similarly, we also postulate that the decision maker is able to evaluate the choice of a given action for different sets of observations.

Both approaches lead to choices which can be observed in experiments as the two-urn Ellsberg example (Ellsberg 1961) shows.

> Consider the following hypothetical experiment. Let us suppose that you confront two urns containing red and black balls from one of which a ball will be drawn at random. To "bet on Red$_I$" will mean that you choose to draw from Urn I; and that you will receive a prize (say, \$100) if you draw a red ball ("if Red$_I$ occurs") and a smaller amount (say \$0) if you draw a black ("if not-Red$_I$ occurs").
>
> You have the following information. Urn I contains 100 red and black balls, but in a ratio entirely unknown to you; there may be from 0 to 100 red balls. In Urn II, you confirm that there are exactly 50 red and 50 black balls. An observer - who, let us say, is ignorant of the state of your information about the urns - sets out to measure your subjective probabilities by interrogating you as to your preferences in the following pairs of gambles:
>
> 1. "Which do you prefer to bet on, Red$_I$ or Black$_I$; or are you indifferent?" That is, drawing a ball from Urn I, on which "event" do you prefer the \$100 stake, red or black: or do you care?
> 2. "Which would you prefer to bet on, Red$_{II}$ or Black$_{II}$?"
> 3. "Which do you prefer to bet on, Red$_I$ or Red$_{II}$?"
> 4. "Which do you prefer to bet on, Black$_I$ or Black$_{II}$?"
>
> Let us suppose that in both the first and the second case, you are indifferent (the typical response). Judging from a large number of responses, under absolutely nonexperimental conditions, your answer to the last two questions are likely to fall into one of three groups. You may still be indifferent within each pair of options. (If so, you may sit back now and watch for awhile.) But if you are in the majority, you will report that you prefer to bet on Red$_{II}$ rather than Red$_I$, and on Black$_{II}$ rather than Black$_I$. The preferences of a small minority run the other way, preferring bets on Red$_I$ to Red$_{II}$, and Black$_I$ to Black$_{II}$. (Ellsberg 1961, pp. 650-651).

Ellsberg's example shows that decision makers do rank acts within the information context of a given urn (Gambles 1 and 2) but may also have preferences over differing information regarding the two urns for a given act (Gambles 3 and 4).

Ellsberg provides differing information about the urns directly by their description. In principle, however, one can imagine the information available to the decision maker to be structured in several different ways. It is exactly the structure of the available information that distinguishes our approach from the one taken in GHTV (2007). In our model, the information arrives in form of data. This allows us to capture situations in which the decision maker has non-aggregated statistical data and tries to make predictions based on this information. In contrast, the approach chosen by GHTV (2007) could be seen as an intermediate stage in this process, at which, e.g., a classical statistician has used the data to generate a set of probabilistic predictions and conveys them to the decision maker.

We illustrate our approach by the following example. Consider a decision maker who faces a bet on a white ball being drawn from an urn with an unknown number of black and white balls. In this situation data-sets could consist of sample draws with replacements. A first criterion for the evaluation of these data-sets may be the frequency of observations. When evaluating the action "betting on white", it seems reasonable if the decision maker would prefer a data-set with ten sample draws of which eight were white over a data-set with ten draws of which only five were white.

A second characteristic which may be relevant for the evaluation of a data-set is the number of observations, which may serve as a proxy for its informativeness or accuracy. In particular, for a given frequency of observations, a longer data-set allows the decision maker to exclude more probability distributions as potential descriptions of the data-generating process. If both the frequencies and the lengths of two data-sets are distinct, the decision maker faces a trade-off: e.g., a data-set with eight out of ten white draws may be preferred to a data-set with just two draws both of which were white, thus indicating that the informativeness of the former is valued higher than the favorable frequency of outcomes in the latter.

Not all decision problems have the simple structure of an urn experiment. In general, not all observations in a data-set will be equally relevant for the evaluation of an action. Differently from the Savage framework, the case-based approach allows us to make this distinction by assigning different weights to information with different degree of relevance, thus capturing the idea that the similarity (or the perceived relevance) of cases may also matter for preferences over data-sets. For instance, a physician, who is trying to evaluate the performance of a specific treatment, would consider a data-set containing observations of similar treatments resulting in favorable outcomes to be more valuable than a set of observations of very different treatments with equally favorable outcomes.

The aim of this paper, is to analyze such preferences over action-data-set pairs and identify conditions under which the representation of beliefs suggested in BGSS (2005) obtains. In particular, we provide conditions under which the Concatenation axiom of BGSS (2005) holds and, hence, the decision maker is indifferent between data-sets with equal frequencies, but distinct length[4]. This appears reasonable if the database is relatively large for the decision problem under consideration. Indeed, BGSS (2005) note that this approach

> "... might be unreasonable when the entire database is very small. Specifically, if there is only one observation, [....] However, for large databases it may be acceptable to assign zero probability to a state that has never been observed." (BGSS (2005), p. 1129)

Hence, differently from GHTV (2007), in this paper we do not focus on the attitude of the decision maker towards information precision[5]. Instead, we concentrate on the frequencies of observations and their relevance for the action under consideration and highlight the role of these two factors for the evaluation of data-sets.

1.1 A short overview of the results

We provide axioms on the preference relation over data-sets D and actions a which characterize the following representation:

$$V(a;D) = \sum_{r \in R} v(r) \frac{\sum_{c \in C} f_D(c)\, s(a;a_c)\, \hat{p}_a^c(r)}{\sum_{c \in C} f_D(c)\, s(a;a_c)} \tag{1}$$

for all probability distributions over outcomes $\hat{p}_a^c$ from a compact and convex sets of probability distributions over outcomes $\hat{P}_a^c$. Here $v(r)$ is the utility of outcome r and $f_D(c)$ denotes the frequency of case c in the data-set D. Moreover, $s(a;a_c)$ is interpreted as the relevance, or similarity, of case c containing action a_c for the evaluation of action a, and $\hat{p}_a^c(r)$ denotes the probability of outcome r in case action a is chosen and the only evidence available is a data-set

[4] This gives rise to a model of learning which is close in spirit to the frequentist approach and very different from Bayesian learning, as axiomatized by Easley and Rustichini (1999). In particular, observing a data-set in which an action pays the highest possible outcome with frequency 1 would lead the decision maker in our model to assign a probability of 1 to this outcome, regardless of the number of observations, making his behavior inconsistent with the one of a Bayesian.

[5] Note, however, that the structure of our model allows for this generalization. We analyze this question in a companion paper, Eichberger and Guerdjikova (2008).

consisting of the single case c.

The representation in Equation (1) can be interpreted in the following way:

(i) For each piece of evidence c and for each action a, the decision maker entertains a probability distribution over outcomes $\hat{p}_a^c$, i.e., a prediction about the outcome which would result from the choice of a, given evidence c. If the information consists of a single case c, he compares actions according to their expected utility with respect to these predictions and a utility function over outcomes, $v\left(r\right)$.

(ii) For more complex data-sets, two additional characteristics of the available information will influence preferences. First, if the data-set consists only of cases containing the same action, say a', then all cases will be considered equally relevant for the prediction at hand. Hence, the decision maker will weigh the prediction $\hat{p}_a^c$ associated with each of the cases by the frequency $f_D\left(c\right)$ with which the case occurs in the data-set. In particular, if the data-set consists only of cases containing action a, it would represent a controlled statistical experiment with respect to the evaluation of a. In this case, the frequency of outcomes in the data-set could serve as a first estimate of the probabilities with which the outcomes occur. In contrast, when the observed action is distinct from a, the probability distribution $\hat{p}_a^{(a';r)}$ need not predict that outcome r will obtain with probability 1. Instead, $\hat{p}_a^{(a';r)}$ will capture the subjectively perceived correlation between outcomes of the distinct actions a and a'.

(iii) If the data-set contains heterogenous cases, i.e. cases containing different actions, then different cases will have different degrees of relevance for the evaluation of a. Hence, the frequencies of cases have to be modified by the similarity weights $s\left(a;a_c\right)$. Note that the similarity weights depend only on the action chosen in case c and not on its outcome. We argue below that this is an attractive feature of the representation, since it excludes predictions biased towards favorable or unfavorable outcomes.

We can rewrite Equation (1) as

$$V\left(a;D\right)=\sum_{r\in R}u\left(r\right)h_a\left(D;r\right)$$

with

$$h_a\left(D;r\right)=:\frac{\sum_{c\in C}f_D\left(c\right)s\left(a;a_c\right)\hat{p}_a^c\left(r\right)}{\sum_{c\in C}f_D\left(c\right)s\left(a;a_c\right)} \tag{2}$$

denoting the probability weight assigned to outcome r conditional on the information contained in D. Hence, our approach provides a behavioral foundation for the representation of BGSS (2005). While BGSS (2005) take the function $h_a\left(D\right)$ as a primitive concept and provide conditions for representing it as similarity-weighted frequencies of outcomes as in Equation (2), we derive the beliefs of the decision maker explicitly from observed preferences.

The rest of the paper is organized as follows. In the next section, we present the preference relation on action-data-set-pairs. Section 3 presents our axioms. Section 4 contains our main representation result for preferences over action-data-set-pairs. Section 5 concludes. All proofs are collected in the Appendix.

2 Preferences on actions and data-sets

Consider a finite set of actions A with a representative element a. It is known that the payoffs of the actions r come from a finite set R with at least three distinct elements. We deviate from the standard frameworks used in the literature to model decision-making under uncertainty. In particular, we assume that the decision maker knows neither the probability distribution of payoffs associated with a specific action a (as in the von-Neumann-Morgenstern model), nor the mapping which describes the state-contingent outcomes of an action as in Savage's framework. In contrast, all the information available to the decision maker is in form of observations. Each observation, i.e. *case*, consists of an action and an outcome generated from this action. We write

$$c = (a; r), a \in A, r \in R$$

for a specific case and $C = A \times R$ for the set of all possible cases. A set of T such observations is referred to as a data-set of length T ($T \geq 1$):

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

The set of all conceivable data-sets is denoted by $\mathbb{D}$. The set of data-sets of length T is denoted by $\mathbb{D}^T$. We will write $|D|$ for the length of D. The frequency of cases in a data-set $D \in \mathbb{D}^T$ is given by

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

δ_r stands for the Dirac measure assigning a mass 1 to $r \in R$.

Remark 2.1 *The proper specification of cases is important. Depending on how one defines the cases, new data-sets may provide more or less precise additional evidence for these cases. For example, two samples of balls drawn from the same urn at different dates could be seen as distinct cases, because they were drawn at different times, or as additional observations of the* *same cases if the difference in timing is viewed as immaterial.* ■

We assume that a decision maker evaluates decision situations consisting of actions a and associated information given by a data-set D. Hence, we assume a preference order $\succsim$ defined on the set of actions and data-sets $A \times \mathbb{D}$. The preference order $(a; D) \succsim (a'; D')$ means that the decision maker prefers action a in a situation where the evidence is given by the data-set D over action a' in a situation described by the data-set D'.
The following two examples provide an illustration of our framework.

Example 2.1 Urn example
Consider a decision maker who has to choose between bets on the color of a ball drawn from an urn with 100 balls which are either black or white. Hence, we can denote the set of actions by $A = \{B, W\}$, where B denotes the bet on a black ball and W the bet on a white ball. Assuming that a bet wins 1, if the respective ball is drawn, and yields 0 otherwise, we have $R = \{0, 1\}$. Suppose that bets and outcomes of past draws are known. Hence, the available information consists of previous cases $c = (a, r) \in \{B, W\} \times \{0, 1\}$. After T rounds, a set of past observations $D = (c_1, ..., c_T)$ is available and defines the situation in which the choice of a bet takes place.
If the decision maker can decide when and on which color to place a bet, then action-data-set pairs must be compared, e.g., (W, D) with (B, D'). The relation $(W, D) \succsim (B, D')$ expresses a preference to bet on W if the data-set is $D = (c_1, ..., c_T)$ compared to a bet on B based on the data $D' = (c'_1, ..., c'_{T'})$. In this example, past information is given by independent draws with replacement from a given urn and, therefore, the order of the cases does not matter. Hence, one can summarize the information of these data-sets by their frequencies f_D and $f_{D'}$ and the number of observations $T = |D|$ and $T' = |D'|$. ■

From example 2.1 it is clear that a preference $(W, D) \succsim (W, D')$ does not mean that the preferred data-set D could be chosen given the action to bet on W. The relation rather indicates that, when betting on W, the decision maker prefers the information contained in data-set D over the information in D'. Or, choosing bet W in a situation with information D is ranked higher than choosing W in a situation D'. Considering preferences over action-situation pairs allows one to study preferences over information without implying that information itself is necessarily an object of choice.

Example 2.1 corresponds to a controlled decision situation as one finds it in organized gambles or statistical experiments. Most real-world decisions take place, however, in far less structured environments. The following example is taken from the context of health economics.

Example 2.2 Hospitalization *(O'Hara & Luce 2003, pp.62-4)*
The decision concerns the cost-effectiveness of a new drug relative to the standard treatment. Cost-effectiveness is measured by the number of days a patient has to stay in hospital.
Data is available from a clinical trial where two treatments with 100 patients each were recorded, $|D| = 200$*. A case* c *is described by the action* a *whether the patient got the new drug* a_d *or not* a_n*,* $A = \{a_d; a_n\}$*, and the outcome* r*, i.e., the number of days the patient spent in hospital. For the group of patients receiving the standard treatment (*a_n*) a total of 25 days in hospital were observed, while for the group receiving the new drug (*a_d*) only 5 days were recorded. Based on this data-set of cases, the decision may be made to have the new drug replace the standard treatment.*
The data base of 200 patients is, however, quite small. Suppose there are also data from a larger study of a similar drug *at another hospital in which the average number of days in hospital was 0.21. This data-set contains cases with the action* $a = a'_d$ *and, as before, as outcomes* r *the days the patients spent in hospital. This may cast doubts on the reliability of the observation of an average time in hospital of 0.05 days for the new drug. Whether one feels persuaded by the new or by the old evidence depends to a large extent on how similar one judges the two situations reflected in the data-sets.* ■

The description of Example 2.2 follows closely the wording of O'Hara & Luce (2003). It is clear from this description that the decision maker is concerned about the lack of data and considers explicitly data which seem similar but are not fully adequate for the choice under consideration. In particular, the decision maker shows a clear preference for "more adequate data".

As these examples illustrate, information contained in data-sets may be quite diverse. Data can be very crisp, as we usually find them in controlled experiments, or, very opaque, if one has to rely on information from similar cases. The representation which we will axiomatize below will allow us to deal with decision problems of the type described in Examples 2.1 and 2.2.

3 Axioms

We now impose conditions on the preference relation $\succsim$ which guarantee that the utility of action a given the informational content of D can be written as:

$$V(a;D) = \frac{\sum_{r\in R} v(r) \sum_{c\in C} f_D(c)\, s(a;a_c)\, \hat{p}_a^c(r)}{\sum_{c\in C} f_D(c)\, s(a;a_c)}. \tag{3}$$

We make extensive use of the following notation: $\mathbb{D}_{a'}$ denotes the set of data-sets containing only cases in which a specific action $a' \in A$ was chosen. Formally,

$$\mathbb{D}_{a'} = \{D \in \mathbb{D} \,|\, f_D(c) > 0 \text{ only if } c = (a';r') \text{ for some } r' \in R\}.$$

We also use $\mathbb{D}_{a'}^T$ to denote the restriction of $\mathbb{D}_{a'}$ to data-sets of length T.

Axiom 1 (Complete Order)

$\succsim$ on $A \times \mathbb{D}$ is complete and transitive.

Axiom 2 (Invariance)

For each $a \in A$ and for any two D and $D' \in \mathbb{D}^T$ such that there exists a one-to-one mapping $\pi : \{1...T\} \to \{1...T\}$, with the property:

$$D = (c_t)_{t=1}^T \text{ and } D' = \left(c_{\pi(t)}\right)_{t=1}^T$$
$$(a;D) \sim (a;D').$$

Axiom 2 states that the order in which information arrives does not influence the evidence that D gives in favor of a given action. This axiom implicitly assumes that the outcome of an action does not depend on the order or on the time period in which different actions are chosen. Also, it implies that each data-set is uniquely characterized by its length and frequency. Hence, from now on, we will think of the elements of $\mathbb{D}$ as multisets and write $D = D'$ whenever the two multisets are equal, i.e. whenever D and D' are equivalent up to a permutation.

We now define the concatenation operator:

Definition 3.1 *Concatenation*
Let D and $D' \in \mathbb{D}$ be given by

$$\begin{aligned} D &= \{(a_1;r_1)\ldots(a_T;r_T)\} \\ D' &= \{(a_1';r_1')\ldots(a_{T'}';r_{T'}')\} \end{aligned}$$

The data-set

$$D \circ D' = \{(a_1;r_1)\ldots(a_T;r_T);(a_1';r_1')\ldots(a_{T'}';r_{T'}')\}$$

is called the concatenation of D and D'.

We write $D^k := \underbrace{D \circ D \circ ... \circ D}_{k\text{-times}}$ for the k^{th} concatenation of the set D with itself.

Axiom 3 (Continuity + Concatenation)

For any $a \in A$, $(a; D) \succ (a; D'')$ implies

$$(a; D) \succ (a; D \circ D'') \succ (a; D'') .$$

Furthermore, for any $a \in A$, if D' satisfies

$$(a; D) \succ (a; D') \succ (a; D'') ,$$

then there are k, l m and $n \in \mathbb{N} \setminus \{0\}$ with $\frac{k}{l} > \frac{m}{n}$, such that

$$\left(a; D^k \circ (D'')^l\right) \succ (a; D') \succ \left(a; D^m \circ (D'')^n\right) .$$

To understand the axiom, note that a preference for $(a; D)$ over $(a; D'')$ indicates that D is considered to contain more favorable evidence for a than D''. Then, the concatenation of D and D'' contains both favorable and less favorable evidence for a. Hence, it is natural to evaluate this set to be worse than D and better than D''. In contrast, suppose that D' is evaluated in between D and D''. Then, a sufficiently large number of replicas of D (k large relative to l) can outweigh the negative evidence contained in D'' and, thus make the data-set $D^k \circ (D'')^l$ better than D'. Since we do not allow l to become 0, the negative evidence in D'' will never be completely eliminated, but its weight in the evaluation of the data-set $D^k \circ (D'')^l$ can be made arbitrarily small, so that for any data-set $(a; D') \prec (a; D)$ the preference $\left(a; D^k \circ (D'')^l\right) \succ (a; D')$ can be obtained. The argument for $(a; D') \succ (a; D^m \circ (D'')^n)$ is analogous.

The second part of Axiom 3 represents a standard continuity assumption defined on the set $\mathbb{D}$. Our next lemma shows that the first part of the axiom is closely related to the Concatenation axiom of BGSS (2005). In particular, Axiom 3 implies that for every $k \in \mathbb{Z}/\{0\}$, the data-sets D and D^k are considered indifferent, which, in our representation implies that they are associated with identical probability distributions over outcomes. We state this as a lemma:

Lemma 3.1 *Under Axioms 1, 2 and 3, for all $a \in A$, all $D \in \mathbb{D}$ and all $k \in \mathbb{N} \setminus \{0\}$,*

$$\left(a; D^k\right) \sim (a; D) .$$

Hence, the beliefs associated with a data-set depend only on the frequency with which cases appear in the data-set, but not on the length of the data-set. A decision maker, who, e.g. values data-sets with a larger number of observations, because he considers them to be more reliable, will in general violate Axiom 3. In particular, when comparing the data-sets D and $D \circ D''$, he will also take into account that $D \circ D''$ is longer, and, hence, (potentially) more reliable than

D. Hence, it is possible that $(a; D \circ D'') \succ (a; D)$ obtains despite the fact that the evidence contained in D'' is considered less favorable for a than the one contained in D. If one thinks of longer data-sets as more reliable, i.e., more precise in the sense that they allow the decision maker to exclude more probability distributions as plausible descriptions of the data generating process, the fact that the decision maker ignores the reliability of the data-set means that he is insensitive to the degree of information precision. Under Axioms 2 and 3, it is possible to identify a data-set D with the frequency of cases f_D it generates.

Axiom 4 (Independence for controlled statistical experiments)

For all a', $a'' \in A$, all D'_1, $D'_2 \in \mathbb{D}_{a'}$ and D''_1, $D''_2 \in \mathbb{D}_{a''}$ and for any natural numbers k, l, m, $n \in \mathbb{N}$ such that $\left|(D'_1)^k\right| = \left|(D''_1)^l\right|$ and $|(D'_2)^m| = |(D''_2)^n|$,

$$(a'; D'_1) \underset{(\prec)}{\succ} (a''; D''_1) \tag{4}$$

$$(a'; D'_2) \underset{(\precsim)}{\succsim} (a''; D''_2)$$

implies:

$$\left(a'; (D'_1)^k \circ (D'_2)^m\right) \underset{(\prec)}{\succ} \left(a''; (D''_1)^l \circ (D''_2)^n\right) \tag{5}$$

and if $(a'; D'_2) \sim (a''; D''_2)$, the two statements, (4) and (5) are equivalent.

Consider first the case of $a' = a'' = a$. The axiom then claims that if the evidence in data-set D'_1 is considered more favorable for a than the evidence in D''_1 and, similarly, the evidence in D'_2 is considered at least as favorable as the one in D''_2, then the combination of D'_1 and D'_2 should be thought at least as favorable as the combination of the evidence contained in D''_1 and D''_2. Furthermore, if the evidence in D'_2 and D''_2 is regarded as equivalent, the preferences between $(D'_1)^k \circ (D'_2)^m$ and $(D''_1)^l \circ (D''_2)^n$ should be determined only by the comparison between D'_1 and D''_1.

The axiom extends this intuition to all actions a' and a'' as long as each action is evaluated in situations in which only cases containing the choice of this particular action are observed. If action a' is preferred to action a'' in two situations: when the evidence for a' is D'_1, while the evidence for a'' is D''_1 and when the evidence for a' is D'_2, while the evidence for a'' is D''_2, then combining the evidence for a' to $(D'_1)^k \circ (D'_2)^m$ should render a' more preferred than a'' under the combined evidence $(D''_1)^l \circ (D''_2)^n$.

It is important to note the restrictions of the axiom. First, it is necessary to control for the length of the data-sets when applying the concatenation operator. Second, it is important that the cases in D_1' and D_2' contain only action a', while the cases in D_1'' and D_2'' contain only action a''. This ensures that all cases in $(D_1')^k \circ (D_2')^m$ will be equally relevant for the evaluation of a', and similarly, for $(D_1'')^l \circ (D_2'')^n$. These two restrictions in the statement of the axiom imply that the evidence contained in D_1' should receive the same weight in $(D_1')^k \circ (D_2')^m$ as does the evidence in D_1'' in the evaluation of $(D_1'')^l \circ (D_2'')^n$, thus motivating the independence property.

The independence property plays a crucial role in the representation, allowing us to separate utility and beliefs.

Axiom 5 (Most favorable and least favorable evidence)

For all $a \in A$, there exist $\bar{r}_a$ and $\underline{r}_a \in R$ such that

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

and for all $D \in \mathbb{D}$,

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

First note that according to Lemma 3.1, the number of repetitions of cases in the data-set is irrelevant. Hence, for all $a \in A$,

$$(a; (a; \bar{r}_a)) \sim (a; (a; \bar{r}_a))^k$$

for any $k \geq 1$, and similarly for $(a; (a; \underline{r}_a))$.

Axiom 5 then asserts that the (repeated) observation of a resulting in the "worst" outcome w.r.t. this action would represent the least favorable evidence and the (repeated) observation of a leading to the "best" outcome w.r.t. a would constitute the most favorable evidence in support of a.

Axioms 1-5 allow us to establish an important intermediate result: preferences on $\{a\} \times \mathbb{D}_a$ can be represented by $V_a(D) = \sum_{r \in R} v_a(r) f_D(a; r)$ for a von-Neumann-Morgenstern utility function $v_a(r)$. Axiom 5 further states that for an *arbitrary data-set* D (not necessarily one in $\mathbb{D}_a$), $(a; D)$ is ranked between $(a; (a; \bar{r}_a))$ and $(a; (a; \underline{r}_a))$. This means that we can approximate the utility of $(a; D)$ by a sequence of data-sets in $\mathbb{D}_a$ and in this way extend the function $V_a(D)$ to all $D \in \mathbb{D}$. In particular, $\succsim$ on $\{a\} \times \mathbb{D}$ can be represented as:

$$V_a(D) = \sum_{r \in R} v_a(r) h(r), \; h(r) \in \mathcal{H}_a(D)$$

where $\mathcal{H}_a(D)$ is the set of frequencies over outcomes of those data-sets D' in $\mathbb{D}_a$ which are considered indifferent to D when evaluating action a, i.e. $(a; D) \sim (a; D')$. Furthermore, we can show that $\mathcal{H}_a(D)$ can be written as:

$$\mathcal{H}_a(D) = \left\{ \frac{\sum_{c \in C} f_D(c)\, s_a(c)\, \hat{p}_a^c(r)}{\sum_{c \in C} f_D(c)\, s_a(c)} \mid \hat{p}_a^c \in \hat{P}_a^c \right\}$$

for some uniquely determined sets $\hat{P}_a^c$ of probability distributions over outcomes and some positive and unique up to a multiplication by a positive constant similarity values $s_a(c)$[6].

We now show how the relative weight assigned to a specific subset of observations can be determined. Consider two data-sets $D \in \mathbb{D}_{a'}^T$ and $D' \in \mathbb{D}_{a''}^T$ such that $(a; D) \succ (a; D')$. To determine the relative weight assigned to D in the concatenation $D \circ D'$, we will use the following construction: suppose that for some $\hat{T}$, there exist data-sets $\tilde{D}_1$ and $\tilde{D}_2 \in \mathbb{D}_a^{\hat{T}}$ such that

$$\begin{aligned} \left(a; \tilde{D}_1\right) &\sim (a; D) \\ \left(a; \tilde{D}_2\right) &\sim (a; D') . \end{aligned} \tag{6}$$

Together with Axiom 3, (6) implies:

$$\left(a; \tilde{D}_1\right) \sim (a; D) \succ (a; D \circ D') \succ (a; D') \sim \left(a; \tilde{D}_2\right).$$

Hence, also by Axiom 3, we can approximate the utility of $(a; D \circ D')$ arbitrarily closely by a sequence of data-sets of the form $\tilde{D}_1^k \circ \tilde{D}_2^n$, k, $n \in \mathbb{N}$. Let κ denote the limit of the ratio $\frac{k}{n}$ for this sequence. By Axiom 4, all cases contained in $\tilde{D}_1$ and $\tilde{D}_2$ are weighted equally in the concatenation $\tilde{D}_1^k \circ \tilde{D}_2^n$. Furthermore, for the evaluation of a, $\tilde{D}_1$ and $\tilde{D}_2$ are equivalent to D and D', respectively. It follows that the weight assigned to D relative to D' is given by κ.

This construction shows how the similarity coefficients can be derived from preferences[7]. It relies, however, on the assumption that data-sets $\tilde{D}_1$ and $\tilde{D}_2$ satisfying condition (6) can be found. Note, however, that by Axioms 3 and 4 we can express κ as:

$$\kappa = \liminf \left\{ \frac{k}{n} \mid \begin{array}{l} \text{for all } \hat{T} \text{ and all } D_1 \text{ and } D_2 \in \mathbb{D}_a^{\hat{T}} \text{ such that } (a; D_1) \succ (a; D) \\ \text{and } (a; D_2) \succ (a; D'), \left(a; D_1^k \circ D_2^n\right) \succ (a; D \circ D') \end{array} \right\}, \tag{7}$$

or, equivalently, as

$$\kappa = \limsup \left\{ \frac{k}{n} \mid \begin{array}{l} \text{for all } \hat{T} \text{ and all } D_1 \text{ and } D_2 \in \mathbb{D}_a^{\hat{T}} \text{ such that } (a; D) \succ (a; D_1) \\ \text{and } (a; D') \succ (a; D_2), (a; D \circ D') \succ \left(a; D_1^k \circ D_2^n\right) \end{array} \right\}. \tag{8}$$

This provides us with an alternative method to derive the coeffiicient κ, which does not rely on the assumption that data-sets satisfying condition (6) exist.

We use this property of the similarity coefficients to formulate our next axiom. We require that

---

6 This result follows from Theorem 1 proved in EG (2007).

7 In particular, setting $D = (a'; r')$, $D' = (a''; r'')$ we obtain $\frac{s_a(a'; r')}{s_a(a''; r'')} = \kappa$, which identifies the similarity weights up to a multiplication by a positive constant.

replacing the sequences of outcomes of actions a' and a'' in the data-sets D and D' with different sequences of outcomes of the same length should leave the sets on the r.h.s. of expressions (7) and (8), and, therefore, also the coefficient κ unchanged. In this way, we ensure that the similarity weights $s_a(c)$ depend only on the action chosen in case c, but not on the observed payoff.

Axiom 6 (Outcome Independence)

Let $D, \hat{D} \in \mathbb{D}^T_{a'}$ and $D', \hat{D}' \in \mathbb{D}^T_{a''}$. For all $a \in A$, if $(a; D) \succsim (a; D')$ and $\left(a; \hat{D}\right) \underset{(\precsim)}{\succsim} \left(a; \hat{D}'\right)$, then for any k and n, and any $\hat{T} \in \mathbb{N}\setminus\{0\}$,

$$\begin{array}{c}\left(a; D_1^k \circ D_2^n\right) \succ (a; D \circ D') \text{ for all } D_1 \text{ and } D_2 \in \mathbb{D}_a^{\hat{T}} \\ \text{such that } (a; D_1) \succ (a; D) \text{ and } (a; D_2) \succ (a; D')\end{array}$$

holds if and only if

$$\left(a; \hat{D}_1^k \circ \hat{D}_2^n\right) \underset{(\prec)}{\succ} \left(a; \hat{D} \circ \hat{D}'\right) \text{ for all } \hat{D}_1 \text{ and } \hat{D}_2 \in \mathbb{D}_a^{\hat{T}}$$

$$\text{such that } \left(a; \hat{D}_1\right) \underset{(\prec)}{\succ} \left(a; \hat{D}\right) \text{ and } \left(a; \hat{D}_2\right) \underset{(\prec)}{\succ} \left(a; \hat{D}'\right)$$

and

$$\begin{array}{c}\left(a; D_1^k \circ D_2^n\right) \prec (a; D \circ D') \text{ for all } D_1 \text{ and } D_2 \in \mathbb{D}_a^{\hat{T}} \\ \text{such that } (a; D_1) \prec (a; D) \text{ and } (a; D_2) \prec (a; D')\end{array}$$

holds if and only if

$$\left(a; \hat{D}_1^k \circ \hat{D}_2^n\right) \underset{(\succ)}{\prec} \left(a; \hat{D} \circ \hat{D}'\right) \text{ for all } \hat{D}_1 \text{ and } \hat{D}_2 \in \mathbb{D}_a^{\hat{T}}$$

$$\text{such that } \left(a; \hat{D}_1\right) \underset{(\succ)}{\prec} \left(a; \hat{D}\right) \text{ and } \left(a; \hat{D}_2\right) \underset{(\succ)}{\prec} \left(a; \hat{D}'\right).$$

In the statement of Axiom 6, the combination of evidence from observations of distinct alternatives is a key feature. The axiom (combined with the preceding argument) implies that the weight on the evidence obtained from the observation of a given action should not depend on the outcome of this action. Hence, this axiom precludes the case of a decision maker who thinks that, say, good outcomes are more relevant for the evaluation of a than bad outcomes. This, of course, does not mean that the evidence from two observations with different outcomes $(a'; r_1)$ and $(a'; r_2)$ is considered to be equally favorable for the evaluation of a, merely that the relative weight this evidence is given when combined with the evidence from a data-set of potentially different observations, D, is identical. Hence, the similarity function $s_a(c)$ derived above can be written as $s_a(a_c)$, with a_c denoting the action chosen in case c.

Axiom 7 (Action Independence)

For all a, $a' \in A$ and all $r \in R$,

$$(a; (a; r)) \sim (a'; (a'; r)).$$

In combination with Axiom 5, Axiom 7 implies that the best outcome $\bar{r}_a$ and the worst outcome, $\underline{r}_a$ are the same for all actions in A. Furthermore, observing an outcome r as a result of action a gives the same evidence in favor (or against) a as does observing the same outcome r as a result of action a' for action a'. Hence, the decision maker is able to order the outcomes w.r.t. their desirability independently of the action from which they result. This allows us to construct a utility function over outcomes $v(r)$, which is independent of the action a, thus leading to the desired representation.

4 The Representation

Axioms 1 — 7 imply the desired representation:

Theorem 4.1 *The preference relation $\succsim$ satisfies Axioms 1 — 7, if and only if there exists a probability correspondence $\hat{P} : A \times C \rightrightarrows \Delta^{|R|-1}$, a family of similarity functions $s_a : A \to \mathbb{R}^+ \setminus \{0\}$ for $a \in A$, and a utility function over outcomes $v : R \to \mathbb{R}$, such that $\succsim$ on $A \times \mathbb{D}$ can be represented by:*

$$V(a; D) = \frac{\sum_{r \in R} v(r) \sum_{c \in C} \hat{p}_a^c(r)\, s_a(a_c)\, f_D(c)}{\sum_{c \in C} s_a(a_c)\, f_D(c)}.$$

Here, a_c denotes the action chosen in case c, $f_D(c)$ is the frequency with which case c is observed in D and $\hat{p}_a^c$ is any probability distribution over outcomes in the image of $\hat{P}(a; c)$. The utility function v is unique up to an affine-linear transformation, the similarity functions s_a are unique up to a multiplication by a positive number and there is a unique probability correspondence $\hat{P}$ such that for all $r \in R$, $\hat{P}(a; (a'; r))$ is maximal with respect to set inclusion when $a' \neq a$ and satisfies: $\hat{P}(a; (a; r)) = \{\delta_r\}$.

It is easily checked that this representation has the following properties:

1. When evaluating a data-set in $\mathbb{D}_{a'}$ with respect to a specific action a, the evidence contained in each of the cases is weighted equally.
2. For controlled statistical experiments w.r.t. the action under consideration, a, the probability distribution over outcomes coincides with the observed frequency of outcomes. However, when the observed action is distinct from a, the probability distribution over outcomes in general depends on both the action and the outcome in the observed case c. Moreover, in this case, the beliefs of the decision maker cannot be uniquely identified and are instead given by a set of probability distributions $\hat{P}_a^c$.
3. For data-sets containing observations of different actions, the weight put on the evidence

contained in a specific case in general depends both on the action under consideration and on the action chosen in this case, but not on the observed outcome.

We now discuss these three properties in turn. Property 1 applies to the case of controlled statistical experiments, in which an action (potentially different from the one being evaluated, a), say a' is observed in each period. Since the choice of action remains unchanged, the relevance of each case in the data-set for the evaluation of a will be identical, i.e. the weight given to the evidence derived from each of the cases in the data-set should be the same. This requires that the decision maker does not differentiate between positive and negative outcomes when deciding on the relevance of a specific case. We think that this is a desirable property of the representation. It excludes decision makers, who discard, say all negative evidence and base their evaluation of a given action only on the "favorable" cases in the data-set.

Property 2 states that the predictions based only on observations of action a have to coincide with the observed frequency of outcomes. This seems natural, given that Axiom 3 implies that the decision maker is not sensitive to the degree of precision of the information — he does not differentiate between longer and shorter data-sets when evaluating the evidence. Our representation allows, however, for different probabilistic predictions when a specific outcome results from the choice of a different action, a'. For instance, observing that a specific drug a'_d was successful (resulted in outcome $\bar{r}$) in the treatment of a specific disease might lead to the prediction that outcome $\bar{r}$ will occur with probability of 1 when using this drug. This need not mean that using a similar, drug a_d, which, e.g. has an additional component, would also be successful. Instead, the physician might know that the additional component suppresses the effect of the active component of drug a'_d. Thus, a probability less than 1 would be assigned to $\bar{r}$ when evaluating action a_d if the only case observed in the data-set is $(a'_d; \bar{r})$.

Property 3 applies to heterogenous data-sets, allowing the evidence contained in different cases to receive different weights depending on the relevance of these cases with respect to a. For instance, cases in which a was chosen will be presumably the most relevant ones. To avoid predictions biased in favor of good or bad outcomes, we require the relevance of a case to be independent of its outcome.

We now explain why it is impossible to uniquely identify the beliefs of the decision maker when the data-set contains observations of actions different from a. In the Savage approach, the possibility to generate betting acts for each state of the world is what allows the unique elicitation of subjective beliefs. In the case-based framework, states are not defined a priori, the cases serving as primitives of the model. In our model, the set of cases is restricted: the decision maker does not observe the outcomes of two distinct actions simultaneously[8]. Hence, regardless of the number of observations, the joint probability distribution over outcomes of a and a' is never observed. E.g., in example 2.2, each patient can be subject to only one of the two possible treatments. In the Ellsberg experiment, it might be that the decision maker learns only the color of the ball drawn from the urn he actually placed a bet on.

To elicit beliefs according to the Savage approach, we must define the relevant state-space, $A^{|R|}$, and construct an extended set of acts, $\mathcal{A} : A^{|R|} \to R$. Preferences are defined on $\mathcal{A}$ and include comparisons among all betting acts on $A^{|R|}$ and, in particular, among bets on the joint outcomes of a and a'. This is a necessary condition to identify $\hat{p}_a^{(a';r')}$. In our framework, the impossibility to bet on and observe the simultaneous performance of two actions precludes this extension from A to $\mathcal{A}$. Hence, the set of actions is not sufficiently rich to uniquely determine $\hat{p}_a^{(a';r')}$.

Therefore, instead of following Savage's approach, we first derive a von-Neumann-Morgenstern representation of preferences on $\{a\} \times \mathbb{D}_a$, where the frequencies of observations serve as probability weights. We then calibrate the utility of $(a; D)$ for arbitrary sets D by using data-sets in $\mathbb{D}_a$. The frequency of a data-set $D' \in \mathbb{D}_a$ such that $(a; D) \sim (a; D')$ can be used as probability distribution over outcomes in the evaluation of D. However, for each $D \in \mathbb{D}$, there will be (in general) many such equivalents in $\mathbb{D}_a$, and hence, many probability distributions consistent with the expressed preferences. Specifically, for $D = (a'; r')$, the set of such equivalents identifies the set of probability distributions $\hat{P}_a^{(a';r')}$ used in the representation.

It is important to note the different reasons for considering multiple priors in EG (2007) as compared to the current paper. There, the motivation for working with sets of probability distributions is that the decision maker considers the information contained in a longer data-set to be more precise. In controlled statistical experiments, the set of priors shrinks as additional data confirm the existing evidence.

In contrast, in the current paper, we consider a decision maker whose preferences do not depend on the length of the data-set. After observing a controlled statistical experiment, he entertains a single probability distribution over outcomes, which coincides with the observed frequencies.

[8] This is true for many real-life situations such as the choice of an investment project, in which only the outcome of the implemented project is observed, or the choice of a medical treatment, where it is impossible to observe the outcome of a treatment that has not been applied, etc.

However, the decision maker faces persistent uncertainty with respect to the joint probability distribution of distinct actions. As a result, he might either, as a Savage subjective expected utility maximizer, reduce this uncertainty to a single probability distribution, or, he might entertain a set of probability distributions, all of which, however, imply the same expected utility of the action under consideration. The structure of the model does not allow us to differentiate between these two scenarios based on observable preferences.

5 Conclusion and Outlook

In this paper, we analyzed preferences on two types of objects: preferences on data-sets containing evidence in favor of a choice of a specific action and preferences over actions for a given set of observations. We stated conditions under which only the frequency of observations in a data-set matters for the choices of a decision maker and derived an expected utility representation for this situation. In particular, we were able to identify the von-Neumann-Morgenstern utility function over outcomes as well as the beliefs associated with each action for a given set of observations. We could also separate the beliefs into three components: the frequency of cases in the data-set, the relevance of each of the cases for the prediction to be made and the probabilistic prediction associated with each case. For the case of controlled statistical experiments, we showed that a decision maker who is insensitive towards the degree of information precision will use the frequency of observed outcomes to make predictions.

The results derived in this paper, however neglect the possibility that the precision of the information might influence the preferences of the decision maker as in the model of GHTV (2007). E.g., 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."

In a companion paper, Eichberger and Guerdjikova (2008), we extend the BGSS (2005) approach by allowing the decision maker to take into account the precision of the data. We illustrate how the Ellsberg paradox can be generalized to apply to different degrees of information precision. By restricting the validity of Axiom 3 to data-sets of equal length (and hence, of equal precision), it is possible to generalize the results derived above to the case of a decision maker who is not indifferent to information precision. In Eichberger and Guerdjikova (2008), we provide an axiomatization for this scenario and identify, for the case of controlled statistical experiments, the degree of perceived imprecision of information, as well as the relative degrees of optimism and pessimism. While the perceived imprecision declines as the number of observations increases, the relative degrees of optimism and pessimism are constant across data-sets and can be interpreted as inherent characteristics of the decision maker.

In controlled statistical experiments, the number of observations can be used as a proxy for the precision of information. However, in less structured environments, data-sets containing more relevant cases for the action under consideration may be considered more precise. Hence, it is conceivable that there exists a connection between the similarity of cases and the perceived precision of information. We plan to address this issue in future research.

6 Appendix

Proof of Lemma 3.1:

Suppose (w.l.o.g.) that $(a; D^n) \succ \left(a; D^k\right) \succ (a; D^m)$ for some distinct n, k and $m \in \mathbb{N}\setminus\{0\}$. Then, by Axiom 3, we have:

$$\begin{aligned}
(a; D^n) &\succ \left(a; D^{n+k}\right) \succ \left(a; D^k\right) \\
(a; D^n) &\succ \left(a; D^{2n+k}\right) \succ \left(a; D^{n+k}\right) \succ \left(a; D^k\right) \\
&\dots \\
(a; D^n) &\succ \left(a; D^{mn+k}\right) \succ \left(a; D^k\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\left(a; D^k\right) &\succ \left(a; D^{k+m}\right) \succ (a; D^m) \\
\left(a; D^k\right) &\succ \left(a; D^{k+m}\right) \succ \left(a; D^{k+2m}\right) \succ (a; D^m) \\
&\dots \\
\left(a; D^k\right) &\succ \left(a; D^{k+mn}\right) \succ (a; D^m),
\end{aligned}$$

implying

$$(a; D^n) \succ \left(a; D^{mn+k}\right) \succ \left(a; D^k\right) \succ \left(a; D^{k+mn}\right) \succ (a; D^m),$$

in contradiction to Axiom 1.

Now suppose that $(a; D^n) \succ \left(a; D^k\right) \sim (a; D^m)$ for all m distinct from n and k. Then, by

Axiom 3,

$$(a; D^n) \succ \left(a; D^{n+k}\right) \succ \left(a; D^k\right),$$

hence, e.g., setting $m = n + k \notin \{n; k\}$, we obtain $(a; D^m) \succ \left(a; D^k\right)$, in contradiction to the assumption above.■

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

It is straightforward to check that the representation satisfies the axioms. Hence, we only prove the existence of the representation and its uniqueness properties. We do this in a sequence of lemmas. Here we sketch the outline of the proof. Lemma 3.1 demonstrates that each data-set can be considered equivalent to the frequency of observed cases it entails and hence, preferences on action-data-set pairs $A \times \mathbb{D}$ can be reduced to preferences on action-frequency pairs, $A \times \Delta^{|C|-1} \cap \mathbb{Q}^{|C|-1}$. For a fixed action a, Lemma 6.1 derives an expected utility representation of preferences $\succsim$ restricted to $\{a\} \times \mathbb{D}_a$, in which beliefs are given by the frequency of cases observed in the data-set under consideration. This Lemma exploits the fact that under Axioms 1-5, preferences on $\{a\} \times \mathbb{D}_a$, or, equivalently, on $\{a\} \times \Delta^{|R|-1} \cap \mathbb{Q}^{|R|-1}$ satisfy the mixture space axioms, which, together with the denseness of $\mathbb{Q}^{|R|-1}$ in $\Delta^{|R|-1}$ and the continuity assumption entailed in Axiom 3 leads to the desired representation. Still for a fixed $a \in A$, Lemma 6.2 determines for each data-set $D \in \mathbb{D}$, the set of data-sets in $\mathbb{D}_a$, which are indifferent to D for the given action a. The frequencies of these sets are denoted by $\mathcal{H}_a(D)$. The expected utility representation of preferences can now be extended to the set $\{a\} \times \mathbb{D}$ by using the utility function over outcomes derived in Lemma 6.1 and using the frequencies in $\mathcal{H}_a(D)$ to represent beliefs for each D. Lemma 6.3 shows that the correspondence $\mathcal{H}_a(D)$ satisfies the main assumption of BGSS (2005), Concatenation. This property is used in Lemma 6.4 together with Theorem 1 in EG (2007) to show that $\mathcal{H}_a(D)$ can be represented as similarity-weighted frequencies of cases observed in D, thus identifying the sets $\hat{P}^c_a$ and the similarity values $s_a(c)$. Lemma 6.5 uses Axiom 6 to establish that the similarity function $s_a(c)$ is independent of the observed outcomes and only depends on the action chosen in the specific case c. Last, Lemma 6.6 shows that under Axiom 7, the same utility function over outcomes can be used in each of the representations derived above for individual values of a. Axiom 5 then implies the desired representation.

Lemma 6.1 *Fix an $a \in A$. Under Axioms A1-A5, there exists a function $v_a(r) : R \to \mathbb{R}$ such* *that for all* $D \in \mathbb{D}_a$

$$V_a(D) = \sum_{r \in R} v_a(r) f_D(a;r) \tag{9}$$

represents $\succsim$ *on* $\{a\} \times \mathbb{D}_a$*. Moreover,* v_a *is unique up to an affine-linear transformation.*

<u>**Proof of Lemma 6.1**</u>:

Lemma 3.1 implies that the preference relation $\succsim$ induces a preference relation over action-frequency pairs $(a; f)$ with $f \in \Delta^{\|C\|-1} \cap \mathbb{Q}^{\|C\|-1}$ defined by

$$(a; D) \succsim (a'; D')$$

iff

$$(a; f_D) \succsim (a'; f_{D'}) .$$

It inherits all properties with which we endow the preference relation $\succsim$. In particular, consider $\succsim$ constrained to $\{a\} \times \mathbb{D}_a$. By Axioms 1-5, it is complete, transitive, continuous, has a largest and a smallest element and satisfies the following independence condition: for all D, $D' \in \mathbb{D}_a^T$,

$$\begin{aligned} (a; f_D) &\succsim (a; f_{D'}) \text{, iff} \\ (a; \alpha f_D + (1-\alpha) f_{D''}) &\succsim (a; \alpha f_{D'} + (1-\alpha) f_{D''}) \end{aligned} \tag{10}$$

for all $\alpha \in (0;1) \cap \mathbb{Q}$ and all $D'' \in \mathbb{D}$. This follows from Axiom 3 by setting $a' = a'' = a$, assuming that $D_1'' = D_2'' = D''$ and noting that

$$\begin{aligned} f_{D \circ D''} &= \frac{T \cdot f_D + T'' \cdot f_{D''}}{T + T''}, \\ f_{D' \circ D''} &= \frac{T \cdot f_{D'} + T'' \cdot f_{D''}}{T + T''} \end{aligned}$$

holds, where T'' is the length of D''. Let $\alpha = \frac{T}{T+T''}$. It is obvious that by choosing the lengths T and T'' in an appropriate way, we can generate any rational-valued α. The result of Lemma 6.1 then follows almost directly from an application of the mixture-space theorem, see Fishburn (1970, p.112). Small modifications have to be made to Fishburn's proof to take into account the fact that we allow only for rational-valued frequencies. In particular, Fishburn's argument requires that for all $D \in \mathbb{D}_a$, there exists a data-set

$$D' = (a; \bar{r}_a)^k \circ (a; \underline{r}_a)^n$$

such that $(a; D') \sim (a; D)$. It is obvious, that in general such natural numbers k and n need not exist. However, the continuity assumption contained in Axiom 3 together with the fact that rational frequencies are dense in the simplex $\Delta^{|R|-1}$ allows us to approximate the utility of

$(a; D)$ by a sequence of data-sets D' of the form $(a; \bar{r}_a)^k \circ (a; \underline{r}_a)^n$ arbitrarily closely. It follows that the utility values of all sets $D \in \mathbb{D}_a$ can be uniquely identified.■

Lemma 6.2 *Fix an $a \in A$. For every $D \in \mathbb{D}$, there exists a maximal non-empty, compact and convex subset of $\Delta^{|R|-1}$,*

$$\mathcal{H}_a(D) = \left\{ h \in \Delta^{|R|-1} \mid \sum_{r \in R} v_a(r)\, h(r) = V_a(D) \right\}$$

such that

$$\begin{array}{rcl} V_a(D) & \geq & V_a(D') \text{ iff} \\ (a; D) & \succsim & (a; D'). \end{array}$$

<u>**Proof of Lemma 6.2**</u>:

To construct the sets $\mathcal{H}_a(D)$ for a given $a \in A$, we first define the following sets:

$$\mathcal{F}_a = \left\{ f \in \Delta^{|C|-1} \cap \mathbb{Q}^{|C|-1} \mid f(a'; r) = 0 \text{ for all } a' \neq a \right\}$$

is the set of rational-valued frequencies which assign a frequency of 0 to all cases not containing action a, and, for a given set D,

$$\begin{array}{rcl} \mathcal{F}_a(D, \succ) & = & : \{ f \in \mathcal{F}_a \mid (a; f) \succ (a; f_D) \} \\ \mathcal{F}_a(D, \prec) & = & : \{ f \in \mathcal{F}_a \mid (a; f_D) \succ (a; f) \} \end{array}$$

denote the sets of frequencies in $\mathcal{F}_a$ which are preferred to, respectively, less preferred than f_D. By Axiom 5, at least one of these sets is non-empty. If both of these sets are non-empty, Axiom 3 implies that:

$$\inf \mathcal{F}_a(D, \succ) = \sup \mathcal{F}_a(D, \prec) =: \mathcal{F}_a(D, \sim)$$

If $\mathcal{F}_a(D, \succ) = \varnothing$, then $(a; D) \sim (a; (a; \bar{r}_a))$ and we define $\mathcal{F}_a(D, \sim)$ to be equal to $\sup \mathcal{F}_a(D, \prec)$ and symmetrically if $\mathcal{F}_a(D, \prec) = \varnothing$.

We now define the set $\mathcal{H}_a(D)$ to be the projection of the set $\mathcal{F}_a(D, \sim)$ to $\Delta^{|R|-1}$, i.e.

$$\mathcal{H}_a(D) = \left\{ h \in \Delta^{|R|-1} \mid \text{for some } f \in \mathcal{F}_a(D, \sim),\, h(r) = f(a; r) \text{ for all } r \in R \right\}.$$

By Lemma 6.1, we know that for all h, $h' \in \mathcal{H}_a(D)$,

$$\sum_{r \in R} v_a(r)\, h(r) = \sum_{r \in R} v_a(r)\, h'(r)$$

Define the value of V_a at D to be:

$$V_a(D) =: \sum_{r \in R} v_a(r)\, h(r)$$

for some (and hence for all) $h \in \mathcal{H}_a(D)$. By the construction of $\mathcal{H}_a(D)$, it is obvious that $V_a(D)$ represents the preference relation $\succsim$ constrained to $\{a\} \times \mathbb{D}$.

To complete the proof, we show that each $h \in \Delta^{|R|-1}$ with the property that

$$\sum_{r \in R} v_a(r)\, h(r) = V_a(D)$$

is in $\mathcal{H}_a(D)$. Since the set of such h is compact and convex, this establishes the result.

Note that for every real-valued h, we can construct two sequences of rational-valued frequencies $\tilde{f}_1 \ldots \tilde{f}_n \ldots \in \mathcal{F}_a$ and $\tilde{f}^1 \ldots \tilde{f}^n \ldots \in \mathcal{F}_a$ such that:

$$\lim_{n\to\infty} \tilde{f}_n(a;r) = \lim_{n\to\infty} \tilde{f}^n(a;r) = h(r) \text{ for all } r \in R.$$

The separating hyperplane theorem ensures that we can choose the two sequences in such a way that:

$$\begin{aligned}
\sum_{r\in R} v_a(r)\, \tilde{f}_n(a;r) &< \sum_{r\in R} v_a(r)\, h(r) = V_a(D) \\
\sum_{r\in R} v_a(r)\, \tilde{f}^n(a;r) &> \sum_{r\in R} v_a(r)\, h(r) = V_a(D)
\end{aligned}$$

for all n and rearrange the elements so that:

$$\begin{aligned}
\sum_{r\in R} v_a(r)\, \tilde{f}_n(a;r) &\leq \sum_{r\in R} v_a(r)\, \tilde{f}_{n+1}(a;r) \\
\sum_{r\in R} v_a(r)\, \tilde{f}^n(a;r) &\geq \sum_{r\in R} v_a(r)\, \tilde{f}^{n+1}(a;r)
\end{aligned}$$

for all n. Then, for each n,

$$\begin{aligned}
\tilde{f}^n &\in \mathcal{F}_a(D, \succ) \\
\tilde{f}_n &\in \mathcal{F}_a(D, \prec)
\end{aligned}$$

Furthermore, for each $f' \in \mathcal{F}_a(D, \succ)$, there is an N such that for all $n > N$, $\tilde{f}^n \prec f'$ and for each $f'' \in \mathcal{F}_a(D, \prec)$, there is an M such that for all $n > M$, $\tilde{f}_n \succ f''$. Hence, the frequency vector f such that $f(a;r) = h(r)$ for all $r \in R$ and $f(a';r) = 0$ for all $a' \neq a$ satisfies

$$f \in \mathcal{F}_a(D, \sim)$$

Therefore, $h \in \mathcal{H}_a(D)$.■

Our next result shows that the set of probability distributions associated with the concatenation of two data-sets is a convex combination of the probability distributions associated with these data-sets:

Lemma 6.3 *For every D, $D' \in \mathbb{D}$,*

$$\mathcal{H}_a\left(D\circ D'\right)=\alpha\mathcal{H}_a\left(D\right)+\left(1-\alpha\right)\mathcal{H}_a\left(D'\right)$$

for some $\alpha\in(0;1)$.

<u>**Proof of Lemma 6.3**</u>:

Let first $(a;D)\succ(a;D')$. By Axiom 3,

$$(a;D)\succ(a;D\circ D')\succ(a;D')\,.$$

Then,

$$h_D\cdot v_a>h_{D\circ D'}\cdot v_a>h_{D'}\cdot v_a$$

holds for all $h_D\in\mathcal{H}_a(D)$, $h_{D'}\in\mathcal{H}_a(D')$ and $h_{D\circ D'}\in\mathcal{H}_a(D\circ D')$. It is then obvious that there is an $\alpha\in(0;1)$ such that:

$$\left[\alpha h_D+\left(1-\alpha\right)h_{D'}\right]\cdot v_a=h_{D\circ D'}\cdot v_a.$$

Hence,

$$\alpha\mathcal{H}_a\left(D\right)+\left(1-\alpha\right)\mathcal{H}_a\left(D'\right)=\mathcal{H}_a\left(D\circ D'\right).$$

Let now $(a;D)\sim(a;D')$. Then, by definition $\mathcal{H}_a(D)=\mathcal{H}_a(D')$ and the result would obtain if we could show that $(a;D\circ D')\sim(a;D)$.

Consider two sequences of data-sets $\left(\tilde{D}_n\right)_{n=1,2...}$ with $\lim_{n\to\infty}f_{\tilde{D}_n}=f_D$ such that $\left(a;\tilde{D}_{n-1}\right)\succ\left(a;\tilde{D}_n\right)\succ(a;D)$ for all n and $\left(\hat{D}'_n\right)_{n=1,2...}$ with $\lim_{n\to\infty}f_{\hat{D}_n}=f_D$ such that $(a;D)\succ\left(a;\hat{D}_n\right)\succ\left(a;\hat{D}_{n-1}\right)$ for all n. By Axiom 3, we have that for all n:

$$\left(a;\tilde{D}_n\circ D'\right)\succ(a;D)\sim(a;D')\succ\left(a;\hat{D}_n\circ D'\right).$$

Hence, for all n and all $\tilde{h}_n\in\mathcal{H}_a\left(\tilde{D}_n\circ D'\right)$, $\hat{h}_n\in\mathcal{H}_a\left(\hat{D}_n\circ D'\right)$, we have

$$v_a\cdot\tilde{h}_n>v_a\cdot h_D=v_a\cdot h_{D'}>v_a\cdot\hat{h}_n.$$

Furthermore, since

$$\lim_{n\to\infty}\tilde{D}_n\circ D'=\lim_{n\to\infty}\hat{D}_n\circ D'=D\circ D',$$
$$\lim_{n\to\infty}v_a\cdot\tilde{h}_n=\lim_{n\to\infty}v_a\cdot\hat{h}_n=v_a\cdot h_{D\circ D'}$$

for some (and hence, for all) $h_{D\circ D'}\in\mathcal{H}_a(D\circ D')$. This implies:

$$\lim_{n\to\infty}v_a\cdot\tilde{h}_n=\lim_{n\to\infty}v_a\cdot\hat{h}_n=v_a\cdot h_{D\circ D'}=v_a\cdot h_D=v_a\cdot h_{D'},$$

or, according to the definition of $\mathcal{H}_a$,

$$(a;D)\sim(a;D\circ D')\sim(a;D')\,.\blacksquare$$

Lemma 6.4 *Fix an $a \in A$. Under Axioms 1 – 5, there exist a probability correspondence $\hat{P}_a : C \rightrightarrows \Delta^{|R|-1}$, a similarity function $s_a : A \times R \to \mathbb{R}^+ \setminus \{0\}$ and a utility function over outcomes $v_a : R \to \mathbb{R}$, such that $\succsim$ on $\{a\} \times \mathbb{D}$ can be represented by:*

$$V_a(D) = \frac{\sum_{r \in R} v_a(r) \sum_{c \in C} \hat{p}_a^c(r) s_a(c) f_D(c)}{\sum_{c \in C} s_a(c) f_D(c)}.$$

Here, $f_D(c)$ is the frequency with which case c is observed in D and $\hat{p}_a^c$ is any probability distribution over outcomes in the image of $\hat{P}_a(c)$. The utility function v is unique up to an affine-linear transformation, the similarity function s_a is unique up to a multiplication by a positive number and there is a unique probability correspondence $\hat{P}_a$ such that $\hat{P}_a(c)$ is maximal with respect to set inclusion when $c = (a'; r)$ with $a' \neq a$ and satisfies $\hat{P}_a(a; r) = \delta_r$.

<u>Proof of Lemma 6.4</u>:

Construct the function v_a as in Lemma 6.1. Lemma 6.2 implies the existence of a convex and compact correspondence $\mathcal{H}_a : \mathbb{D} \rightrightarrows \Delta^{|R|-1}$ such that

$$V_a(D) = \sum_{r \in R} h(r) v_a(r) \text{ with } h(r) \in \mathcal{H}_a(D)$$

represents $\succsim$ on $\{a\} \times \mathbb{D}$. From $\mathcal{H}_a$, we now construct a new correspondence $\tilde{\mathcal{H}}_a$ in the following way: let

$$\tilde{\mathcal{H}}_a(D) = \left\{ h(r) = (f_D(a; r))_{r \in R} \right\}$$

for all $D \in \mathbb{D}_a$. For $D = (a'; r)$, $a' \neq a$, let

$$\tilde{\mathcal{H}}_a(D) = \mathcal{H}_a(D).$$

For all other data-sets, D such that $D \in \mathbb{D}^T$ for some T, let $\alpha_i(D) \in (0; 1)$, $i \in \{1...T\}$, $\sum_{i=1}^{T} \alpha_i(D) = 1$ be such that:

$$\mathcal{H}_a(D) = \sum_{i=1}^{T} \alpha_i(D) \mathcal{H}_a(c_i).$$

Such coefficients exist by an iterative application of Lemma 6.3. Using these coefficients, define

$$\tilde{\mathcal{H}}_a(D) =: \sum_{i=1}^{T} \alpha_i(D) \tilde{\mathcal{H}}_a(c_i).$$

It is easily checked that the correspondence $\tilde{\mathcal{H}}_a$ is non-empty, convex and compact and $\tilde{\mathcal{H}}_a(D) \subseteq \mathcal{H}_a(D)$ for all $D \in \mathbb{D}$. Hence, $V_a(D)$ can be written equivalently as:

$$V_a(D) = \sum_{r \in R} h(r) v_a(r) \text{ with } h(r) \in \tilde{\mathcal{H}}_a(D).$$

Furthermore, $\tilde{\mathcal{H}}_a$ inherits the property of $\mathcal{H}_a$:

$$\tilde{\mathcal{H}}_a(D \circ D') = \alpha \tilde{\mathcal{H}}_a(D) + (1 - \alpha) \tilde{\mathcal{H}}_a(D')$$

for some $\alpha \in (0; 1)$.

To prove the Lemma, we have to show that $\tilde{\mathcal{H}}_a$ can be written as

$$\tilde{\mathcal{H}}_a(D) = \left\{ \frac{\sum_{c\in C} \hat{p}_a^c(r)\, s_a(c)\, f_D(c)}{\sum_{c\in C} s_a(a_c)\, f_D(c)} \mid \hat{p}_a^c \in \hat{P}_a(c) \right\}$$

for some compact and convex subsets of $\Delta^{|R|-1}$, $\hat{P}_a(c)$ and some positive numbers $s_a(c)$.

We set

$$\hat{P}_a(c) =: \tilde{\mathcal{H}}_a(c)$$

for all $c \in C$.

As shown in EG (2007), for a given compact and convex-valued correspondence $\tilde{\mathcal{H}}_a(D)$, the existence of a similarity function unique up to a multiplication by a positive number is guaranteed under the following conditions:

1. *(Invariance)* $\tilde{\mathcal{H}}_a(D)$ depends only on the frequency f_D and the length of D.

2. *(Concatenation)* Consider a data-set $\hat{D} \in \mathbb{D}$ with $\left|\hat{D}\right| = T$ and, for some $n \in \mathbb{Z}_+$, let $D_1 ... D_n \in \mathbb{D}^T$ be such that $D_1 \circ ... \circ D_n = \hat{D}^n$. Then, there exists a vector $(\lambda_1 ... \lambda_{n-1}) \in int(\Delta^{n-1})$ such that, for every $k \in \mathbb{Z}^+$,

$$\sum_{i=1}^{n} \lambda_i \tilde{\mathcal{H}}_a(D_i^k) = \tilde{\mathcal{H}}_a\left(\hat{D}^k\right).$$

3. *(Linear Independence)* For every $T \in \mathbb{Z}^+$, the data-sets $(c_1)^T$,...,$\left(c_{|C|}\right)^T$ satisfy the following condition:

   There are at least three distinct i, j, $k \in \{1...|C|\}$, such that $\tilde{\mathcal{H}}_a\left((c_i)^T\right)$, $\tilde{\mathcal{H}}_a\left((c_j)^T\right)$ and $\tilde{\mathcal{H}}_a\left((c_k)^T\right)$ are:

   – either singletons

$$\tilde{\mathcal{H}}_a\left((c_m)^T\right) = \left\{h_a\left((c_m)^T\right)\right\} \text{ for } m \in \{i; j; k\}$$

   and $h_a\left((c_i)^T\right)$, $h_a\left((c_j)^T\right)$ and $h_a\left((c_k)^T\right)$ are non-collinear,

– or polyhedra with a non-empty interior such that no three of their extreme points are collinear.

Condition 1, Invariance, is satisfied by Axiom 2, with the additional property that under Axiom 3, $\tilde{\mathcal{H}}_a(D)$ does not depend on the length of the data-set D, but only on its frequency, f_D, Condition 2, Concatenation, is implied by Lemma 6.3. To see this, take sets $D_1 ... D_n$ and $\hat{D}$ as in the statement of condition 2 and apply Axiom 3 iteratively:

$$\tilde{\mathcal{H}}_a(D_1 \circ D_2) = \alpha_1 \tilde{\mathcal{H}}_a(D_1) + (1-\alpha_1)\tilde{\mathcal{H}}_a(D_2)$$

$$\begin{aligned} \tilde{\mathcal{H}}_a(D_1 \circ D_2 \circ D_3) &= \alpha_2 \tilde{\mathcal{H}}_a(D_1 \circ D_2) + (1-\alpha_2)\tilde{\mathcal{H}}_a(D_3) = \\ &= \alpha_2\left[\alpha_1 \tilde{\mathcal{H}}_a(D_1) + (1-\alpha_1)\tilde{\mathcal{H}}_a(D_2)\right] + (1-\alpha_2)\tilde{\mathcal{H}}_a(D_3) = \\ &= \alpha_1\alpha_2 \tilde{\mathcal{H}}_a(D_1) + (1-\alpha_1)\alpha_2\tilde{\mathcal{H}}_a(D_2) + (1-\alpha_2)\tilde{\mathcal{H}}_a(D_3) \end{aligned}$$

$$\ldots$$

$$\begin{aligned} \tilde{\mathcal{H}}_a(D_1 \circ \ldots \circ D_n) &= \prod_{i=1}^{n-1}\alpha_i \tilde{\mathcal{H}}_a(D_1) + (1-\alpha_1)\prod_{i=2}^{n-1}\alpha_i\tilde{\mathcal{H}}_a(D_2) \\ &\quad + (1-\alpha_2)\prod_{i=3}^{n-1}\alpha_i\tilde{\mathcal{H}}_a(D_2) \ldots + (1-\alpha_{n-1})\tilde{\mathcal{H}}_a(D_n) \end{aligned}$$

Set $\lambda_1 =: \prod_{i=1}^{n-1}\alpha_i$, $\lambda_2 =: (1-\alpha_1)\prod_{i=2}^{n-1}\alpha_i \ldots$ $\lambda_n =: (1-\alpha_{n-1})$ and note that $\sum_{i=1}^{n}\lambda_i = 1$ and $\lambda_i \in (0;1)$ for all $i \in \{1 \ldots n\}$.

$$\tilde{\mathcal{H}}_a(D_1 \circ \ldots \circ D_n) = \tilde{\mathcal{H}}_a\left(\hat{D}^n\right),$$

but by Lemma 3.1, for every $n \in \mathbb{Z}^+ \setminus \{0\}$, we have that

$$f_{\hat{D}^n} \sim_a f_{\hat{D}},$$

and hence

$$\tilde{\mathcal{H}}_a\left(\hat{D}\right) = \tilde{\mathcal{H}}_a\left(\hat{D}^n\right) = \sum_{i=1}^{n}\lambda_i\tilde{\mathcal{H}}_a(D_i).$$

Furthermore, also by Lemma 3.1 , for any $k \in \mathbb{Z}^+ \setminus \{0\}$

$$\tilde{\mathcal{H}}_a\left(D^k\right) = \tilde{\mathcal{H}}_a(D).$$

Therefore,

$$\tilde{\mathcal{H}}_a\left(\hat{D}^k\right) = \tilde{\mathcal{H}}_a\left(D_1^k \circ \ldots \circ D_n^k\right) = \sum_{i=1}^{n}\lambda_i\tilde{\mathcal{H}}_a\left(D_i^k\right) = \sum_{i=1}^{n}\lambda_i\tilde{\mathcal{H}}_a(D_i)$$

for the same values of λ_i as above.

To show that Condition 3, Linear Independence is satisfied, consider three cases $c_1 = (a; r_1)$, $c_2 = (a; r_2)$ and $c_3 = (a; r_3)$ with three distinct outcomes r_1, r_2, r_3. $\tilde{\mathcal{H}}_a(c_1)$, $\tilde{\mathcal{H}}_a(c_2)$ and $\tilde{\mathcal{H}}_a(c_3)$ consist of single non-collinear points δ_{r_1}, δ_{r_2} and δ_{r_3}, respectively. Hence, the Linear Independence condition is satisfied.

Theorem 1 in EG (2007) then ensures the existence of a similarity function $s_a : A \times R \to \mathbb{R}^+ \setminus \{0\}$ unique up to a multiplication by a positive number. The specific values of s can be derived as in Lemma 5 in EG (2007). In particular, let $h_a : \mathbb{D} \to \Delta^{|R|-1}$ be a function such that

$h_a(D) \in \tilde{\mathcal{H}}_a(D)$ for all $D \in \mathbb{D}$ such that whenever

$$\begin{aligned}\tilde{\mathcal{H}}_a(D \circ D') &= \alpha\tilde{\mathcal{H}}_a(D) + (1-\alpha)\tilde{\mathcal{H}}_a(D'), \\ h_a(D \circ D') &= \alpha h_a(D) + (1-\alpha) h_a(D').\end{aligned}$$

Lemma A4 in EG (2007) shows how $\tilde{\mathcal{H}}_a$ can be represented as a collection of functions h_a satisfying this property. Define $\hat{p}_a^c =: h_a(c)$. The system of equations:

$$\frac{\sum_{c\in C}\frac{1}{|C|}s_a(c)\hat{p}_a^c}{\sum_{c\in C}s_a(c)} = \sum_{c\in C}\alpha_c h_a(c), \tag{11}$$

where the coefficients $\alpha_c \in (0;1)$ with $\sum_{c\in C}\alpha_c = 1$ are such that for the data-set D^C, in which each case in C is observed exactly once,

$$\tilde{\mathcal{H}}_a\left(D^C\right) = \sum_{c\in C}\alpha_c\tilde{\mathcal{H}}_a(c).$$

The proof in EG (2007) demonstrates that a solution to the equation exists, is unique up to a multiplication by a positive number and does not depend on the choice of the function h_a as long as it satisfies the condition above. Hence, a similarity function with the desired properties exists and we can write

$$\tilde{\mathcal{H}}_a(D) = \left\{\frac{\sum_{c\in C}\hat{p}_a^c s_a(c) f_D(c)}{\sum_{c\in C}s_a(c) f_D(c)} \mid \hat{p}_a^c \in \hat{P}_a(c)\right\}$$

for all $D \in \mathbb{D}$, where $\hat{P}_a(c) = \tilde{\mathcal{H}}_a(c)$, thus leading to the desired representation.■

The next lemma shows that under Axiom 6, the similarity function derived above is independent of outcomes.

Lemma 6.5 *Under Axioms 1- 6, the similarity weights derived in Lemma 6.4, $s_a(a;r)$ are independent of r. Hence, for any $a \in A$,*

$$V_a(D) = \frac{\sum_{r\in R}v_a(r)\sum_{c\in C}\hat{p}_a^c(r)\,s_a(a^c)\,f_D(c)}{\sum_{c\in C}s_a(c)\,f_D(c)},$$

where a^c denotes the action chosen in case c, represents $\succsim$ restricted to $\{a\} \times \mathbb{D}$.

Proof of Lemma 6.5:

For any two cases $(a_1;r_1)$ and $(a_2;r_2)$, $s_a(a_1;r_1)$ and $s_a(a_2;r_2)$ satisfy:

$$\begin{aligned}&\frac{s_a(a_1;r_1)\,\hat{p}_a(a_1;r_1) + s_a(a_2;r_2)\,\hat{p}_a(a_2;r_2)}{s_a(a_1;r_1) + s_a(a_2;r_2)} = \\ &= \tilde{\mathcal{H}}_a((a_1;r_1);(a_2;r_2)) = \\ &= \alpha\tilde{\mathcal{H}}_a(a_1;r_1) + (1-\alpha)\tilde{\mathcal{H}}_a(a_2;r_2)\end{aligned} \tag{12}$$

for some $\alpha \in (0;1)$. To ensure that s_a can be written only as a function on A, we have to show that for any outcomes $r_1' \neq r_1$ and $r_2' \neq r_2$, the same similarity values can be applied, i.e.,

$$\begin{aligned}\frac{s_a(a_1;r_1)\,\hat{p}_a(a_1;r_1') + s_a(a_2;r_2)\,\hat{p}_a(a_2;r_2')}{s_a(a_1;r_1)+s_a(a_2;r_2)} &= \\ = \tilde{\mathcal{H}}_a\left((a_1;r_1');(a_2;r_2')\right) &= \\ = \alpha'\tilde{\mathcal{H}}_a(a_1;r_1') + (1-\alpha')\,\tilde{\mathcal{H}}_a(a_2;r_2')&.\end{aligned} \tag{13}$$

This would be immediately implied, if we could show that $\alpha = \alpha'$, or, more generally that for any D, $\hat{D} \in \mathbb{D}^T_{a'}$ and D', $\hat{D}' \in \mathbb{D}^T_{a''}$,

$$\tilde{\mathcal{H}}_a(D \circ D') = \alpha\tilde{\mathcal{H}}_a(D) + (1-\alpha)\,\tilde{\mathcal{H}}_a(D')$$

if and only if

$$\tilde{\mathcal{H}}_a\left(\hat{D}\circ\hat{D}'\right) = \alpha\tilde{\mathcal{H}}_a\left(\hat{D}\right) + (1-\alpha)\,\tilde{\mathcal{H}}_a\left(\hat{D}'\right).$$

To show this, we write Axiom 6 in terms of frequencies: let f_1, f_2, $\hat{f}_1$ and $\hat{f}_2$ denote the frequencies of D_1, D_2, $\hat{D}_1$ and $\hat{D}_2$, respectively and let $(a;D) \succsim (a;D')$ and $\left(a;\hat{D}\right) \underset{(\precsim)}{\succsim} \left(a;\hat{D}'\right)$. Then, for $\alpha \in (0;1)\cap\mathbb{Q}$

$$\alpha f_1 + (1-\alpha)\,f_2 \in \mathcal{F}_a(D\circ D';\succ) \text{ for all } f_1 \in \mathcal{F}_a(D;\succ) \text{ and all } f_2 \in \mathcal{F}_a(D';\succ)$$

holds if and only if

$$\alpha\hat{f}_1 + (1-\alpha)\,\hat{f}_2 \in \mathcal{F}_a\left(\hat{D}\circ\hat{D}';\underset{(\prec)}{\succ}\right) \text{ for all } \hat{f}_1 \in \mathcal{F}_a\left(\hat{D};\underset{(\prec)}{\succ}\right) \text{ and all } \hat{f}_2 \in \mathcal{F}_a\left(\hat{D}';\underset{(\prec)}{\succ}\right)$$

and

$$\alpha f_1 + (1-\alpha)\,f_2 \in \mathcal{F}_a(D\circ D';\prec) \text{ for all } f_1 \in \mathcal{F}_a(D;\prec) \text{ and all } f_2 \in \mathcal{F}_a(D';\prec)$$

holds if and only if

$$\alpha\hat{f}_1 + (1-\alpha)\,\hat{f}_2 \in \mathcal{F}_a\left(\hat{D}\circ\hat{D}';\underset{(\succ)}{\prec}\right) \text{ for all } \hat{f}_1 \in \mathcal{F}_a\left(\hat{D};\underset{(\succ)}{\prec}\right) \text{ and all } \hat{f}_2 \in \mathcal{F}_a\left(\hat{D}';\underset{(\succ)}{\prec}\right).$$

Recall that

$$\begin{aligned}\inf\mathcal{F}_a(D;\succ) &= \sup\mathcal{F}_a(D;\prec) = \tilde{\mathcal{H}}_a(D)\\ \inf\mathcal{F}_a(D\circ D';\succ) &= \sup\mathcal{F}_a(D\circ D';\prec) = \tilde{\mathcal{H}}_a(D\circ D')\end{aligned}$$

Let $\bar{\alpha}$ be such that:

$$\tilde{\mathcal{H}}_a(D\circ D') = \bar{\alpha}\tilde{\mathcal{H}}_a(D) + (1-\bar{\alpha})\,\tilde{\mathcal{H}}_a(D')$$

and $\hat{\alpha}$ be such that:

$$\tilde{\mathcal{H}}_a\left(\hat{D}\circ\hat{D}'\right) = \hat{\alpha}\tilde{\mathcal{H}}_a\left(\hat{D}\right) + (1-\hat{\alpha})\,\tilde{\mathcal{H}}_a\left(\hat{D}'\right).$$

Such $\bar{\alpha}$ and $\hat{\alpha} \in (0;1)$ exist according to Lemma 6.3. Since $(a;D) \succsim (a;D')$,

$$\alpha f_1 + (1-\alpha)\,f_2 \in \mathcal{F}_a(D\circ D';\succ) \text{ for all } f_1 \in \mathcal{F}_a(D;\succ) \text{ and all } f_2 \in \mathcal{F}_a(D';\succ)$$

implies $\alpha > \bar{\alpha}$. For $\left(a; \hat{D}\right) \underset{(\precsim)}{\succsim} \left(a; \hat{D}'\right)$, we have

$\alpha \hat{f}_1 + (1-\alpha) \hat{f}_2 \in \mathcal{F}_a \left(\hat{D} \circ \hat{D}'; \underset{(\prec)}{\succ}\right)$ for all $\hat{f}_1 \in \mathcal{F}_a \left(\hat{D}; \underset{(\prec)}{\succ}\right)$ and all $\hat{f}_2 \in \mathcal{F}_a \left(\hat{D}'; \underset{(\prec)}{\succ}\right)$,

only if $\alpha \underset{(<)}{>} \hat{\alpha}$. According to Axiom 6, these two statements are equivalent, hence $\alpha > \bar{\alpha}$ holds if and only if $\alpha > \hat{\alpha}$, implying that $\hat{\alpha} = \bar{\alpha}$.

Hence, the coefficients α and α' in expressions (12) and (13) are equal, implying that

$$s(a_1; r_1) = s(a_1; r_1')$$

for all $a_1 \in A$ and all $r_1, r_1' \in R$. Similarity, therefore, does not depend on the outcome and can be written as:

$$s_a : A \to \mathbb{R}. \blacksquare$$

Lemma 6.6 *For $a \in A$, let $v_a : R \to \mathbb{R}$ be the family of utility functions over outcomes, $s_a : A \to \mathbb{R}$ be the family of similarity functions and $\hat{P}_a : C \rightrightarrows \Delta^{|R|-1}$ be the family of probability correspondences derived in Lemmas 6.4 and 6.5. Under Axioms 1-7, there exist positive affine-linear transformations of $(v_a)_{a \in A}$, $(\tilde{v}_a)_{a \in A}$ with*

$$\tilde{v}_a = A_a v_a + B_a \ (A_a > 0,\ B_a \in \mathbb{R})$$

such that

$$\tilde{v}_a(r) = v(r)$$

for all $a \in A$, and all $r \in R$ and such that

$$V(a; D) = \frac{\sum_{r \in R} v(r) \sum_{c \in C} \hat{p}_a^c(r)\, s_a(a^c)\, f_D(c)}{\sum_{c \in C} s_a(c)\, f_D(c)}$$

represents $\succsim$ on $A \times \mathbb{D}$.

Proof of Lemma 6.6:

According to Lemma 6.1, all v_a's are unique up to a positive affine-linear transformation. Hence, $V(a; D)$ will represent preferences constrained to the set $\{a\} \times \mathbb{D}$ for any $\tilde{v}_a = A_a v_a + B_a$ with $A_a > 0$ and $B_a \in \mathbb{R}$. According to Axiom 7,

$$(a; (a; r)) \sim (a'; (a'; r))$$

for all a, $a' \in A$ and all $r \in R$. Hence, we rescale all v_a so that:

$$\tilde{v}_a(\bar{r}) = 1$$

and

$$\tilde{v}_a(\underline{r}) = 0,$$

where $\bar{r} = \bar{r}_a$ and $\underline{r} = \underline{r}_a$ for all $a \in A$. Note that this rescaling can be done in a unique way and that it implies:

$$V(a;(a;r)) = V(a';(a';r))$$

for all a, $a' \in A$ and all $r \in R$. Hence, it remains to show that $V(a;D)$ indeed represents $\succsim$. Take two arbitrary action-data-set-pairs: $(a;D)$ and $(a';D')$. Suppose that

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

Let $h \in \tilde{\mathcal{H}}_a(D)$ and $h' \in \tilde{\mathcal{H}}_{a'}(D')$ and suppose that $\tilde{D} \in \mathbb{D}_a$ with frequency $f_{\tilde{D}}(a;r) = h(r)$ for all $r \in R$ and $\tilde{D}' \in \mathbb{D}_{a'}$ with frequency $f_{\tilde{D}'}(a';r) = h'(r)$ for all $r \in R$ exist (i.e., h and h' are rational-valued). Then,

$$\begin{aligned}(a;D) &\sim \left(a;\tilde{D}\right)\\ (a';D') &\sim \left(a';\tilde{D}'\right)\end{aligned}$$

Let $\hat{D} \in \mathbb{D}_{a'}$ have the property that:

$$f_{\hat{D}}(a';r) = f_{\tilde{D}}(a;r) \text{ for all } r \in R.$$

We now show that $\left(a';\hat{D}\right) \sim \left(a;\tilde{D}\right)$. We do this by an induction argument on the supports of $f_{\tilde{D}}$ and $f_{\hat{D}}$. If $|supp(f_{\hat{D}})| = 1$, the statement immediately follows from Axiom 7. Assume that the statement is true for all $f_{\tilde{D}}$ and $f_{\hat{D}}$ such that $|supp(f_{\hat{D}})| \leq N-1$. Now let $|supp(f_{\hat{D}})| = N$ and pick an $\hat{r} \in R$ such that $f_{\hat{D}}(a';\hat{r}) > 0$. Let D_1 and D_1' be two data-sets of equal length with frequencies $\frac{f_{\tilde{D}}(a;r)}{\sum_{r\neq\hat{r}} f_{\tilde{D}}(a;r)}$ and $\frac{f_{\hat{D}}(a';r)}{\sum_{r\neq\hat{r}} f_{\hat{D}}(a';r)}$ for all $r \neq \hat{r}$, respectively. Let $\frac{m}{q} = \frac{\sum_{r\neq\hat{r}} f_{\hat{D}}(a';r)}{f_{\hat{D}}(a';\hat{r})}$ and note that both $\hat{D}$ and $\tilde{D}$ can be represented as concatenations using such D_1 and D_1':

$$\begin{aligned}\tilde{D} &= (D_1)^{pm} \circ (a;\hat{r})^{pq}\\ \hat{D} &= (D_1')^{km} \circ (a';\hat{r})^{kq}\end{aligned}$$

for some natural numbers p and k We know that $(a;(a;\hat{r})) \sim (a';(a';\hat{r}))$ and, since $|supp(f_{D_1})| = \left|supp\left(f_{D_1'}\right)\right| = N-1$, by the induction hypotheses, we have

$$(a;D_1) \sim (a';D_1').$$

Hence, by Axiom 5

$$(a;(D_1)^m \circ (a;\hat{r})^q) \sim \left(a';(D_1')^m \circ (a';\hat{r})^q\right),$$

or, according to Lemma 3.1, $\left(a;\tilde{D}\right) \sim \left(a;\hat{D}\right)$.

It follows that for the sets D, $\hat{D}$ and $\tilde{D}$ chosen above, we have:

$$(a;D) \sim \left(a;\tilde{D}\right) \sim \left(a';\hat{D}\right).$$

Hence, the comparison between $(a; D)$ and $(a'; D')$ reduces to the one between $\left(a'; \hat{D}\right)$ and $\left(a'; \tilde{D}'\right)$ and we have

$$\left(a'; \hat{D}\right) \succsim \left(a'; \tilde{D}'\right).$$

We already know from Lemma 6.5 that these preferences are represented by

$$\begin{aligned} V_{a'}\left(\hat{D}\right) &= \frac{\sum_{r\in R} v_{a'}(r) \sum_{c\in C} f_{\hat{D}}(c)\, s(a'; a_c)\, \hat{p}^c_{a'}(r)}{\sum_{c\in C} f_{\hat{D}}(c)\, s(a'; a_c)} \\ &\geq \frac{\sum_{r\in R} v_{a'}(r) \sum_{c\in C} f_{\tilde{D}'}(c)\, s(a'; a_c)\, \hat{p}^c_{a'}(r)}{\sum_{c\in C} f_{\tilde{D}'}(c)\, s(a'; a_c)} = V_{a'}\left(\tilde{D}'\right) \end{aligned}$$

Since $v_{a'}(r) = v(r)$ and since $\hat{D}$ and $\tilde{D}' \in \mathbb{D}_{a'}$, this expression reduces to:

$$\sum_{r\in R} v(r)\, f_{\hat{D}}(a'; r) \geq \sum_{r\in R} v(r)\, f_{\tilde{D}'}(a'; r).$$

By construction,

$$\sum_{r\in R} v(r)\, f_{\hat{D}}(a'; r) = \sum_{r\in R} v(r)\, f_{\tilde{D}}(a; r)$$

and, by Lemma 6.5,

$$\begin{aligned} \sum_{r\in R} v(r)\, f_{\tilde{D}}(a; r) &= \frac{\sum_{r\in R} v(r) \sum_{c\in C} f_D(c)\, s(a; a_c)\, \hat{p}^c_a(r)}{\sum_{c\in C} f_D(c)\, s(a; a_c)} \\ \sum_{r\in R} v(r)\, f_{\tilde{D}'}(a; r) &= \frac{\sum_{r\in R} v(r) \sum_{c\in C} f_{D'}(c)\, s(a'; a_c)\, \hat{p}^c_{a'}(r)}{\sum_{c\in C} f_{D'}(c)\, s(a'; a_c)}. \end{aligned}$$

We, therefore, conclude that

$$\begin{aligned} V(a; D) &= \frac{\sum_{r\in R} v(r) \sum_{c\in C} f_D(c)\, s(a; a_c)\, \hat{p}^c_a(r)}{\sum_{c\in C} f_D(c)\, s(a; a_c)} \\ &\geq \frac{\sum_{r\in R} v(r) \sum_{c\in C} f_{D'}(c)\, s(a'; a_c)\, \hat{p}^c_{a'}(r)}{\sum_{c\in C} f_{D'}(c)\, s(a'; a_c)} = V(a'; D'). \end{aligned}$$

Similarly, starting with the assumption that $V(a; D) \geq V(a'; D')$, and repeating the same arguments in reverse order, one can show that $(a; D) \succsim (a'; D')$. For the case, in which $\tilde{D}$ and $\tilde{D}'$ as defined above do not exist, we can choose sequences of data-sets which approximate the corresponding frequencies $h \in \tilde{\mathcal{H}}_a(D)$ and $h' \in \tilde{\mathcal{H}}_{a'}(D')$ and using the continuity assumption in Axiom 3 show that the same result applies. Hence,

$$V(a; D) = \frac{\sum_{r\in R} v(r) \sum_{c\in C} f_D(c)\, s(a; a_c)\, \hat{p}^c_a(r)}{\sum_{c\in C} f_D(c)\, s(a; a_c)}$$

represents $\succsim$ on $A \times \mathbb{D}$.■

7 References

EASLEY,. D., RUSTICHINI, A. (1999). "Choice Without Beliefs", *Econometrica 67,* 1157-1184.

EICHBERGER, J., AND GUERDJIKOVA, A. (2007). "Multiple Priors and Similarity-Weighted Frequencies", *mimeo,* Cornell University.

EICHBERGER, J., AND GUERDJIKOVA, A. (2008). "Learning from Data with the Best and the Worst in Mind", *mimeo,* Cornell University.

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

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", *fortchcoming in Journal of Economic Theory.*

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

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

O'HARA, A., AND LUCE, B.R. (2003). *A Primer on Bayesian Statistics in Health Economics and Outcomes Research.* MEDTAP International, Inc.

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

Nr.	Author	Title
08-40	Florian Kutzner Tobias Vogel Peter Freytag Klaus Fiedler	Pseudocontingencies in stereotype formation: extending illusory correlations
08-39	Wendelin Schnedler Adam Dominiak	Uncertainty Aversion and Preference for Randomization
08-38	Susanne Abele Sandra I. Vaughan-Parsons Garold Stasser	Information Flow and Influence during Collective Search, Discussion, and Choice
08-37	Peter Freytag Tobias Vogel Florian Kutzner Klaus Fiedler	The reproduction of base-rates promotes pseudocontingencies
08-36	Klaus Fiedler Peter Freytag Thorsten Meiser	Pseudocontingencies: An Integrative Account of an Intriguing Cognitive Illusion
08-35	Alexander Unger Dagmar Stahlberg	The Influence of Self-Regulation on Decision Making
08-34	Dirk Simons Nicole Zein	Kosten aus einer asymmetrischen Informationsverteilung zwischen Abschlussprüfer und Mandant
08-33	Michael Bremert Axel Schulten	The Impact of Supervisory Board Characteristics on Firm Performance
08-32	Jürgen Eichberger Ani Guerdjikova	Case-Based Expected Utility: Preferences over Actions and Data
08-31	Jürgen Eichberger Simon Grant Jean-Philippe Lefort	Neo-additive capacities and updating
08-30	Vladimir I. Danilov Gleb A. Koshevoy Christine Lang	Equilibria with indivisible goods and package-utilities

Nr.	Author	Title
08-19	Jana Janßen Patrick A. Müller Rainer Greifeneder	A field study on the role of ease-of-retrieval in procedural justice judgments
08-18	Rainer Greifeneder Benjamin Scheibehenne	When choosing is difficult: Complexity and choice-overload
08-17	Paul Grout Wendelin Schnedler	Non-Profit Organizations in a Bureaucratic Environment
08-16	Clemens Kroneberg Isolde Heintze Guido Mehlkop	On shoplifting and tax fraud: An action-theoretic analysis of crime
08-15	Hermann Jahnke Dirk Simons	A rationale for the payback criterion
08-14	Peter Dürsch Jörg Oechssler Radovan Vadovic	Sick Pay Provision in Experimental Labor Markets
08-13	Carsten Schmidt Martin Strobel Henning Oskar Volkland	Accuracy, Certainty and Surprise - A Prediction Market on the Outcome of the 2002 FIFA World Cup
08-12	Mathias Sommer	Understanding the trends in income, consumption and
08-11	Hans Gersbach Hans Haller	Club Theory and Household Formation
08-10	Michael F. Meffert Thomas Gschwend	Strategic Voting in Multiparty Systems: A Group Experiment
08-09	Jens Wüstemann Jannis Bischof	Ausweis von Finanzinstrumenten in europäischen Bankbilanzen nach IFRS: Normative Erkenntnisse empirischer Befunde
08-08	Jürgen Eichberger David Kelsey	Are the Treasures of Game Theory Ambiguous?
08-07	Jürgen Eichberger Ani Guerdjikova	Multiple Priors as Similarity Weighted Frequencies