



**HAL**  
open science

## Cases and States \*

Ani Guerdjikova, Jürgen Eichberger

► **To cite this version:**

| Ani Guerdjikova, Jürgen Eichberger. Cases and States \*. 2023. hal-03962412

**HAL Id: hal-03962412**

**<https://hal.science/hal-03962412>**

Preprint submitted on 30 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Cases and States\*

Jürgen Eichberger<sup>†</sup>      Ani Guerdjikova<sup>‡</sup>

January 20, 2023

## Abstract

In this paper, we provide a novel framework for decision making under uncertainty based on information available in the form of a data set of cases. A case contains information about an action taken, an outcomes obtained, and other circumstances that were recorded with the action and the outcome. The set of actions, the set of outcomes and the set of possibly relevant recorded characteristics are derived from the cases in the data set. The information from the data set induces a belief function over outcomes for each action. From a decision maker's preferences over belief functions one can derive a representation evaluating outcomes according to the  $\alpha$ -max min criterion. New data affects behavioral parameters, such as awareness, ambiguity and ambiguity attitude, and may suggest a classifications of data into states.

**Keywords**      partial information, case-based decisions, data, objective ambiguity, subjective ambiguity attitudes

**JEL Classification: D81**

---

\*We would like to thank Ithaca Gilboa and Stefania Mirandi for helpful discussions in the early stages of this paper. We also benefited from comments by Adam Dominiak, Edi Karni, John Quah, Jacob Sagi, and Marie-Louise Vierø. Moreover, we gratefully acknowledge numerous helpful comments from the audiences when we presented the paper at RUD 2021 (online), a Research Seminar 2022 (UGA, Grenoble), at the Theoretischen Ausschuss 2022 (Berlin), T-DEA 2022 (Paris), the Annual Meeting of Verein für Socialpolitik 2022 (Basel), and, in particular, at the Conference in honor of Adi Karni 2022 (Johns Hopkins).

<sup>†</sup>Alfred Weber Institute, Department of Economics, University of Heidelberg, Bergheimer Straße 58, 69115 Heidelberg, Germany. Email: juergen.eichberger@awi.uni-heidelberg.de.

<sup>‡</sup>Grenoble Applied Economics Lab, University of Grenoble Alpes, FRANCE. Email: ani.guerdjikova@univ-grenoble-alpes.fr.

# 1 Introduction

Decision theory under uncertainty in the tradition of Savage (1954) is purely subjective. Uncertainty concerns unknown states of the world and probabilities are derived from a subjective preference order over state-contingent outcomes (acts). The set of states are exogenously given and may differ for different decision makers. In economic models, one often reasonably assumes that all agents agree on the state space, possibly because all observe the same data. Such an assumption is however hardly ever made explicit.

In this paper, we suggest a fundamentally different approach to modeling uncertainty. In the spirit of Gilboa and Schmeidler (2001), we consider a data set of cases observed in the past as basis of our theory. Cases in the data set will record actions, outcomes and circumstances of decisions observed in the past. For each action, this data will induce a mass distribution (belief function) over sets of outcome distributions. Subjective preferences over such mass distributions will induce a representation of preferences over actions as in Jaffray (1989). In general, data allows only for an incomplete description of the outcomes of an action leaving room for a subjective evaluation of ambiguity. If agents use the same data set their models of uncertainty will be anchored in the same fundamental data-based framework. In general, however, data will not allow for a description of states in the sense of Savage (1954) that resolve all uncertainty regarding the outcomes of actions. Such a detailed description of circumstances may only arise with sufficiently large and complete data sets.

To illustrate the approach suggested in this paper, consider the *portfolio choice* problem of an investor who chooses a portfolio of two assets  $a = (a^1, a^2)$ , and assume that the investor has no other information regarding the (monetary) payoffs of these assets than those observed in a data set  $D$  of cases,  $t = 1, \dots, T$ . A case  $(a_t, r_t, x_t)$  registers the quantity of an *asset* purchased  $a_t$ , the *outcome* obtained  $r_t$ ,<sup>1</sup> and some *characteristics* of the decision environment  $x_t$ . For example, a characteristic feature recorded in the data of a portfolio choice problem could be the *inflationary environment*, high inflation  $H$  or low inflation  $L$ , respectively.<sup>2</sup> Disregarding private information of the investors, we assume data in  $D$  to be equally available for all agents.

Given a (finite) set  $D$  of cases, for each action  $a$  and each characteristic  $x$  recorded in the data, a set  $\mathbf{R}_D(a, x)$  of outcome distributions can be associated.

For pairs of actions and characteristics, either a frequency distribution  $\rho(a, x)$  over outcomes  $r$  obtained with this action will be recorded in the data set  $D$  or no information will be available. In the former case, the set of outcome distributions  $\mathbf{R}_D(a, x) = \{\rho(a, x)\}$  is a singleton, in the latter case, there is complete *ambiguity*

---

<sup>1</sup>E.g., the return rate over a given period.

<sup>2</sup>The features of a decision situation that may be recorded as a characteristic may consist of either unintentionally collected data or deliberately chosen observations, e.g., reflecting a theory about the factors influencing the payoffs. We will discuss this issue in more detail below.

	characteristic, frequency	
actions	$H, f_H$	$L, f_L$
$a^1$	$\mathbf{R}_D(a^1, H)$	$\mathbf{R}_D(a^1, L)$
$a^2$	$\mathbf{R}_D(a^2, H)$	$\mathbf{R}_D(a^2, L)$

Table 1: Outcome distributions for action-characteristic pairs  $(a, x)$

regarding the outcome distribution for  $(a, x)$ . In this case, a *subjective set* of possible outcome distributions  $\mathbf{R}_D(a, x)$  can be assigned. As a default, we will assume that all outcome distributions observed in  $D$  will be considered possible.<sup>3</sup> Notice that outcome distributions observed in  $D$  may be degenerate and based on few observations.<sup>4</sup>

A table of outcome distributions such as Table 1 induces a mass distribution (belief function) for each action in the data  $D$ . Assuming subjective preferences over the set of mass distributions (belief functions) induced by the actions for a given data set, we can deduce a preference representation that combines subjective elements such as risk and ambiguity attitudes of a decision maker with objective elements, i.e., data-based elements, such as the frequency information in the given data set. Decision makers with the same data set share the data-based common objective elements.

In this paper, we provide axioms inspired by those advanced in Jaffray (1989) yielding an  $\alpha$ -max min representation:<sup>5</sup>

$$V_D(a) = \sum_{x \in X_D} f_D(x) \left[ \alpha \max_{\rho \in \mathbf{R}_D(a, x)} \sum_r u(r) \rho(r) + (1 - \alpha) \min_{\rho \in \mathbf{R}_D(a, x)} \sum_r u(r) \rho(r) \right].$$

While probabilities of outcomes are largely based on data in this representation, some probabilities will be based on a large amount of evidence, while others will be supported by few observations only.<sup>6</sup> This motivates us to extend this model by allowing for unforeseen contingencies in the spirit of Karni and Vierø (2013); Karni and Vierø (2017) in order to model a subjective response to a lack of data. The following example is inspired by Dow and da Costa Werlang (1992) and will illustrate how a lack of data may induce reasoning about other, so far unobserved contingencies, that may change the investor's behavior.

<sup>3</sup>In particular, all Dirac measures in the support of outcome distributions observed in  $D$ ,  $\mathbf{R}_D(a, x) = \{\delta_{r_t} | (a_t, r_t, x_t) \in D\}$  will be assumed to be possible. We will discuss other assumptions regarding outcome distributions in Section 2.

<sup>4</sup>Indeed, if supported by many observations, degenerate distributions  $\delta_r$  yielding a particular outcome  $r$  with probability one for some characteristic may suggest this characteristic as a state in the sense of Savage (1954).

<sup>5</sup>As in state-based decision theory, different representations of preferences can be deduced from different systems of axiom, e.g., a smooth representation could be deduced from axioms as in Eichberger and Pasichnichenko (2021).

<sup>6</sup>In Eichberger and Guerdjikova (2013), we studied this problem in the context of the case-based decision model of Gilboa and Schmeidler (2002) and Billot *et al.* (2005).

**Example.** For an explicit example, consider portfolios of two assets  $a = (s, b)$ , a stock and a corporate bond, and consider a data set  $D$  containing observations for 155 cases of trade in these assets.

- For the stock, the risky security, the data set  $D$  contains 20 cases of a high (real) payoff  $r^1$  and 50 cases of a low (default) payoff  $r^0$  in a low inflation environment, while, for high inflation, no observations are available.
- For the corporate bond  $b$ , the same payoff  $r_b$  has been observed in 5 cases of high inflation and 80 cases of low inflation.<sup>7</sup>

This data can be summarized in the following table.

	frequency of characteristics $\{H, L\}$	
actions	$f_H = \frac{5}{155}$	$f_L = \frac{150}{155}$
$s$	$\mathbf{R}_D(s, H) = \{\delta_{r^1}, \delta_{r^0}\} = \{(r_s^1, 1), (r_s^0, 1)\}$	$\mathbf{R}_D(s, L) = \{\rho(s, L)\} = \{(\frac{20}{70}, r^1; \frac{50}{70}, r^0)\}$
$b$	$\mathbf{R}_D(b, H) = \{\delta_{r_b}\} = \{(r_b, 1)\}$	$\{\mathbf{R}_D(b, L) = \{\rho(b, L)\}\} = \{\delta_{r_b}\} = \{(r_b, 1)\}$

Notice that the data in  $D$  do not provide an outcome distribution for the stock  $s$  in case of high inflation  $H$ . As a default, we consider distributions over outcomes observed for the stock in the low inflation scenario  $L$ , in particular the best and worst outcome distribution for the stock in this situation. Hence, ignorance about the return distribution for the stock in the high inflation scenario  $H$  is modeled by a set of outcome distributions  $\mathbf{R}_D(s, H) := \{\delta_{r^1}, \delta_{r^0}\}$ .<sup>8</sup> Choosing w.l.o.g. the

bond as the numeraire, suppose that the assets trade at prices  $q_S$  for the stock and  $q_B = 1$  for the bond. Moreover, suppose that short-sales of each of the two assets are restricted to  $\bar{d}$ . Assume that the decision maker takes the information provided in the data  $D$  as objective and uses the available frequencies as probabilities. For the sake of simplicity of our argument, assume that the investor is risk-neutral,  $u(r) = r$ , and if faced with missing data ambiguity-averse,  $\alpha = 0$ . Thus, a portfolio is evaluated by

$$V_D(s, b) = \left[ f_H \min_{r_s \in \{r^0, r^1\}} (sr^s + br_b) + f_L \left[ s \left( \frac{5}{7}r^0 + \frac{2}{7}r^1 \right) + br_b \right] \right].$$

The investor is assumed to choose a portfolio  $(s, b)$  that maximizes  $V_D(s, b)$  subject to a budget constraint  $q_S s + b \leq w$  and the two short-sale constraints  $s \geq -\bar{d}$  and  $b \geq -\bar{d}$ . Standard arguments, similar to those in Dow and da Costa Werlang (1992), then yield the solution:

$$(s, b) = \begin{cases} (-\bar{d}, w + q_S \bar{d}) & \text{for } q_S > \bar{Q}, \\ (0, w) & \text{for } q_S \in (\underline{Q}, \bar{Q}), \\ (\frac{w + \bar{d}}{q_S}, -\bar{d}) & \text{for } q_S < \underline{Q}, \end{cases}$$

<sup>7</sup>For notational simplicity, we assume that only two outcomes  $r^1, r^0$  have been recorded for the stock and a single outcome  $r_b$  for the corporate bond,  $r^1 > r_b > r^0$ .

<sup>8</sup>In our general approach below, we will allow for more general outcome distributions. For simplicity of the example, we will consider only the extreme distributions.

$$\text{with } \underline{Q} = \frac{f_H r^0 + f_L \left(\frac{5}{7}r^0 + \frac{2}{7}r^1\right)}{r^b} \text{ and } \overline{Q} = \frac{f_H r^1 + f_L \left(\frac{5}{7}r^0 + \frac{2}{7}r^1\right)}{r^b}$$

Given the lack of data for the high inflation scenario, a decision maker may wonder whether a hyperinflation - a characteristic so far not observed in the data - might not affect bond rates negatively, due to increased corporate defaults or increased nominal costs that cannot be passed on to consumers. The investor may take into account such concerns about the outcomes in situations not covered by the data in  $D$  as "other" characteristics  $o$ , to which she assigns a subjective weight of  $\gamma$ , a *degree of unawareness*.

The weight assigned to the information in the original data  $D$  is thus scaled down by  $(1 - \gamma)$ , the degree of confidence in the data.

	frequency of characteristics $\{H, L\}$ and $o$		
<b>actions</b>	$(1 - \gamma) f_H = (1 - \gamma) \frac{5}{155}$	$(1 - \gamma) f_L = (1 - \gamma) \frac{150}{155}$	$\gamma$
$s$	$\mathbf{R}_D(s, H) = \{\delta_{r^1}, \delta_{r^0}\}$	$\mathbf{R}_D(s, L) = \left\{\left(\frac{20}{70}, r^1; \frac{50}{70}, r^0\right)\right\}$	$\mathbf{R}_D(s, o) = \{\delta_{r^1}, \delta_{r^0}\}$
$b$	$\mathbf{R}_D(b, H) = \{\delta_{r_b}\}$	$\mathbf{R}_D(b, L) = \{\delta_{r_b}\}$	$\mathbf{R}_D(b, o) = \{\delta_{r_b}, \delta_{\tilde{r}}\}$

With a degree of unawareness  $\gamma$ , suppose that the investor considers the same set of outcome distributions  $\{\delta_{r^1}, \delta_{r^0}\}$  for the stock given the characteristic  $o$  (hyperinflation) as for the characteristic  $H$  (high inflation), while for the corporate bond under characteristic  $o$  a second outcome distribution concentrated on a low return  $\tilde{r}$ ,  $\tilde{r} < r^0$ , is now taken into account,  $\{\delta_{r_b}, \delta_{\tilde{r}}\}$ .

Given the ambiguity regarding the bad consequences for the corporate bond in a situation of "hyperinflation", i.e., for the characteristic  $o$ , a portfolio is now evaluated as

$$V_D(s, b) = (1 - \gamma) \left[ f_H \min_{r_s \in \{r^0, r^1\}} \{sr_s + br^b\} + f_L \left[ s \left( \frac{5}{7}r^0 + \frac{2}{7}r^1 \right) + br_b \right] \right] \\ + \gamma \min_{\substack{r_s \in \{r^0, r^1\} \\ r_b \in \{r_b, \tilde{r}\}}} \{sr_s + br_b\}$$

while the budget constraint and the short-sale constraints remain the same.

If the decision maker is pessimistic ( $\alpha = 0$ ) and believes that holding corporate bonds may result in a loss if a hyperinflation occurs,<sup>9</sup> then the investor may never invest a strictly positive amount into the corporate bond. Moreover, the investor will may not want to short-sell the corporate bond either. Hence,  $b = 0$  will be optimal. Thus, the corporate bond market may collapse due to worries about a situation of hyperinflation, i.e., an unforeseen contingency that is not supported by the data available in  $D$ .

The example illustrates how a decision-maker's concern about "other" unforeseen characteristics, i.e., characteristics not observed in the data, may change the evaluation of assets and may make arbitrage based solely on observed data risky. Notice that, in contrast to an exogenous prespecified state space, a data set determines both a set of possible returns for the assets and the likelihood of the characteristics. Hence, missing or incomplete data may justify a decision maker's

<sup>9</sup>More precisely, if unawareness  $\gamma$  is sufficiently large and if a potential loss  $\tilde{r}$  in case of a hyperinflation is sufficiently high, i.e., if  $(1 - \gamma) r_b + \gamma \tilde{r} < r_b$  holds, then  $b = 0$  will be optimal..

consideration of other characteristics of a decision situation influencing the outcomes of an action. Therefore, the axiomatization of preferences over outcome distributions that are partially supported by objective, data-based evidence will not only allow for ambiguity attitudes and risk attitudes as subjective properties of the preference representation but also subjective concerns about unforeseen contingencies.

In general, factors influencing outcome distributions of actions are restricted to circumstances recorded in the data. However, circumstances considered in the data do not resolve uncertainty completely as states in the framework of Savage (1954) do. Hence, decision makers are likely to be aware of “other” circumstances that may be relevant for determining the outcome distributions of actions. “Other”, so far unobserved, characteristics may influence the choice of actions and may induce exploratory actions generating cases for which the so far unobserved characteristics may occur.<sup>10</sup>

In this paper, we will derive a subjective probability  $\gamma$  for the possibility that an “other” so far unobserved characteristic may occur from preferences of the decision maker over actions. The set of outcome distributions of an action in case of other characteristics reflects on the one hand the ignorance of the decision maker due to a lack of data-based information but, on the other hand, it may also reflect “theories” in the sense of Karni (2022) about possible outcome distributions.

Notice that both the probability  $\gamma$  and the set of outcome distributions for actions in case of an unforeseen “other” characteristic are subjective for a given data set at a given point in time, but may become data-based objective information at a later point in time, once data with new characteristics have become available. The careful distinction between objective, i.e., data-based, information and subjective theories based on preferences provides a framework for studying the evolution of subjective information in the light of new data in which Bayesian learning appears as a special but well-defined case.

For a given decision situation, we take the data set as given. Over time and in response to the actions chosen in a period the data set will change. New data may simply increase the number of observations of cases already considered. In contrast to the state-based approach however, new characteristics may be recorded and appear as potential factors influencing the outcome distributions of actions.

A special feature of the decision model in this paper is the learning process generated by changing data. Learning of new categories of characteristics induces a permanent revision process as new actions, outcomes, or circumstances appear. Pairs of action and characteristics induce outcome distributions similar to theories introduced in Karni (2022). New data of cases may make some of the action-circumstance pairs less relevant leading to vanishing unconfirmed theories as in the case of theories. In contrast, however, new cases may suggest new action-circumstance pairs thus inducing potential new states. Delaying choices at the expense of outcome distributions provides a natural notion of costly exploration. Characteristics may suggest new explanation in cases of unforeseen contingencies.

---

<sup>10</sup>This is similar to unimaginable consequences or actions of which the decision maker is unaware of but that determine new conceivable states in Karni and Vierø (2017, p. 304).

## 1.1 Relation to the Savage framework

Decision theory in the tradition of Savage (1954) considers a set of states of the world and an independent set of outcomes as primitive concepts of the theory. Actions, also called “state-contingent outcomes” or “acts”, map states to outcomes. Observing a state is assumed to resolve all uncertainty regarding the outcome of an action. The subjective expected utility theory of Savage (1954) derives an expected utility representation with a subjective probability distribution over states and a subjective utility function evaluating outcomes from the subjective preferences of a decision maker over state-contingent outcomes.

In this setup, the set of states, the set of outcomes as well as the mapping from states to outcomes have to be exogenous specified by the analyst while the probability distribution over states as well as the evaluation of outcomes are deduced from preferences alone. Factual information regarding the set of states and the outcomes of actions as well as the frequencies of states do not enter the description of a decision maker’s choice situation. Experimental evidence suggests, however, that information about the set of states and the frequency of observations influences choice behavior.<sup>11</sup>

In a recent paper, Gilboa *et al.* (2020) study the tension between the requirement for states to provide a complete description of all circumstances relevant for determining the outcomes of an action and the observability of actions and outcomes. Re-assessing the Savage paradigm, Gilboa *et al.* (2020) write:

“Models of decision making under uncertainty gain much of their power from the specification of states so as to resolve all uncertainty. However, this specification can undermine the presumed observability of preferences on which axiomatic theories of decision making are based. We introduce the notion of a contingency. Contingencies need not resolve all uncertainty, but preferences over functions from contingencies to outcomes are (at least in principle) observable. In sufficiently simple situations, states and contingencies coincide. In more challenging situations, the analyst must choose between sacrificing observability in order to harness the power of states that resolve all uncertainty, or preserving observability by working with contingencies.”

In our model, we will distinguish and identify from the data of cases: *(i) actions*, *(ii) outcomes*, and *(iii) characteristics* of a situation. Characteristics are factors that influence or determine the set of outcomes of an action. They can be recorded as part of legal requirements, medical history, statistical data, etc. They are observable and can be retrieved from data: they can thus serve as empirical proxies for states.

However, characteristics differ from the Savage (1954) concept of states, in three important aspects. First, a characteristic may fail to specify an outcome

---

<sup>11</sup>Indeed, one may interpret the famous paradoxes of Ellsberg (1961) as showing that partial information about the probability of events (the proportion of colors in the urns) substantially influences subjects’ choices.



for all available actions, e.g., because it has never been observed in combination with this action. Second, the set of characteristics observed in the data need not be exhaustive: it may capture only a subset of all relevant contingencies. Third, characteristics may provide a coarse description of the underlying uncertainty with a single characteristic corresponding to a set of states. Taking characteristics as a primitive for our approach means that the relevant factors for determining the outcomes of an action need not and cannot be specified ex-ante before the analysis can begin.

Thus, while data sets of cases in the spirit of Gilboa and Schmeidler (2002) provide objective information both about the structure of the uncertainty and about empirical frequencies, they typically do not uniquely identify the relevant state space.<sup>12</sup> Subjective factors will influence the decision maker’s evaluation of actions. In particular: (i) predictions about counterfactuals; (ii) perception of bounded awareness: existence of “other, yet unobserved” characteristics, or of yet unidentified but relevant categories, and (iii) attitude to the indeterminacy of predictions given such unawareness. In the first part of the paper, we derive a preference representation from axioms over the preference order for a given data set. This representation combines the objective information in the data with the subjective characteristics of the decision maker and identifies his perception of bounded awareness, as well as his attitude towards unawareness.

Changes in a data set may provide new evidence about the frequency of characteristics and outcomes. They may lead to the observations of new characteristics, similar in spirit though not in detail, to the approaches by (Karni and Vierø, 2017) and by Gilboa *et al.* (2017). They can also make the decision maker aware of the existence of factors that are relevant for the outcome of an action, but unobserved in the data. The second part of the paper describes by means of an example how these three types of learning can occur. While statistical frequency-based learning leaves the perception of the relevant state space unchanged, the discovery of new characteristics leads to an expansion of the perceived state space. The discovery of new categories can provide a refinement for an initially coarse perception of the state space. Our framework thus provides a way of modeling dynamic awareness of the state space which evolves with the available data and might approach the ideal of a Savagean state space.

## 2 Basic notation and leading examples

In this paper, we will not assume a priori known sets of actions and consequences. In contrast to most of the literature, we will derive these sets from a data set of previously observed cases. Before discussing our leading examples, however, we need to introduce a common notation.

---

<sup>12</sup>A decision situation where there is complete information about the relevant states mapping actions into outcomes is the exception rather than the rule. Lotteries or bets are the prototypes of acts in the Savage approach: they are the special case of actions describing the link between the states of a random device and a monetary outcome directly. We discuss this special case in Section 2.2.2 below.

## 2.1 Cases and states: the basic model

The primitive concept of our approach is a case  $c = (a, x, r)$  that records an *action*  $a$ , an *outcome*  $r$ , and a vector of *characteristics*  $x$  listing possibly relevant context variables. *Information* available at the point of decision making is a finite data set of cases that have been observed and recorded in previous decision situations:

$$D = \left\{ (a_n, x_n, r_n)_{n=1}^N \right\}.$$

Note that the same case  $c = (a, x, r)$  may have been observed several times.<sup>13</sup>

Given a data set of cases  $D$ , the set of observed actions is given by:

$$A_D = \{a \mid (a_n = a; x_n; r_n) \in D \text{ for some } n \in \{1 \dots N\}\}.$$

The set of observed outcomes is:

$$R_D = \{r \mid (a_n; x_n; r_n = r) \in D \text{ for some } n \in \{1 \dots N\}\}.$$

Characteristics recording the circumstances of a decision may be classified in categories. For example, a medical doctor who has recorded the case of a patient with a particular treatment usually also notes some biometric characteristics of the patient. We will refer to the type of biometric data recorded, such as blood pressure, temperature, weight, etc, as categories and to the entries in these categories as characteristics. Hence, *categories* classify characteristics. The data set  $D$  identifies a set of categories  $T$  and, for each category  $t \in T$ , the set of observed characteristics:

$$X_D^t = \{x^t \mid (a_n; x_n = (x_n^1, \dots, x_n^t = x^t, \dots, x_n^T); r_n) \in D \text{ for some } n \in \{1 \dots N\}\}.$$

The set of all characteristics is obtained as the Cartesian product of  $X_D^t$ :

$$X_D = \prod_{t=1}^T X_D^t.$$

When we refer to characteristics without mentioning a category, we mean the vector  $x \in X_D$  with components  $x^t$  for all categories  $t \in T$ .

The data set  $D$  also specifies for each characteristic  $x \in X_D$  the frequency with which this characteristic has been observed

$$f_D(x) = \frac{|\{(a_n, x_n, r_n) \in D \mid (a_n, x_n, r_n) = (a_n, x, r_n)\}|}{N}.$$

Let  $X^t$  be the set of characteristics of category  $t$  and denote by  $T$  the set of categories distinguished in the data. The set of characteristics  $X = \prod_{t \in T} X^t$  is the Cartesian product of the categories of characteristics.<sup>14</sup>

<sup>13</sup>In principle, there may be missing entries for some cases. In our basic model, however, we will assume that records of a case are complete.

<sup>14</sup>Over time, with changing data, both the set of categories  $T$  and the set of characteristics observed within a category,  $X^t$ , may change.

Typically, in a data set  $D$ , the outcome observed from an action  $a \in A_D$  in combination with characteristic  $x \in X_D$  will not be unique. Thus, we associate with a pair  $(a, x)$  the conditional frequency  $\rho_D(r | a, x)$  of an outcome  $r \in R_D$  when action  $a \in A_D$  is chosen and characteristic  $x \in X_D$  has been realized:

$$\rho_D(r | a, x) = \frac{|\{(a_n, x_n, r_n) \in D \mid (a_n, x_n, r_n) = (a, x, r)\}|}{|\{(a_n, x_n, r_n) \in D \mid (a_n, x_n) = (a, x)\}|}.$$

Denoting by  $\Delta(Z)$  the simplex of all probability distributions over a finite set  $Z$ , we have  $f_D \in \Delta(X_D)$  and  $\rho_D(\cdot | a, x) \in \Delta(R_D)$ . We will denote by  $\mathfrak{R}_D$  the set of all finite subsets of  $\Delta(R_D)$ .

*Remark 1.* We emphasize that, for a given point in time, the decision maker's information is fully summarized by the data set  $D$ , which we take as given. In particular, there is no prior information about the set of possible categories, their relevance, or the number of characteristics within each category. Over time with new data, however, new actions, new outcomes, new categories, or new characteristics may be discovered, as discussed in Section 4.

### 2.1.1 Ambiguity: Uncertainty about outcome distributions

While a state identifies the outcome of each available action, this need not be the case for a characteristic. If an action-characteristic combination  $(a, x)$  has not been observed<sup>15</sup> in  $D$ , the corresponding frequency of outcomes  $\rho_D(\cdot | a, x)$  is not well defined. The decision maker will thus have to make a subjective prediction about the outcome of  $a$  when characteristic  $x$  occurs.

Different methods can be used to arrive at such predictions: statistical methods, logical inference, analogy or similarity. In certain situations, such a method might uniquely identify the distribution of outcomes. In general, however, a set of possible distributions will obtain,<sup>16</sup> see the examples in Section 2.2.

We denote by  $\mathbf{R}_D(a, x) \subset \Delta(R_D)$  the set of possible outcome distributions the decision maker associates with  $(a, x)$ , and assume that this set is finite, i.e.  $\mathbf{R}_D(a, x) \in \mathfrak{R}_D$ . For observed  $(a, x)$ -combinations, this set is a singleton and comprises the observed frequency of outcomes  $\mathbf{R}_D(a, x) = \{\rho_D(\cdot | a, x)\}$ .

Sometimes the context of a decision situation may suggest a particular set of outcome distributions for an unobserved action-characteristic combination. The following sets of outcome distributions appear to be natural candidates for unobserved action-characteristics pairs  $(a, x)$ : (i) the (finite) set of all frequencies over outcomes in the data  $D$ ,  $\mathbf{R}_D = \cup_{(a,x) \in A_D \times X_D} \mathbf{R}_D(a, x)$ ; (ii) the set of all frequencies over outcomes observed in combination with a particular action  $a$ ,

<sup>15</sup>Missing counterfactuals may not only be a problem of small data sets  $D$ . In some decision situations, counterfactuals cannot be observed, because of practical, legal or moral constraints on actions. For instance, testing certain treatments on pregnant women might be considered too risky for the fetus and, thus, the outcomes of such treatments might never be observed.

<sup>16</sup>E.g., a non-parametric model might be only partially identified; the decision maker may decide to use the confidence interval of a parametric estimation instead of the estimate itself; there might be uncertainty about the correct analogy, etc.

$\mathbf{R}_D(a) = \cup_{x \in X_D} \mathbf{R}_D(a, x)$ ; (iii) the set of all frequencies over outcomes observed in combination with a particular characteristic  $x$ ,  $\mathbf{R}_D(x) = \cup_{a \in A_D} \mathbf{R}_D(a, x)$ . Note that the sets of outcome distributions  $\mathbf{R}_D, \mathbf{R}_D(a), \mathbf{R}_D(x)$ , are all finite subsets of  $\Delta(R_D)$ .

### 2.1.2 Awareness of unawareness: Other characteristics

While the set of Savage states is exhaustive, i.e., describes all relevant contingencies, the set of observed characteristics need not be: there might be a state which does not correspond to any of the characteristics observed in the data,  $X_D$ . It may be that such characteristics cannot occur in the context in which the data were collected or that there have not yet been sufficient observations. Recording categories of decision-relevant characteristics may make a decision maker aware of the fact that some category could contain characteristics which have not yet been observed.

We model the awareness of “other, so far unobserved, characteristics” by extending the set of characteristics  $X_D$  with a (place holder) characteristic “ $x_o$ ” (for “other characteristics”). The augmented set of characteristics is  $\hat{X}_D = X_D \cup \{x_o\}$ .

From the data set  $D$ , neither a frequency for such “other” characteristics nor an outcome distributions  $\rho_D(\cdot \mid a, x_o)$  can be deduced. A decision maker thus faces ambiguity given this lack of information and may again associate a set of distributions  $\mathbf{R}_D(a, x_o)$  with the occurrence of  $x_o$ . A natural such candidate is the set of all outcome distributions that have been observed for an action in  $D$ ,  $\mathbf{R}_D(a, x_o) = \cup_{x \in X_D} \mathbf{R}_D(a, x)$ .<sup>17</sup>

The decision maker is assumed to attribute a subjective weight, interpreted as his degree of unawareness,  $\gamma_D$  to this unobserved characteristic.  $(1 - \gamma_D)$  is then the degree of confidence assigned to the information in the data set and in particular to the frequency of observed characteristics  $f_D(x)$ . This degree of unawareness is purely subjective and will be derived from the decision maker’s preferences.<sup>18</sup>

### 2.1.3 Awareness of unawareness: Other categories

Each state uniquely identifies the outcome of an action. In contrast, as we saw above, an action  $a$  in combination with a given characteristic  $x$  can result in several distinct outcomes. In some situation, such variation can be considered as noise and the decision maker might reasonably use the observed frequency of outcomes

<sup>17</sup>In general, however, one may want to allow also for other sets of outcome distributions. Chung *et al.* (2018) provide a method of estimating the outcomes on “unknown unknowns” in the context of machine learning.

<sup>18</sup>Although unknown in the data set at a particular point in time, a sequence of data sets may reveal information about the frequency of unobserved characteristics. Over time, a decision maker may learn about the reliability of the data in  $D$  and the likelihood of observing a so far unknown characteristic. Thus, the degree of unawareness can be related to the frequency of new characteristics observed over time. We will study such learning in Section 4. See also Chung *et al.* (2018) for an econometric method for estimating the weight assigned to “unknown unknowns” in the context of classification.

$\rho_D(\cdot | a, x)$  as a unique prediction. In other contexts, the decision maker might infer that the observed variability is due to some underlying, but so far unobserved (latent) factor. He would thus be aware that the characteristic  $x$  corresponds to a set of states rather than to a single state and hypothesize the existence of a yet unobserved category, the characteristics of which would correspond to the underlying states grouped in  $x$ . This can result in ambiguity in the immediate sense of the word.<sup>19</sup> The identification of such a category would explain the variability of outcomes of  $a$  in combination with  $x$ . We discuss this type of unawareness in Section 4.3.2.

## 2.2 Leading examples

Before proceeding to the analysis of choice under uncertainty in this framework, we will illustrate our suggested approach by three examples, ranging from the case of no information about characteristics, over the classic situation of betting on an urn, to the more applied situation of a medical decision.

### 2.2.1 No information about characteristics

Consider first the special case where cases record only actions and outcomes. Hence, cases cannot be distinguished according to characteristics. We refer to such a decision situation as *no information about characteristics*.<sup>20</sup>

**Example 1** (no characteristics). Consider the data set

$$D = \{(a_1, r_1), (a_3, r_2), (a_1, r_4), (a_2, r_1), (a_3, r_4), \\ (a_2, r_1), (a_3, r_4), (a_1, r_3), (a_2, r_3), (a_3, r_2), (a_1, r_1)\}.$$

From the data, one obtains a single outcome distribution  $\rho_D(a) \in \Delta(R_D)$  for each action  $a$  as recorded in the left column of Table 2. Depending on the size and quality of the data set, a decision maker may have limited confidence in these outcome distributions and, therefore, may attach a weight  $\gamma$  to some other outcome distributions she considers possible although she is not aware of them in the sense of not having observed them in the data set. This is illustrated in the right column of Table 2.

For large data sets, frequencies of outcomes will approximate the probabilities of the outcomes. Hence, with no information about characteristics, choice over actions corresponds to the choice of a lottery over outcomes as in (von Neumann and Morgenstern, 1944). There is little room for unawareness and ambiguity if cases are recorded properly. Ambiguity can arise only due to incomplete records or small data sets (see Eichberger and Guerdjikova, 2010) and will vanish as the number of observations increases.

<sup>19</sup>Cicero writes “ex ambiguous controversial nascitur, cum res in unam sententiam scripta duas aut plures sententias significat“.(Short, 2018, p.3)

<sup>20</sup>This is similar to the early models of case-based decision making (Gilboa and Schmeidler, 2002), where cases only record actions and outcomes  $c = (a, r)$  for a given decision problem. A data set  $D$  contains information only about the frequency distributions  $\rho_D(a)$ .

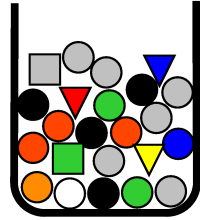
	$(1 - \gamma)$	$\gamma$
$a_1$	$\rho_D(a_1) = (\frac{2}{4}, \frac{0}{4}, \frac{1}{4}, \frac{1}{4})$	$\mathbf{R}_D(a_1, x_o)$
$a_2$	$\rho_D(a_2) = (\frac{2}{3}, \frac{0}{3}, \frac{1}{3}, \frac{0}{3})$	$\mathbf{R}_D(a_2, x_o)$
$a_3$	$\rho_D(a_3) = (\frac{0}{4}, \frac{2}{4}, \frac{0}{4}, \frac{2}{4})$	$\mathbf{R}_D(a_3, x_o)$

Table 2: No information from characteristics

### 2.2.2 Urn with unknown content

Consider an urn containing unknown objects. Sequentially, objects are drawn from the urn. For each object a list of properties (characteristics) is recorded in the data  $D$ . Such characteristics could include

- **color:** red, blue, yellow, ....
- **shape:** ball, cube, pyramid,.....
- **material:** wood, iron, glass, .....
- **weight** (grams): 20,10,50, ....



Characteristics are classified into categories: color  $c$ , shape  $s$ , material  $m$ , and weight  $w$ . A category is a set of characteristics of the same type, such as a set of colors or a set of shapes. Hence, a characteristic registered in a case may be a quadruple  $(x^c, x^s, x^m, x^w)$  indicating the color, the shape, the material, and the weight of the object drawn from the urn in this case.

Actions are bets on characteristics of the next object drawn from the urn. Outcomes are monetary payments  $r \in \mathbb{R}$ . All information of the agent is given by a data set of  $N$  observed cases:

$$D = \left\{ (a_n, x_n, r_n)_{n=1}^N \right\}.$$

**Example 2.** Table 3 shows a data set  $D = \{(c_n)_{n=1}^{20}\}$  of 20 cases with four actions,  $A_D = \{a_1, a_2, a_3, a_4\}$ , two outcomes,  $R_D = \{0, 1\}$  and a single category, the color of the objects,  $X_D = \{R, B, Y\}$ . Table 4 organizes these cases in a matrix listing the observed outcome distributions with respect to the actions and the characteristics. The first table records the first 10 cases and the second table all 20 cases. A tuple  $(x, y)$  records the frequency of outcome  $r = 0$  by  $x$  and the frequency of outcome  $r = 1$  by  $y$ . If an action-characteristic pair  $(a, x)$  has not been observed yet, then there is ambiguity about the outcome distribution with a set of possible distributions  $\mathbf{R}_D(a, x)$ . The more cases are contained in the data set the fewer cells of the matrix will be left open. We have also included a column for so far unobserved colors, that is for characteristics  $x_o$  the decision maker might be unaware of.

For a bet where the winning condition is recorded as a characteristic, e.g.  $a_2$  means “betting on  $B$ ” and  $B$  is a characteristic in  $X_D$ , it appears natural to assume that outcome distributions are concentrated, i.e.,  $\rho_D(a_2, B) = (0, 1)$  and  $\rho_D(a_2, x) = (1, 0)$  for  $x \neq B$ . If the action is, however, not specified completely

$D$	$A$	$X^1$	$R$
$c_1$	$a_2$	$Y$	1
$c_2$	$a_1$	$B$	0
$c_3$	$a_3$	$Y$	1
$c_4$	$a_4$	$Y$	1
$c_5$	$a_2$	$B$	1
$c_6$	$a_3$	$R$	1
$c_7$	$a_4$	$B$	1
$c_8$	$a_4$	$B$	1
$c_9$	$a_1$	$Y$	0
$c_{10}$	$a_2$	$Y$	0

$D$	$A$	$X^1$	$R$
$c_{11}$	$a_2$	$B$	1
$c_{12}$	$a_1$	$Y$	0
$c_{13}$	$a_3$	$B$	0
$c_{14}$	$a_2$	$B$	1
$c_{15}$	$a_1$	$R$	1
$c_{16}$	$a_3$	$Y$	1
$c_{17}$	$a_4$	$R$	0
$c_{18}$	$a_1$	$R$	1
$c_{19}$	$a_3$	$B$	0
$c_{20}$	$a_2$	$Y$	0

Table 3: Data set:  $N = 20$

	$(1 - \gamma)f_D(R) = \frac{(1-\gamma)}{10}$	$(1 - \gamma)f_D(B) = \frac{4(1-\gamma)}{10}$	$(1 - \gamma)f_D(Y) = \frac{5(1-\gamma)}{10}$	$\gamma$
$a_1$	$\mathbf{R}_D(a_1, R)$	$\{(1, 0)\}$	$\{(1, 0)\}$	$\mathbf{R}_D(a_1, x_o)$
$a_2$	$\mathbf{R}_D(a_2, R)$	$\{(0, 1)\}$	$\{(\frac{1}{2}, \frac{1}{2})\}$	$\mathbf{R}_D(a_2, x_o)$
$a_3$	$\{(0, 1)\}$	$\mathbf{R}_D(a_3, B)$	$\{(0, 1)\}$	$\mathbf{R}_D(a_3, x_o)$
$a_4$	$\mathbf{R}_D(a_4, R)$	$\{(0, 1)\}$	$\{(0, 1)\}$	$\mathbf{R}_D(a_4, x_o)$

$N = 10$

	$(1 - \gamma)f_D(R) = \frac{4(1-\gamma)}{20}$	$(1 - \gamma)f_D(B) = \frac{8(1-\gamma)}{20}$	$(1 - \gamma)f_D(Y) = \frac{8(1-\gamma)}{20}$	$\gamma$
$a_1$	$\{(0, 1)\}$	$\{(1, 0)\}$	$\{(1, 0)\}$	$\mathbf{R}_D(a_1, x_o)$
$a_2$	$\mathbf{R}_D(a_2, R)$	$\{(0, 1)\}$	$\{(\frac{2}{3}, \frac{1}{3})\}$	$\mathbf{R}_D(a_2, x_o)$
$a_3$	$\{(0, 1)\}$	$\{(1, 0)\}$	$\{(0, 1)\}$	$\mathbf{R}_D(a_3, x_o)$
$a_4$	$\{(1, 0)\}$	$\{(0, 1)\}$	$\{(0, 1)\}$	$\mathbf{R}_D(a_4, x_o)$

$N = 20$

Table 4: Data sets:  $N = 10$  and  $N = 20$

in regard to the characteristics in the data, for example, if action  $a_2$  was a “bet on the color  $Y$  and the shape cube, while the set of characteristics  $X_D$  records only colors, then outcome frequencies need not be concentrated on  $r = 0$  or  $r = 1$ . Instead, non-degenerate outcome frequencies  $(\frac{2}{3}, \frac{1}{3})$  will occur, reflecting the missing category of “shapes”. Outcome distributions that are not Dirac measures might make the decision maker aware of missing categories.

Example 2 shows the distinction between our framework and typical experiments in statistics which rely on the specification of states. The latter consider an urn for which it is known that all objects are balls that are distinct only in color. Furthermore, the set of possible colors is specified. It is thus known that there is a single category “color” with an exhaustive list of possible characteristics. The only unknown aspect concerns the frequency of the colors in the urn.

The only acts in this scenario are bets on colors which are known to be in the urn. Learning the color of the ball drawn from the urn resolves all uncertainty

and determines uniquely the outcome for all acts. Each characteristic, i.e., each “color”, is a state in the sense of (Savage, 1954).

Observing repeated drawings from this urn, one learns the frequencies (and thus, the probabilities) of the characteristics “color”. Bayesian learning is thus sufficient to resolve all existing uncertainty.

### 2.2.3 The history of bone marrow transplants

At the end of the 1960’s bone-marrow transplantation was still in its infancy. The HLA-typing, which is the standard compatibility test between donor and recipient nowadays, was yet to be discovered and implemented. The most common distinction in studies and in practice was that of a syngeneic, i.e., genetically equivalent (identical twin) donor and an allogeneic, i.e., genetically distinct, one. Furthermore, the use of immunosuppressants prior to the transplantation was also not commonly practiced.

**Example 3.** In 1967, a physician facing a patient in need of a bone-marrow transplant (HCT)<sup>21</sup> as a result of an *Aplastic Anemia* (AA) or *Severe Combined Immunodeficiency* (SCI), might have had at his disposal the following data set<sup>22</sup> The set of recorded characteristics distinguishes two categories: (i) the type of

Disease	Donor	Action	Outcome	# cases
AA	allogeneic	HCT	improvement	1
AA	allogeneic	steroids + HCT	improvement	4
AA	allogeneic	HCT	no improvement	6
AA	allogeneic	steroids + HCT	no improvement	2
AA	allogeneic	HCT	death	39
AA	allogeneic	steroids+HCT	death	17
AA	syngeneic	HCT	recovery	3
AA	syngeneic	steroids+HCT	death	2
AA	syngeneic	steroids+HCT	recovery	2
SCI	allogeneic	HCT	death	9

Table 5: Data on bone-marrow transplantations in 1967

disease where AA is denoted by  $x_1^1$  and SCI by  $x_2^1$ , and (ii) the type of donor where  $x_1^2$  is used for allogeneic and  $x_2^2$  for syngeneic. Hence, the set of all characteristics is

$$X_D = X_D^1 \times X_D^2 = \{x_1^1, x_2^1\} \times \{x_1^2, x_2^2\}$$

<sup>21</sup>HCT stands for *Hematopoietic Cell Transplantation*, i.e., bone marrow transplant.

<sup>22</sup>Both AA and SCI are diseases for which bone-marrow transplant is currently used. The data set below is loosely based on Bortin (1970). Bortin’s data also include malignant disease and leukemia. Certain cases contain further specifications for the allogeneic donor (family relationship, ABO-compatibility), and more details concerning outcomes, (e.g., whether grafting occurred). For the purposes of the example, the table below abstracts from these details. It also assumes that the two successful transplants for SCI recorded by Bortin and first published in 1969 were not yet available as data points.



and the set of observed actions is  $A_D = \{a_1, a_2\}$  with  $a_1$  denoting bone-marrow transplant (HCT), and  $a_2$  an bone-marrow transplant preceded by a steroid treatment. The set of possible outcomes  $R_D = \{r_1 \dots r_4\}$  contains recovery,  $r_1$ , improvement,  $r_2$ , no improvement (but alive),  $r_3$ , and death,  $r_4$ . The number of cases of each type is indicated in the last column. The frequency of characteristics is

$$f_D(x) = \left( f_D(x_1^1, x_1^2) = \frac{69}{85}, f_D(x_1^1, x_2^2) = \frac{7}{85}, f_D(x_1^2, x_1^2) = \frac{9}{85}, f_D(x_1^2, x_2^2) = 0 \right).$$

Notice that the data specifies the frequency of outcomes for a given characteristic  $(x_1^1, x_1^2)$ , i.e., AA with an allogeneic donor, and a given action  $a_1$ , i.e., HCT, as

$$\rho_D(a_1, (x_1^1, x_1^2)) = \left( 0, \frac{1}{46}, \frac{3}{23}, \frac{39}{46} \right).$$

The data set does not include all combinations of actions and characteristics, e.g., there is no observation for  $(a_1, (x_2^1, x_2^2))$ , i.e., a transplant from a syngeneic donor in a case of SCI.

A physician treating such a patient might thus consider the outcomes of other cases as relevant, e.g., transplants with syngeneic donors for AA and transplants with allogeneic donors for SCI, yielding a set of outcome distributions:

$$\mathbf{R}_D(a_1, (x_2^1, x_2^2)) = \{(1, 0, 0, 0), (0, 0, 0, 1)\}.$$

Consider now the director of a medical training program who has to decide whether to include HCT into the curriculum. The available data suggest that the benefits of HCT are rather limited. However, the possibility of observing "other" characteristics, i.e., other diseases which can potentially be treated with HCT need to be considered. For example, the chronic granulomatous disease, a rare immunodeficiency condition in infants, was discovered only in 1959. In the late 1960's no cases of HCT had been performed yet for this disease.<sup>23</sup>

Although no data was available for the treatment of this disease, a possible assignment of outcome distributions to yet unobserved action-characteristics pairs could contain all frequencies observed in the data for the specific action,  $a_1$  (HCT):

$$\mathbf{R}_D(a_1, x^o) = \left\{ \left( 0, \frac{1}{46}, \frac{3}{23}, \frac{39}{46} \right), (1, 0, 0, 0), (0, 0, 0, 1) \right\}.$$

### 3 Decisions for a given set of data

In this section, we will show that the data-based framework that we introduced generates a belief function over outcome distributions for each action. Hence, one can derive a representation of preferences over these belief functions similar to Jaffray (1989). In addition, we will provide axioms in order to characterize a subjective degree of unawareness regarding potential other characteristics.

---

<sup>23</sup>The first HCT for the treatment of this disease was performed in 1973. By 2011 a total of 99 bone marrow transplants for this disease were performed resulting in survival rates of 95-99%, see Kang *et al.* (2011).

### 3.1 From data to choice options

At a given point in time, a decision maker knows the data in a set  $D$ . As described in Section 2, for each pair of actions and characteristics  $(a, x) \in A_D \times X_D$ , there is a finite set of outcome distributions  $\mathbf{R}_D(a, x) \subset \Delta(R_D)$ . Recall that  $\mathbf{R}_D := \cup_{a \in A_D, x \in X_D} \mathbf{R}_D(a, x)$  is the finite set of all frequency distributions over outcomes in  $D$ . We will assume that  $\mathbf{R}_D$  contains all degenerate outcome distributions,  $\delta_r$ , for  $r \in R_D$ . We denote by  $\mathcal{R}_D$  the set of all subsets of  $\mathbf{R}_D$  and by  $\Delta(\mathcal{R}_D)$  the set of all probability distributions on  $\mathcal{R}_D$ .

The following table summarizes the primitive concepts derived from data in  $D$  and the relevant notation:

**Summary of basic notation:**

A finite data set of cases:  $c = (a, x, r) \in D$ , induces

- a set of actions:  $a \in A_D$ ,
- a set of characteristics:  $x \in X_D$ ,
  - a frequency distribution over characteristics:  $f_D \in \Delta(X_D)$ ,
- extended set of characteristics:  $\hat{X}_D = X_D \cup \{x_o\}$ 
  - degree of unawareness:  $\gamma_D$
- a set of outcomes:  $R_D$ ,
  - a frequency distribution over outcomes:  $\rho \in \Delta(R_D)$ ,
  - the set of all finite subsets of  $\Delta(R_D)$ :  $\mathfrak{R}_D$
  - for each  $(a, x) \in A_D \times X_D$ , a finite set of frequencies over outcomes:  $\mathbf{R}_D(a, x) \in \mathfrak{R}_D$
- the set of all frequency distributions over outcomes in  $D$ :  $\mathbf{R}_D := \cup_{a \in A_D, x \in X_D} \mathbf{R}_D(a, x)$ 
  - $\delta_r \in \mathbf{R}_D$  for all  $r \in R_D$
  - the set of all subsets of  $\mathbf{R}_D$ :  $\mathcal{R}_D$
  - the set of all probability distributions on  $\mathcal{R}_D$ :  $\Delta(\mathcal{R}_D)$

### 3.2 From actions to mass distributions

In this section, we concentrate on the specification of actions on the set of observed characteristics,  $X_D$ , disregarding for the time being  $x_o$  (we will return to it in section 3.3).

For a given action  $a \in A_D$ , consider the mapping:  $a : X_D \rightarrow \mathcal{R}_D$  which to every characteristic  $x$  assigns the predicted set of outcomes of action  $a$  for this characteristic,  $a(x) = \mathbf{R}_D(a, x) \in \mathcal{R}_D$ . Note that the frequency of characteristics

in the data  $f_D(x)$  gives a probability distribution over these predictions, assigning a probability of  $f_D(x)$  to  $\mathbf{R}_D(a, x)$ . Thus, for a given action  $a$ , the observed frequency of characteristics generates a probability distribution over the power set of  $\mathbf{R}_D$ ,  $\mathcal{R}_D$ , given by:

$$m_a^{f_D}(R) = \sum_{\{x \in X_D \mid \mathbf{R}_D(a, x) = R\}} f_D(x).$$

By definition,  $m_a^{f_D}(R) \geq 0$  for all  $R \in \mathcal{R}_D$  and  $\sum_{R \in \mathcal{R}_D} m_a^{f_D}(R) = 1$ .

A probability distribution over the elements of a power set,  $m_a^{f_D} \in \Delta(\mathcal{R}_D)$ , however, is a mass distribution that defines a belief function Grabisch (2016, p. 380). The fact that the outcomes of actions on the set of observed characteristics can be represented as a mass distribution allows us to use the seminal approach by Jaffray (1989) to characterize preferences.

The set of actions  $A_D$  together with the realized frequency  $f_D$  generates a finite set of mass distributions  $\{m_a^{f_D} : a \in A_D\}$ . Similarly to Savage (1954), we assume that the set of hypothetical actions  $\mathbf{A}_D$  which the decision maker can conceive is larger than  $\{m_a^{f_D} : a \in A_D\}$  and includes all mappings from characteristics to sets of observed outcome distributions in  $\mathcal{R}_D$ :

$$\mathbf{A}_D = \{a : X_D \rightarrow \mathcal{R}_D\}.$$

We call an action unambiguous when  $a(x)$  is a singleton for all  $x$ . Such actions induce a probability distribution over outcomes  $m_a(r) = \sum_{x \in X_D} f_D(x) \rho_D(r \mid a, x)$ .

Note that the specification of an action combines the (set-valued) consequences of actions with the information contained in the mass distribution. If this information suffices to associate a probability distribution with each action (when all actions are unambiguous) then preferences will be over probability distributions as in von Neumann and Morgenstern (1944).

The following example illustrates this construction.

**Example 4.** Consider a data set  $D$  with two characteristics  $X_D = \{x_1, x_2\}$  yielding two outcome distributions  $\mathbf{R}_D = \{\rho_1, \rho_2\}$  with the power set  $\mathcal{R}_D = \mathcal{P}(\{\rho_1, \rho_2\}) = \{\{\rho_1\}, \{\rho_2\}, \{\rho_1, \rho_2\}\}$ . The set of all basic actions is  $\mathbf{A}_D = \{a : X_D \rightarrow \mathcal{R}_D\}$ . Given a probability (frequency) distribution over the characteristics  $\{x_1, x_2\}$ , say  $(f_1, f_2)$ , each action  $a \in \mathbf{A}_D$  induces a mass distribution  $m_a^f$  in  $\Delta(\mathcal{R}_D)$ . Given the distribution  $f$  on  $X_D$ , the nine acts in  $\mathbf{A}_D$  induce nine mass distributions  $m_a^f \in \Delta(\mathcal{R}_D)$  as illustrated in Table 6.

As Example 4 illustrates, the set of mass distributions  $m_a^{f_D}$  induced by the actions  $a \in A_D$  together with the frequency distribution  $f_D$  observed in a data set  $D$  will be a small subset of all mass distributions  $\Delta(\mathcal{R}_D)$ . Allowing for mixtures of acts in  $\mathbf{A}_D$ , however, will extend the set of mass distributions on  $\mathbf{R}_D$  considerably.

For  $\lambda \in [0, 1]$  and two actions  $a_1, a_2 \in \mathbf{A}_D$ , denote by  $\lambda a_1 + (1 - \lambda) a_2$  the lottery over elements of  $\mathcal{R}_D$  which associates with each  $x \in X_D$  the set of outcome

	$a \in \mathbf{A}_D$	$m_a^f \in \Delta(\mathcal{R}_D)$		
	$(a(x_1), a(x_2))$	$m_a^f(\{\rho_1\})$	$m_a^f(\{\rho_2\})$	$m_a^f(\{\rho_1, \rho_2\})$
$a_1$	$(\{\rho_1\}, \{\rho_1\})$	1	0	0
$a_2$	$(\{\rho_1\}, \{\rho_2\})$	$f_1$	$f_2$	0
$a_3$	$(\{\rho_1\}, \{\rho_1, \rho_2\})$	$f_1$	0	$f_2$
$a_4$	$(\{\rho_2\}, \{\rho_1\})$	$f_2$	$f_1$	0
$a_5$	$(\{\rho_2\}, \{\rho_2\})$	0	1	0
$a_6$	$(\{\rho_2\}, \{\rho_1, \rho_2\})$	0	$f_1$	$f_2$
$a_7$	$(\{\rho_1, \rho_2\}, \{\rho_1\})$	$f_2$	0	$f_1$
$a_8$	$(\{\rho_1, \rho_2\}, \{\rho_2\})$	0	$f_2$	$f_1$
$a_9$	$(\{\rho_1, \rho_2\}, \{\rho_1, \rho_2\})$	0	0	1

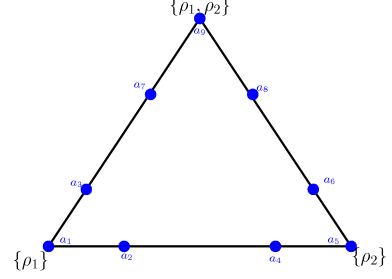


Table 6: Induced mass distributions:  $m_a^f \in \Delta(\mathcal{R}_D)$

distributions  $a_1(x)$  with probability  $\lambda$  and the set of outcome distributions  $a_2(x)$  with probability  $(1 - \lambda)$ . The resulting mass distribution is:

$$m_{\lambda a_1 + (1-\lambda)a_2}^f = \lambda m_{a_1}^f + (1 - \lambda) m_{a_2}^f. \quad (1)$$

Mixtures of acts in  $\mathbf{A}_D$  are elements of the simplex  $\Delta(\mathcal{R}_D)$ . Given the frequency distribution  $f_D$  on  $X_D$  and the set of basic actions  $\mathbf{A}_D$ , let

$$\mathfrak{M}(\mathbf{A}_D, f_D) = \left\{ m_a^f = \sum_k \lambda_k m_{a_k}^f \mid a_k \in \mathbf{A}_D, k \in N \right\}$$

be the set of all mass distributions induced by mixed actions in  $\mathbf{A}_D$ .

The following Lemma 1 shows that the set of mass distributions  $\mathfrak{M}(\mathbf{A}_D, f_D)$  obtained from all mixtures of actions in  $\mathbf{A}_D$  equals the set of all mass distributions on  $\mathcal{R}_D$ ,  $\Delta(\mathcal{R}_D)$ , provided that  $\mathcal{R}_D$  contains all Dirac measures over outcomes in  $R_D$ .

**Lemma 1.**  $\mathfrak{M}(\mathbf{A}_D, f_D) = \Delta(\mathcal{R}_D)$ .

### 3.3 Actions on the extended set of characteristics

As argued in Section 2, a decision maker may also consider the possibility of characteristics  $x_o$  that have not been recorded in the data  $D$ . In this section, we will extend the specification of actions to  $x_o$ .

Assume that a decision maker associates a finite set of possible outcome distribution  $R_a^o \in \mathfrak{R}_D$  with  $x_o$  for every act  $a \in A_D$ . Allowing for  $a(x_o) = R_a^o \in \mathfrak{R}_D$  amounts to assuming that the decision maker can in principle envision hypothetical outcome distributions that have not been observed before.<sup>24</sup>

<sup>24</sup>Outcome distributions  $\rho \in R_a^o$  must, however, be elements of  $\Delta(R_D)$ , i.e., must only consider outcomes observed in  $D$ . In this paper, we do not consider the possibility of "new outcomes" and the resulting "new actions" as Karni and Vierø (2017) do. Such an extension is, however, possible.

Given the extended set of characteristics  $\hat{X}_D = X_D \cup \{x_o\}$ , we consider actions in the extended action set  $\mathbf{A}_D^o = \mathbf{A}_D \times \mathfrak{R}_D$ . As shown in the previous subsection, each basic actions in  $\mathbf{A}_D$  induces a mass distribution  $m_a^{f_D} \in \Delta(\mathcal{R}_D)$  which is then combined with the set of outcome distributions  $R_a^o = a(x_o) \in \mathfrak{R}_D$  associated with  $x_o$ . Omitting the index  $f_D$ , an extended action  $a$  can be written as  $(m_a, R_a) \in \Delta(\mathcal{R}_D) \times \mathfrak{R}_D$ .

As in Subsection 3.2, we allow for mixtures on the extended set of actions  $a \in \mathbf{A}_D^o$ . Given two actions  $a_1 = (m_{a_1}, R_{a_1}^o)$  and  $a_2 = (m_{a_2}, R_{a_2}^o)$  and any  $\lambda \in [0, 1]$ , define the convex combination of the two actions  $\lambda a_1 + (1 - \lambda) a_2$  as the action  $a = (m_{\lambda a_1 + (1-\lambda)a_2}; R_{\lambda a_1 + (1-\lambda)a_2}^o)$  where  $m_{\lambda a_1 + (1-\lambda)a_2}$  is the mixture of  $m_{a_1}$  and  $m_{a_2}$  as defined above in (1) and  $R_{\lambda a_1 + (1-\lambda)a_2}^o$  is the set of outcome distributions,

$$R_{\lambda a_1 + (1-\lambda)a_2}^o = \lambda R_{a_1}^o + (1 - \lambda) R_{a_2}^o = \{ \lambda \rho_1 + (1 - \lambda) \rho_2 \mid \rho_1 \in R_{a_1}^o, \rho_2 \in R_{a_2}^o \}.$$

By construction,  $m_{\lambda a_1 + (1-\lambda)a_2} \in \Delta(\mathcal{R}_D)$  and  $R_{\lambda a_1 + (1-\lambda)a_2}^o \in \mathfrak{R}_D$ . That is, all convex combinations of extended actions  $a = (m_a, R_a^o)$  will be elements of  $\Delta(\mathcal{R}_D) \times \mathfrak{R}_D$ .

We thus consider the set of actions  $\mathcal{A}_D = \Delta(\mathcal{R}_D) \times \mathfrak{R}_D$ . It is easy to check that  $\mathcal{A}_D$  is a mixture set.

**Lemma 2.**  $\mathcal{A}_D = \Delta(\mathcal{R}_D) \times \mathfrak{R}_D$  is a mixture set.

### 3.4 Preferences and Suggested Representation

Denote by  $\succsim$  on  $\Delta(\mathcal{R}_D) \times \mathfrak{R}_D$ , the preference order of the decision maker on the set of actions  $\mathcal{A}_D$ . We assume that the decision maker can rank all actions in this set. Similarly to Savage (1954) this amounts to the ability to rank the consequences of the actions associated with the different characteristics. We will present axioms that imply that any action  $a = (m_a, R_a^o) \in \mathcal{A}_D$  is evaluated by the following functional:

$$V_D(a) = (1 - \gamma_D) \sum_{R \in \mathcal{R}_D} m_a(R) \left[ \alpha_D \max_{\rho \in R} \sum_r u(r) \rho(r) + (1 - \alpha_D) \min_{\rho \in R} \sum_r u(r) \rho(r) \right] \\ + \gamma_D \left[ \alpha_D^o \max_{\rho \in R_a^o} \sum_r u(r) \rho(r) + (1 - \alpha_D^o) \min_{\rho \in R_a^o} \sum_r u(r) \rho(r) \right] \quad (2)$$

where

- $u : R_D \rightarrow \mathbb{R}$  is a von Neumann-Morgenstern utility function over outcomes (unique up to a positive-affine transformation);
- $\alpha_D \in [0, 1]$  is the parameter of optimism,  $(1 - \alpha_D)$  is the degree of pessimism regarding ambiguity in the outcome distributions given the observed data  $D$ ,

and  $\alpha_D^o \in [0, 1]$  and  $(1 - \alpha_D^o)$  are the degrees of optimism and pessimism with regard to the outcome distributions for other, yet unobserved, characteristics. The degrees of optimism/ pessimism  $\alpha_D$  and  $\alpha_D^o$  may, but need not, coincide;

- $\gamma_D \in (0, 1)$  is the degree of unawareness, the subjective weight assigned to  $x_o$ .

In particular, any actually observed action  $a \in A_D$  can be evaluated using the observed frequencies of characteristics  $f_D(x)$  and the outcome predictions  $\mathbf{R}_D(a, x)$  derived from the data:

$$V_D(a) = (1 - \gamma_D) \sum_{x \in X_D} f_D(x) \left[ \alpha_D \max_{\rho \in \mathbf{R}_D(a, x)} \sum_r u(r) \rho(r) + (1 - \alpha_D) \min_{\rho \in \mathbf{R}(a, x)} \sum_r u(r) \rho(r) \right] \\ + \gamma_D \left[ \alpha_D^o \max_{\rho \in \mathbf{R}(a, x_o)} \sum_{r \in R_o} u(r) \rho(r) + (1 - \alpha_D^o) \min_{\rho \in \mathbf{R}(a, x_o)} \sum_r u(r) \rho(r) \right] \quad (3)$$

To illustrate the representation reconsider Example 3.

**Example** (Example 3 resumed). Given  $D$  the evaluation of action  $a_1$ , (HCT) can be written as:

$$V_D(a_1) = (1 - \gamma_D) \left[ \frac{69}{85} \left( \frac{1}{46} u(r_2) + \frac{3}{23} u(r_3) + \frac{39}{46} u(r_4) \right) + \frac{7}{85} u(r_1) + \frac{9}{85} u(r_4) \right] \\ + \gamma_D [\alpha_D^o u(r_1) + (1 - \alpha_D^o) u(r_4)].$$

The evaluation thus combines the observed outcomes of  $a_1$  for each of the observed characteristics with the subjective evaluation for the case of “other” characteristics. In the latter case, only the extreme distributions in  $\mathbf{R}_D(a_1, x^o)$ ,  $(1, 0, 0, 0)$  and  $(0, 0, 0, 1)$  are taken into account, weighted by the corresponding degrees of optimism and pessimism for unobserved characteristics.

Limited confidence in the data (high  $\gamma_D$ ) combined with sufficient optimism,  $\alpha_D^o$ , might lead to the conclusion that including HCT into training programs is the right decision, especially if the cost of doing so are relatively low. As it was, given the limited success of the early attempts at HCT, by 1967, the procedure was declared a “dead-end”, Granot and Strob, (2020, p. 2717), implying that the medical field felt rather pessimistic about future attempts.

### 3.5 Axiomatization

We now provide an axiomatization of preferences over the set of actions  $\mathcal{A}_D = \Delta(\mathcal{R}_D) \times \mathfrak{R}_D$ ,  $a = (m_a, R_a^o)$ . Our axiomatization builds on the approach in Jaffray (1989).

**Axiom 1** The preference order  $\succsim$  on  $\mathcal{A}_D$  is complete, transitive and non-trivial in the following sense: there is an  $R^o \in \mathcal{R}_D$  and  $m_{a_1}, m_{a_2} \in \Delta(\mathcal{R}_D)$  such that for  $a_1 = (m_{a_1}, R^o)$  and  $a_2 = (m_{a_2}, R^o) \in \mathcal{A}_D$ ,  $(m_{a_1}, R^o) \succ (m_{a_2}, R^o)$ .

The non-triviality condition in Axiom 1 is somewhat stronger than usual. In particular, it requires that there is some outcome  $R^o$  associated with the “other characteristics” which is a subset of the actually observed probability distributions in the data and for which the decision maker is not fully indifferent among all extended actions. This non-triviality condition requires that the decision maker be not indifferent among all mass functions for at least some set of probability distributions associated with the “other” characteristics.

**Axiom 2** For all  $(m_{a_1}, R_{a_1}^o), (m_{a_2}, R_{a_2}^o), (m_{a_3}, R_{a_3}^o) \in \mathcal{A}_D$  and all  $\lambda \in [0; 1]$ ,

$$\begin{aligned} (m_{a_1}, R_{a_1}^o) &\succsim (m_{a_2}, R_{a_2}^o) \\ \Leftrightarrow \lambda (m_{a_1}, R_{a_1}^o) + (1 - \lambda) (m_{a_3}, R_{a_3}^o) &\succsim \lambda (m_{a_2}, R_{a_2}^o) + (1 - \lambda) (m_{a_3}, R_{a_3}^o). \end{aligned}$$

**Axiom 3** For all  $(m_{a_1}, R_{a_1}^o), (m_{a_2}, R_{a_2}^o), (m_{a_3}, R_{a_3}^o) \in \mathcal{A}_D$  such that  $(m_{a_1}, R_{a_1}^o) \succ (m_{a_2}, R_{a_2}^o) \succ (m_{a_3}, R_{a_3}^o)$ , there are  $\lambda, \mu \in (0; 1)$  such that

$$\lambda (m_{a_1}, R_{a_1}^o) + (1 - \lambda) (m_{a_3}, R_{a_3}^o) \succ (m_{a_2}, R_{a_2}^o) \succ \mu (m_{a_1}, R_{a_1}^o) + (1 - \mu) (m_{a_3}, R_{a_3}^o).$$

*Remark 2.* The three Axioms imply that preferences are separable across the two dimensions,  $m$  and  $R^o$ , see Proposition 1 in the Appendix.

The following corollary obtains:

**Corollary 1.** *Axioms 1–3 imply that for any  $m_a, m_b \in \Delta(\mathcal{R}_D)$ , and any  $R_a^o, R_b^o \in \mathfrak{R}_D$ ,  $(m_a, R_a^o) \succsim (m_b, R_a^o)$  iff  $(m_a, R_b^o) \succsim (m_b, R_b^o)$  and  $(m_a, R_a^o) \succsim (m_a, R_b^o)$  iff  $(m_b, R_a^o) \succsim (m_b, R_b^o)$ .*

*Remark 3.* Note that in general  $m$  and  $R$  are different objects,  $m$  is a probability distribution on  $\mathcal{R}_D$ , whereas  $R$  is an element of  $\mathfrak{R}_D \supset \mathcal{R}_D$ . Nevertheless, each  $m$  which assigns a probability of 1 to a single set  $R \in \mathcal{R}_D$  can be uniquely identified with an  $R^o \in \mathfrak{R}_D$  such that  $R^o = R$ . We could thus use the subset of actions for which the mass distributions have a singleton support,  $\Delta(\mathcal{R}_D)^C \times \mathfrak{R}_D$ , as well as the subset of actions for which the set assigned to  $x_o$  is an element of  $\mathcal{R}_D$ ,  $\Delta(\mathcal{R}_D) \times \mathcal{R}_D$  to formulate an Anscombe-Aumann-type axiom of state-independence of preferences. More specifically, to allow for different degrees of optimism in regard to observed and unobserved contingencies and thus for the possibility that two sets  $R$  and  $R'$  can be ranked when associated with the already observed or the unobserved characteristics. Thus, our state-independence axiom is imposed only on singleton sets.

**Axiom 4** For any  $\lambda \in [0, 1]$ , any  $\{\rho_1\}, \{\rho_2\}$  and  $\{\rho\} \in \mathcal{R}_D$ , and  $m$  and  $m' \in \Delta(\mathcal{R}_D)$  such that  $m(\{\rho_1\}) = \lambda$ ,  $m(\{\rho_2\}) = 1 - \lambda$ ,  $m'(\{\rho\}) = 1$ ,

$$\begin{aligned} (\tilde{m}, \lambda \{\rho_1\} + (1 - \lambda) \{\rho_2\}) &\succsim (\tilde{m}, \{\rho\}) \text{ for some } \tilde{m} \in \Delta(\mathcal{R}_D) \text{ holds iff} \\ (m, R^o) &\succsim (m', R^o) \text{ holds for some } R^o \in \mathfrak{R}_D. \end{aligned}$$

While our last Axiom 4 concentrated on preferences with respect to singletons, we now turn to preferences regarding sets with multiple elements. Consider  $R$  and  $R' \in \mathfrak{R}_D$ . We will write  $R \succsim^o R'$  iff  $(m, R) \succsim (m, R')$  for some and thus, by Corollary 1, for all  $m \in \Delta(\mathcal{R}_D)$ . We will write  $R \succsim^d R'$  iff for  $m_a(R) = 1$  and  $m_b(R') = 1$ ,  $(m_a, R'') \succsim (m_b, R'')$  for some and thus, by Corollary 1, for all  $R'' \in \mathfrak{R}_D$ . Axiom 4 then implies that these two relations coincide for singletons  $\{\rho\} \in \mathcal{R}_D$ :  $\{\rho\} \succsim^o \{\rho'\}$  iff  $\{\rho\} \succsim^d \{\rho'\}$  for  $\{\rho\} \in \mathcal{R}_D$ , while for  $\{\rho\}$  or  $\{\rho'\} \in \mathfrak{R}_D \setminus \mathcal{R}_D$ , the preference  $\succsim^d$  is not defined and the comparison of the two sets is determined by  $\succsim^o$ . In both cases, with a slight abuse of notation, we write  $\rho \succsim \rho'$ . Axiom 4 thus implies a well-defined preference order over the singleton sets, regardless of whether they are associated with observed or unobserved characteristics. This in turn allows us to define for each set of outcome distributions  $R \in \mathfrak{R}_D$ , a "best" and "worst" outcome distribution  $\underline{\rho}_R, \bar{\rho}_R \in \Delta(R_D)$ . The following axiom is an adaptation of the axiom introduced in Jaffray (1989):

**Axiom 5** For all  $R, R'$ , if  $\underline{\rho}_R \succsim \underline{\rho}_{R'}$  and  $\bar{\rho}_R \succsim \bar{\rho}_{R'}$ , then  $R \succsim^o R'$  and  $R \succsim^d R'$ .

Axiom 5 implies that the comparison between any two sets of outcome distributions only depends on their best and worst elements. In contrast to Jaffray (1989), however, preferences between sets of outcome distributions may depend on whether they are associated with an already observed characteristic, or a yet unobserved, "other" characteristic, i.e.,  $\succsim^d$  and  $\succsim^o$  might differ on non-singleton sets.

Axioms 1–5 allow a representation similar to 2, but for the fact that the degrees of optimism in general depend on the best and worst outcomes in the set. To obtain  $\alpha_D$  and  $\alpha_D^o$  that are independent of the set, we impose two additional axioms, see Proposition 3 in the Appendix.

To understand the axiom, suppose that we compare two actions  $a$  and  $b$  with identical mass distributions  $m_a = m_b =: m$ . One of the actions attributes a set with two outcome distributions to  $x_o$ ,  $R_a = \{\rho, \rho'\}$  with  $\rho \succ \rho'$ , while, for the second action  $b$ , the set of outcome distributions on  $x_o$  contains only the mixture  $R_b = \{\alpha\rho + (1 - \alpha)\rho'\}$ . Axioms 1–5 imply that there exists a unique  $\alpha \in [0, 1]$  such that  $R_a \sim^o R_b$ . This  $\alpha$  is the weight assigned to the best outcome of the set  $R_a$ , i.e., the degree of optimism relative to unobserved characteristics with respect to this set. Axiom 6 postulates that the so-determined  $\alpha$  is independent of the set  $R_a$  under consideration.

**Axiom 6** For any  $\rho, \rho', \rho'', \rho''' \in \Delta(R_D)$ , such that  $\rho \succ \rho'$  and  $\rho'' \succ \rho'''$  and any  $m \in \Delta(\mathcal{R}_D)$ , let  $R_a = \{\rho, \rho'\}$ ,  $R_c = \{\rho'', \rho'''\}$  and for some  $\alpha \in [0, 1]$ ,

$$\begin{aligned} R_b &= \{\alpha\rho + (1 - \alpha)\rho'\} \\ R_d &= \{\alpha\rho'' + (1 - \alpha)\rho'''\} \end{aligned}$$

Then  $(m, R_a) \sim (m, R_b)$  iff  $(m, R_c) \sim (m, R_d)$ .

The final Axiom is analogous to Axiom 6, but imposed on the set  $\Delta(\mathcal{R}_D)$ , i.e., on the mass functions associated with the observed characteristics  $X_D$ . It implies that the degree of optimism relative to observed characteristics does not depend on the specific set under consideration.



**Axiom 7** For any  $\{\rho\}, \{\rho'\}, \{\rho''\}, \{\rho'''\} \in \mathcal{R}_D$ , such that  $\rho \succ \rho'$  and  $\rho'' \succ \rho'''$  and any  $R \in \mathfrak{R}_D$ , let<sup>25</sup>  $m_a(\{\rho, \rho'\}) = 1$ ,  $m_c(\{\rho'', \rho'''\}) = 1$  and for some  $\alpha \in [0, 1]$ ,

$$\begin{aligned} m_b(\{\rho\}) &= \alpha, \quad m_b(\{\rho'\}) = (1 - \alpha) \\ m_d(\{\rho''\}) &= \alpha, \quad m_d(\{\rho'''\}) = (1 - \alpha) \end{aligned}$$

Then  $(m_a, R) \sim (m_b, R)$  iff  $(m_c, R) \sim (m_d, R)$ .

Axioms 1–7 are necessary and sufficient to obtain our desired representation:

**Theorem 1.** *The preference order  $\succsim$  on  $\mathcal{A}_D = \Delta(\mathcal{R}_D) \times \mathfrak{R}_D$  satisfies Axioms 1–7, iff there is a representation*

$$\begin{aligned} V_D(m, R^o) &= \sum_{R \in \mathcal{R}_D} (1 - \gamma_D) m(R) \left[ \alpha_D \max_{\rho \in R} \sum_{r \in R_D} u(r) \rho(r) + (1 - \alpha_D) \min_{\rho \in R} \sum_{r \in R_D} u(r) \rho(r) \right] \\ &\quad + \gamma_D \left[ \alpha_D^o \max_{\rho \in R^o} \sum_{r \in R_D} u(r) \rho(r) + (1 - \alpha_D^o) \min_{\rho \in R^o} \sum_{r \in R_D} u(r) \rho(r) \right] \end{aligned} \quad (5)$$

where  $u$  is a unique (up to a positive-affine transformation) von-Neumann-Morgenstern utility function over outcomes,  $\gamma_D \in (0, 1)$  is a unique parameter describing the perception of unawareness and  $\alpha_D, \alpha_D^o \in [0, 1]$  are unique parameters of optimism relevant to the set of observed, respectively, unobserved, characteristics.

A special case of the representation is the one in which the coefficient of optimism does not depend on the type of characteristics under consideration, i.e.,  $\alpha_D = \alpha_D^o$ . Such a representation can be easily obtained by replacing Axiom 7 with the following one:

**Axiom 7'** For any  $\{\rho\}, \{\rho'\} \in \mathcal{R}_D$ , such that  $\rho \succ \rho'$  and any  $R \in \mathfrak{R}_D$ , let  $m_a(\{\rho, \rho'\}) = 1$ , and for some  $\alpha \in [0, 1]$ ,

$$m_b(\{\rho\}) = \alpha, \quad m_b(\{\rho'\}) = (1 - \alpha)$$

Then  $(m_a, R) \sim (m_b, R)$  iff for some  $m \in \Delta(\mathcal{R}_D)$ ,

$$(m, \{\rho, \rho'\}) \sim (m, \{\alpha\rho + (1 - \alpha)\rho'\}). \quad (6)$$

Axiom 7' can be seen as an extension of Axiom 4 to sets in  $\mathcal{R}_D$  containing two elements. It requires that the mass function concentrated on a two element set  $\{\rho, \rho'\}$  is considered indifferent to a mass function which mixes the two singletons in proportions  $\alpha$  (for the better one) and  $(1 - \alpha)$  for the worse one iff the same two element set, but assigned on the "other" characteristics is considered indifferent

<sup>25</sup>Recall that  $\mathcal{R}_D$  is the set of all subsets of observed frequencies in the data,  $\mathbf{R}_D$ . Thus, if  $\mathcal{R}_D$  contains the singletons  $\{\rho\}, \{\rho'\}, \{\rho''\}$  and  $\{\rho'''\}$ , then it also contains the sets  $\{\rho, \rho'\}$  and  $\{\rho'', \rho'''\}$  and vice-versa.

to the singleton set  $\{\alpha\rho + (1 - \alpha)\rho'\}$  which is a mixture of its two elements in the same proportions  $\alpha$  and  $(1 - \alpha)$ . Since by Axiom 6, the coefficient  $\alpha$  does not depend on the choice of  $\rho$  and  $\rho'$ , we obtain that the  $\alpha_D^o$  identified in the proof of Proposition 1 applies also to all two-element sets  $\{\rho'', \rho'''\} \in \mathcal{R}_D \subset \mathfrak{R}_D$ . We can thus set  $\alpha_D^o = \alpha_D$  and obtain the following Corollary:

**Corollary 2.** *The preference order  $\succsim$  on  $\mathcal{A}_D$  satisfies Axioms 1–6 and 7', iff there is a representation*

$$V_D(m, R^o) = \sum_{R \in \mathcal{R}_D} (1 - \gamma_D) m(R) \left[ \alpha_D \max_{\rho \in R} \sum_{r \in R_D} u(r) \rho(r) + (1 - \alpha_D) \min_{\rho \in R} \sum_{r \in R_D} u(r) \rho(r) \right] + \gamma_D \left[ \alpha_D \max_{\rho \in R^o} \sum_{r \in R_D} u(r) \rho(r) + (1 - \alpha_D) \min_{\rho \in R^o} \sum_{r \in R_D} u(r) \rho(r) \right] \quad (7)$$

where  $u$  is a unique (up to a positive-affine transformation) von-Neumann-Morgenstern utility function over outcomes,  $\gamma_D \in (0, 1)$  is a unique parameter describing the perception of unawareness and  $\alpha_D$  is a unique parameter of optimism relevant both to the set of observed and the unobserved, characteristics.

## 4 Acquiring new data

So far, we have kept the data set,  $D$ , and thus, the corresponding sets of characteristics,  $X_D$ , outcomes,  $R_D$  and observed actions,  $A_D$ , fixed. As a result, we were able to identify: (i) the vNM utility function on outcomes,  $u$ , (ii) the weight assigned on "other characteristics",  $x_o$ ,  $\gamma_D$ , which can be interpreted as the degree of unawareness, (iii) the parameters of optimism for the known characteristics,  $\alpha_D$  and the "other" characteristics,  $\alpha_D^o$ . In general, these three components of the representation depend on the available data set  $D$ .

Assume now that the decision maker obtains access to a different data set  $D'$ . This data set may take the form of a "continuation" of the history recorded in  $D$ , i.e.,  $D \subseteq D'$ . Alternatively,  $D'$  may be a hypothetical data set, disjoint from  $D$ , containing different cases with an alternative set of circumstances and different outcomes that the decision maker may face. We will write  $D = \{(a_n, x_n, r_n)_{n=1}^N\}$  and  $D' = \{(a'_n, x'_n, r'_n)_{n=1}^{N'}\}$  for the two data sets in question. For the purposes of the following discussion, we will assume that as long as the set of outcomes remains constant,  $R_D = R_{D'}$ , the decision maker uses the same utility function  $u$  to evaluate lotteries over outcomes.

We will use Example 3 from Section 2 to illustrate how new information may affect the parameters of the model and, thus, the decision maker's predictions.

## 4.1 Statistical learning

In the case of statistical learning, the set of characteristics,  $X_D$ , the set of actions,  $A_D$ , and the set of outcomes,  $R_D$ , do not change with the arrival of new information. The data sets  $D$  and  $D'$  differ only in terms of their length,  $N$  and  $N'$  respectively, and of the observed frequencies of outcomes for each action-characteristic combination  $(a, x)$ :  $\rho_D(\cdot | a, x)$ ,  $\rho_{D'}(\cdot | a, x)$ . Learning concerns only the frequency of outcomes and possibly, as the number of observations increases, the perception of unawareness.

Reconsider the data in Table 5 of Example 3 where the value of including the BMT procedure into a medical program was computed in Equation (4). Suppose that the director of the program obtains additional data,  $\tilde{D}$  of size  $\tilde{N}$  consisting of the same characteristics, actions and outcomes. This information can be pooled

to yield a new data set  $D' = \left( \underbrace{(a_1, x_1, r_1) \dots (a_N, x_N, r_N)}_D, \underbrace{(\tilde{a}_1, \tilde{x}_1, \tilde{r}_1) \dots (\tilde{a}_N, \tilde{x}_N, \tilde{r}_N)}_{\tilde{D}} \right)$  with corresponding new frequencies of characteristics  $f_{D'}(x)$  and outcomes  $\rho_{D'}(a, x)$ .

In a special, but particularly relevant case,  $\tilde{D}$  exactly replicates the information in  $D$ ,  $\tilde{D} = D$ , so that  $f_{D'}(x) = f_D(x)$  and  $\rho_D(a, x) = \rho_{D'}(a, x)$  remain unchanged for all  $x$  and  $a$ . Assuming also that the outcomes of "other" categories remain unchanged,  $\mathbf{R}_D(a_1, x_o) = \mathbf{R}_{D'}(a_1, x_o)$ , the evaluations  $V_{D'}(a_1)$  and  $V_D(a_1)$  can only differ because of differences in the subjective parameters of unawareness  $\gamma$  and the degree of optimism,  $\alpha^o$ . Since the new data confirm the already available evidence, the decision maker's attitude towards ambiguity should not change,  $\alpha_{D'}^o = \alpha_D^o$ . Furthermore, since the additional data do not contain observations of "other" characteristics, the decision maker perceives less unawareness and more confidence in the light of the larger data set  $D'$ ,  $\gamma_{D'} < \gamma_D$ .

Based on the data set  $D$ , a sufficiently optimistic program director may well believe that HCT should be part of the curriculum, yet the additional data in  $\tilde{D}$ , confirming the limited rate of success of this treatment, leaves less room for the possibility that the treatment of some other, so far unobserved diseases might benefit from the procedure. Hence, although the director's optimism remains unchanged, the decrease in perceived unawareness due to the additional evidence,  $\gamma_{D'} < \gamma_D$ , may reduce the evaluation of HCT as treatment for some other illness sufficiently to lead to its exclusion from the curriculum. Formally,  $V_{D'}(a_1) < V_D(a_1)$  obtains, whenever unawareness decreases,  $\gamma_D > \gamma_{D'}$  and optimism  $\alpha_D^o$  is sufficiently large.<sup>26</sup>

## 4.2 Learning new characteristics

The set of characteristics relevant for the outcome of an action may not be contained in a given data set ex-ante. First, we will consider a scenario in which the

<sup>26</sup>That is, if the max-min utility assigned to "other" characteristics,  $\alpha_D^o u(r_1) + (1 - \alpha_D^o) u(r_4)$ , exceeds the average utility in the data  $\left[ \frac{69}{85} \left( \frac{1}{46} u(r_2) + \frac{3}{23} u(r_3) + \frac{39}{46} u(r_4) \right) + \frac{7}{85} u(r_1) + \frac{9}{85} u(r_4) \right]$ .

set of observed categories  $T$  remains unchanged but where a new characteristic is observed within a given category.

Returning to Example 3, in 1968 a 2-year old boy was diagnosed with Wiskott-Aldrich syndrome (an immune-deficiency disorder).<sup>27</sup> His sister served as an allogeneic donor for an HCT and the boy's condition substantially improved. A physician with access to the data set in table 5 and having read the article by Bach *et al.* (1968) reporting the new case, would add a new case  $(a_1, (x_3^1 = \text{Wiskott-Aldrich syndrome}, x_1^2), r_2)$  to his data set. In particular, the new case contains an unforeseen so far unobserved characteristic in the category "disease",  $x^{new} = (x_3^1, x_1^2)$ . Let  $D' = D \cup \{(x^{new}, a_1, r_2)\}$  be the new data set and the sets of categories of characteristics evolve as  $X_{D'}^1 = X_D^1 \cup \{x_3^1\}$ ,  $X_{D'}^2 = X_D^2$ , whereas the sets actions and outcomes remain unchanged. Three effects may follow as a result of obtaining such new additional evidence.

Firstly, the number of unobserved action-characteristics pairs increases: there are no observations for the cases  $(a_1, (x_3^1, x_2^2))$ ,  $(a_2, (x_3^1, x_1^2))$ ,  $(a_2, (x_3^1, x_2^2))$  in  $D'$ . Unable to associate a single outcome distribution based on data in  $D$  perceived ambiguity will increase. This effect will manifest itself in the specification of  $\mathbf{R}_D(a, (x_3^1, x^2))$  for  $a \neq a_1$ , and  $x^2 \neq x_1^2$ .

Secondly, the observation of a new, previously unobserved characteristic in  $D'$  will constitute a "surprise", since it may serve as evidence for the existence of further "other" characteristics and reinforce perceived unawareness,  $\gamma_{D'} > \gamma_D$ . For example, the fact that a so-far unobserved disease could be successfully treated with an HCT might provide credence for the hypothesis that other, yet unobserved diseases might be also treatable in this way.

Thirdly, the new data set  $D'$  also records the outcome  $r_2$  for the treatment by HCT realized for the new characteristic  $x_3^1$ . We will call the new observation  $(a_1, (x_3^1, x_1^2), r_2)$  in  $D'$  a "positive surprise" ("negative surprise") if the action  $a_1$  results in strictly higher (lower) utility in case of the new characteristic compared to the evaluation of action  $a_1$  given the data recorded in  $D$ :  $u(r_2) > (<) V_D(a_1)$ . Positive and negative surprises may effect expectations about future potential outcomes for other yet unobserved characteristics making them more optimistic or pessimistic,  $\alpha_{D'}^o > (<) \alpha_D^o$ , respectively.<sup>28</sup>

Notice that the initial evaluation of HCT combines the negative evidence from HCTs in  $D$  with its evaluation for unobserved characteristics  $x_o$ ,  $\alpha_D^o u(r_1) + (1 - \alpha_D^o) u(r_4)$ . The observation of the new characteristic constitutes a "positive surprise" provided that either the initial confidence  $(1 - \gamma_D)$  is sufficiently large or the degree of optimism  $\alpha_D^o$  is sufficiently small.

Our framework allows one to formally model learning as observed in the shift of attitude of physicians reported in Granot and Strob (2020, p. 2718). Physicians

---

<sup>27</sup>The Wiskott-Aldrich symptom is a rare genetic disease which was first discovered in 1937 and later described in 1954. In 1968, a physician might have been unaware of the possibility of using HCT for the treatment of this disease.

<sup>28</sup>This defines the notions of a positive or negative surprise relative to the global evaluation of the action given the data. Other benchmarks for  $x_o$ , such as observed average utility or max-min utility, are possible. Which one will be relevant for the adjustment of the optimism parameter is an empirical question that can be studied using our framework.

view of bone-marrow transplants in 1967 was rather pessimistic, corresponding to low levels of  $\alpha_D^o$  a low evaluation  $V_D(a_1)$ . The success recorded in the new case can be viewed a positive surprise,  $u(r_2) > V_D(a_1)$  that provided new impetus for the further development of HCT procedures. Even with the limited amount of data available at this time, a single successful case would not significantly affect the recorded rates of success for this treatment. It is therefore likely that the overall increase in the positive evaluation of HCT reported by Granot and Strob (2020) can be attributed to an increase in optimism with respect to other characteristics,  $\alpha_{D'}^o > \alpha_D^o$ , combined with an increase in perceived unawareness, the total effect might be sufficiently strong so as to reverse the initially negative assessment.

HCT is by now part of the standard treatment for a variety of diseases but conditions (characteristics) for which the treatment might be beneficial continue to be identified. An example is the CTLA-4 insufficiency, a rare genetic immune disorder first identified in 2014. Treatment options for this disease are being investigated shed light on the limited success of the initial transplants., see Egg and et al. (2021) with HCT being one such option applied successfully in the cases of 13 out of 18 patients.

### 4.3 Learning new categories

New data may contain not only information about new characteristics but may also reveal the existence of new relevant categories for characteristics. In the late 1960’s researchers gained a better understanding of the human immune system that also shed new light on the limited success of HCT. Researchers discovered the HLA system responsible for the immune responses of the body and its role in graft rejection and in graft-versus-host disease. This insight allowed the set of relevant characteristics of patients to be refined such as to include HLA-compatibility of the donor.<sup>29</sup> This refinement of the categories substantially increased the success rate of future HCT.

Formally, let  $t^{new}$  stand for the new category “HLA-compatibility” and let the components of this category be given by  $X^{t^{new}} := X^3 = \{x_1^3, x_2^3, x_3^3\}$ , where  $x_1^3$  stands for HLA-matched,  $x_2^3$  for HLA-haploidentical and  $x_3^3$  for neither matched nor haploidentical.

#### 4.3.1 Reconstructing missing data

When a new relevant category is identified, it may be that measurements of the characteristics corresponding to this category are available also for past observations and, thus, existing data in  $D$  can be completed to take the new category into account. In this case, it is straightforward to rewrite the valuation in Equation

---

<sup>29</sup>An HLA-haploidentical donor is one who shares, by common inheritance, exactly one HLA haplotype with the recipient and is mismatched for a variable number of HLA genes, ranging from zero to six (HLA-A, -B, -C, -DRB1, DQB1, and -DPB1), on the unshared haplotype. Potential HLA-haploidentical donors include biological parents, biological children, full or half siblings, and even extended family donors.

(5) for the expanded set of categories. Since the total number of characteristics increases, the number of combinations  $(a, x)$  so far unobserved in the data will also increase. Hence, the decision-maker needs to specify potentially ambiguous payoff distributions for these scenarios and perceived unawareness with respect to observed characteristics is likely to increase.

In general, data for characteristics of categories either unknown or deemed irrelevant in the past will not be available. This is the case for the data on HCT assembled by Bortin (1970). For syngeneic donors (identical twins), HLA-matching is guaranteed. Since most physicians were unaware of the HLA-system, however, only very few cases of allogeneic donors contain information about their HLA-compatibility.<sup>30</sup> This raises the question of how one should interpret data reconstructed from past cases.

Consider the data set  $D$  restricted only to the set of HCTs  $(a_1)$  for aplastic anemia with an an allogeneic donor  $D = \left\{ \underbrace{(a_1, (x_1^1, x_1^2), r_2)}_{1 \text{ time}}, \underbrace{(a_1, (x_1^1, x_1^2), r_3)}_{3 \text{ times}}, \underbrace{(a_1, (x_1^1, x_1^2), r_4)}_{39 \text{ times}} \right\}$ . For

this data, the initial set of characteristics is a singleton:  $X_D = \{(x_1^1, x_1^2)\}$  and the corresponding outcome frequency is  $\rho_D(a_1, (x_1^1, x_1^2)) = (0, \frac{1}{46}, \frac{3}{23}, \frac{39}{46})$ . Suppose now that the decision maker learns of additional data for which HLA-compatibility has been recorded,  $\tilde{D} = \{(a_1, (x_1^1, x_1^2, x_1^3), r_1), (a_1, (x_1^1, x_1^2, x_1^3), r_2), (a_1, (x_1^1, x_1^2, x_1^3), r_3), (a_1, (x_1^1, x_1^2, x_1^3), r_4)\}$ . Combining  $D$  and  $\tilde{D}$  into a new data set  $D'$ , yields the new set of characteristics  $X_{D'} = \{(x_1^1, x_1^2, x_1^3), (x_1^1, x_1^2, x_1^2), (x_1^1, x_1^2, x_1^3)\}$ . The data set  $D'$ , however, contains some observations (those in  $D$ ) for which only the categories,  $x^1$  and  $x^2$ , have been recorded.

Hence, the additional information about a new category  $X^{t^{new}}$  generates additional ambiguity due to missing observations.<sup>31</sup> The decision-maker does not know which HLA-characteristic,  $x_1^3$ ,  $x_2^3$  or  $x_3^3$ , to associate with the observations  $(a_1, (x_1^1, x_1^2))$ . As far as one is concerned with the evaluation of an action across all possible characteristics, i.e., all possible HLA-types of donors, such ambiguity need not be of concern, because the outcome distribution for each characteristic can be evaluated according to with its own category. In our example, the decision-maker could consider the set of characteristics  $\tilde{X}_{D'} = \{(x_1^1, x_1^2), (x_1^1, x_1^2, x_1^3), (x_1^1, x_1^2, x_1^2), (x_1^1, x_1^2, x_1^3)\}$  and replace  $X_D$  by  $\tilde{X}_{D'}$  in Equation (3) in order to obtain an evaluation of  $a_1$ . Such an approach may be appropriate if the number of observations containing characteristics of the new category is small.

A physician, however, facing a patient with SCI who has a haplo-identical donor will find such ambiguity highly relevant, since information about the HLA-type is recorded in only two cases, while in 46 cases the record is missing. As more evidence

<sup>30</sup>Instead, in some cases the family relationship with the donor is mentioned, while others contain the ABO-compatibility with the donor. In some cases the HLA-classification can be recovered since a parent-donor is by definition HLA-haploidentical. In contrast, a sibling is identical in 25%, haplo-identical in 50% and neither in 25% of the cases.

<sup>31</sup>Arad and Gayer (2012) also address ambiguity which arises from incomplete data and provide experimental evidence. Dominiak and Guerdjikova (2021 forthcoming) discuss the issue of increase in ambiguity resulting from the discovery of new finer contingencies in a Savagean framework.

is accumulated, however, the relative frequency of observations with missing data,  $f(x)$  will become negligibly small and thus of little impact for the evaluation of the actions, regardless of the predictions  $\mathbf{R}_{D'}(a, x)$ .

### 4.3.2 Contradictory evidence and new categories

When observable characteristics fail to uniquely predict the outcome of an action, the decision maker may perceive data as contradictory. Consider HCT for the case of SCI and an allogeneic donor. In  $D$ , all 9 cases of such transplants resulted in the death as the unique prediction of outcome for the patient,  $\rho_D(r_4 | (a_1, (x_2^1, x_1^2))) \sim 1$ . According to Granot and Strob (2020, p. 2718), two new cases published in 1968 documented success of HCT in patients with SCI and allogeneic donors resulting in a new data set  $D' = \left\{ \underbrace{(a_1, (x_2^1, x_1^2), r_4)}_{9 \text{ times}}, \underbrace{(a_1, (x_2^1, x_1^2), r_1)}_{2 \text{ times}} \right\}$ .

Such evidence may appear contradictory and suggest that some unobserved underlying factor may influence the result of the treatment as illustrated by the HLA-classification).<sup>32</sup> Thus, rather than averaging out the evidence and using  $\rho_{D'}(r_1 | (a_1, (x_2^1, x_1^2))) = \frac{2}{11}$  and  $\rho_{D'}(r_4 | (a_1, (x_2^1, x_1^2))) = \frac{9}{11}$  for evaluating the actions, the physician may consider a set of outcome distributions  $\mathbf{R}_{D'}(a_1, (x_2^1, x_1^2)) = \{\eta\delta_{r_1} + (1 - \eta)\rho_D(a_1, (x_2^1, x_1^2)), \eta\delta_{r_4} + (1 - \eta)\rho_D(a_1, (x_2^1, x_1^2))\}$ . This set combines the data-based information  $\rho_D(a_1, (x_2^1, x_1^2))$  with the extreme distributions regarding the extreme observed outcomes  $r_1$  or  $r_4$ . These extreme distributions would correspond to potential discoveries of a characteristic for which the outcome of  $a_1$  is  $r$  for sure. Each of these distributions is an “ $\eta$ -distance” away from the observed frequency. The parameter  $\eta$  can be interpreted as the subjective relevance assigned to the identification of a category which would allow to differentiate between the outcomes for action  $a_1$ . Special cases of these beliefs are  $\eta = 0$ , for which  $\mathbf{R}_{D'}(a_1, (x_2^1, x_1^2)) = \{\rho_{D'}(a_1, (x_2^1, x_1^2))\}$ , and the likelihood of identifying the relevant category is null. For  $\eta = 1$ , the decision maker considers only the Dirac measures concentrated on the extreme outcomes,  $\mathbf{R}_{D'}(a_1, (x_2^1, x_1^2)) = \{\delta_{r_1}, \delta_{r_4}\}$ . This corresponds to the case where the relevance of discovering the new category is maximal and for the evaluation of action  $a_1$  the largest.

Assuming that both the disease as well as the type of donor are identified and, thus, no weight is assigned to other characteristics  $x_o$ , the evaluation of a planned HCT then becomes

$$V_{D'}(a_1|\eta) = (1 - \eta) \left[ \frac{2}{11}u(r_1) + \frac{9}{11}u(r_4) \right] + \eta [\alpha_{D'}u(r_1) + (1 - \alpha_{D'})u(r_4)].$$

Depending on  $\eta$ , as well as on his degree of optimism the doctor may issue a recommendation in favor or against the treatment.<sup>33</sup>

<sup>32</sup>Similar problems arise when randomized controlled trials are conducted in different countries and document varying levels of success of the policy intervention studied. A simple aggregation of the results without taking into account locally specific factors might significantly bias the results, see Deaton and Cartwright (2018).

<sup>33</sup>Such hypothetical reasoning can be applied before the relevant category is identified, allowing

### 4.3.3 The value of new categories

One can use this approach to determine the value of identifying a new category. To do so, suppose that a data set  $\tilde{D}'$  with a discriminatory category  $\tilde{X}^3$  is discovered such that the outcome of  $(a_1, (x_2^1, x_1^2, \tilde{x}_1^3))$  is always  $r_1$ , whereas the outcome of

$(a_1, (x_2^1, x_1^2, \tilde{x}_3^3))$  is always  $r_4$  with frequencies  $\tilde{D}' = \left\{ \underbrace{(a_1, (x_2^1, x_1^2, \tilde{x}_3^3), r_4)}_{9 \text{ times}}, \underbrace{(a_1, (x_2^1, x_1^2, \tilde{x}_1^3), r_1)}_{2 \text{ times}} \right\}$ .

According to the data in  $\tilde{D}'$ , the recorded frequency of  $\tilde{x}_1^3$  would be  $\frac{2}{11}$ , whereas that of  $\tilde{x}_3^3$  is  $\frac{9}{11}$ . The evaluation of  $a_1$  given  $(x_2^1, x_1^2, \tilde{x}_1^3)$  and  $(x_2^1, x_1^2, \tilde{x}_3^3)$ , respectively, aggregated across the realizations of  $\tilde{X}^3$  becomes  $V_{\tilde{D}'}(a_1 | (x_2^1, x_1^2)) = \frac{2}{11}u(r_1) + \frac{9}{11}u(r_4)$ . The difference

$$VNC(a_1, \eta) := V_{\tilde{D}'}(a_1 | (x_2^1, x_1^2)) - V_{D'}(a_1 | \eta, (x_2^1, x_1^2))$$

captures the value of identifying the new category  $\tilde{X}^3$ . It represents the decision maker's willingness to pay in order to get access to a new category which perfectly explains the variability of outcomes of action  $a_1$ . For  $\eta \neq 0$ , it is positive (negative) and depends positively (negatively) on the degree of relevance of the unobserved category,  $\eta$ , whenever the decision maker is sufficiently pessimistic (optimistic), i.e.,

$$\frac{2}{11}u(r_1) + \frac{9}{11}u(r_4) \geq (\leq) \alpha_{D'}u(r_1) + (1 - \alpha_{D'})u(r_4).$$

Once the willingness to pay has been identified, it can be compared to the actual cost  $c$  of identifying and applying the additional categorization, such as research cost, the cost of measuring and recording the relevant characteristics, storing the data, as well as complexity cost related to the increased amount of information to be processed by the decision maker.

This discussion shows that our framework allows us to also formally model the perception of “other” categories, as well as the perception of “other” characteristics. This perception relies on beliefs about the nature and the structure of the data, which cannot however be inferred from the data set itself. Thus, just as with the perception of unawareness about other characteristics, the perception of unawareness about other categories will be an individual subjective feature of the decision-maker and may depend on the context in which learning is taking place. From a formal point of view, however, our representation derived in Proposition 1 can be used to evaluate actions given such unawareness and can also capture the fact that unawareness will disappear, once the relevant categories have been identified.<sup>34</sup>

---

the decision maker to determine the willingness to pay for research given the classification costs associated with it.

<sup>34</sup>Grant *et al.* (2020) model the perception of ambiguity due to unawareness of propositions, as well as the process by which such ambiguity diminishes as the decision-maker's awareness increases.



## 5 Concluding remarks: From characteristics to states

In an ideal Savagean world, data would be perfectly adapted to the description of uncertainty by a set of states of the world. In particular, the decision maker would be aware and know all contingencies and there would be no unobserved categories, or characteristics. Notably, the following three conditions would be satisfied:

- (i) The set  $X_D$  corresponds exactly to the Savagean state-space  $S$ ,  $X_D = S$ .
- (ii) Actions are functions from states to outcomes:  $a : S \rightarrow R$ . For each action,  $a \in A_D$ , and each characteristic,  $x \in X_D$ , exactly one outcome is observed in the data, i.e.,  $\text{supp}(\rho_D(\cdot | a, x))$  is a singleton for all  $a$  and  $x$ .
- (iii) There are no redundant categories and characteristics, i.e., for each  $t \in T$ ,  $x^t \neq \tilde{x}^t$  implies that there is an  $a \in A_D$  and  $x^{-t} \in \Pi_{\tau \neq t} X_D^\tau$  such that  $\rho_D(\cdot | a, (x^t, x^{-t})) \neq \rho_D(\cdot | a, (\tilde{x}^t, x^{-t}))$ .

Suppose that the number of relevant contingencies is finite and that within a finite number of observations, all combinations  $(a, x)$  are observed in the data. If conditions (i)–(iii) are satisfied, arrival of new data corresponds to the case of statistical learning as described in Section 4.1. Since all possible characteristics have already been observed, no surprises occur. Thus, as the number of observations goes to  $\infty$ ,  $\gamma_D \rightarrow 0$  and the decision maker eventually behaves like an expected utility maximizer w.r.t. the state space  $S = X_D$  with probabilities informed by the limit frequencies recorded in the data.

In general there will be few decision situations which would satisfy the conditions listed above. Thus, a decision maker who wishes to be a Savagean, but is faced with empirical data that do not perfectly fit the desiderata will have to learn the best approximation of such a model given the available evidence.

The first type of learning was discussed in Section 4.2 and concerns becoming aware of new characteristics. We can model awareness of such unawareness by using a placeholder characteristic  $x_o$  which is taken into account for the evaluation of actions. As explained above, such learning increases the set of relevant characteristics, while at the same time increasing the degree of unawareness concerning the existence of "other" yet unobserved characteristics.

The second type of learning concerns the learning of new categories discussed in Section 4.3. Indeed, for a Savagean decision maker, an action which results in two distinct outcomes for a given state entails a contradiction and signals that the state-space is not well-specified. Call a data set  $D$  consistent if for each  $a$  and  $x$ ,  $a$  has resulted in a single outcome in combination with  $x$  and thus,  $\text{supp}(\rho_D(\cdot | a, x))$  is a singleton. Otherwise, we call the data set inconsistent. Let

$$X_D^C = \{x | \text{supp}(\rho_D(\cdot | a, x)) \text{ is not a singleton for some } a\}$$

be the set of characteristics for which such indeterminacy of outcomes has been generated. From the point of view of a Savagean decision maker, these are the

characteristics in need of refinement, if they were to represent states of the world. The existence of such characteristics signals the decision maker’s awareness that he is unaware of some relevant categories, without knowing explicitly what those could be. Such awareness of unawareness, in a natural way, also leads to ambiguity: the decision maker assigns multiple payoffs to an already observed  $(a, x)$  combination, rather than using the generated frequency of outcomes in the data.

In turn, observing a relevant category  $t^{new}$  such that  $supp(\rho_D(\cdot | a, (x, x^{new})))$  is a singleton for each realization of the characteristic  $x^{new}$ , restores consistency of the data set, but might generate ambiguity if the relevant characteristics have not been measured for past observations.

As the decision maker learns new categories and thus, the elements of the support of  $\rho_D(\cdot | a, x)$  can be attributed to distinct characteristics, and as sufficient observations of the so-refined characteristics are gathered, so that the number of missing observations becomes negligible, the two types of ambiguity related to categories also vanish.

Finally, new measurement methods might lead to the observation of new categories, even though these might appear redundant given the empirical information available. In particular, if  $\rho_D(\cdot | a, (x^t, x^{-t})) = \rho_D(\cdot | a, (\tilde{x}^t, x^{-t}))$  holds for all  $x^t, \tilde{x}^t \in X^t$ , all  $x^{-t} \in \Pi_{\tau \neq t} X_D^\tau$  and all  $a \in A_D$ , the decision maker may decide that the relevant state-space  $S = \Pi_{\tau \neq t} X_D^\tau$  is a sufficient description of the underlying uncertainty, all be it coarser than the one suggested by the data set,  $X_D$ .

Whether or not such coarsening of the state-space is warranted will be an empirical question. As data accumulate, such coarsening might need to be reversed, as new observations might result in an inconsistent data set signaling that category  $t$  is not redundant after all. The process described above would then repeat.

The preceding discussion is closely related to the literature on unawareness. Notably, the type of learning described in Section 4.2 corresponds to the decision maker initially perceiving a reduction of the actual state-space, which then expands to take into account new contingencies, see e.g., Grant and Quiggin (2013a,b); Grant *et al.* (2017). In contrast, learning new categories (Section 4.3) models an initial situation of coarsening of the state-space, which is sequentially refined, see, e.g., Grant and Quiggin (2006). The works of Karni and Vierø (2013); Karni and Vierø (2017) and Vierø (2018) also model expansion of the state-space, though one related to the discovery of new acts or new outcomes, as opposed to learning new characteristics. In this paper, we provide a unified framework which allows us to capture these phenomena and relate them to empirical data.

In Eichberger and Guerdjikova (2022), we will study in detail how new data affect the subjective parameters of the representation as well as the relation of our approach to machine learning.

## 6 Appendix: Proofs

### Proof of Lemma 1:

Note that for each set of outcome distributions  $R \in \mathfrak{R}_D$ , there is a constant action  $\bar{a}_R$  which has  $R$  as the set of outcomes for each  $x \in X_D$ ,  $a(x) \equiv R$ . The

Dirac measures  $\delta_R \in \Delta(\mathcal{R}_D)$  are elements of  $\mathfrak{M}(\mathbf{A}_D, f_D)$  since for each  $R \in \mathfrak{R}_D$ ,  $\bar{a}_R \in \mathbf{A}_D$ .

Moreover, the set of Dirac measures  $\{\delta_R \mid R \in \mathfrak{R}_D\}$  are the extreme points of the simplex  $\Delta(\mathcal{R}_D)$ . By Carathéodory's theorem, every  $m \in \Delta(\mathcal{R}_D)$  can be obtained as a convex combination of these extreme points.  $\square$

We prove the results of Theorem 1 using a sequence of Propositions. Combining the result of Jaffray (1989) with the implications of the first three axioms in the Anscombe-Aumann framework, see Kreps (1988, p. 102), we obtain:

**Proposition 1.** *Preferences  $\succsim$  on  $\mathcal{A}_D$  satisfy Axioms 1 – 3 iff there exist functions  $U : \mathcal{R}_D \rightarrow \mathbb{R}$  and  $U_o : \mathfrak{R}_D \rightarrow \mathbb{R}$  such that for  $a = (m_a, R_a^o)$  and  $b = (m_b, R_b^o)$*

$$(m_a, R_a^o) \succsim (m_b, R_b^o) \text{ iff } \sum_{R \in \text{supp}(m_a)} m_a(R) U(R) + U_o(R_a^o) \geq \sum_{R \in \text{supp}(m_b)} m_b(R) U(R) + U_o(R_b^o).$$

Furthermore,  $U_o$  is affine and  $U$  and  $U_o$  are unique up to a positive-affine transformation with a common multiplication factor  $z_1 > 0$ .

**Lemma 3.** *Assume that preferences  $\succsim$  satisfy Axioms 1–4.*

(i) For some singleton sets  $\{\bar{\rho}\}$  and  $\{\underline{\rho}\} \in \mathcal{R}_D$  and for any  $\tilde{m} \in \Delta(\mathcal{R}_D)$ ,

$$(\tilde{m}, \{\bar{\rho}\}) \succsim (\tilde{m}, \{\rho\}) \succsim (\tilde{m}, \{\underline{\rho}\}) \quad (8)$$

holds for all singleton sets  $\{\rho\} \in \mathcal{R}_D$ .

(ii) For  $\bar{m}$  and  $\underline{m}$  satisfying  $\bar{m}(\{\bar{\rho}\}) = 1$  and  $\underline{m}(\{\underline{\rho}\}) = 1$ , and for any  $R^o \in \mathfrak{R}_D$ ,

$$(\bar{m}, \{R^o\}) \succsim (m, R^o) \succsim (\underline{m}, R^o),$$

holds for any  $m$  with  $m(\{\rho\}) = 1$  for some singleton  $\{\rho\} \in \mathcal{R}_D$ .

(iii) If

$$(\bar{m}, \{\bar{\rho}\}) \succ (\underline{m}, \{\underline{\rho}\})$$

there exists a unique  $\gamma_D \in [0, 1]$  such that

$$(\underline{m}, \{\bar{\rho}\}) \sim (\gamma_D \bar{m} + (1 - \gamma_D) \underline{m}, \gamma_D \{\bar{\rho}\} + (1 - \gamma_D) \{\underline{\rho}\}).$$

**Proof of Lemma 3:**

(i) Note that according to Corollary 1, by the finiteness of  $\mathcal{R}_D$  and the fact that  $\{\delta_r\} \in \mathcal{R}_D$  for each  $r \in \mathfrak{R}_D$ , we have that there exist singleton sets  $\{\bar{\rho}\}$  and  $\{\underline{\rho}\} \in \mathcal{R}_D$  such that for some (and thus, for any)  $\tilde{m} \in \Delta(\mathcal{R}_D)$ ,

$$(\tilde{m}, \{\bar{\rho}\}) \succsim (\tilde{m}, \{\rho\}) \succsim (\tilde{m}, \{\underline{\rho}\})$$

holds for all singleton sets  $\{\rho\} \in \mathcal{R}_D$ . We refer to  $\{\bar{\rho}\}$  and  $\{\underline{\rho}\}$  as the best and the worst singleton element of  $\mathcal{R}_D$ .

(ii) Given the statement of part (i), the implication of Axiom 4 is that for  $\bar{m}$  and  $\underline{m}$  satisfying  $\bar{m}(\{\bar{\rho}\}) = 1$  and  $\underline{m}(\{\underline{\rho}\}) = 1$ , and any  $m$  with  $m(\{\rho\}) = 1$  for some singleton  $\{\rho\} \in \mathcal{R}_D$ , we have for some (and thus for any)  $R^\circ \in \mathfrak{R}_D$ ,

$$(\bar{m}, \{R^\circ\}) \succsim (m, R^\circ) \succsim (\underline{m}, R^\circ),$$

i.e.,  $\bar{m}$  and  $\underline{m}$  are the best and the worst elements in  $\Delta(\mathcal{R}_D)$  among those assigning full mass to singleton sets.

(iii) To show part (iii), consider next the extended action defined by  $(\underline{m}, \{\bar{\rho}\})$ . We have, by Corollary 1,

$$(\bar{m}, \{\bar{\rho}\}) \succsim (\underline{m}, \{\bar{\rho}\}) \succsim (\underline{m}, \{\underline{\rho}\}).$$

If  $(\bar{m}, \{\bar{\rho}\}) \succ (\underline{m}, \{\underline{\rho}\})$ , by continuity, Axiom 3, there exists a unique  $\gamma_D \in [0, 1]$  such that

$$(\underline{m}, \{\bar{\rho}\}) \sim (\gamma_D \bar{m} + (1 - \gamma_D) \underline{m}, \gamma_D \{\bar{\rho}\} + (1 - \gamma_D) \{\underline{\rho}\})$$

Using Axioms 1–4 we obtain the following result.

**Proposition 2.** *Suppose that  $\succsim$  satisfy Axioms 1–4. If*

$$(\bar{m}, \{\bar{\rho}\}) \succ (\underline{m}, \{\bar{\rho}\}) \succ (\underline{m}, \{\underline{\rho}\}), \quad (9)$$

then  $\gamma_D$  satisfies  $\gamma_D \in (0, 1)$ . Furthermore, there exist functions  $U : \mathcal{R}_D \rightarrow \mathbb{R}$  and  $U_O : \mathfrak{R}_D \rightarrow \mathbb{R}$  such that for  $a = (m_a, R_a^\circ)$  and  $b = (m_b, R_b^\circ)$

$$\begin{aligned} (m_a, R_a^\circ) \succsim (m_b, R_b^\circ) \text{ iff} \\ (1 - \gamma_D) \sum_{R \in \text{supp}(m_a)} m_a(R) U(R) + \gamma_D U_O(R_a^\circ) \\ \geq (1 - \gamma_D) \sum_{R \in \text{supp}(m_b)} m_b(R) U(R) + \gamma_D U_O(R_b^\circ), \end{aligned}$$

where  $U$  is the function identified in Proposition 1 and there is a unique  $\gamma_D$  such that  $U_O(\{\rho\}) = U(\{\rho\})$  for all singleton sets  $\{\rho\} \in \mathcal{R}_D$  and  $U_O(R) = \frac{1 - \gamma_D}{\gamma_D} U_O(R)$  for any  $R \in \mathfrak{R}_D$ .

$U_O$  is affine and  $U$  and  $U_O$  are unique up to a positive-affine transformation with a common multiplication factor  $z_1 > 0$ .

### Proof of Proposition 2:

Using  $\gamma_D$  identified in part (iii) of Lemma 3 and the representation from Proposition 1, we have:

$$\begin{aligned} U(\{\underline{\rho}\}) + U_o(\{\bar{\rho}\}) &= \gamma_D U(\{\bar{\rho}\}) + (1 - \gamma_D) U(\{\underline{\rho}\}) \\ &\quad + \gamma_D U_o\{\bar{\rho}\} + (1 - \gamma_D) U_o\{\underline{\rho}\} \\ (1 - \gamma_D) [U_o(\{\bar{\rho}\}) - U_o(\{\underline{\rho}\})] &= \gamma_D [U(\{\bar{\rho}\}) - U(\{\underline{\rho}\})] \end{aligned} \quad (10)$$

When condition (9) holds, we have that  $\gamma_D \notin \{0, 1\}$  and we can rewrite (10) as:

$$[U_o(\{\bar{\rho}\}) - U_o(\{\underline{\rho}\})] = \frac{\gamma_D}{1 - \gamma_D} [U(\{\bar{\rho}\}) - U(\{\underline{\rho}\})] \quad (11)$$

By continuity, Axiom 3, we have that for any singleton  $\{\rho\} \in \mathcal{R}_D$ , there is a unique coefficient  $\lambda_\rho \in [0, 1]$  such that for any  $\tilde{m} \in \Delta(\mathcal{R}_D)$

$$(\tilde{m}, \lambda_\rho \{\bar{\rho}\} + (1 - \lambda_\rho) \{\underline{\rho}\}) \sim (\tilde{m}, \{\rho\})$$

and by Axiom 4, this is equivalent to the statement that for  $m$  such that  $m(\{\rho\}) = 1$  and any  $R^o \in \mathfrak{R}_D$ ,

$$(\lambda_\rho \tilde{m} + (1 - \lambda_\rho) \underline{m}, R^o) \sim (m, \{R^o\}).$$

Hence, normalizing, w.l.o.g.  $U(\{\underline{\rho}\}) = U_o(\{\underline{\rho}\}) = 0$  and  $U(\{\bar{\rho}\}) = 1$  and thus, by (11)  $U_o(\{\bar{\rho}\}) = \frac{\gamma_D}{1 - \gamma_D}$ , we obtain that for any  $\{\rho\} \in \mathcal{R}_D$ ,

$$\begin{aligned} U(\{\rho\}) &= \lambda_\rho \\ U_o(\{\rho\}) &= \frac{\gamma_D}{1 - \gamma_D} \lambda_\rho. \end{aligned}$$

Using the representation in Proposition 1, we thus obtain that for  $m_a$  and  $m_b$  which put their entire mass on singleton sets, i.e.,  $\text{supp}(m_a), \text{supp}(m_b) \subseteq \{\{\rho\} \in \mathcal{R}_D\}$  and  $R_a^o = \{\rho_a\}$ ,  $R_b^o = \{\rho_b\}$  for some singletons  $\{\rho_a\}, \{\rho_b\} \in \mathcal{R}_D$ , we have

$$\begin{aligned} (m_a, R_a^o) \succeq (m_b, R_b^o) &\text{ iff} \\ &\sum_{\{\rho\} \in \text{supp}(m_a)} m_a(\{\rho\}) U(\{\rho\}) + \frac{\gamma_D}{1 - \gamma_D} U(\{\rho_a\}) \\ &\geq \sum_{\{\rho\} \in \text{supp}(m_b)} m_b(\{\rho\}) U(\{\rho\}) + \frac{\gamma_D}{1 - \gamma_D} U(\{\rho_b\}) \\ &\text{ iff} \\ &(1 - \gamma_D) \sum_{\{\rho\} \in \text{supp}(m_a)} m_a(\{\rho\}) U(\{\rho\}) + \gamma_D U(\{\rho_a\}) \\ &\geq (1 - \gamma_D) \sum_{\{\rho\} \in \text{supp}(m_b)} m_b(\{\rho\}) U(\{\rho\}) + \gamma_D U(\{\rho_b\}). \end{aligned}$$

Note further than since  $U_o$  is affine, we have that for any  $\rho \in \Delta(R_D)$ ,

$$U_o(\{\rho\}) = \sum_{r \in R_D} \rho(r) U_o(\{\delta_r\}).$$

Since  $\{\delta_r\} \in \mathcal{R}_D$  for every  $r \in R_D$ , this implies, that we can set  $\underline{\rho} = \delta_{\underline{r}}$  and  $\bar{\rho} = \delta_{\bar{r}}$ , where  $\underline{r}$  is the "worst" and  $\bar{r}$ , the "best" outcome in  $R_D$ . It is then obvious that the two inequalities in (8) hold for all singleton sets  $\{\rho\} \in \mathfrak{R}_D$ , and we can

thus refer to  $\{\bar{\rho} = \delta_{\bar{r}}\}$  and  $\{\underline{\rho} = \delta_{\underline{r}}\}$  as the best and the worst singleton element of  $\mathfrak{R}_D$ . Thus, we can define  $u(r) = U_O(\{\delta_r\}) = \frac{1-\gamma_D}{\gamma_D} U_o(\{\delta_r\}) = U(\{\delta_r\})$  for every  $r \in R_D$  so as to obtain for any  $\rho \in \Delta(R_D)$

$$U_o(\{\rho\}) = \frac{\gamma_D}{1-\gamma_D} \sum_{r \in R_D} \rho(r) U(\{\delta_r\}) = \frac{\gamma_D}{1-\gamma_D} \sum_{r \in R_D} \rho(r) u(r) \quad (12)$$

and any  $\{\rho\} \in \mathfrak{R}_D$

$$U(\{\rho\}) = \sum_{r \in R_D} \rho(r) U(\{\delta_r\}).$$

Define the function  $U_O : \mathfrak{R}_D \rightarrow \mathbb{R}$  as follows. Let

$$U_O(\{\rho\}) = U(\{\rho\}) = \sum_{r \in R_D} \rho(r) U(\{\delta_r\}) = \sum_{r \in R_D} \rho(r) u(r) \quad (13)$$

for all  $\rho \in \Delta(R_D)$ . Provided that  $\gamma_D \neq 0$ , for any  $R \in \mathfrak{R}_D$ , we can define  $U_O(R) = \frac{1-\gamma_D}{\gamma_D} U_o(R)$  (note that by (12) and (12), this equality also holds for singletons  $\{\rho\} \in \mathfrak{R}_D$ ). The so-defined  $U_O(R) = \frac{1-\gamma_D}{\gamma_D} U_o(R)$  is a positive-affine transformation of  $U_o$  determined in Proposition 1. Indeed, we obtain that for  $a = (m_a, R_a^o)$  and  $b = (m_b, R_b^o)$

$$\begin{aligned} (m_a, R_a^o) \succeq (m_b, R_b^o) & \text{ iff} \\ & \sum_{R \in \text{supp}(m_a)} m_a(R) U(R) + U_o(R_a^o) \\ & \geq \sum_{R \in \text{supp}(m_b)} m_b(R) U(R) + U_o(R_b^o) \end{aligned}$$

iff

$$\begin{aligned} & \sum_{R \in \text{supp}(m_a)} m_a(R) U(R) + \frac{\gamma_D}{1-\gamma_D} U_O(R_a^o) \\ & \geq \sum_{R \in \text{supp}(m_b)} m_b(R) U(R) + \frac{\gamma_D}{1-\gamma_D} U_O(R_b^o) \end{aligned}$$

iff

$$\begin{aligned} & (1-\gamma_D) \sum_{R \in \text{supp}(m_a)} m_a(R) U(R) + \gamma_D U_O(R_a^o) \\ & \geq (1-\gamma_D) \sum_{R \in \text{supp}(m_b)} m_b(R) U(R) + \gamma_D U_O(R_b^o). \square \end{aligned}$$

A consequence of the last part of the proof of Proposition 2 is that  $\{\bar{\rho}\}$  and  $\{\underline{\rho}\}$  are also the best and the worst singleton elements of  $\mathfrak{R}_D$  and that each of them can be taken to be a Dirac measure on a single outcome,  $\bar{\rho} = \delta_{\bar{r}}$ ,  $\underline{\rho} = \delta_{\underline{r}}$ , where  $\bar{r}$  and  $\underline{r}$ , are respectively the best and the worst outcome in  $R_D$ .

**Corollary 3.** *The two inequalities in (8) hold for all singleton sets  $\{\rho\} \in \mathfrak{R}_D$ . Furthermore, one can set  $\underline{\rho} = \delta_{\underline{r}}$  and  $\bar{\rho} = \delta_{\bar{r}}$ , for some  $\underline{r}$  and  $\bar{r} \in R_D$ . Finally, there exists a utility function over outcomes  $u : R_D \rightarrow \mathbb{R}$  which is unique up to a positive-affine transformation and satisfies:*

$$u(r) = U(\{\delta_r\}) = U_O(\{\delta_r\})$$

for any  $r \in R_D$  and

$$\begin{aligned} U(\{\rho\}) &= \sum_{r \in R_D} u(r) \rho(r) \text{ for all } \{\rho\} \in \mathfrak{R}_D \\ U_O(\{\rho\}) &= \sum_{r \in R_D} u(r) \rho(r) \text{ for all } \{\rho\} \in \mathfrak{R}_D. \end{aligned}$$

A straightforward adaptation of Jaffray (1989)'s result yields the following proposition.

**Proposition 3.** *The preference order  $\succsim$  on  $\mathcal{A}_D$  satisfies Axioms 1–5 iff there exist a non-constant functions  $w^d : \Delta(R_D) \times \Delta(R_D) \rightarrow \mathbb{R}$  non-decreasing w.r.t. the order  $\succsim^d$  in its arguments, a non-constant function  $w^o : \Delta(R_D) \times \Delta(R_D) \rightarrow \mathbb{R}$  non-decreasing w.r.t. the order  $\succsim^o$  in its arguments, and a unique weight  $\gamma_D \in (0, 1)$  such that*

$$\begin{aligned} (m_a; R_a) \succsim (m_b; R_b) \text{ iff} & \tag{14} \\ & \sum_{R \in \mathcal{R}_D} (1 - \gamma_D) m_a(R) w^d(\underline{\rho}_R, \bar{\rho}_R) + \gamma_D w^o(\underline{\rho}_{R_a}, \bar{\rho}_{R_a}) \\ & \geq \sum_{R \in \mathcal{R}_D} (1 - \gamma_D) m_b(R) w^d(\underline{\rho}_R, \bar{\rho}_R) + \gamma_D w^o(\underline{\rho}_{R_b}, \bar{\rho}_{R_b}). \end{aligned}$$

where  $w^d(\rho, \rho) = w^o(\rho, \rho)$  for all  $\rho$  such that  $\{\rho\} \in \mathfrak{R}_D$  and  $w^o(\rho, \rho) = \sum_{r \in R_D} u(r) \rho(r)$ , where  $u$  is the utility function over outcomes. The weights  $w^o$  and  $w^d$  are unique up to a positive-affine transformation with a common factor  $z_1 > 0$ .

### Proof of Proposition 3:

Since the comparison between any two sets,  $R$  and  $R'$  depends only on their best and worst elements, since the ordering of the best and the worst elements is the same as that of the singletons and coincides on the set  $\mathfrak{R}_D$  and since the best and the worst singletons on both  $\mathfrak{R}_D$  and  $\mathfrak{R}_D$  are given by  $\{\delta_{\bar{r}}\}$  and  $\{\delta_{\underline{r}}\}$ , we have that for any  $(m, R^o) \in \mathbf{A}_D^o$ ,

$$(m(\{\delta_{\bar{r}}\}) = 1, \{\delta_{\bar{r}}\}) \succsim (m, R^o) \succsim (m(\{\delta_{\underline{r}}\}) = 1, \{\delta_{\underline{r}}\}),$$

or in the notation of Lemma 3,

$$(\bar{m}, \{\bar{\rho}\}) \succsim (m, R^o) \succsim (\underline{m}, \{\underline{\rho}\})$$

By Corollary 1, we further have:

$$(\bar{m}, R^o) \succsim (m, R^o) \succsim (\underline{m}, R^o) \tag{15}$$

while by the non-triviality condition in Axiom 1, we have that there is an  $(m, R^o)$  for which either:

$$\begin{aligned} (m, R^o) &\succ (\underline{m}, R^o) \text{ or} \\ (\bar{m}, R^o) &\succ (m, R^o). \end{aligned}$$

If  $(m, R^o) \succ (\underline{m}, R^o)$ , we have by Corollary 1 that

$$(\bar{m}, \{\bar{\rho}\}) \succ (\underline{m}, R^o) \succeq (\underline{m}, \{\underline{\rho}\})$$

and thus,  $(\bar{m}, \{\bar{\rho}\}) \succ (\underline{m}, \{\underline{\rho}\})$ , whereas if  $(\bar{m}, R^o) \succ (m, R^o)$ , we have

$$(\bar{m}, \{\bar{\rho}\}) \succeq (\bar{m}, R^o) \succ (\underline{m}, \{\underline{\rho}\})$$

and thus, again  $(\bar{m}, \{\bar{\rho}\}) \succ (\underline{m}, \{\underline{\rho}\})$ . It follows that the  $\gamma_D$  identified in Lemma 3 is unique.

Next, observe that

$$(\bar{m}, \{\bar{\rho}\}) \sim (\underline{m}, \{\bar{\rho}\})$$

would contradict the non-triviality assumption imposed by Axiom 1. Indeed, it would imply, by Corollary 1 and by equation (15),

$$(\bar{m}, R^o) \sim (m, R^o) \sim (\underline{m}, R^o)$$

for all  $m \in \Delta(\mathcal{R}_D)$  and all  $R^o \in \mathfrak{R}_D$ , in contradiction to Axiom 1. It follows that  $(\bar{m}, \{\bar{\rho}\}) \succ (\underline{m}, \{\bar{\rho}\})$ .

Next, assume that there is an  $m \in \Delta(\mathcal{R}_D)$  and  $R$  and  $R' \in \mathfrak{R}_D$  such that  $(m, R) \succ (m, R')$ . Thus,

$$(\underline{m}, \{\bar{\rho}\}) \succ (\underline{m}, \{\underline{\rho}\}),$$

which by Lemma 3 implies that  $\gamma_D \in (0, 1)$  and, thus, the representation of Proposition 2.

Thus, we can set  $U(R) = w^d(\underline{\rho}_R, \bar{\rho}_R)$  and  $U_O(R) = w^o(\underline{\rho}_R, \bar{\rho}_R)$ . By Proposition 2, these functions coincide on singleton sets  $\{\rho\} \in \mathcal{R}_D$ , i.e.,

$$w^d(\rho, \rho) = w^o(\rho, \rho)$$

and by Corollary 3, we can thus set  $w^o(\rho, \rho) = \sum_{r \in R_D} u(r) \rho(r)$ , where  $u$  is the utility function over outcomes identified in the Corollary. The uniqueness of the functions  $w^o$  and  $w^d$  follows from the respective uniqueness of  $U$ ,  $U_O$  and  $u$ .

Hence the result of the Proposition obtains, provided that there is an  $m \in \Delta(\mathcal{R}_D)$  and  $R$  and  $R' \in \mathfrak{R}_D$  such that  $(m, R) \succ (m, R')$  holds.

To complete the proof thus, suppose in a manner of contradiction that there are no  $m \in \Delta(\mathcal{R}_D)$ ,  $R$  and  $R' \in \mathfrak{R}_D$  such that  $(m, R) \succ (m, R')$ . We then have

$$(\underline{m}, \{\bar{\rho}\}) \sim (\underline{m}, \{\underline{\rho}\}),$$

resulting in  $\gamma_D = 0$ . Furthermore,  $U_o(\{\delta_{\bar{r}}\}) = U_o(\{\delta_r\})$ . Axiom 5 then implies,  $U_o(R^o) = U_o(\{\delta_r\})$  for every  $R^o \in \mathfrak{R}_D$ . But by Axiom 4, we then obtain that for



any  $\lambda \in [0, 1]$ , any  $\{\rho\} \in \mathcal{R}_D$ , and  $m$  and  $m'$  such that  $m(\{\rho_1\}) = \lambda$ ,  $m(\{\rho_2\}) = 1 - \lambda$ ,  $m'(\{\rho\}) = 1$ ,

$$(\bar{m}, R^o) \sim (\underline{m}, R^o).$$

for some and thus, by Corollary 1, for any  $R^o \in \mathfrak{R}_D$ . At the same time, Axiom 5 gives us:

$$(\bar{m}, R^o) \sim (m, R^o) \sim (\underline{m}, R^o)$$

for any  $m \in \Delta(\mathcal{R}_D)$  for any  $R^o \in \mathfrak{R}_D$  in contradiction to the non-triviality assumption in Axiom 1.  $\square$

### **Proof of Theorem 1**

By representation (14), we can write for any  $R \in \mathfrak{R}_D$ ,

$$U_O(R) = \alpha_D^o \left( \underline{\rho}_R, \bar{\rho}_R \right) w^o(\bar{\rho}_R, \bar{\rho}_R) + \left( 1 - \alpha_D^o \left( \underline{\rho}_R, \bar{\rho}_R \right) \right) w^o(\underline{\rho}_R, \underline{\rho}_R) = U_O \left( \left\{ \underline{\rho}_R, \bar{\rho}_R \right\} \right) \quad (16)$$

and for any  $R \in \mathcal{R}_D$ ,

$$U(R) = \alpha_D \left( \underline{\rho}_R, \bar{\rho}_R \right) w^o(\bar{\rho}_R, \bar{\rho}_R) + \left( 1 - \alpha_D \left( \underline{\rho}_R, \bar{\rho}_R \right) \right) w^o(\underline{\rho}_R, \underline{\rho}_R) = U \left( \left\{ \underline{\rho}_R, \bar{\rho}_R \right\} \right) \quad (17)$$

It is thus sufficient to determine  $U_O$  and  $U$  for sets with two elements (the case  $\underline{\rho}_R = \bar{\rho}_R$  has already been discussed above).

Consider thus  $\rho, \rho' \in \Delta(R_D)$  with  $\rho \succ \rho'$ , the corresponding set of these two outcome distributions  $R_a = \{\rho, \rho'\}$  and the singleton set  $R_b = \{\alpha\rho + (1 - \alpha)\rho'\}$  for some  $\alpha \in [0, 1]$ . By continuity, Axiom 3, there exists a unique  $\alpha$  such that for some (and then all)  $m \in \Delta(\mathcal{R}_D)$ ,  $(m, R_a) \sim (m, R_b)$ . By Proposition 3, we then have:

$$\begin{aligned} U_O(R_a) &= \alpha(\rho', \rho) \sum_{r \in R_D} u(r) \rho(r) + (1 - \alpha(\rho', \rho)) \sum_{r \in R_D} u(r) \rho'(r) \\ &= \alpha \sum_{r \in R_D} u(r) \rho(r) + (1 - \alpha) \sum_{r \in R_D} u(r) \rho'(r) = U_O(R_b) \end{aligned}$$

and, thus,  $\alpha_D^o(\rho', \rho) = \alpha$ .

By Axiom 6, for any  $\rho'', \rho''' \in \Delta(R_D)$  with  $\rho'' \succ \rho'''$ , the corresponding set of these two outcome distributions  $R_c = \{\rho'', \rho'''\}$  and the singleton set  $R_d = \{\alpha\rho'' + (1 - \alpha)\rho'''\}$  we have  $(m, R_c) \sim (m, R_d)$ . We thus obtain  $\alpha_D^o(\rho''', \rho'') = \alpha_D^o(\rho', \rho) = \alpha$  for any  $\rho, \rho', \rho''$  and  $\rho'''$ . It follows that

$$\alpha_D^o \left( \underline{\rho}_R, \bar{\rho}_R \right) = \alpha$$

for all  $\underline{\rho}_R$  and  $\bar{\rho}_R$  and thus, for all  $R \in \mathfrak{R}_D$ . Setting  $\alpha_D^o = \alpha$  thus implies that for any  $R \in \mathfrak{R}_D$ ,

$$U_O(R) = \alpha_D^o \max_{\rho \in R} \sum_{r \in R_D} u(r) \rho(r) + (1 - \alpha_D^o) \min_{\rho \in R} \sum_{r \in R_D} u(r) \rho(r) \quad (18)$$

and thus, the optimism parameter for unobserved characteristics  $x_o$  does not depend on the set of outcomes  $R$ .

The argument that Axiom 7 implies that there exists an  $\alpha_D \in [0, 1]$  such that for any  $R \in \mathcal{R}_D$ ,

$$U(R) = \alpha_D \max_{\rho \in R} \sum_{r \in R_D} u(r) \rho(r) + (1 - \alpha_D) \min_{\rho \in R} \sum_{r \in R_D} u(r) \rho(r) \quad (19)$$

is analogous and thus omitted.

Combining the representation in Proposition 2 with the expressions in (18) and (19) gives the desired representation.  $\square$

## References

- ARAD, A. and GAYER, G. (2012). Imprecise Data Sets as a Source of Ambiguity: A Model and Experimental Evidence. *Management Science*, **58**, 188–202.
- BACH, F., ALBERTINI, R., JOO, P., ANDERSON, J. and BORTIN, M. (1968). Bone-marrow transplantation in a patient with the Wiskott-Aldrich syndrome. *Lancet*, **7583** (2), 1366–1369.
- BILLOT, A., GILBOA, I., SAMET, D. and SCHMEIDLER, D. (2005). Probabilities as Similarity-Weighted Frequencies. *Econometrica*, **73**, 1125–1136.
- BORTIN, M. M. (1970). A Compendium of Reported Human Bone Marrow Transplants. *Transplantation*, **9** (6), 571 – 587.
- CHUNG, Y., HAAS, P. J., UPFAL, E. and KRASKA, T. (2018). Unknown examples & machine learning model generalization. *ArXiv*, **abs/1808.08294**.
- DEATON, A. and CARTWRIGHT, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, **210** (C), 2–21.
- DOMINIAK, A. and GUERDJIKOVA, A. (2021 forthcoming). Pessimism and optimism towards new discoveries. *Theory and Decision*.
- DOW, J. and DA COSTA WERLANG, S. R. (1992). Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica*, **60** (1), 197–204.
- EGG, D. and ET AL. (2021). Therapeutic options for CTLA-4 insufficiency. *The Journal of allergy and clinical immunology*, p. DOI:<https://doi.org/10.1016/j.jaci.2021.04.039>.
- EICHBERGER, J. and GUERDJIKOVA, A. (2010). Case-based belief formation under ambiguity. *Mathematical Social Sciences*, **60**, 161–177.
- and — (2013). Ambiguity, Data and Preferences for Information - A Case-Based Approach. *Journal of Economic Theory*, **148**, 1433–1462.
- and — (2022). Data and learning. *Working Paper*.

- and PASICHNICHENKO, I. (2021). Decision-making with partial information. *Journal of Economic Theory*, **198**, 105369.
- ELLSBERG, D. (1961). Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics*, **75**, 643–669.
- GILBOA, I., MINARDI, S. and SAMUELSON, L. (2017). Cases and Scenarios in Decision under Uncertainty.
- , —, — and SCHMEIDLER, D. (2020). States and contingencies: How to understand savage without anyone being hanged. *Revue économique*, **71** (2), 365–386.
- and SCHMEIDLER, D. (2001). Reaction to Price Changes and Aspiration Level Adjustments. *Review of Economic Design*, **6**, 215–223.
- and — (2002). *A Theory of Case-Based Decisions*. Cambridge,: Cambridge University Press.
- GRABISCH, M. (2016). *Set Functions, Games and Capacities in Decision Making, Theory and Decision Library C*, vol. 46. Springer, game theory, social choice, decision theory, and optimization edn.
- GRANOT, N. and STROB, R. (2020). History of hematopoietic cell transplantation: challenges and progress. *Haematologica*, **105** (12), 2716–2729.
- GRANT, S., GUERDJIKOVA, A. and QUIGGIN, J. (2020). Ambiguity and awareness: A coherent multiple prior model. *The B.E. Journal of Theoretical Economics*.
- , MENEGHEL, I. and TOURKY, R. (2017). Learning under unawareness. *Working Paper*.
- and QUIGGIN, J. (2006). Learning and discovery. *Risk and Sustainable Management Group Working Papers 151174*.
- and — (2013a). Bounded awareness, heuristics and the precautionary principle. *Journal of Economic Behavior & Organization*, **93**, 17–31.
- and — (2013b). Inductive reasoning about unawareness. *Economic Theory*, **54** (3), 717–755.
- JAFFRAY, J.-Y. (1989). Linear Utility Theory for Belief Functions. *Operations Research Letters*, **8**, 107–112.
- KANG, E., MARCIANO, B. E., DERAVIN, S., ZAREMBER, K., HOLLAND, S. M. and MALECH, H. L. (2011). Chronic Granulomatous Disease: Overview and Hematopoietic Stem Cell Transplant. *The Journal of allergy and clinical immunology*, **127**, 1319–1326.
- KARNI, E. (2022). A theory-based decision model. *Journal of Economic Theory*, **201**.

- and VIERØ, M.-L. (2013). Reverse Bayesianism: A choice-based theory of growing awareness. *American Economic Review*, **103**, 2790–2810.
- and VIERØ, M.-L. (2017). Awareness of Unawareness: A Theory of Decision Making in the Face of Ignorance. *Journal of Economic Theory*, **168**, 301–328.
- KREPS, D. M. (1988). *Underground Classics in Economics*, Bolder: Westview Press.
- SAVAGE, L. J. (1954). *Foundations of Statistics*. New York: Wiley.
- SHORT, W. M. (2018). The spacial metaphors of ambiguity in roman culture. In M. Fontaine, C. McNamara and W. M. Short (eds.), *Quasi Labor Intus: Ambiguity in Latin Literature*, Leiden: Brill, p. 1–25.
- VIERØ, M.-L. (2018). An intertemporal model of growing awareness. *Working Paper*.
- VON NEUMANN, J. and MORGENSTERN, O. (1944). *The Theory of Games and Economic Behavior*. New Jersey: Princeton University Press.