The Behavioral Implications of Statistical Decision Theory

Bruno Furtado*

Department of Economics, Columbia University

Job Market Paper
Latest version here

November 22, 2022

Abstract

Statistical decision theory (SDT), pioneered by Abraham Wald, models how decision makers use data for decisions under uncertainty. Despite its prominence in information economics and econometrics, SDT has not been given formal choice-theoretic or behavioral foundations. This paper axiomatizes preferences over decision rules and experiments for a broad class of SDT models. The axioms show how certain seemingly-natural decision rules are incompatible with this broad class of SDT models. Using my representation result, I develop a methodology to translate axioms from classical decision-theory, a la Anscombe and Aumann (1963), to the SDT framework. I illustrate its usefulness by translating various classical axioms to refine my baseline SDT framework into more specific SDT models, some of which are novel to SDT. I also discuss foundations for SDT under other kinds of choice data.

JEL classification: D01

Keywords: statistical decision theory, uncertainty, decision rules, Blackwell experiments

1 Introduction

Statistical decision theory (henceforth SDT) models decisions under uncertainty as an optimization problem. Nature selects a true *parameter*, or state of the world. Meanwhile, the decision maker (DM) chooses *decision rules* (strategies) and designs *experiments* (information structures) to maximize her objective function, without knowing which parameter Nature chose. A signal is then drawn according to a distribution determined by the parameter and the experiment. The DM observes the signal and takes the *action* prescribed by the decision rule. The ultimate pay-off for the DM is based on the realised action and Nature's chosen parameter. Models of this kind are ubiquitous in information economics. A clear example is the rational inattention literature, where a DM designs an experiment subject to constraints

*Please send any questions and comments to bruno.furtado@columbia.edu.
I am indebted to Navin Kartik, Mark Dean and Jose Luis Montiel Olea for invaluable guidance during the whole process of writing this paper. Elliot Lipnowski, Mauricio Couri, Gil Riella, Laura Doval, Evan Sadler and Andrew Olenski provided precious and detailed feedback. I would also like to thank the attendees of the Microeconomic Theory Colloquium of Columbia University and the 2022 SMYE, especially Yeon-Koo Che, Qingmin Liu, Duarte Gonçalves, Yu Fu Wong and Humberto Moreira, for useful questions and constructive criticism. Any remaining errors are mine.

on its informativeness, simultaneously to choosing a decision rule. Moreover, many econometric problems can be formulated in statistical decision theoretic terms.

Given their importance, surprisingly little is known about the behavioral assumptions implicit in various statistical decision theoretic models. That is, there are no formal results describing which choices over decision rules and experiments are consistent with the predictions of SDT models. The main goal of this paper is to set SDT on a rigorous axiomatic — and behaviorally falsifiable — foundation. This is done by axiomatically characterizing *preferences* over the choice objects of SDT: pairs of decision rules and experiments. Formally, an experiment is a collection of probability distributions over signals, indexed by the parameters, and a decision rule is a function assigning an action to each possible signal arising from the experiment. The preferences over these objects, taken here as behavioral primitives, are assumed to have been elicited by observing the DM's choices from different feasible sets, following the revealed preference principle.

Although much work has been done to behaviorally characterize various models of decisions under uncertainty, existing results overwhelmingly focus on preferences (or choices) over *acts* – functions mapping the unknown states of the world directly to final consequences. This is the framework of Anscombe and Aumann (1963) and Savage (1954). SDT differs from traditional decision theory under uncertainty in that observable information – in the form of signals from experiments – is treated separately from the parameters, which determine final outcomes. Therefore, unlike the traditional decision theoretic framework, SDT considers information acquisition as an integral part of the model description. This makes it a more natural setting to model situations in which the DM expects to receive partial information about the parameter before making a decision.

To fix ideas, suppose our DM is a policymaker choosing whether to implement a costly social program (action a_1) or not (action a_0). The program should be implemented if and only if its mean value, θ, exceeds a known fixed cost c. Although the true value of the parameter θ is unknown, the DM can observe a sample of size n from a randomized controlled trial designed to assess the effectiveness of the program. Each sample is drawn independently from a normal distribution with mean θ and known variance. The full set of n samples is the observed signal, and the experiment can be characterized by a collection of random vectors, indexed by θ, having independent normally distributed components. An example of a decision rule could be to take action a_1 whenever the empirical mean of the signal is greater than c, and action a_0 otherwise. Suppose the DM can choose between two combinations of decision rules and experiments. The first pairs a sample of size 50 with the decision rule previously described, based on the empirical mean. The second pairs a sample of size 10 with a decision rule in which the policy is not implemented for any signal. A preference relation between these pairs could be, for instance, that the first combination is always chosen over the second one.

I impose axioms on the preference relation that fully characterize a DM who chooses as if using a general SDT framework, which subsumes most models in applications. This establishes a tight link between the predictions of such a model and behaviorally falsifiable constraints (axioms) on preferences, which in turn allows us to tell what sets of choices are compatible with the SDT framework.

Preferences that abide by such a framework are characterized by a state-dependent utility function u and a parameter aggregation functional I. The utility function $u : A \times \Theta \to \mathbb{R}$ — where Θ and A are the sets of parameters and actions respectively — determines the pay-offs of the actions at each parameter. For example, if A and Θ are subsets of $\mathbb{R}^n$, a common utility function is given by the negative of the *squared error*: $u(a, \theta) = -\|a - \theta\|_2$, where $\|\cdot\|_2$ denotes the Euclidian norm.

Using u, one can calculate the agent's risk functions, which define, for each parameter value, the expected utility implied by a decision rule–experiment pair. Formally, given a decision rule $\delta : X \to A$ – where X is the signal space – and an experiment $P = \{P_\theta : \theta \in \Theta\}$, I define the DM's risk function as the mapping $\theta \mapsto r_u(\delta, P)(\theta) \equiv \int_X u(\delta(x), \theta)\mathrm{d}P_\theta(x)$. Risk functions relate to lotteries, i.e., prospects involving objective probabilities. They represent taking the expected utility whenever the DM faces an objective distribution, while remaining completely agnostic about the subjective uncertainty captured by the parameters.

The second part of the representation, given by the aggregator I, summarizes the agent's attitude towards ambiguity, and can be viewed as an ordinal utility on the space of risk functions. It captures the way in which DMs deal with subjective uncertainty, i.e., the fact that they do not know the true parameter.

Formally, I characterize a DM who chooses decision rules and experiments to maximize the functional

$$V(\delta, P) = I(r_u(\delta, P)) = I\left(\left(\int_X u(\delta(x), \theta)\mathrm{d}P_\theta(x)\right)_{\theta\in\Theta}\right). \tag{1}$$

The aggregator I is only required to be continuous and monotone – basic properties that are satisfied by many models used in applications. Considering its particular form, I call this the *monotone risk aggregation* (MRA) model. Specific objective functions are obtained by specifying particular functional forms for I that satisfy these two properties.

Consider, for instance, two of the most well-known models in statistics and decision theory: the subjective expected utility (SEU) and the maximin expected utility (MEU) models. An SEU agent is assumed to have a prior belief π, which is a probability distribution on the set of parameters. Their ex-ante utility from decision rule δ and experiment P is $\int_\Theta \int_X u(\delta(x), \theta)\mathrm{d}P_\theta(x)\mathrm{d}\pi(\theta) = \int_\Theta r_u(\delta, P)\mathrm{d}\pi$. In the language of the MRA framework, I is the expectation operator with respect to the prior belief. On the other hand, an MEU agent maximizes the expected utility assuming the true parameter is the worst possible one for whichever decision rule they choose. Their ex-ante utility is given by $\min_{\theta\in\Theta} \int_X u(\delta(x), \theta)\mathrm{d}P_\theta(x) = \min_{\theta\in\Theta} r_u(\delta, P)(\theta)$, that is, $I(\cdot) = \min_{\theta\in\Theta}(\cdot)$. In both cases, I is monotone and continuous, thus SEU and MEU specialize the MRA model. While SEU and MEU models are widely applied in SDT, their exact behavioral implications were previously unknown. I provide characterizations of these and other models in Section 5.

The model in (1) is characterized by a set of six axioms on preferences. Two of those axioms are standard: they impose that preferences are rational (complete and transitive) and continuous. The remaining axioms are specific to my setting. The Consequentialism axiom states that the DM ultimately cares only about the parameter contingent probability distributions over actions induced by decision rules and experiments, not about the rules and experiments themselves. Independence of Irrelevant Parameters imposes that if one fixes the outcomes of two alternatives to be the same for every parameter but one, then preferences are completely determined by conditioning on the single parameter where they may differ. Taken together, these two axioms allow the ex-post utility u to be a function exclusively of actions and parameters, rather than depend on the particular decision rule and experiment being evaluated. Monotonicity says that if the DM prefers alternative 1 to alternative 2 conditional on any parameter being the truth, then she prefers alternative 1 unconditionally as well. This guarantees that the aggregator I is monotone. Finally, Conditional Mixture Independence requires that whenever the DM knows the true parameter, she chooses as if maximizing expected utility.

To see how the axioms can help us empirically test whether preferences satisfy the MRA model, recall the policymaker example above. For simplicity, suppose there are only two possible values of the parameter, $\theta_1 > c$ or $\theta_0 < c$. In principle, the DM can choose as her decision rule any signal-contingent distribution over the two possible actions. Suppose the decision rule that is strictly preferred by the DM is based on a likelihood ratio test: she specifies a significance level α and acts as if $\theta > c$ if, and only if, she rejects the null hypothesis that $\theta < c$. Consequently, she takes action a_1 (implements the program) whenever the null hypothesis is rejected, and takes action a_0 otherwise. This is an example of

(a) δ_1^*

$P_1^*(a\|\theta)$	a_0	a_1
θ_0	$1-\alpha_1$	α_1
θ_1	β_1	$1-\beta_1$

(b) δ_2^*

$P_2^*(a\|\theta)$	a_0	a_1
θ_0	$1-\alpha_2$	α_2
θ_1	β_2	$1-\beta_2$

(c) $\hat{\delta}_1$

$\hat{P}_1(a\|\theta)$	a_0	a_1
θ_0	$1-\alpha_2$	α_2
θ_1	β_1	$1-\beta_1$

(d) $\hat{\delta}_2$

$\hat{P}_2(a\|\theta)$	a_0	a_1
θ_0	$1-\alpha_1$	α_1
θ_1	β_2	$1-\beta_2$

Table 1: Conditional action distributions under different decision rules

an inference-based decision rule – for further discussion of such rules, see Manski (2021). Furthermore, the DM has different preferred significance levels α_i for different sample sizes, perhaps feeling that more informative experiments (larger n) permit more stringent standards for implementing the policy. Each of these significance levels induce different decision rules δ_i^*, which are assumed to be strictly preferred to any other rule that could be paired with experiment i.

It is a priori unclear whether the apparently reasonable choice behavior just described is compatible with any objective function in the mold of eq. (1). As it turns out, a simple argument shows that such preferences violate the combination of Consequentialism and Independence of Irrelevant Parameters, making them incompatible with the MRA model. Indeed, suppose we observe that the DM strictly prefers significance level α_1 for her likelihood ratio test when the sample size is n_1, and α_2 if it is n_2. Tables 1a and 1b show the conditional action distributions P_i^*, for $i = 1, 2$, resulting from decision rules δ_i^*, where β_i is the probability of type II error (i.e., the probability of wrongfully rejecting the policy). It is straightforward to construct alternative (mixed) decision rules $\hat{\delta}_1$ and $\hat{\delta}_2$, for sample sizes n_1 and n_2 respectively, such that $P_i^*(\cdot|\theta_1) = \hat{P}_i(\cdot|\theta_1)$ and $P_i^*(\cdot|\theta_0) = \hat{P}_{-i}(\cdot|\theta_0)$, as shown in tables 1c and 1d.

Since δ_i^* is the strictly preferred decision rule for experiment i, it is in particular strictly preferred to $\hat{\delta}_i$. Now note that the action distributions induced by each pair of decision rules, δ_i^* and $\hat{\delta}_i$, differ only on parameter θ_0. Moreover, the action distributions conditional on θ_0 induced by δ_1^* and δ_2^* are mirror images of each other. Therefore, either the DM cares about something other than parameter contingent action distributions, or preferences conditional on θ_0 depend on the action distribution at θ_1. In other words, if the DM satisfies Consequentialism, so that all that matters to her are the parameter contingent action distributions, then she must violate Independence of Irrelevant Parameters. The fact that we can not pinpoint exactly which of the axioms is violated in this example, is due to the fact that we only have data on two choice problems, rather than the full preference relation.

To obtain the MRA representation, I first recover the DM's utility function u from a particular incomplete binary relation that is implied by the DM's preferences. The aggregator I can then be viewed as a standard utility function on the space of the risk functions defined by u. In fact, using a direct analogy with consumer theory, risk functions can be interpreted as consumption bundles, with different parameters representing the different goods. The expected utility of a risk function at parameter $\theta \in \Theta$ is analogous to the quantity of good θ in the bundle. Therefore, showing the existence of an aggregator I is formally equivalent to proving the existence of a utility function in a generalized space of consumption bundles. This allows one to draw an analogy between a preference for hedging – where the DM intrinsically prefers risk functions that have a more balanced utility profile across parameters – to the concept of complementarity between different goods.

SEU agents act as if they can quantify the uncertainty about parameters with a single probability measure – in the terminology of Ellsberg (1961), they show no *ambiguity aversion*. Looking at it through the lens of consumer theory, they have additive utility across parameters, thus perceiving parameters as perfect substitutes. At the opposite end of the spectrum, MEU agents act as if they have no reason to believe one state is more likely than any other, and "play it safe" by planning for the worst case scenario. They have maximal ambiguity aversion. In the language of consumer theory, MEU agents have a parameter aggregator I that is similar to Leontieff preferences, implying perfect complements. This underscores the analogy between complementarity and ambiguity aversion that will be formalized in Section 4.

Characterizing the MRA model is an important step towards putting SDT on a sound axiomatic foundation. However, applications are usually couched on more structured models, which impose further constraints on DM's preferences. Fortunately, the representation of the MRA model is a key piece of a methodology that can be used to import representation results from the Anscombe-Aumann framework to SDT. Indeed, eq. (1) indicates that one can interpret risk functions as acts (mappings from states of the world to consequences), and that preferences over decision rule–experiment pairs induce a preference relation over such acts. Therefore, once I have characterized the MRA model, obtaining a representation of preferences in the SDT setting reduces to obtaining an Anscombe-Aumann representation of the corresponding preferences over risk functions. Using this technique, one can translate axioms on preferences over acts into axioms on preferences over decision rules and experiments. Hence if a model has a representation satisfying the analogue of eq. (1) in the Anscombe-Aumann setting (Cerreia-Vioglio et al., 2011), I can easily obtain its corresponding SDT representation.

The second piece needed to establish this toolkit is Lemma 3, which connects properties of the risk functionals r_u to those of the decision rules and experiments that define them. The lemma contains three separate statements. First, the risk function arising from a convex mixture of decision rules is the convex mixture of the risk functions induced by each individual decision rule. Second, the lemma describes, in terms of the decision rule and experiment that induces it, the risk function obtained by substituting the value of one risk function, on a given set of parameters, by the corresponding values of another risk function. Most axioms in the Anscombe-Aumann framework are stated in terms of these two operations on acts. Finally, the lemma also characterizes the set of decision rule–experiment pairs yielding constant risk functions. Constant acts also tend to be important ingredients for Anscombe-Aumann axioms. In summary, Lemma 3 translates the main ingredients of Anscombe-Aumann representations into the SDT framework.

From a purely technical perspective, Consequentialism transforms the problem of representing a preference over decision rules and experiments into representing a preference over parameter contingent action distributions. Much like I did with risk functions, I can also identify such parameter contingent distributions with Anscombe-Aumann acts. But while risk functions map parameters directly to parameter-independent consequences (utility units), the utility of an action distribution itself depends on the true parameter. Therefore, viewed through the lens of traditional decision theory, my methodology assists in generating representations of state-dependent preferences (over action distributions) from state-independent ones (over risk functions). This is a useful technique, because state-independent representations have been studied much more thoroughly than their state-dependent counterparts. The rationale behind developing such a methodology when studying SDT is that parameter-dependence is often the whole point of statistical applications — e.g., when performing inference.

After establishing the formal connection between the two frameworks, I apply my methodology to obtain behavioral characterizations in SDT, of some models for which axiomatizations already exist in the Anscombe-Aumann setting. These applications illustrate how my main results provide powerful tools than can be used to obtain the exact behavioral implications of a wide array of SDT models. Specifically, I provide characterizations of statistical decision theoretic versions of subjective expected utility (Anscombe and Aumann, 1963); multiple priors and maximin expected utility (Gilboa and Schmeidler, 1989; Stoye, 2011); and variational and multiplier preferences (Maccheroni et al., 2006; Hansen and Sargent, 2001; Strzalecki, 2011).

In some applications, data on preferences over decision rule–experiment pairs is not forthcoming. Hence, to expand the scope of my main results, I characterize the MRA model for two alternative types of behavioral data.

First, I consider data in the form of parameter dependent stochastic choices. This consists of action probabilities conditional on every parameter, and can be obtained by calculating action frequencies from repeated observations of choices from different decision problems. Such data is commonplace in the psychometric literature, and has recently received some attention in economics (Caplin and Martin, 2015; Caplin and Dean, 2015). I characterize when a parameter dependent stochastic choice function can be rationalized by a preference on decision rule–experiment pairs satisfying the MRA representation. This allows one to test whether choices are compatible with the MRA model without having to observe preferences over decision rules and experiments. I also discuss the identification problem using such data. That is, when can the particular decision rule and experiment chosen by the DM be recovered by observing only parameter dependent stochastic choice data. This turns out to be a problem of statistical identifiability of mixture models, for which answers are available in the econometrics literature.

Second, I examine the case where data comes in the form of two collections of preferences. One defines a preference over decision rules for each fixed experiment. This describes a DM who takes the experiment as given, and chooses the decision rule accordingly. The other collection consists of a preference over experiments for each menu M of decision rules. This models an agent, whom I call the Experimenter, who chooses an experiment, assuming that a decision rule will subsequently be chosen from M according to some choice procedure known to her in advance. I separately characterize the MRA model for each of these decision problems. These can be viewed as two separate DMs — one choosing the information structure, the other choosing the decision rule — or as two cross sections of the same DM's preferences over decision rule–experiment pairs.

It turns out that the axioms characterizing the MRA model for the DM who chooses over decision rules are closely related to those characterizing the same model for decision rule–experiment pairs. The same is true for the Experimenter, apart from two extra axioms. Consistency implies that the Experimenter is forward looking and correctly anticipates the decision rule that will be subsequently chosen from the feasible menu. Optimism says that the Experimenter assumes that if multiple decision rules might be chosen in the second stage of the decision process, the ultimate choice will be in her favor.

The remainder of the paper is structured as follows. Section 2 contextualizes the contributions of the present paper within the existing literature. Section 3 presents the decision theoretical setting and introduces some notation. Section 4 contains the main result: a representation of the MRA model. Section 5 develops a methodology that can be used to apply existing behavioral foundations of classic decision theory to SDT, and illustrates it with some applications. In Section 6 I characterize the MRA model for alternative data sets. All proofs are in the appendix.

2 Related Literature

Statistical decision theory was pioneered by Abraham Wald, who first applied it to the optimal choice of decision rules (Wald, 1939), and then to the design of experiments (Wald, 1947b). Wald himself framed his theory as a particular case of John von Neumann's theory of games, with Nature and the statistician as the players. As such, in SDT the statistician is endowed with an utility function that maps actions and unobserved parameters to ex-post pay-offs. Since the sampling distributions of experiments at every parameter are assumed to be known in advance, the statistician uses the expected utility criterion when assessing the ex-ante payoff at each parameter. Payoff aggregation across the parameter space — on which there is no objectively given probability distribution — can then be done in different ways. For example, Wald favored the maximin criterion: maximizing utility conditional on the worst-case parameter.[1]

This canonical model of statistical decision making has since provided the conceptual framework for many important results in mathematical statistics, such as David Blackwell's equivalency result for the economic comparison of experiments (Blackwell, 1953; Marschak and Miyasawa, 1968; Crémer, 1982), and Wald's complete class theorem (Wald, 1947a; Kuzmics, 2017). A comprehensive review of theoretical results and applications of SDT to mathematical statistics can be found in Inoue (2009).

SDT has also become one of the preferred languages of information economics. Among the many influential economics models couched in the SDT framework are the Bayesian persuasion literature initiated by Gentzkow and Kamenica (2011), and the rational inattention and costly information acquisition models studied by Sims (2003), Caplin and Dean (2015) and Matêjka and McKay (2015), among others. As for other areas of economic inquiry, Manski (2021) argues for the application of SDT to econometrics, provides a good summary of the relatively recent efforts in this direction, and outlines the remaining obstacles to this approach.

As was mentioned in the Introduction, models in SDT make predictions in terms of choices over decision rules and experiments. For example, Sims (2003) models a Bayesian agent who jointly chooses a decision rule and an experiments, subject to a constraint on the mutual information between the chosen experiment and the prior distribution over states of the world. The present paper characterizes the exact behavioral implications of this and many other SDT models, by axiomatizing preferences over decision rule–experiment pairs.

A different strand of the literature, which I will simply call decision theory, has sought to derive the behavioral implications of models of choice under uncertainty, usually by axiomatizing a DM's preferences over acts, i.e., functions from states of the world to final consequences. The early results most relevant for this paper were presented by Savage (1954) and Anscombe and Aumann (1963). Both characterize SEU agents, but unlike Savage, Anscombe and Aumann assume that acts' consequences consist of lotteries with known probability distributions. This introduces both objective and subjective probabilities into the model. In Section 5, I show that there is a natural formal connection between the SDT setting and the Anscombe and Aumann framework of preferences over acts. I then leverage this connection to develop a methodology that generates representation results in the SDT setting by importing analogous results from the Anscombe and Aumann framework.

A large literature has sought to characterize different models of choice under uncertainty by modifying the set of axioms put forth by Anscombe and Aumann (1963). I briefly cite a few that are specially relevant in the context of this paper. Gilboa and Schmeidler (1989) characterize the multiple priors expected utility (MPEU) model, where the DM has a set of prior beliefs over the states of the world, and picks the act yielding the best outcome according to the worst prior in this set. Maccheroni et al. (2006) generalize MPEU by characterizing preferences that can be represented by an elementary variational problem. Strzalecki (2011) provides a representation theorem for the multiplier preferences model proposed by Hansen and Sargent (2001). This is a special case of variational preferences, and can interpreted as modelling a DM who cares about the robustness of choices to deviations from a prior distribution. Cerreia-Vioglio et al. (2011) axiomatize a very general class of preferences, which subsumes all models cited in this paragraph and can be viewed as a decision theoretic analogue of the MRA model.

Some decision theory papers have studied statistical models from the perspective of the Anscombe and Aumann framework. One approach is to take the standard view that both signals and parameters of the SDT framework are contained in the state space of decision theory, and then proceed by proving results

[1] When the objective is to minimize loss rather than maximize utility, as in Wald's original formulation, maximin becomes minimax, since loss is the negative of utility.

in the Anscombe and Aumann setting. With this interpretation, any decision theoretic representation can be directly applied to statistical decision problems. However, this leaves out an important feature of SDT, which is the natural decomposition of the states into an informative but payoff irrelevant part (signals) and a payoff relevant but unobservable component (parameters). Cerreia-Vioglio et al. (2020) and Amarante (2009), for example, specifically use statistical applications to motivate representation theorems in the Anscombe and Aumann setting.

A second approach, exemplified by Epstein and Seo (2010), Cerreia-Vioglio et al. (2013) and Al-Najjar and De Castro (2014), also takes acts as primitives, and provides conditions on preferences under which the DM can be thought as having a parametric representation of the state space. In SDT terms, under such conditions preferences over acts can be used to elicit the DM's subjective statistical model (the experiment).

A different approach is taken by Stoye (2011, 2012), who characterizes versions of many widely used SDT models by essentially taking risk functions (also called utility acts) as decision theoretic primitives. As will become clear in Section 5, such an approach is complementary to the one I take in the present paper, and can be combined with my results to obtain novel representation theorems in the SDT framework.

Finally, the decision theory paper that comes closest to modelling choices between decision rules and experiments as I do here, is due to Jakobsen (2021). In it, the author axiomatizes the decision problems of two agents. The first chooses between experiments, knowing that for each realized signal an act will later be chosen by the second agent. The preferences of each agent depend on both agents' choices. This is similar to the approach taken in Section 6.2, except that Jakobsen (2021) works with acts rather than decision rules, focuses on Bayesian representations for both agents, and considers only the case of state-independent utility.

After separately characterizing each agent, Jakobsen (2021) gives conditions under which the utility functions and prior beliefs in the representations are identified. He also provides the conditions for when both representations coincide, which can be interpreted as the two decision problems describing the same agent. Such a case can thus be viewed as a DM who chooses an act-experiment pair, albeit under the constraint that each experiment must always be paired with the act that is optimal for it. By assuming richer observable data, I am able to characterize a single agent's preferences over decision rule–experiment pairs without any such constraints.

3 Setting and notation

Experiments. Consider a set Θ of parameters, assumed to be a separable topological space, with a σ-algebra Σ that includes the singletons. The set of all experiments the decision maker may be asked to choose from is denoted by $\mathcal{P}$. Each element $P \in \mathcal{P}$ is a function $\theta \mapsto P_\theta$ from parameters to probability distributions over a standard Borel signal space, i.e., a Polish space (separable complete metric space) endowed with its Borel σ-algebra. This can be viewed as the parameter contingent distribution of a random variable, denoted by X_P. Since all signal spaces are standard Borel, there is no loss of generality in assuming that all $P \in \mathcal{P}$ are defined on a common Polish *sample space* X, and I do so for the remainder of the paper. To guarantee that the sample space is rich enough to incorporate signals from a variety of distributions, I assume that X is uncountable.

For any Polish space Y with the usual topology, denote by $\Delta(Y)$ the set of probability distributions on its Borel σ-algebra $\mathcal{Y}$. This is a convex space, with mixture operation defined by $(\alpha p+(1-\alpha)q)(E) = \alpha p(E) + (1-\alpha)q(E)$ for all $E \in \mathcal{Y}$, $p, q \in \Delta(Y)$ and $\alpha \in [0, 1]$. I endow $\Delta(Y)$ with the topology of weak convergence of measures. That is, a sequence $(p_n)_{n\geq 1} \in \Delta(Y)$ converges to p if $\int_Y f \mathrm{d}p_n \to \int_Y f \mathrm{d}p$ for every bounded continuous function $f : Y \to \mathbb{R}$. By Prokhorov's theorem, this topology induces a metric on $\Delta(Y)$.

The class of all functions $\Theta \ni \theta \mapsto P_\theta \in \Delta(Y)$ will be denoted by $\Delta(Y)^\Theta$. This is also a convex space, with mixture operation defined point-wise on parameters. That is, for all $P, Q \in \Delta(Y)^\Theta$ and $\alpha \in [0,1]$, $\alpha P + (1-\alpha)Q = (\alpha P_\theta + (1-\alpha)Q_\theta)_{\theta\in\Theta} \in \Delta(Y)^\Theta$. For any $E \in \mathcal{Y}$ and $P \in \Delta(Y)^\Theta$, denote $P(E) = (P_\theta(E))_{\theta\in\Theta}$. Convergence is also defined point-wise, equipping $\Delta(Y)^\Theta$ with the product topology. When Θ is finite, this topology is equivalent to the one induced by Euclidean distance. I assume that $\mathcal{P}$ is a convex subspace of $\Delta(X)^\Theta$, in the sense that if $P, Q \in \mathcal{P}$, then $\alpha P + (1-\alpha)Q \in \mathcal{P}$ for all $\alpha \in [0,1]$.

Not knowing realized signal $x \in X$, conditional on $\theta \in \Theta$, is called *risk*, while lack of knowledge about the parameter itself is referred to as *ambiguity*. The former is objectively quantifiable, since the probabilities for a given $\theta \in \Theta$ are deemed objectively given, while the latter is not. Assume throughout that there exists at least one $P^* \in \mathcal{P}$ with the *full information property*: $(\operatorname{supp} P_\theta) \cap (\operatorname{supp} P_{\theta'}) = \emptyset$ for all $\theta \neq \theta'$.[2] This experiment's signal realizations perfectly reveal the true parameter, so call it the *fully informative* experiment. On the other hand, I call any $P^0 \in \mathcal{P}$ *uninformative* if $P^0_\theta = P^0_{\theta'}$ for all $\theta, \theta' \in \Theta$. Since the likelihood of different signals of an uninformative experiment does not depend on the parameter, observing a signal from P^0 does not provide any information about the true parameter value.

Decision rules. The decision maker observes a signal $x \in X$ coming from some experiment $P \in \mathcal{P}$ and then chooses an *action* from a set A. I assume $(A, \mathcal{A})$ is a compact Polish space with its Borel σ-algebra $\mathcal{A}$. A decision rule $\rho = \{\rho_P\}_{P\in\mathcal{P}}$ is a family of Markov kernels from X to A, indexed by the experiments.[3] Decision rules can be viewed as mixed strategies, one for each experiment, assigning a distribution over actions to each signal realization. Any decision rule defined by $\rho_P(x, \cdot) = \chi_{d_P(x)}(\cdot)$, where $d_P : X \to A$ for every $P \in \mathcal{P}$ and $x \in X$, and χ_a denotes the degenerate distribution with full mass on $a \in A$, is called a *pure* decision rule. These are decision rules that assign each signal to a single action with certainty. Denote by $\mathcal{D}$ the set of all decision rules, and by $\mathcal{D}^\chi$ the set of all pure decision rules. I assume that $\mathcal{D}$ is the class of all families of Markov kernels from X to A, indexed by the set $\mathcal{P}$. In other words, I require the DM to have preferences over all experiment- and signal-contingent distributions over actions.

A decision rule $\rho \in \mathcal{D}$ is called *invariant* if $\rho_P = \rho_{P'}$ for all $P, P' \in \mathcal{P}$. Examples of invariant decision rules in statistics include the sample mean and the least squares estimators. On the other hand, estimators which make use of the specification of the statistical model, such as maximum likelihood, are not invariant. Let $\overline{\mathcal{D}}$ be the set of all *pure and invariant* decision rules. Thus $\overline{\mathcal{D}} \subset \mathcal{D}^\chi \subset \mathcal{D}$. When it is clear from context, I slightly abuse notation and omit P subscripts from ρ_P. When it does not lead to confusion, I may also call ρ_P, for a fixed $P \in \mathcal{P}$, a decision rule.

Parameter contingent action distribution. For any $\rho \in \mathcal{D}$, the *parameter contingent action distribution* induced by ρ and P, ρP, is the mapping $\theta \mapsto \rho P_\theta \equiv \int_X \rho_P(x, \cdot) \mathrm{d}P_\theta(x)$. In particular, for $\delta \in \mathcal{D}^\chi$, we have that $\delta P(F) = (P_\theta \circ \delta_P^{-1}(F))_{\theta\in\Theta}$ are the (parameter contingent) pushforward measures of δ_P acting on P. Let $\mathcal{F} \equiv \Delta(A)^\Theta$ and note that $\rho P \in \mathcal{F}$ for all ρ and P. For any function $v : A \to \mathbb{R}$ and $\rho \in \mathcal{D}$, I denote by $v(\rho_P) : X \to \mathbb{R}$ its expectation under ρ_P, i.e., $v(\rho_P(x)) = \int_A v(a)\rho_P(x, \mathrm{d}a)$ for each $x \in X$.

Preferences. Let $\mathcal{S} = \mathcal{D} \times \mathcal{P}$ be the set of all decision rule–experiment pairs, with typical element $\sigma \in \mathcal{S}$ — which I simply call a *pairing*. I model the decision maker's preferences as a binary relation

[2] For example, if $\Theta \subseteq X$, then P^* such that $P^*_\theta(\{\theta\}) = 1$ for every $\theta \in \Theta$ is fully informative.

[3] That is, for every $P \in \mathcal{P}$, the mapping $\rho_P : X \times \mathcal{A} \to [0,1]$ satisfies: (i) $\rho_P(x, \cdot)$ is a probability for every $x \in X$; and (ii) $\rho_P(\cdot, E)$ is measurable for every $E \in \mathcal{A}$.

$\succsim$ on $\mathcal{S}$. As usual, $\succ$ and $\sim$ denote the asymmetric and the symmetric parts of $\succsim$, respectively. The relation $\succsim$ is called *trivial* if $\sigma \succ \sigma'$ for no $\sigma, \sigma' \in \mathcal{S}$. By taking a single preference relation as a primitive, I am ruling out SDT models where preferences depend on the choice problem being considered. One prominent such model is minimax regret.

Example 1 (Minimax Regret): Let Θ be compact and $u : A \times \Theta \to \mathbb{R}$ be a continuous utility function. The minimax regret choice criterion C^R selects, for any choice problem $D \times \Gamma \subseteq \mathcal{S}$ such that $\{\rho P \in \mathcal{F} : \rho \in D,\ P \in \Gamma\}$ is compact,

$$C^R(D \times \Gamma) = \operatorname*{arg\,min}_{(\rho,P)\in D\times\Gamma} \max_{\theta\in\Theta} \left[\max_{a\in A_D} u(a,\theta) - \int_X u(\rho,\theta)\mathrm{d}P_\theta \right],$$

where $A_D = \{\rho(X) : \rho \in D\}$. In words, the DM chooses the alternative that minimizes, for the worst-case scenario, the difference between the optimal choice with perfect information and the expected utility of the pairing.

Clearly, the ranking of pairings under the minimax regret criterion may depend on the specific choice problem being considered, hence can not be represented by a single preference relation $\succsim$. $\|$

Let f and g be functions defined on some set Y. For any $E \subseteq Y$, let $f_{(E)}g$ denote the function given by $f_{(E)}g(y) = f(y)$ if $y \in E$ and $f_{(E)}g(y) = g(y)$ if $y \in Y \setminus E$. Recalling the definition of the fully informative experiment P^*, denote $S_\theta = \operatorname{supp} P^*_\theta$ and note that $S_\theta \cap S_{\theta'} = \emptyset$ for all $\theta \neq \theta'$. Moreover, define $S_T \equiv \bigcup_{\theta\in T} S_\theta$, for any $T \subseteq \Theta$. To economize on notation, let $\rho_{(S_T)}\tau \equiv \rho_T\tau$, for any $\rho, \tau \in \mathcal{D}$ and $T \subseteq \Theta$. Also denote by $\mathbf{1}_E$ the indicator function of $E \subseteq Y$, that is, $\mathbf{1}_E(y) = 1$ if $y \in E$ and $\mathbf{1}_E(y) = 0$ otherwise.

Risk functions. Given a utility function $u : A \times \Theta \to \mathbb{R}$ which is continuous in the first argument, define the *risk function* of the pairing $(\rho, P) \in \mathcal{S}$ under u as the function $r_u(\rho, P) : \Theta \to \mathbb{R}$ given by

$$r_u(\rho,P)(\theta) = \int_X u(\rho,\theta)\mathrm{d}P_\theta \quad \text{for every } \theta \in \Theta. \tag{2}$$

In words, the risk function describes the parameter contingent expected utility of a pairing. Define $\mathcal{R}_u \equiv \{r_u(\rho,P) \in \mathbb{R}^\Theta : (\rho,P) \in \mathcal{S}\}$, the set of all possible risk functions under u. Finally, a functional $I : \mathcal{R}_u \to \mathbb{R}$ is called *monotone* if $r(\theta) \geq r'(\theta)$ for all $\theta \in \Theta$ implies $I(r) \geq I(r')$; it is called *continuous* if it is continuous in the topology of point-wise convergence.

4 Main results

In this section, I present axioms that characterize a general model of statistical decisions. This will provide a foundation to obtain representations of many models of interest, including SDT versions of subjective expected utility (SEU), maximin expected utility (MEU) and multiplier preferences. We are interested in models of the following form.

Definition 1 (Monotone Risk Aggregation Representation). A binary relation $\succsim$ on $\mathcal{S}$ has *a monotone risk aggregation (MRA) representation* (u, I) if there exists a utility function $u : A \times \Theta \to \mathbb{R}$, continuous in the first argument, and a monotone and continuous functional $I : \mathcal{R}_u \to \mathbb{R}$, such that for all $\rho, \tau \in \mathcal{D}$ and $P, Q \in \mathcal{P}$,

$$(\rho,P) \succsim (\tau,Q) \iff I(r_u(\rho,P)) \geq I(r_u(\tau,Q)), \tag{3}$$

where $r_u : \mathcal{S} \to \mathbb{R}^\Theta$ is defined by eq. (2). $\diamondsuit$

A decision maker whose preferences are compatible with the MRA model evaluates prospects according to expected utility when probabilities are known, and aggregates unknown parameters in a way that favors point-wise improvements of the risk function on the parameters. Importantly for statistical applications, the ex-post utility u is allowed to be parameter-*dependent*. This is in contrast to most of the decision theory under uncertainty literature, where the Bernoulli utility is defined on ultimate consequences, and thus is assumed to be state-independent. In fact, the model of Definition 1 can be viewed as a state-dependent, statistical version of the monotone, Bernoullian and Archimedean (MBA) preferences studied by Cerreia-Vioglio et al. (2011).

Unsurprisingly given its generality, most objective functions commonly used in statistical decision problems fit into this basic framework, differing only in the choice of utility function u and aggregator I.

Example 2 (Bayesian parameter estimation): A Bayesian statistician wants to estimate a parameter $\theta \in \Theta$ with the least possible mean squared error. Assume she has a prior belief $\pi \in \Delta(\Theta)$ over possible parameter values. In terms of the MRA framework, we have $A = \Theta$, and $u(a, \theta) = -(a - \theta)^2$ and $I(\cdot) = \mathbb{E}_\pi(\cdot)$. For a fixed experiment P, the statistician's preference is given by $(\rho, P) \succsim (\tau, P)$ if, and only if, $\mathbb{E}_\pi(\int_X (\rho - \theta)^2 \mathrm{d}P_\theta) \leq \mathbb{E}_\pi(\int_X (\tau - \theta)^2 \mathrm{d}P_\theta)$. ‖

Example 3 (Ellsberg preferences): A decision maker is asked to place a bet on which color ball will be drawn from an urn containing 30 blue and 60 green or yellow balls. We can model this decision problem by setting $\Theta = \{(n_g, n_y) : n_g = 0, \ldots, 60,\ n_y = 60 - n_g\}$, $X = \{b, g, y\}$ and $A = \{0, 1\}$. The interpretation of an action $a \in A$ is of receiving a dollars. Thus, the decision rule $\delta_P = (\delta_P(b) = 1, \delta_P(g) = 0, \delta_P(y) = 1)$ corresponds to betting on either a blue or a yellow ball being drawn from experiment $P \in \mathcal{P}$.

Let $P \in \mathcal{P}$ be the experiment where the balls are drawn according to an uniform distribution. Then $P_\theta(b) = 1/3$ and $P_\theta(g) = \theta/90 = 2/3 - P_\theta(y)$. Typical Ellsberg choices for this decision problem imply $((1,0,0)_P, P) \succ ((0,1,0)_P, P)$ and $((0,1,1)_P, P) \succ ((1,0,1)_P, P)$, where $(b, g, y)_P$ denotes any decision rule $d \in \mathcal{D}^X$ such that $\delta_P = (b, g, y)$. Such choice data is not consistent with any version of SEU preferences, but can be rationalized by MEU preferences: $(\delta, P) \succsim (\delta', Q)$ if, and only if, $\min_{\theta \in \Theta} \int_X u(\delta) \mathrm{d}P_\theta \geq \min_{\theta \in \Theta} \int_X u(\delta') \mathrm{d}Q_\theta$ for an increasing, state-independent utility function u. Here, the aggregator $I(\cdot)$ corresponds to $\min_{\theta \in \Theta}(\cdot)$. ‖

First, I impose a basic rationality postulate on the preference relation.

Axiom 1 (Weak Order): The preference relation $\succsim$ is complete and transitive.

The next axiom needed to characterize the MRA model captures a notion of the DM being a consequentialist. It states that the decision maker's preferences ultimately depend only on the parameter contingent action distribution induced by the pairings, not on the particular decision rule and experiment that generate this distribution.

Axiom 2 (Consequentialism): For all $(\rho, P), (\tau, Q) \in \mathcal{S}$: if $\rho P = \tau Q$, then $(\rho, P) \sim (\tau, Q)$.

Consequentialism is not a completely innocuous axiom. For instance, it may not hold in models where choosing different experiments entail different subjective costs.

Example 4: Consider a rationally inattentive DM of the kind postulated by Matêjka and McKay (2015). Suppose $A = \{-1, 1\}$, $\Theta = \{-1, 0, 1\}$ and $X = \{x_{-1}, x_0, x_1\}$. The DM is a Bayesian, with a uniform prior belief $\pi \in \Delta(\Theta)$, and utility function $u(a, \theta) = \theta a$. She also faces a cost to choosing experiment $P \in \mathcal{P}$ given by

$$k(P) = \sum_{\theta \in \Theta} \pi(\theta) \sum_{x \in X} P_\theta(x) \left[\log P_\theta(x) - \log \sum_{\theta \in \Theta} \pi(\theta) P_\theta(x) \right].$$

The DM chooses a decision rule δ and experiment P to maximize

$$V(\delta, P) = \sum_{\theta \in \Theta} \sum_{x \in X} u(\delta(x), \theta) P_\theta(x) \pi(\theta) - k(P).$$

If the DM can obtain conclusive information about whether $\theta = -1$, further telling $\theta = 0$ apart from $\theta = 1$ is costly, but does not affect the optimal decision.

For instance, take the fully informative experiment $P^*_j(x_j) = 1$ for $j = -1, 0, 1$, and compare it to P' such that $P'_{-1}(x_{-1}) = 1$ and $P'_0(x_0) = P'_1(x_0) = P'_0(x_1) = P'_1(x_1) = 1/2$. Then $k(P^*) > k(P')$. It can be easily verified that if δ' and δ^* denote optimal decision rules given P' and P^* respectively, then the DM strictly prefers (δ', P') to (δ^*, P^*), although they induce the same parameter contingent action distributions.

If the DM had instead been modelled as in Sims (2003), where she faces a hard constraint on the mutual information between the experiment and the prior and solves $\max_{\delta, P} V(\delta, P)$ subject to $k(P) \leq K$, $K > 0$, then Consequentialism would be satisfied. $\|$

The sets of pairings with the same induced distribution partition $\mathcal{S}$ into equivalence classes. For any $\sigma \in \mathcal{S}$, denote by $[\sigma]$ the equivalence class to which σ belongs, and let $\mathcal{S}/_*$ be the set of all such equivalence classes. The space of decision rules is rich enough that every equivalence class in $\mathcal{S}/_*$ has a representative involving the fully informative experiment. When given any pairing, the DM can mimic the parameter dependent action distribution arising from any garbling of the given experiment by appropriately randomizing the decision rule. This allows the DM to achieve any parameter contingent action distribution when given the fully informative experiment, since any any other experiment is a garbling of it (Blackwell, 1951). This is stated formally as follows.

Lemma 1. For all $[\sigma] \in \mathcal{S}/_*$, there exists $\rho^\sigma \in \mathcal{D}$ such that $(\rho^\sigma, P^*) \in [\sigma]$.

Axiom 2 guarantees that $\mathcal{S}/_*$ is a refinement of $\mathcal{S}/_\sim$, the quotient space of $\sim$. In other words, if two pairings induce the same distribution, they are deemed indifferent, but the converse is not necessarily true. Therefore, transitivity of $\succsim$ implies that, to obtain a characterization of the full preference relation, it suffices to characterize the restriction of $\succsim$ to a single representative of each member of $\mathcal{S}/_*$, since the DM is indifferent between all pairings within the same equivalence class. Now, Lemma 1 states that any parameter contingent distribution induced by an element of $\mathcal{S}$ can be achieved by some some decision rule acting on P^*. Therefore, in the presence of Axiom 2 and transitivity, $\succsim$ is completely determined by comparisons of pairings involving the fully informative experiment.

When the DM is provided the fully informative experiment P^*, then there is effectively no uncertainty about the parameter. Therefore, when comparing decision rules, she can focus on their induced actions at each parameter separately. This suggests that two decision rules that coincide on S^c_θ should be ranked in the same way, regardless of what specific actions each rule prescribes for this set of signals. This intuitive notion is captured by the next axiom.

Axiom 3 (Independence of Irrelevant Parameters): For all $\theta \in \Theta$ and $\rho, \tau, \gamma, \gamma' \in \mathcal{D}$:

$$(\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*) \implies (\rho_{\{\theta\}}\gamma', P^*) \succsim (\tau_{\{\theta\}}\gamma', P^*).$$

In the presence of the fully informative experiment, Independence of Irrelevant Parameters (IIP) allows us to interpret $\{\rho_{\{\theta\}}\gamma : \rho \in \mathcal{D}\}$, for any $\gamma \in \mathcal{D}$, as the set of decision rules conditional on θ, since it implies that the ranking of such rules does not depend on what actions they prescribe on S^c_θ. Axiom 3 is the SDT formulation of a weaker version of the "sure thing principle", presented by Savage (1954). The main conceptual difference is that Savage's postulate holds when conditioning the decision rules on any subset $T \in \Sigma$, while IIP is only required to hold for singletons $\{\theta\} \in \Sigma$.

The following is a monotonicity axiom: if one decision rule is preferred to another after conditioning on every parameter, then it is preferred unconditionally.

Axiom 4 (Monotonicity): For all $\rho, \tau \in \mathcal{D}$: if $(\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*)$ for every $\gamma \in \mathcal{D}$ and $\theta \in \Theta$, then $(\rho, P^*) \succsim (\tau, P^*)$.

If the DM is choosing between decision rules to pair with the fully informative experiment, the problem reduces to picking a terminal distribution over actions conditional on each parameter being the truth. Now suppose the actions induced by the available decision rules coincide on every parameter except one. Then only the objectively given probabilities, conditional on the relevant parameter, should matter. Since this choice only involves quantifiable risk, modifying the distributions conditional on the relevant parameter in the same way, for all decision rules under consideration, does not provide hedging value. Therefore, such a modification should not change preferences. This gives an intuitive interpretation to the following axiom.

Axiom 5 (Conditional Mixture Independence): For every $\rho, \tau \in \mathcal{D}$ and $\theta \in \Theta$: if $(\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*)$, then $(\alpha\rho_{\{\theta\}}\gamma + (1-\alpha)\kappa_{\{\theta\}}\gamma, P^*) \succsim (\alpha\tau_{\{\theta\}}\gamma + (1-\alpha)\kappa_{\{\theta\}}\gamma, P^*)$ for all $\alpha \in (0, 1]$ and $\kappa \in \mathcal{D}$.

Conditional Mixture Independence (CMI) implies that the DM views randomization of actions conditional on a given parameter as objective risk. Therefore, she is able to evaluate distributions conditional on each parameter via expected utility.

Finally, I impose a form of continuity on the preference relation. This guarantees that the preference order is not reversed for pairings with arbitrarily similar induced distributions.

Axiom 6 (Continuity): The set $\{(\rho P, \tau Q) \in \mathcal{F}^2 : (\rho, P) \succsim (\tau, Q)\}$ is closed in $\mathcal{F}^2$.

Note that, while axioms 3 to 5 are essentially properties of preferences on decision rules alone, the other axioms act on both decision rules and experiments. Therefore, choice data that includes simultaneous variation of decision rules and experiments is needed to falsify axioms 1, 2 and 6. In Section 6, I show how to characterize a dominance representation from preferences defined exclusively on either decision rules or experiments.

For the remainder of this section, I show that if $\succsim$ satisfies the axioms above, then one can describe a sub-relation $\hat{\succsim} \subseteq \succsim$ which captures the decision maker's ex-post utility function u. This in turn reveals the set $\mathcal{R}_u$ of all risk functions available to the DM. The aggregator I in Definition 1 can then be viewed as an ordinal utility on the space of risk functions.

4.1 The dominance sub-relation

The first step in obtaining an MRA representation is to characterize a behavioral version of the concept of *dominance*, made precise in the following definition.

Definition 2 (Dominance Representation). A binary relation $\succsim$ on $\mathcal{S}$ has a *dominance representation* if there exists a utility function $u : A \times \Theta \to \mathbb{R}$, continuous in the first argument, such that for all $\rho, \tau \in \mathcal{D}$ and $P, Q \in \mathcal{P}$,

$$(\rho, P) \succsim (\tau, Q) \iff \int_X u(\rho, \theta)\mathrm{d}P_\theta \geq \int_X u(\tau, \theta)\mathrm{d}Q_\theta \text{ for all } \theta \in \Theta. \quad (4)$$

If $\succsim$ has a dominance representation, we call it a dominance relation. $\diamondsuit$

When an (incomplete) preference has a dominance representation, a pairing is preferred to another if and only if it yields higher expected utility for every parameter. If one views a statistical decision problem as a game where Nature chooses the true unknown parameter, as in the tradition of Wald (1949), then Definition 2 corresponds to (weak) dominance in mixed strategies. Furthermore, if one fixes an experiment and looks at the corresponding preferences over decision rules induced by a dominance representation, one obtains the statistical concept of *admissibility*.

Consider the following two axioms: one is the converse of Monotonicity, and the other is a weakening of completeness.

Axiom 7 (Admissibility): For all $\rho, \tau \in \mathcal{D}$: $(\rho, P^*) \succsim (\tau, P^*) \implies (\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*)$ for every $\gamma \in \mathcal{D}$ and $\theta \in \Theta$.

Axiom 8 (Conditional Completeness): For all $\rho, \tau, \gamma \in \mathcal{D}$ and $\theta \in \Theta$: $(\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*)$ or $(\tau_{\{\theta\}}\gamma, P^*) \succsim (\rho_{\{\theta\}}\gamma, P^*)$.

The following result characterizes the dominance representation in terms of preferences, and presents its uniqueness properties.

Theorem 1. *A binary relation $\succsim$ on $\mathcal{S}$ is a transitive and reflexive preference that satisfies Consequentialism, Independence of Irrelevant Parameters, Monotonicity, Conditional Mixture Independence, Continuity, Admissibility, and Conditional Completeness if, and only if, it has a dominance representation with utility function $u : A \times \Theta \to \mathbb{R}$.*

Furthermore, u is parameter-wise cardinally unique: if u' also represents $\succsim$, then there exist $\{(b_\theta, c_\theta) : \theta \in \Theta\}$ with $b_\theta > 0$ and $c_\theta \in \mathbb{R}$, such that $u'(\cdot, \theta) = b_\theta u(\cdot, \theta) + c_\theta$ for all $\theta \in \Theta$.

Dominance relations are of independent interest, as they behaviorally characterize the important concepts of strategic dominance and admissibility. However, my main purpose in proving Theorem 1 is to find a particular subrelation of the DM's full preferences that is a dominance relation, and thus can behaviorally elicit the DM's utility, and consequently their risk functions.

I say that σ is unanimously preferred to σ' if, conditional on each parameter, the outcome from σ is deemed preferred to that of σ'. Formally:

Definition 3. Let $\sigma, \sigma' \in \mathcal{S}$ and $\succsim$ be a preference. Then σ *is unanimously preferred to* σ', denoted by $\sigma \mathrel{\hat{\succsim}} \sigma'$, if there exist $(\rho, P^*) \in [\sigma]$ and $(\tau, P^*) \in [\sigma']$ such that $(\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*)$ for all $\theta \in \Theta$ and $\gamma \in \mathcal{D}$. $\diamondsuit$

Let $\succsim$ satisfy axioms 1 to 6. A straightforward consequence of Lemma 1 is that $\hat{\succsim}$ is reflexive, therefore non-empty. It is also usually incomplete, but retains many of the properties of $\succsim$.

Proposition 1. *If $\succsim$ satisfies axioms 1 to 6, the following statements hold:*

1. *If $\sigma \mathrel{\hat{\succsim}} \sigma'$, then $\sigma \succsim \sigma'$. Moreover, $\hat{\succsim}$ is transitive.*

2. *$\hat{\succsim}$ is a dominance relation and, for every other dominance sub-relation $\succsim' \subseteq \succsim$, we have $\succsim' \subseteq \hat{\succsim}$.*

3. *$\hat{\succsim}$ satisfies the sure-thing principle: for all $\rho, \tau, \gamma, \gamma'$ and $T \in \Sigma$,*

$$(\rho_T\gamma, P^*) \mathrel{\hat{\succsim}} (\tau_T\gamma, P^*) \iff (\rho_T\gamma', P^*) \mathrel{\hat{\succsim}} (\tau_T\gamma', P^*).$$

4. *$\hat{\succsim}$ satisfies mixture independence: for all $\rho, \tau, \gamma \in \mathcal{D}$, $P \in \mathcal{P}$ and $\alpha \in (0, 1]$,*

$$(\rho, P) \mathrel{\hat{\succsim}} (\tau, P) \iff (\alpha\rho + (1-\alpha)\gamma, P) \mathrel{\hat{\succsim}} (\alpha\tau + (1-\alpha)\gamma, P).$$

5. *If $\hat{\succsim}$ is trivial, so is $\succsim$.*

The following result characterizes the unanimously preferred relation.

Lemma 2. Let $\succsim$ satisfy axioms 1 to 6. Then the unanimously preferred relation $\hat{\succsim}$ has a dominance representation with a utility $u : A \times \Theta \to \mathbb{R}$ that is continuous in the first argument. Moreover, u is parameter-wise cardinally unique.

In other words, the unanimously preferred relation $\hat{\succsim}$ allows us to elicit the DM's risk functions. In Section 4.2, I will use this fact to obtain a representation of the MRA model.

The representation in Lemma 2 also implies that, given the fully informative experiment, there exists unambiguously best and worst decision rules. Fix $\overline{a}_\theta \in \arg\max_{a\in A} u(a,\theta)$ and $\underline{a}_\theta \in \arg\min_{a\in A} u(a,\theta)$, and set $\overline{\rho}(x,\cdot) = \chi_{\overline{a}_\theta}$ and $\underline{\rho}(x,\cdot) = \chi_{\underline{a}_\theta}$ for all $x \in S_\theta$, $\theta \in \Theta$. Then $(\overline{\rho}, P^*) \mathrel{\hat{\succsim}} \sigma \mathrel{\hat{\succsim}} (\underline{\rho}, P^*)$, for all $\sigma \in \mathcal{S}$. I can then provide the following characterization, which will be useful going forward.

Definition 4. For any dominance relation $\hat{\succsim}$,

$$\mathcal{K}(\hat{\succsim}) = \left\{\sigma \in \mathcal{S} : \sigma \mathrel{\hat{\sim}} (\alpha\overline{\rho} + (1-\alpha)\underline{\rho}, P^*),\ \alpha \in [0,1]\right\} \tag{5}$$

defines its set of *constant-risk-equivalent* (CRE) pairings. ◇

The set $\mathcal{K}(\hat{\succsim})$ amounts to a behavioral characterization of constant-utility pairings, since members of $\mathcal{K}(\hat{\succsim})$ are behaviorally equivalent to pairings with a constant risk function. With state-independent utility, the role of $\mathcal{K}(\hat{\succsim})$ is played by pairings which induce parameter contingent distributions over actions that are constant across parameters. Since I am working within the more general framework of parameter-dependent utility, CRE pairings must be elicited from preferences. The interpretation of eq. (5) as a characterization of constant-utility acts will be formalized in Section 5, Lemma 3.

4.2 Attitudes toward ambiguity

I now construct the decision maker's risk functions from her preferences. Let $\succsim$ satisfy axioms 1 to 6 and fix a utility function u that represents its unanimously preferred relation $\hat{\succsim}$. Consider the mapping $r_u : \mathcal{S} \to \mathbb{R}^\Theta$, defined by eq. (2), and the corresponding set $\mathcal{R}_u$. Define a preference $\succeq^u$ on $\mathcal{R}_u$ as follows:

$$\forall r, r' \in \mathcal{R}_u :\ r \succeq^u r' \iff \exists \sigma \succsim \sigma' \text{ such that } r = r_u(\sigma) \text{ and } r' = r_u(\sigma'). \tag{6}$$

Let $\simeq^u$ and $\succ^u$ denote the symmetric and asymmetric parts of $\succeq^u$, respectively. By construction, if $r \geq r'$, then $r \succeq^u r'$, where $\geq$ denotes the usual order on $\mathbb{R}^\Theta$. Further, since $\succsim$ is complete, so is $\succeq^u$. Also note that $\sigma' \in [\sigma]$ implies $r_u(\sigma) = r_u(\sigma')$.

The following is the main result of the paper.

Theorem 2. *A preference $\succsim$ satisfies Consequentialism, Weak Order, Independence of Irrelevant Parameters, Monotonicity, Conditional Mixture Independence and Continuity if, and only if, its unanimously preferred relation $\hat{\succsim}$ has a dominance representation with utility $u : A \times \Theta \to \mathbb{R}$ and there exists a monotone, continuous functional $I : \mathcal{R}_u \to \mathbb{R}$ such that, for all $\sigma, \sigma' \in \mathcal{S}$,*

$$r_u(\sigma) \succeq^u r_u(\sigma') \iff I(r_u(\sigma)) \geq I(r_u(\sigma')). \tag{7}$$

Therefore, $\succsim$ has an MRA representation (u, I) if, and only if, it satisfies axioms 1 to 6. Moreover, u is parameter-wise cardinally unique and there exists a representation $(\tilde{u}, \tilde{I})$ such that $\tilde{u}(\cdot, \theta)$ is constant for no $\theta \in \Theta$.

Theorem 2 characterizes preferences that are compatible with the MRA model, and shows that the functional I in Definition 2 is essentially a utility on the space of risk functions. Indeed, the proof boils down to applying a general utility representation theorem, found in Herden (1989), to the preference $\succeq^u$ on $\mathcal{R}_u$.

Attitudes toward uncertainty can then be viewed as patterns of complementarity and substitutability between risk under different parameters. Decision makers who want to hedge act as if risk is *complementary* across parameters: they favor pairings that yield more balanced pay-off profiles across parameters over those that have a particularly large expected utility on any specific event $T \in \Sigma$, but not on others. The following example formally illustrates this intuition.

Example 5 (Uncertainty attitudes as parameter complementarity): Take two preferences $\succsim_1$ and $\succsim_2$ such that $\hat{\succsim}_1 = \hat{\succsim}_2 = \hat{\succsim}$. Following Ghirardato and Marinacci (2002), I say that $\succsim_1$ is more averse to ambiguity than $\succsim_2$ if for all $\kappa \in \mathcal{K}(\hat{\succsim})$, $\sigma \succsim_1 \kappa$ implies $\sigma \succsim_2 \kappa$. In other words, whenever $\succsim_1$ prefers an uncertain pairing σ to a CRE pairing κ, so does $\succsim_2$. This is a widely accepted way of comparing preferences in terms of ambiguity aversion (Maccheroni et al., 2006; Cerreia-Vioglio et al., 2011).

Now consider a family of preferences $\{\succsim_s : s \in \mathbb{R} \cup \{-\infty\}\}$ with the same unanimously preferred relation — i.e., $\hat{\succsim}_s = \hat{\succsim}_{s'} = \hat{\succsim}$ for all $s, s' \in \mathbb{R} \cup \{-\infty\}$. Without loss of generality, normalize the representations $\{u_s\}$ of $\{\hat{\succsim}_s\}$ so that $\mathcal{R}_{u_s} = [0,1]^\Theta$ for all s. Under this normalization, $u_s = u$ for all s, and the risk functions associated with members of $\mathcal{K}(\hat{\succsim})$ are constant. Assume that $\{\succsim_s\}$ has a representation (I_s, u) by a Dixit and Stiglitz (1977) constant elasticity of substitution (CES) function, with (full support) shares $\mu \in \Delta(\Theta)$ and substitution parameter s:

$$I_s(r_u(\sigma)) = \begin{cases} \left(\int_\Theta r_u(\sigma)^s \mathrm{d}\mu\right)^{\frac{1}{s}}, & s \notin \{0, -\infty\} \\ \exp\left(\int_\Theta \ln r_u(\sigma) \mathrm{d}\mu\right), & s = 0 \\ \inf_{\theta \in \Theta} r_u(\sigma)(\theta), & s = -\infty \end{cases}$$

for all $\sigma \in \mathcal{S}$. The elasticity of substitution between any two goods (or parameters) of a CES utility function is given by $\varepsilon = 1/(1-s)$. Thus, $s = 1$ corresponds to unbounded elasticity of substitution, while $s = -\infty$ implies perfect complementarity. Moreover, it can be shown that $\lim_{s \to s^*} I_s = I_{s^*}$ for $s^* \in \{0, -\infty\}$. CES utility is a particular case of second order expected utility, which was axiomatized for the setting of Savage (1954) by Neilson (2010).

If $t > s$ and $t, s \neq 0$, we have, for all $\sigma \in \mathcal{S}$,

$$I_t(r_u(\sigma)) = \left(\int_\Theta [r_u(\sigma)^s]^{\frac{t}{s}} \mathrm{d}\mu\right)^{\frac{s}{ts}} \leq \left(\int_\Theta r_u(\sigma)^s \mathrm{d}\mu\right)^{\frac{s}{ts}\frac{t}{s}} = I_s(r_u(\sigma)),$$

where I have used Jensen's inequality. Since for all $\kappa \in \mathcal{K}(\hat{\succsim})$ and $t, s \in \mathbb{R} \cup \{-\infty\}$, we have $I_t(r_u(\kappa)) = I_s(r_u(\kappa))$, then $\sigma \succsim_s \kappa$ implies $\sigma \succsim_t \kappa \iff I_s(r_u(\sigma)) \geq I_t(r_u(\sigma)) \iff s \leq t$. Thus, ambiguity aversion in inversely related to the elasticity of substitution. In particular, an SEU agent ($s = 1$) shows no ambiguity aversion, treating parameters as perfect substitutes. On the other hand, an MEU agent ($s = -\infty$) has maximal ambiguity aversion, and thus perfect complementarity across parameters. ‖

5 Applications

In the Anscombe-Aumann framework, an *act* is a measurable function from some set Ω of states of the world to a convex space of outcomes Y. We can thus identify each risk function of SDT with an Anscombe-Aumann act from $\Theta \equiv \Omega$ to $\mathcal{R}_u \equiv Y$. The representation in Theorem 2 implies that if $\succsim$ satisfies axioms 1 to 6, then $\succeq^u$ satisfies the following essential axioms of decision theory under uncertainty:

- **Weak Order**: the preference $\succeq^u$ is complete and transitive.
- **A-A Monotonicity**: for all $r, r' \in \mathcal{R}_u$, if $r \geq r'$, then $r \succeq^u r'$.
- **Risk Independence:** for all constant risk functions $r, r', q \in \mathcal{R}_u$ and $\alpha \in [0,1]$, $r \succeq^u r'$ implies $\alpha r + (1-\alpha)q \succeq^u \alpha r' + (1-\alpha)q$.
- **Mixture Continuity:** the sets $\{\alpha \in [0,1] : \alpha r + (1-\alpha)r' \succeq^u q\}$ and $\{\alpha \in [0,1] : q \succeq^u \alpha r + (1-\alpha)r'\}$ are closed for all $r, r', q \in \mathcal{R}_u$.

Most models of complete preferences in the Anscombe-Aumann framework satisfy these same properties, and impose extra axioms to obtain more structure on I. Thus there is hope that, by relating properties of $\succeq^u$ with the corresponding axioms on $\succsim$, I can import many of the existing representations from the Anscombe-Aumann framework to SDT. The following lemma is an important step towards establishing this connection.

Lemma 3. Suppose $\hat{\succsim}$ has a dominance representation with utility function u. The following statements hold:

1. $r_u(\alpha\rho + (1-\alpha)\tau, P) = \alpha r_u(\rho, P) + (1-\alpha) r_u(\tau, P)$ for all $\tau \in \mathcal{D}$ and $\alpha \in [0,1]$.
2. $r_u(\rho_T\gamma, P^*) = r_u(\rho, P^*)_{(T)} r_u(\gamma, P^*)$ for all $\rho, \gamma \in \mathcal{D}$ and $T \in \Sigma$.
3. There exists $\tilde{u}$ also representing $\hat{\succsim}$ such that $r_{\tilde{u}}(\sigma)$ is constant if, and only if, $\sigma \in \mathcal{K}(\hat{\succsim})$.

The usefulness of Lemma 3 stems from the fact that most axioms in the decision theory under uncertainty paradigm are stated in terms of (i) mixture of acts, as in the first statement; (ii) combinations of acts, i.e., substituting an act's consequences on a set of parameters for the consequences of a different act, as in statement 2; or (iii) properties of preferences over mixtures or combinations with constant-utility acts, as characterized by $\mathcal{K}(\hat{\succsim})$.

Next, I illustrate this methodology by applying Lemma 3 to a variety of existing axiomatizations of preferences in the Anscombe-Aumann framework, to obtain SDT versions of their representations. Throughout this section, I maintain the assumption that every experiment in $\mathcal{P}$ is measurable with respect to Σ, which makes every risk function also measurable.

5.1 Subjective expected utility

I begin by obtaining a SDT representation of the subjective expected utility (SEU) model. Recall that, in the Anscombe-Aumann setting, state-independent SEU is characterized by preferences over acts satisfying Weak Order, A-A Monotonicity, Mixture Continuity, and the following strengthening of Risk-Independence (which I state here in terms of $\succeq^u$):

- **Independence:** for all $r, r', q \in \mathcal{R}_u$, if $r \succeq^u r'$, then $\alpha r + (1-\alpha)q \succeq^u \alpha r' + (1-\alpha)q$ for all $\alpha \in (0,1)$.

These four axioms imply the existence of a representation by SEU with *finitely additive* prior probability. To guarantee that the prior is *countably additive*, one needs an additional postulate, due to Arrow (1971).

- **Monotone Continuity:** if $r, r', q \in \mathcal{R}_u$, q is constant, $\{T_n\}_{n\geq 1} \in \Sigma$ with $T_1 \supseteq T_2 \supseteq \cdots$ and $\bigcap_{n\geq 1} T_n = \emptyset$, then $r \succ^u r'$ implies that there exists $n_0 \geq 1$ such that $q_{(T_{n_0})} r \succ^u r'$.

Given statement 1 in Lemma 3, it is straightforward to translate Independence of $\succeq^u$ into an axiom on $\succsim$:

Axiom 9 (Mixture Independence): For every $\rho, \tau, \gamma \in \mathcal{D}$ and $P \in \mathcal{P}$: if $(\rho, P) \succsim (\tau, P)$, then $(\alpha\rho + (1-\alpha)\gamma, P) \succsim (\alpha\tau + (1-\alpha)\gamma, P)$ for all $\alpha \in (0, 1]$.

With the aid of Lemma 1 and statements 2 and 3 of Lemma 3, it is also easy to obtain Monotone Continuity of $\succeq^u$ from preferences over pairings.

Axiom 10 (Monotone Continuity*): If $\rho, \tau, \kappa \in \mathcal{D}$, with $(\kappa, P^*) \in \mathcal{K}(\hat{\succsim})$, and $\{T_n\}_{n\geq 1} \in \Sigma$ with $T_1 \supseteq T_2 \supseteq \cdots$ and $\bigcap_{n\geq 1} T_n = \emptyset$, then $(\rho, P^*) \succ (\tau, P^*)$ implies that there exists $m \geq 1$ such that $(\kappa_{T_m}\rho, P^*) \succ (\tau, P^*)$.

We are ready to characterize SEU in the SDT framework. In what follows, we refer to axioms 1 to 6 as the MRA axioms.

Proposition 2 (SEU representation). *A preference $\succsim$ on $\mathcal{S}$ satisfies the MRA axioms, Mixture Independence and Monotone Continuity* if, and only if, there exists a bounded utility function $u : A\times\Theta \to \mathbb{R}$, continuous in the first argument, and a probability distribution $\pi \in \Delta(\Theta)$ such that, for all $\rho, \tau \in \mathcal{D}$ and $P, Q \in \mathcal{P}$,*

$$(\rho, P) \succsim (\tau, Q) \iff \int_\Theta \int_X u(\rho(x), \theta)\mathrm{d}P_\theta(x)\mathrm{d}\pi(\theta) \geq \int_\Theta \int_X u(\tau(x), \theta)\mathrm{d}Q_\theta(x)\mathrm{d}\pi(\theta).$$

Moreover, u is parameter-wise cardinally unique and, for a given u, π is unique.

Preferences that satisfy SEU are also called *Bayesian*, since for any $P \in \mathcal{P}$ and $\rho \in \mathcal{D}$, we have

$$\int_\Theta \int_X u(\rho(x), \theta)\mathrm{d}P_\theta(x)\mathrm{d}\pi(\theta) = \int_X \int_\Theta u(\rho(x), \theta)\mathrm{d}\pi_x(\theta)\mathrm{d}p(x),$$

where $p = \int_\Theta P_\theta \mathrm{d}\pi(\theta)$ and $\pi_x \in \Delta(\Theta)$ is the posterior distribution, conditional on $x \in X$: $\pi_x(T) = \mathbb{E}_\pi[\mathbf{1}_T \,|X_P = x]$ for all $T \in \Sigma$. When $p(x) > 0$, the posterior can be obtained via Bayes' rule.

5.2 Maximin expected utility

Maximin expected utility (MEU) — along with its close cousin, multiple priors expected utility (MPEU) — is perhaps the most thoroughly studied model of decision making under uncertainty, after SEU. Gilboa and Schmeidler (1989) provided a characterization of MPEU for the Anscombe-Aumann setting that relies on Weak Order, A-A Monotonicity, Mixture Continuity, and two extra axioms:

- **Uncertainty Aversion:** for all $r, r' \in \mathcal{R}_u$ and $\alpha \in (0, 1)$, $r \simeq^u r'$ implies $\alpha r + (1-\alpha)r' \succeq^u r$.
- **Certainty Independence:** for all $r, r', q \in \mathcal{R}_u$, with constant q: if $r \succeq^u r'$, then $\alpha r + (1-\alpha)q \succeq^u \alpha r' + (1-\alpha)q$ for all $\alpha \in (0, 1)$.

Uncertainty Aversion and Certainty Independence on the space of risk functions are induced by the following axioms on our primitive $\succsim$.

Axiom 11 (Hedging): For all $\rho, \tau \in \mathcal{D}$, $P \in \mathcal{P}$ and $\alpha \in (0, 1)$: if $(\rho, P) \sim (\tau, P)$, then $(\alpha\rho + (1-\alpha)\tau, P) \succsim (\rho, P)$.

Axiom 12 (CRE-Independence): For all $\rho, \tau, \kappa \in \mathcal{D}$ and $P \in \mathcal{P}$ such that $(\kappa, P) \in \mathcal{K}(\hat{\succsim})$: if $(\rho, P) \succsim (\kappa, P)$, then $(\alpha\rho + (1-\alpha)\kappa, P) \succsim (\alpha\tau + (1-\alpha)\kappa, P)$ for all $\alpha \in (0, 1]$.

Axiom 11 implies a preference for hedging: for any fixed experiment, the DM weakly prefers decision rules that provide a more balanced pay-off profile across parameters. Axiom 12 states that for any two pairings sharing the same experiment, mixing the decision rules with a third constant-risk-equivalent pairing does not change preferences.

I can now state the SDT version of the MPEU representation:

Proposition 3 (MPEU representation). *A preference $\succsim$ on $\mathcal{S}$ satisfies the MRA axioms, Hedging, and CRE-Independence if, and only if, there exists a bounded utility function $u : A \times \Theta \to \mathbb{R}$, continuous in the first argument, and a convex family of probability distributions $\Pi \subseteq \Delta(\Theta)$ such that, for all $\rho, \tau \in \mathcal{D}$ and $P, Q \in \mathcal{P}$,*

$$(\rho, P) \succsim (\tau, Q) \iff \inf_{\pi \in \Pi} \int_\Theta \int_X u(\rho(x), \theta) \mathrm{d}P_\theta(x) \mathrm{d}\pi(\theta) \geq \inf_{\pi \in \Pi} \int_\Theta \int_X u(\tau(x), \theta) \mathrm{d}Q_\theta(x) \mathrm{d}\pi(\theta).$$

Moreover, u is parameter-wise cardinally unique and, for a given u, Π is unique.

An MPEU agent can be interpreted as a pessimist, or as extremely cautious. He acts as if whichever pairing is chosen, the true parameter was drawn from the worst possible distribution among all the prior distributions he considers plausible.

The MEU representation can be seen as the special case of MPEU when $\Pi = \Delta(\Theta)$. Therefore, an MEU agent is maximally pessimistic, since he always consider the worst case for any distribution of the parameters. This model was axiomatized in the space of risk functions by Stoye (2012). The axiom required for the representation is translated to the SDT setting below.

Axiom 13 (Symmetry): Let $T, F \in \Sigma$ be such that $T \cap F = \emptyset$, and consider $\rho, \tau, \kappa, \gamma \in \mathcal{D}$ with $(\kappa, P^*), (\gamma, P^*) \in \mathcal{K}(\hat{\succsim})$. If $(\rho_T \kappa_F \gamma, P^*) \succsim (\tau_T \kappa_F \gamma, P^*)$, then $(\rho_T \gamma_F \kappa, P^*) \succsim (\tau_T \gamma_F \kappa, P^*)$.

Axiom 13 requires that the DM considers all events $T, F \in \Sigma$ as equally plausible: from an ex-ante perspective, reallocating constant pay-offs of different magnitudes across events does not change preferences. I can now state the following:

Proposition 4 (MEU representation). *A preference $\succsim$ on $\mathcal{S}$ has a MPEU representation and satisfies Symmetry if, and only if, there exists a bounded utility function $u : A \times \Theta \to \mathbb{R}$, continuous in the first argument, such that for all $\rho, \tau \in \mathcal{D}$ and $P, Q \in \mathcal{P}$,*

$$(\rho, P) \succsim (\tau, Q) \iff \inf_{\theta \in \Theta} \int_X u(\rho, \theta) \mathrm{d}P_\theta \geq \inf_{\theta \in \Theta} \int_X u(\tau, \theta) \mathrm{d}Q_\theta.$$

Moreover, u is parameter-wise cardinally unique.

5.3 Variational and multiplier preferences

Variational preferences models, put forth by Maccheroni et al. (2006), have also received considerable attention in the decision theoretic literature. In that setting, its characterization hinges on the same axioms as the MPEU model, except with a weakened version of C-Independence, which I translate into the SDT framework as follows:

Axiom 14 (Weak CRE-Independence): For all $\rho, \tau, \kappa, \kappa' \in \mathcal{D}$ and $P \in \mathcal{P}$ such that $(\kappa, P), (\kappa', P) \in \mathcal{K}(\hat{\succsim})$: if $(\alpha\rho + (1-\alpha)\kappa, P) \succsim (\alpha\tau + (1-\alpha)\kappa, P)$, then $(\alpha\rho + (1-\alpha)\kappa', P) \succsim (\alpha\tau + (1-\alpha)\kappa', P)$ for all $\alpha \in (0, 1]$.

Along with some previously defined axioms, Weak CRE-Independence suffices to characterize variational preferences in SDT. Before stating the result, we need the definition of a mathematical property: a function $c : \Delta(\Theta) \to [0, \infty)$ is said to be *grounded* if $\inf_{\pi \in \Delta(\Theta)} c(\pi) = 0$.

Proposition 5 (Variational representation). *A preference $\succsim$ on $\mathcal{S}$ satisfies the MRA axioms, Monotone Continuity*, Hedging and Weak CRE-Independence if, and only if, there exists a utility function $u : A \times \Theta \to \mathbb{R}$, continuous in the first argument, a convex set of probabilities $\Pi \subseteq \Delta(\Theta)$, and a grounded, convex and lower semicontinuous function $c : \Pi \to [0, \infty)$ such that, for all $\rho, \tau \in \mathcal{D}$ and $P, Q \in \mathcal{P}$,*

$$(\rho, P) \succsim (\tau, Q) \iff \inf_{\pi \in \Pi} \{\mathbb{E}_\pi[r_u(\rho, P)] + c(\pi)\} \geq \inf_{\pi \in \Pi} \{\mathbb{E}_\pi[r_u(\tau, Q)] + c(\pi)\}.$$

A DM whose preferences have a variational representation entertains a set of possible prior beliefs, but has different degrees of confidence in each of them, which is represented by the function c. Variational preferences generalize the MPEU model, which in turn generalizes SEU. This can be readily seen from the representation, by setting $c(\pi) = 0$ for all $\pi \in \Pi$ to get MPEU, and setting $\Pi = \{\pi\}$ to obtain SEU. It can also be seen from the axioms, since Weak CRE-Independence is weaker than CRE-Independence, which in turn is weaker than Mixture Independence.

A special case of this model, called *multiplier preferences* by Hansen and Sargent (2001), was characterized by Strzalecki (2011). This is done by imposing Savage's sure thing principle on a preference that has a variational representation.

Axiom 15 (Sure Thing Principle): For all $\rho, \tau, \gamma, \gamma' \in \mathcal{D}$ and $T \in \Sigma$: $(\rho_T \gamma, P^*) \succsim (\tau_T \gamma, P^*)$ implies $(\rho_T \gamma', P^*) \succsim (\tau_T \gamma', P^*)$.

Axiom 15 is clearly a stronger version of Independence of Irrelevant Parameters, and has a similar interpretation. It states that once the DM conditions the decision rule on a particular event $T \in \Sigma$, the action distributions on the remaining parameters $\theta \in T^c$ are inconsequential.

For any $\pi, \mu \in \Delta(\Theta)$, let $\pi \ll \mu$ denote that π is absolutely continuous of with respect to μ, i.e., $\mu(T) = 0$ implies $\pi(T) = 0$ for all $T \in \Sigma$. One can now obtain a representation by multiplier preferences:

Proposition 6 (Multiplier representation). *A preference $\succsim$ on $\mathcal{S}$ has a variational representation and satisfies the Sure Thing Principle if, and only if, there exists a utility function $u : A \times \Theta \to \mathbb{R}$, continuous in the first argument, a probability distribution $\mu \in \Delta(\Theta)$, and $k \in (0, \infty]$ such that, for all $\rho, \tau \in \mathcal{D}$ and $P, Q \in \mathcal{P}$,*

$$(\rho, P) \succsim (\tau, Q) \iff \min_{\pi \in \Delta(\Theta)} \{\mathbb{E}_\pi[r_u(\rho, P)] + kD(\pi \| \mu)\} \geq \min_{\pi \in \Delta(\Theta)} \{\mathbb{E}_\pi[r_u(\tau, Q)] + kD(\pi \| \mu)\},$$

where

$$D(\pi \| \mu) = \begin{cases} \int_\Theta \log \frac{\mathrm{d}\pi}{\mathrm{d}\mu} \mathrm{d}\pi, & \text{if } \pi \ll \mu \\ \infty, & \text{otherwise} \end{cases}$$

is the Kullback-Leibler divergence and $\frac{\mathrm{d}\pi}{\mathrm{d}\mu}$ is the Radon-Nikodym derivative of π with respect to μ.

Multiplier preferences describe a Bayesian DM who is worried about misspecification of the prior distribution. Though the DM's best guess for the parameter distribution is μ, she also entertains other specifications. Alternative candidates for the prior become less plausible as they diverge from the benchmark distribution μ. The degree of confidence in μ being correctly specified is regulated by the constant k, with larger values of k indicating more confidence in the benchmark distribution.

6 Alternative data sets

The previous sections focused on preferences over pairings. This is the most appropriate choice environment if the goal is to better understand the behavior of agents in different SDT models. It is also often observable as empirical data. For example, an econometrician who writes down a statistical model and chooses a particular estimator is ultimately revealing their choice of decision rule and experiment.

However, there are situations in which other types of choice data are forthcoming. In this section, I explore the behavioral implications of the MRA model to three alternative data sets. In Section 6.1, I consider state dependent stochastic choice data as assumed by Caplin and Martin (2015) and Caplin and Dean (2015). This is the standard data set in much of the psychometric literature, and keeps track of the probabilities of choosing each action conditional on the parameters.

In Section 6.2, I characterize the MRA model using choice data in the form of two different collections of preferences, each of which can be viewed as a cross-section of $\succsim$. The first describes, for each experiment $P \in \mathcal{P}$, a preference $\succcurlyeq_P$ on the set of decision rules. This corresponds to a DM who takes the experiment as given an chooses only among available decision rules. The second collection of preferences ascribes to each compact menu of decision rules $M \subseteq \mathcal{D}$, a preference $\succcurlyeq^M$ on the set of all experiments. This models a DM who chooses an experiment, assuming that in a second stage of the decision making process it will be paired with a particular decision rule.

6.1 Parameter dependent stochastic choice

Recall that the grand set of all parameter contingent action distributions is given by $\mathcal{F} = \Delta(A)^\Theta$. Each $f \in \mathcal{F}$ can be understood as a parameter dependent stochastic choice, i.e., the probability (or frequency) of choosing each action $a \in A$ for every parameter $\theta \in \Theta$. Consider a data set where one observes the DM choosing parameter dependent stochastic choices for a finite collection of *decision problems* $\Phi = \{F_1, \dots, F_n : F_i \subseteq \mathcal{F}\}$. Each $F \in \Phi$ represents the set of parameter contingent action distributions available to the DM in a particular choice situation. I call $(\mathcal{F}, \Phi)$ — where Φ consists of a finite collection of closed subsets of $\mathcal{F}$ — a choice space.

Formally, parameter dependent stochastic choice data consists of a choice function $c : \Phi \to \mathcal{F} \setminus \{\emptyset\}$ such that $c(F) \in F$ for every $F \in \Phi$. Although the formalism is given in terms of a single observed choice for each feasible set, actual data is often better understood as repeatedly observing, for each decision problem $F \in \Phi$, the DM's choice of action for every parameter, and calculating action frequencies. To avoid questions of how to elicit multiple selections from the same decision problem in a setting where we only ever observe randomized choices, I assume that c is a function rather than a correspondence.

Given a parameter dependent stochastic choice function c, the goal is to determine whether it can be rationalized by a preference relation $\succsim$ on $\mathcal{S}$, in the following sense.

Definition 5. Given a choice space $(\mathcal{F}, \Phi)$, we say that a preference relation $\succsim$ on $\mathcal{S}$ rationalizes $c : \Phi \to \mathcal{F} \setminus \{\emptyset\}$, if the preference $\succeq$ on $\mathcal{F}$ defined, for all $f, g \in \mathcal{F}$, by

$$f \succeq g \iff \exists (\rho, P), (\tau, Q) \in \mathcal{S} \text{ such that } f = \rho P,\ g = \tau Q \text{ and } (\rho, P) \succsim (\tau, Q), \tag{8}$$

is such that for all $F \in \Phi$,

$$c(F) = \{f \in F : f \succeq g\ \forall g \in F\}. \qquad \diamondsuit$$

Definition 5 has the interpretation that the DM is actually choosing between pairings, but we (the modellers) only observe the induced parameter contingent action distributions. If c can be rationalized by $\succsim$, then $c(F)$ is the parameter contingent action distribution induced by the most preferred pairing (according to $\succsim$), among all pairings that can induce the distributions in F. Note that if the DM does not satisfy Consequentialism, then there is no hope of deducing properties of the preferences over pairings by observing only parameter dependent stochastic choices. Indeed, a failure of Consequentialism would imply that the preference $\succeq$ on $\mathcal{F}$ given by (8) is not well defined.

An alternative is directly revealed preferred to another if it is chosen from a decision problem where the other is available. If there is a sequence of alternatives such that f_1 is directly revealed preferred to f_2 and so on, until f_{k-1} is directly revealed preferred to f_k, then I simply say that f_1 is *revealed preferred* to f_k. This is formalized in the definition below.

Definition 6. For any $f_1, f_k \in \mathcal{F}$, we say that f_1 is revealed preferred to f_k, and denote it by $f_1 \trianglerighteq f_k$, if there exist $\{F_1, \ldots, F_{k-1}\} \subseteq \Phi$ such that $\{f_i, f_{i+1}\} \subseteq F_i$ and $f_i = c(F_i)$ for all $i = 1, \ldots, k-1$. $\diamondsuit$

Consider the following axioms on the parameter dependent stochastic choice data set c.

Axiom C1 (GARP): For all $f, g \in \mathcal{F}$: if $f \trianglerighteq g$, then $g \neq c(F)$ for all $F \in \Phi$ such that $f \in F$.

The Generalized Axiom of Revealed Preference (GARP) is a basic rationality postulate, stating that the revealed preference relation is acyclic. That is, the DM's choices are required to be coherent, in the sense that if f is revealed to be strictly preferred to g, then g can not be chosen when f is available. This is the revealed preference analogue of Weak Order.

For each $\theta \in \Theta$, define a binary relation $\triangleright_\theta$ on $\Delta(A)$ by $f(\theta) \triangleright_\theta g(\theta)$ if, and only if, $f \triangleright g$ and $f(\theta') = g(\theta')$ for all $\theta' \neq \theta$. This can be interpreted it as a directly revealed preference conditional on θ, since $f(\theta) \triangleright_\theta g(\theta)$ when f is chosen over g in a situation where the only relevant parameter is θ.

Let $\trianglerighteq_\theta$ be the convex hull of the transitive closure of $\triangleright_\theta$. That is, $p \trianglerighteq_\theta q$ if, and only if, there exist $\{(r_j, s_j)\}_{j=1}^k \subseteq \Delta(A)^2$ and $\{\lambda_j\}_{j=1}^k \subseteq \mathbb{R}_+$ with $\sum_{j=1}^k \lambda_j = 1$, such that $r_j \triangleright_\theta s_j$ for all $j = 1, \ldots, k$, $p = \sum_{j=1}^k \lambda_j r_j$ and $q = \sum_{j=1}^k \lambda_j s_j$. In more intuitive terms, $p \trianglerighteq_\theta q$ if p and q are convex combinations, with the same weights, of $\{r_j\}_{j=1}^k$ and $\{s_j\}_{j=1}^k$ respectively, and moreover each r_j is directly revealed preferred to s_j conditional on θ. Note that $\trianglerighteq_\theta$ can be elicited from choice data, by first taking the transitive closure and then the convex hull of $\triangleright_\theta$. This relation was first presented by Clark (1993), who also formulated the following axiom.

Axiom C2 (C-LARP): For all $f, g \in \mathcal{F}$, and $\theta \in \Theta$: $f(\theta) \trianglerighteq_\theta g(\theta)$ implies $g(\theta) \not\triangleright_\theta f(\theta)$.

The conditional linear axiom of revealed preference (C-LARP) says that in situations where there is a single relevant parameter, mixing action distributions on that parameter using the same weights does not reverse their conditional preference ranking. Thus, it is a parameter dependent stochastic choice analogue of Conditional Mixture Independence. Moreover, it implies that $\triangleright_\theta$ is antisymmetric. That is, if $f(\theta) \neq g(\theta)$ and $f_{(\theta)}h \trianglerighteq g_{(\theta)}h$, then $g_{(\theta)}h' \trianglerighteq f_{(\theta)}h'$ does not hold for any $h' \in \mathcal{F}$. Therefore, C-LARP also encodes a version of Independence of Irrelevant Parameters.

Axiom C3 (RP-Monotonicity): If $f = c(F)$ and $g \in F \setminus \{f\}$, then there exists $\theta \in \Theta$ such that $g(\theta) \trianglerighteq_\theta f(\theta)$ does not hold.

RP-Monotonicity states that if f is directly revealed preferred to g, then there exists a parameter for which g is not conditionally revealed preferred to f. It can be considered a revealed preference version of Monotonicity.

Finally, we impose a continuity axiom on the conditional revealed preference relations.

Axiom C4 (Conditional Continuity): For all $\theta \in \Theta$, if $(p_n), (q_n) \in \Delta(A)$ are sequences such that $p_n \to p$, $q_n \to q$ and $p_n \trianglerighteq_\theta q_n$ for all $n \in \mathbb{N}$, then $q \not\trianglerighteq_\theta p$.

We can now state the main representation theorem of this section.

Theorem 3. *The parameter dependent stochastic choice function $c : \Phi \to \mathcal{F} \setminus \{\emptyset\}$ satisfies GARP, C-LARP, RP-Monotonicity and Conditional Continuity if, and only if, there exists a preference $\succsim$ on $\mathcal{S}$*

which rationalizes c and satisfies Consequentialism, Weak Order, Independence of Irrelevant Parameters, Monotonicity, Conditional Mixture Independence and Continuity.

Moreover, if Θ is countable, so that $\mathcal{F}$ is metrizable, then $\succsim$ has an MRA representation (u, I) such that, for all $F \in \Phi$,

$$c(F) = \left\{ \rho P \in F : \operatorname*{arg\,max}_{\{(\tau,Q)\in\mathcal{S}:\tau Q\in F\}} I(r_u(\tau, Q)) \right\}.$$

Theorem 3 characterizes when parameter dependent stochastic choice data can be rationalized by postulating that the DM is actually facing a statistical decision problem, and making choices that maximize preferences compatible with the MRA framework. This kind of behavioral data differs from preferences over pairings in two main ways. First, it does not assume that we observe experiments and decision rules, only their final consequences — the induced parameter contingent action distributions. Second, it only assumes that choices from a finite number of decision problems — rather than the whole array of preferences — is observed. Therefore, Theorem 3 allows us to test whether the DM's choices are compatible with the MRA model using a data set that is often easier to obtain. It is thus natural to ask whether this type of data set can also be used to identify preferences over pairings, at least partially. That is, can the DM's choices over pairings be recovered by only looking at parameter dependent stochastic choice data? The following example shows that, at least under some conditions, this is indeed possible.

Example 6: Let A, Θ and X be finite, with $|X| \leq \min\{|\Theta|, |A|\}$. The DM chooses pairings from an array of decision problems $\Psi = \{S_1, \ldots, S_n : S_j \subseteq \mathcal{S}\}$ so as to maximize a preference $\succsim$ on $\mathcal{S}$. Assume that the set of feasible decision rules remains fixed at some $D \subseteq \mathcal{D}$ for every $S \in \Psi$. Therefore, each decision problem only differs by the set of experiments available to the DM.

We do not directly observe the DM's choice of decision rule or experiment, only the realized action frequencies for each parameter $\theta \in \Theta$ and decision problem $S \in \Psi$. Let $\mathbb{P}(a|\theta, S)$ denote the probability of choosing action $a \in A$ from decision problem $S \in \Psi$, when the parameter is $\theta \in \Theta$. Letting $F(S) = \{\rho P : (\rho, P) \in S\}$ denote the set of parameter contingent action distributions that can be induced by members of S, we can express the parameter dependent stochastic choice data set as

$$c(F(S))(\theta) = \begin{pmatrix} \mathbb{P}(a_1|\theta, S) \\ \vdots \\ \mathbb{P}(a_{|A|}|\theta, S) \end{pmatrix} \quad \text{for all } \theta \in \Theta \text{ and } S \in \Psi.$$

Given the DM's postulated choice procedure, the realized action frequencies must satisfy

$$\mathbb{P}(a|\theta, S_j) = \sum_{x \in X} \rho^j(x, a) P^j_\theta(x) \quad \text{for all } \theta \in \Theta \text{ and } j \in \{1, \ldots, n\}, \tag{9}$$

subject to $(\rho, P^j) \succsim (\tau, Q)$ for all $(\tau, Q) \in S_j$ — where I write $\rho^j \equiv \rho_{P^j}$ to simplify notation. With parameter dependent stochastic choice data, we only observe the left hand side of eq. (9): neither the chosen (ρ, P^j), nor the realized signals are deemed observable.

Note that for each $S_j \in \Psi$, varying the parameter changes the optimal signal distribution, P^j_θ, but leaves the decision rule unchanged. Thus, eq. (9) describes a finite mixture model where varying the parameter shifts the mixture weights $P^j_\theta(x)$, but leaves the mixture distributions $\rho(x, \cdot)$ unchanged. If the model in (9) is point identified in the statistical sense, it is possible to recover the chosen decision rule and experiment from each $S \in \Psi$, given enough data.

Henry et al. (2014) thoroughly studied conditions on P^j and ρ^j under which such models are partially identified. Adams (2016) provides necessary and sufficient conditions, again on P^j and ρ^j, for point identification. Such results are of limited applicability, since P^j and ρ^j are not directly observable.

I have an ongoing research project (Furtado, 2022) where I present a sufficient condition on the observable parameter dependent stochastic choice data $\mathbb{P}$ that guarantees point identification of such mixture models. $\|$

6.2 Cross sectional preferences

In sections 4 and 5, I focused on preferences $\succsim$ over pairs of decision rules *jointly with* experiments. However, some applications are not directly concerned with preferences over pairings, but rather with one of two sets of preferences embedded in $\succsim$. The first and most ubiquitous kind of application involves finding an optimal decision rule for a given experiment. That is, given $P \in \mathcal{P}$, one wishes to characterize a preference $\succcurlyeq_P$ defined by $\rho \succcurlyeq_P \tau$ if, and only if, $(\rho, P) \succsim (\tau, P)$.

The second type of application involves ranking experiments, assuming that each will be paired with a decision rule that will be chosen in a second stage. Formally, given a menu of feasible decision rules $M \subseteq \mathcal{D}$ and a collection of choices $\{C_P(M) : P \in \mathcal{P}\}$ with $C_P(M) \subseteq M \setminus \{\emptyset\}$, I wish to characterize the relation $P \succcurlyeq^M Q$ defined by $(\rho^P, P) \succsim (\rho^Q, Q)$ for every $P, Q \in \mathcal{P}$, assuming $\rho^j \in C^j(M)$, $j = P, Q$. This is the setting of the design of experiments literature.

In what follows, I will separately characterize the MRA model for these two collections of preferences.

6.2.1 Preferences over decision rules

Suppose choice data comes as a collection of preferences on decision rules $\succcurlyeq_{\boldsymbol{.}} = \{\succcurlyeq_P \subseteq \mathcal{D}^2 : P \in \mathcal{P}\}$, one for each (fixed) experiment. In the Bayesian persuasion literature, an agent described by such choice data is called a *Receiver*, since she draws inference from information received from a source beyond her control. I will still call this agent a DM, since she has preferences over *decision* rules.

Before presenting a representation theorem with this alternative data set, I define *conditional preferences*. Given P^*, the preference $\succcurlyeq_\theta$ conditional on θ is defined by $\rho \succcurlyeq_\theta \tau$ if, and only if, $\rho_{\{\theta\}}\gamma \succcurlyeq_{P^*} \tau_{\{\theta\}}\gamma$ for all $\gamma \in \mathcal{D}$. The following axiom imposes straightforward adaptations of axioms 1 to 6 for $\succcurlyeq_{\boldsymbol{.}}$, with similar interpretations:

Axiom D1: For every experiment $P \in \mathcal{P}$ and all decision rules $\rho, \tau \in \mathcal{D}$:

1 If $\rho \succcurlyeq_P \tau$, $\rho' P' = \rho P$ and $\tau' P' = \tau P$, then $\rho' \succcurlyeq_{P'} \tau'$.

2 $\succcurlyeq_P$ is complete and transitive.

3 $\rho_{\{\theta\}}\gamma \succcurlyeq_{P^*} \tau_{\{\theta\}}\gamma$ implies $\rho_{\{\theta\}}\gamma' \succcurlyeq_{P^*} \tau_{\{\theta\}}\gamma'$, for all $\gamma, \gamma' \in \mathcal{D}$ and $\theta \in \Theta$.

4 If $\rho \succcurlyeq_\theta \tau$ for all $\theta \in \Theta$, then $\rho \succcurlyeq_{P^*} \tau$.

5 If $\rho \succcurlyeq_\theta \tau$, then $\alpha\rho + (1-\alpha)\gamma \succcurlyeq_\theta \alpha\tau + (1-\alpha)\gamma$ for all $\alpha \in (0, 1]$ and $\gamma \in \mathcal{D}$.

6 The set $\{(\rho P, \tau P) \in \mathcal{F}^2 : \rho \succcurlyeq_P \tau\}$ is closed.

I can now state the following representation theorem.

Theorem 4. *The system of preferences $\succcurlyeq_{\boldsymbol{.}} = \{\succcurlyeq_P \subseteq \mathcal{D}^2 : P \in \mathcal{P}\}$ satisfies Axiom D1 if, and only if, there exists a utility function $v : A \times \Theta \to \mathbb{R}$, continuous in the first argument, and a monotone, continuous functional $J : \mathcal{R}_v \to \mathbb{R}$ such that, for all $P \in \mathcal{P}$ and $\rho, \tau \in \mathcal{D}$,*

$$\rho \succcurlyeq_P \tau \iff J(r_v(\rho, P)) \geq J(r_v(\tau, P)). \tag{10}$$

We then say that $\succcurlyeq_{\boldsymbol{.}}$ has an MRA representation (v, J). Moreover, v is parameter-wise cardinally unique.

Theorem 4 characterizes the MRA model when behavioral data comes in the form of preferences over decision rules, for experiments that are exogenous from the DM's standpoint. This is useful because it provides a method to verify whether choices are compatible with the SDT framework, without having to vary decision rules and experiments simultaneously. For instance, in laboratory settings such data can be elicited by exogenously varying the experiment and available actions, while keeping track of action frequencies for each realized signal.

6.2.2 Preferences over experiments

Now turn to an agent who chooses an information structure for each menu M of available decision rules knowing that, in a second stage, a decision rule will be chosen for her (and, potentially, by her) from M. To differentiate this agent from the DM, who chooses decision rules, I call her the *Experimenter*.[4]

Formally, identify $\mathcal{D}$ with the space of measurable functions from X to $\Delta(A)$ and, as usual, endow it with the topology of point-wise convergence. A menu $M \in \mathcal{M}$ is some closed (thus compact) subset of $\mathcal{D}$. I now assume that the primitive is a system of preferences $\succcurlyeq^{\cdot} = \{\succcurlyeq^M \subseteq \mathcal{P}^2 : M \in \mathcal{M}\}$. I want to characterize which such preference systems are compatible with the MRA model.

Recall that a decision rule $\rho \in \mathcal{D}$ is called invariant if $\rho_P = \rho_{P'}$ for all $P, P' \in \mathcal{P}$. Denote by $\overline{\mathcal{M}}$ the class of all menus consisting only of invariant decision rules. Throughout this section, I assume that $\mathcal{P} = \Delta(X)^{\Theta}$. When this is the case, appendix A.1 shows that there exists an invariant decision rule $\delta^* \in \overline{\mathcal{D}}$ such that for all $(\rho, P) \in \mathcal{S}$, we have $\delta^* P' = \rho P$ for some $P' \in \mathcal{P}$.

First I state some postulates involving only preferences over experiments. These are versions of axioms 1 and 3 to 5 in such a space.

Axiom E1: For all menus $M \in \mathcal{M}$, $\overline{M} \in \overline{\mathcal{M}}$ and experiments $P, Q \in \mathcal{P}$:

1. **Weak Order:** $\succcurlyeq^M$ is complete and transitive.
2. **IIP:** $P_{(\theta)}R \succcurlyeq^{\overline{M}} Q_{(\theta)}R$ implies $P_{(\theta)}R' \succcurlyeq^{\overline{M}} Q_{(\theta)}R'$, for all $R, R' \in \mathcal{P}$ and $\theta \in \Theta$.
3. **Monotonicity:** if $P_{(\theta)}R \succcurlyeq^{\overline{M}} Q_{(\theta)}R$ for all $\theta \in \Theta$ and $R \in \mathcal{P}$, then $P \succcurlyeq^{\overline{M}} Q$.
4. **Experiment CMI:** if $P_{\theta'} = Q_{\theta'} = R_{\theta'}$ for every $\theta' \neq \theta \in \Theta$ and $P \succcurlyeq^{\overline{M}} Q$, then $\alpha P + (1-\alpha)R \succcurlyeq^{\overline{M}} \alpha Q + (1-\alpha)R$ for all $\alpha \in (0, 1]$.
5. **Continuity:** the set $\{(\delta P, \delta Q) \in \mathcal{F}^2 : P \succcurlyeq^{\{\delta\}} Q\}$ is closed for all $\delta \in \mathcal{D}$.

Now consider a family of *choice correspondences* $C = \{C_P : P \in \mathcal{P}\}$, where

$$\begin{aligned} C_P : \mathcal{M} &\to 2^{\mathcal{D}} \setminus \{\emptyset\} \\ M &\mapsto C_P(M) \subseteq M. \end{aligned} \tag{11}$$

The inclusion $\rho \in C_P(M)$ means that, in a second stage, ρ is chosen from the menu M of feasible decision rules, when the experiment is P. The following axiom requires that information structures be chosen taking into account the decision rule selected by C, and that the Experimenter is a consequentialist, caring only about induced action distributions.

Axiom E2 (Consistency): If $P \succcurlyeq^M P'$, and for all $\rho \in C_P(M)$, $\rho' \in C_{P'}(M)$, $\tau \in C_Q(N)$ and $\tau' \in C_{Q'}(N)$ we have $\rho P = \tau Q$ and $\rho' P' = \tau' Q'$, then $Q \succcurlyeq^N Q'$.

Axioms E1 and E2 are sufficient to characterize the Experimenter's preferences given menus from which a single decision rule is chosen by C for every experiment (for instance, singleton menus). For all

[4]In Bayesian persuasion literature, such an agent is called a Sender.

other menus, we need to elicit the Experimenter's tie-breaking rule from her choices. Since in this paper I am mostly concerned a single sophisticated DM, it is natural to assume that indifference is decided in favor of the Experimenter. The next axiom guarantees just that, but first I introduce another bit of notation: for any $M \in \mathcal{M}$ and $\tau \in C_P(M)$, let $P^\tau \in \mathcal{P}$ satisfy $\delta^* P^\tau = \tau P$.

Axiom E3 (Optimism): If $\rho \in C_P(M)$ and $P^\rho \succcurlyeq^{\{\delta^*\}} P^\gamma$ for all $\gamma \in C_P(M)$, then $P \succcurlyeq^M Q$ ($Q \succcurlyeq^M P$) if, and only if, $P \succcurlyeq^{\tilde{M}} Q$ ($Q \succcurlyeq^{\tilde{M}} P$), where $\tilde{M} = \{\delta \in \mathcal{D} : \delta_P = \rho_P \text{ and } \forall R \neq P,\ \exists \gamma \in M \text{ s.t. } \delta_R = \gamma_R\}$.

Axiom E3 is conceptually straightforward. If decision rule τ induces an action distribution under P which is deemed worse by the Experimenter than that induced by ρ, then effectively removing from the menu the possibility of pairing τ with P does not change preferences. In other words, the Experimenter is an optimist, who believes the decision rule chosen in the second stage will be the best possible one, among the alternatives that are deemed "choosable" by C.

I can now characterize the Experimenter's preferences.

Theorem 5. *A system of preferences $\succcurlyeq^{\cdot} = \{\succcurlyeq^M \subseteq \mathcal{P}^2 : M \in \mathcal{M}\}$ satisfies axioms E1 to E3 if, and only if, there exists a utility function $w : A \times \Theta \to \mathbb{R}$, continuous in the first argument, and a monotone, continuous functional $H : \mathcal{R}_w \to \mathbb{R}$ such that, for all $P, Q \in \mathcal{P}$ and $M \in \mathcal{M}$,*

$$P \succcurlyeq^M Q \iff H(r_w(\rho, P)) \geq H(r_w(\tau, Q)) \quad \text{subject to } \rho \in C_P(M),\ \tau \in C_Q(M). \tag{12}$$

We then say that $\succcurlyeq^{\cdot}$ has an MRA representation (w, H). Moreover, w is parameter-wise cardinally unique.

Note that Theorem 5 makes no assumptions about the properties of the choice correspondence. In particular, the decision rule that is ultimately chosen, for any given experiment, could differ from the one the Experimenter would prefer. The Experimenter can thus be seen as a sophisticated, but possibly dynamically inconsistent, agent. She correctly anticipates – but does not necessarily control – which decision rule will be ultimately chosen, and designs the information structure accordingly. Of course, if C takes the form

$$C_P(M) = \underset{\delta \in M}{\arg\max}\, H(r_w(\delta, P)) \quad \forall M \in \mathcal{M} \text{ and } P \in \mathcal{P}, \tag{13}$$

then the Experimenter's problem is equivalent to her sequentially choosing an experiment and then a decision rule for herself. To spell out sufficient conditions when (13) is true could be an interesting avenue for future research, but is beyond the scope of the current paper.

7 Discussion and conclusion

In this paper, I have built an axiomatic foundation that can be used to obtain representation theorems for many statistical decision theoretic models. These theorems provide the full behavioral implications of said models, in terms of preferences over pairings of decision rules and experiments. Such representation results serve two main purposes in positive economics. First, they provide applied theorists working within the SDT framework with a deeper understanding of the behavior of agents in their models. Second, they allow empiricists to gauge the extent to which observed choice behavior deviates from the behavioral predictions of different SDT models. As SDT is the setting of much of information economics, results of this kind may find many uses in the wider economics literature.

Representation theorems such as the ones presented in sections 4 and 5 also serve a normative purpose. For instance, Manski (2021) advocates for the use of statistical decision theory in econometrics, but leaves open the issue of what objective function the econometrician should use. In this context, axioms can serve as a guide for the econometrician's choice of utility function and aggregator, when formulating an SDT model. That is, the econometrician should choose a statistical decision theoretic objective function that leads to desirable constraints on preferences, formalized as normatively appealing axioms.

For this paper, I have mainly focused on obtaining the behavioral implications of a canonical SDT model in which the DM takes the expected utility conditional on each parameter, and aggregates across parameters using a monotone functional. Therefore, although the methodology I presented in sections 4 and 5 is flexible enough to provide characterizations of a wide array of SDT models, it is by no means exhaustive. There are at least three particularly interesting classes of models which can be considered as belonging to the SDT family, but are ruled out by the approach taken here.

The first pertains to models in which not only choices, but DM's preferences themselves, depend on the menu of available options. The canonical example of this is regret-based preferences, as discussed in example 1. Second, by maintaining the Consequentialism axiom throughout the analysis, I am precluding models where the utility function depends not only on actions and parameters, but directly on decision rules and experiments. As discussed in example 4, one such model is the classic rational inattention framework with cognitive costs. A third class of SDT models not contemplated here is one in which, when the true parameter is known, the DM evaluates prospects using a criterion other than maximizing expected utility. For example, Manski and Tetenov (2014) evaluate the performance of a treatment rule based on maximizing a utility quantile, rather than its expectation, conditional on each parameter. Expanding the framework outlined here to any one of these classes of models presents a fruitful avenue for future decision theoretic research.

Since there are situations in which data on preferences over pairings is not easy to come by, in Section 6 I expanded the scope of applicability of my main results by characterizing the MRA model assuming other data sets. In Section 6.1, I considered a data set consisting of action probabilities conditional on each parameter. This type of data is standard in psychometrics, and has recently received attention in economics. I showed when such data can be rationalized by a preference over pairings that has an MRA representation. Moreover, example 6 discusses sufficient conditions for preferences over decision rules and experiments to be identifiable from parameter dependent stochastic choice. This opens the possibility of extending the results of this paper to the case when the DM faces information costs, and thus building on the work of Caplin and Dean (2015) by (i) not assuming the existence of a utility function in their representations; and (ii) extending their results beyond Bayesian DMs.

Section 6.2 provides an MRA characterization for two systems of preferences. The first describes a DM's preferences over decision rules for each given experiment. The second describes a DM who correctly anticipates which decision rule will be chosen from a menu and ranks experiments accordingly. These preferences naturally arise in many applications, and can be viewed as two cross sections of the preference over pairings. Another interesting topic for future research will be to provide sufficient conditions for the MRA representations of these two systems of preferences to coincide, so that they can be interpreted as modelling the same agent.

A Proofs

A.1 Proof of Lemma 1

We will prove the following, stronger result.

Lemma 4. For all $[\sigma] \in \mathcal{S}/_*$,

1. There exists $\rho^\sigma \in \mathcal{D}$ such that $(\rho^\sigma, P^*) \in [\sigma]$.
2. If $\mathcal{P} = \Delta(X)^\Theta$, then there exists an invariant $\delta^* \in \overline{\mathcal{D}}$ such that $(\delta^*, P^\sigma) \in [\sigma]$ for some $P^\sigma \in \mathcal{P}$.

Proof: Fix any $\sigma = (\rho, P) \in \mathcal{S}$. Recall that $\rho P_\theta(E) \equiv \int_X \rho_P(x, E) \mathrm{d}P_\theta(x)$ for all $\theta \in \Theta$ and $E \in \mathcal{A}$.

1. Construct ρ^σ by setting $\rho^\sigma_{P^*}(x, E) = \rho P_\theta(E)$ for all $E \in \mathcal{A}$, $\theta \in \Theta$ and $x \in S_\theta$. Then,

$$\rho^\sigma P^*_\theta(E) = \int_X \rho^\sigma_{P^*}(x, E) \mathrm{d}P^*_\theta(x) = \int_{S_\theta} \rho P_\theta(E) \mathrm{d}P^*_\theta(x) = \rho P_\theta(E) \quad \forall E \in \mathcal{A}.$$

2. First assume that A is countable and take any surjective decision rule $\delta^* \in \overline{\mathcal{D}}$. For all $a \in A$, let $F_a = \{x \in X : \delta^*(x) = a\}$, and define $P^\sigma_\theta(F_a) = \rho P_\theta(\{a\})$ for all $\theta \in \Theta$ and $a \in A$. Then, $\delta^* P^\sigma(\{a\}) = P^\sigma(\delta^{*-1}(\{a\})) = P^\sigma_\theta(F_a) = \rho P_\theta(\{a\})$.

   Now suppose A is uncountable. Since A is standard Borel, there exists a bijective measurable $\delta^* : X \to A$ such that its inverse is also measurable. This implies that the σ-algebra generated by δ^* coincides with $\mathcal{X}$, hence setting $P^\sigma_\theta(\delta^{*-1}(E)) = \rho P_\theta(E)$, for all $E \in \mathcal{A}$ and $\theta \in \Theta$, defines an experiment on $\mathcal{X}$. By construction, $\delta^* P^\sigma = \rho P$. $\square$

A.2 Proof of Theorem 1

From Axiom 2, we know that $\mathcal{S}/_*$ is finer than $\mathcal{S}/_\sim$. Therefore, in the presence of Axiom 2, providing a representation of $\succsim \subseteq \mathcal{S}^2$ is equivalent to characterizing the preference $\succeq$ on $\mathcal{F}$ defined in eq. (8). Denote the symmetric and asymmetric parts of $\succeq$ by $\simeq$ and $\succ$, respectively. It is easy to see that if $\succsim$ is transitive, so is $\succeq$. By Lemma 1, for all $f \in \mathcal{F}$, there exists $\rho \in \mathcal{D}$ such that $f = \rho P^*$. Therefore, completeness and reflexivity of $\succsim$ imply that the same properties hold for $\succeq$.

Consider the following axioms on $\succeq$:

Axiom A3: For all $f, g, h, h' \in \mathcal{F}$ and $\theta \in \Theta$: if $f_{(\theta)}h \succeq g_{(\theta)}h$, then $f_{(\theta)}h' \succeq g_{(\theta)}h'$.

Axiom A4: For all $f, g \in \mathcal{F}$, if $f_{(\theta)}h \succeq g_{(\theta)}h$ for every $\theta \in \Theta$ and $h \in \mathcal{F}$, then $f \succeq g$.

Axiom A5: For all $f, g, h \in \mathcal{F}$ and $\theta \in \Theta$, if $f(\theta') = g(\theta') = h(\theta')$ for all $\theta' \neq \theta$, then $f \succeq g$ implies $\alpha f + (1-\alpha)h \succeq \alpha g + (1-\alpha)h$ for all $\alpha \in (0, 1]$.

Axiom A6: The set $\{(f, g) \in \mathcal{F}^2 : f \succeq g\}$ is closed.

Axiom A7: For any $f, g \in \mathcal{F}$, $f \succeq g \implies f_{(\theta)}h \succeq g_{(\theta)}h$ for every $\theta \in \Theta$ and $h \in \mathcal{F}$.

Axiom A8: For any $f, g \in \mathcal{F}$ and $\theta \in \Theta$, if $f(\theta') = g(\theta')$ for all $\theta' \neq \theta$, then $f \succeq g$ or $g \succeq f$.

We have the following result linking the axioms above to axioms 3 to 7.

Lemma 5. Let $\succsim \subseteq \mathcal{S}^2$ satisfy Consequentialism. Then, the following statements hold:

1. If $\succsim$ satisfies Axiom 3, then $\succeq$ satisfies Axiom A3.
2. If $\succsim$ satisfies Axiom 4, then $\succeq$ satisfies Axiom A4.
3. If $\succsim$ satisfies Axiom 5, then $\succeq$ satisfies Axiom A5.
4. If $\succsim$ satisfies Axiom 6, then $\succeq$ satisfies Axiom A6.
5. If $\succsim$ satisfies Axiom 7, then $\succeq$ satisfies Axiom A7.
6. If $\succsim$ satisfies Axiom 8, then $\succeq$ satisfies Axiom A8.

Proof:

1. Take any $f,g,h,h' \in \mathcal{F}$ such that $f_{(\theta)}h \succeq g_{(\theta)}h$. By Lemma 1, there exists $\rho,\tau,\gamma,\gamma' \in \mathcal{D}$ such that $f=\rho P^*,\ g=\tau P^*,\ h=\gamma P^*$ and $h'=\gamma' P^*$. Fix any $\theta \in \Theta$. Then,

$$(\rho_{\{\theta\}}\gamma)P^*_{\theta'}(\cdot) = \int_{S_{\theta'}} \rho_{\{\theta\}}\gamma_{P^*}(x,\cdot)\mathrm{d}P^*_{\theta'}(x) = \begin{cases} \rho P^*_{\theta}(\cdot), & \text{if } \theta'=\theta \\ \gamma P^*_{\theta'}(\cdot), & \text{if } \theta' \neq \theta, \end{cases}$$

hence $f_{(\theta)}h = (\rho_{\{\theta\}}\gamma)P^*$, and similarly for $g_{(\theta)}h = (\tau_{\{\theta\}}\gamma)P^*$, $f_{(\theta)}h' = (\rho_{\{\theta\}}\gamma')P^*$, and $g_{(\theta)}h' = (\tau_{\{\theta\}}\gamma')P^*$. By Consequentialism and $f_{(\theta)}h \succeq g_{(\theta)}h$, we have $(\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*)$. By Axiom 3, this implies $(\rho_{\{\theta\}}\gamma', P^*) \succsim (\tau_{\{\theta\}}\gamma', P^*)$, and the definition of $\succeq$ gives us $f_{(\theta)}h' \succeq g_{(\theta)}h'$.

2. Let $f,g \in \mathcal{F}$ be such that $f_{(\theta)}h \succeq g_{(\theta)}h$ for all $\theta \in \Theta$ and $h \in \mathcal{F}$. Lemma 1 implies that for all $h \in \mathcal{F}$, there exists $\gamma \in \mathcal{D}$ such that $h=\gamma P^*$. Let $\rho P^* = f$ and $\tau P^* = g$. Now, an argument analogous to the one in part 1 implies that, for all $\theta \in \Theta$ and $h \in \mathcal{F}$, there exists $\gamma \in \mathcal{D}$ with $f_{(\theta)}h = (\rho_{\{\theta\}}\gamma)P^*$ and $g_{(\theta)}h = (\tau_{\{\theta\}}\gamma)P^*$. Consequentialism and $f_{(\theta)}h \succeq g_{(\theta)}h$ for all $h \in \mathcal{F}$ now imply that $(\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*)$ for all $\gamma \in \mathcal{D}$. By Axiom 4, we have $(\rho, P^*) \succsim (\tau, P^*)$, hence $f \succeq g$.

3. Fix $\theta \in \Theta$ and take any $f,g,h \in \mathcal{F}$ such that $f(\theta')=g(\theta')=h(\theta')$ for all $\theta' \neq \theta$. Assume $f \succeq g$. By Lemma 1, there exists $\rho,\tau,\gamma,\kappa \in \mathcal{D}$ such that $f=(\rho_{\{\theta\}}\kappa)P^*$, $g=(\tau_{\{\theta\}}\kappa)P^*$ and $h=(\gamma_{\{\theta\}}\kappa)P^*$. By Consequentialism, $f \succeq g$ implies $(\rho_{\{\theta\}}\kappa, P^*) \succsim (\tau_{\{\theta\}}\kappa, P^*)$. Applying Axiom 5 we get $(\alpha\rho_{\{\theta\}}\kappa + (1-\alpha)\gamma_{\{\theta\}}\kappa, P^*) \succsim (\alpha\tau_{\{\theta\}}\kappa + (1-\alpha)\gamma_{\{\theta\}}\kappa, P^*)$ for all $\alpha \in (0,1]$. The result follows from how ρ, τ, γ and κ were defined.

4. This follows immediately from the definition of $\succeq$.

5. Take $f,g \in \mathcal{F}$ and suppose $f \succeq g$. Then there exists $\rho,\tau \in \mathcal{D}$ such that $f=\rho P^*$ and $g=\tau P^*$. By Axiom 2, $(\rho, P^*) \succsim (\tau, P^*)$, thus by Axiom 7, $(\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*)$ for all $\theta \in \Theta$ and $\gamma \in \mathcal{D}$. Lemma 1 and Axiom 2 then imply that $f_{(\theta)}h \succeq g_{(\theta)}h$ for every $\theta \in \Theta$ and $h \in \mathcal{F}$.

6. Fix any $\theta \in \Theta$ and suppose $f,g \in \mathcal{F}$ are such that $f(\theta')=g(\theta')$ for all $\theta' \neq \theta$. By a now familiar argument, there exist $\rho,\tau,\gamma \in \mathcal{D}$ such that $f=(\rho_{\{\theta\}}\gamma)P^*$ and $g=(\tau_{\{\theta\}}\gamma)P^*$. Axiom 8 implies that $(\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*)$ or $(\tau_{\{\theta\}}\gamma, P^*) \succsim (\rho_{\{\theta\}}\gamma, P^*)$. The result follows. $\square$

We are now ready to prove Theorem 1. It is routine to check that the representation satisfies the axioms, so we only prove sufficiency. Let $\succsim$ be a reflexive and transitive preference satisfying axioms 2 to 8, and define $\succeq$ as in (8). If $\succsim$ is trivial, it is clear that it has a dominance representation via a constant utility function u, and we are done. So assume that there exists $\sigma,\sigma' \in \mathcal{S}$ such that $\sigma \succ \sigma'$. By Lemma 5, $\succeq$ is reflexive, transitive and satisfies axioms A3 to A8.

For each $\theta \in \Theta$, define the conditional preference $\succeq_\theta$ on $\Delta(A)$ as follows:

$$\forall f(\theta), g(\theta) \in \Delta(A),\ f(\theta) \succeq_\theta g(\theta) \iff f \succeq g \text{ and } f(\theta')=g(\theta')\ \forall \theta' \neq \theta. \tag{14}$$

Axiom A3, and reflexivity and transitivity of $\succeq$ guarantee that $\succeq_\theta$ is well-defined, transitive and reflexive for every $\theta \in \Theta$, while Axiom A8 implies that it is also complete. By Axiom A5, $\succeq_\theta$ satisfies *Independence*: if $p \succeq_\theta q$, then $\alpha p + (1-\alpha)s \succeq_\theta \alpha q + (1-\alpha)s$ for any $\alpha \in (0,1]$ and $s \in \Delta(A)$. Take sequences $(p_n),(q_n) \in \Delta(A)$ such that $p_n \to p$, $q_n \to q$ and $p_n \succeq_\theta q_n$ for all $n \geq 1$. Axiom A6 implies that $p,q \in \Delta(A)$ and $p \succeq_\theta q$. Since $\Delta(A)$ is a metric space, this implies that $\{(p,q) \in \Delta(A)^2 : p \succeq_\theta q\}$ is closed. We can thus apply the classic expected utility representation theorem for compact prize spaces – see, e.g., Kreps (2018, Chapter 5) – to find, for every $\theta \in \Theta$, a continuous utility function $u_\theta : A \to \mathbb{R}$

such that, for all $f(\theta), g(\theta) \in \Delta(A)$,

$$f(\theta) \succeq_\theta g(\theta) \iff \int_A u_\theta \mathrm{d} f(\theta) \geq \int_A u_\theta \mathrm{d} g(\theta). \tag{15}$$

Now take any $f, g \in \mathcal{F}$ such that $f \succeq g$. By axioms A4 and A7, we must have $f(\theta) \succeq_\theta g(\theta)$ for all $\theta \in \Theta$. Considering the representations in (15), this implies that for all $f, g \in \mathcal{F}$,

$$f \succeq g \iff \int_A u_\theta \mathrm{d} f(\theta) \geq \int_A u_\theta \mathrm{d} g(\theta) \quad \forall \theta \in \Theta.$$

In view of (8) and Axiom 2, this implies that, for all $(\rho, P), (\tau, Q) \in \mathcal{S}$,

$$(\rho, P) \mathrel{\hat{\succsim}} (\tau, Q) \iff \int_A u_\theta \mathrm{d}\rho P_\theta \geq \int_A u_\theta \mathrm{d}\tau Q_\theta \quad \forall \theta \in \Theta \iff \int_X u_\theta(\rho) \mathrm{d} P_\theta \geq \int_X u_\theta(\tau) \mathrm{d} Q_\theta \quad \forall \theta \in \Theta.$$

The representation obtains by defining $u(\cdot, \theta) = u_\theta$ for each $\theta \in \Theta$. Its uniqueness properties derive from cardinal uniqueness of u_θ for every $\theta \in \Theta$. That is, if u'_θ also represents $\succeq_\theta$, then there exist $b_\theta > 0$ and $c_\theta \in \mathbb{R}$ such that $u'_\theta = b_\theta u_\theta + c_\theta$. This is guaranteed by the expected utility theorem.

Consider the following definition.

Definition 7. We call the parameters in $\Theta^I \equiv \{\theta \in \Theta : (\rho_{\{\theta\}}\gamma, P^*) \mathrel{\hat{\sim}} (\tau_{\{\theta\}}\gamma, P^*) \ \forall \rho, \tau \in \mathcal{D}\}$ *irrelevant.* $\diamondsuit$

Irrelevant parameters get their name from the fact that changing the action distributions conditional on Θ^I does not affect the DM's valuation of a pairing. For future reference, note that $\succeq_\theta$ is trivial for all $\theta \in \Theta^I$. This then implies that any utility representation u_θ of $\succeq_\theta$, with $\theta \in \Theta^I$, is a constant function.

A.3 Proof of Lemma 2

Let $\succsim$ be a preference satisfying axioms 1 to 6 and define its unanimously preferred sub-relation $\hat{\succsim}$ as in Definition 3. We start by proving Proposition 1:

Proof of Proposition 1:

1. The inclusion $\hat{\succsim} \subseteq \succsim$ follows immediately. To prove transitivity, take $\sigma \mathrel{\hat{\succsim}} \sigma'$ and $\sigma' \mathrel{\hat{\succsim}} \tilde{\sigma}$. Then there exist $\rho, \tau, \gamma \in \mathcal{D}$, with $(\rho, P^*) \in [\sigma]$, $(\tau, P^*) \in [\sigma']$ and $(\gamma, P^*) \in [\tilde{\sigma}]$, such that $(\rho_{\{\theta\}}\kappa, P^*) \succsim (\tau_{\{\theta\}}\kappa, P^*)$ and $(\tau_{\{\theta\}}\kappa, P^*) \succsim (\gamma_{\{\theta\}}\kappa, P^*)$ for all $\kappa \in \mathcal{D}$ and $\theta \in \Theta$, . By transitivity of $\succsim$, we have $(\rho_{\{\theta\}}\kappa, P^*) \succsim (\gamma_{\{\theta\}}\kappa, P^*)$, for every $\kappa \in \mathcal{D}$ and $\theta \in \Theta$. Thus, by definition, $\sigma \mathrel{\hat{\succsim}} \tilde{\sigma}$.

2. We want to show that $\hat{\succsim}$ satisfies axioms 2 to 8. First note that since $\succsim$ satisfies axioms 3 and 4, we have that $\hat{\succsim}$ is reflexive, thus non-empty. Axiom 2 follows directly from $\hat{\succsim} \subseteq \succsim$. Axioms 3 and 5 are implied by the Sure Thing Principle (STP) and Mixture Independence (MI) respectively, which we show in parts 3 and 4, below.

   Now, note that for any $T \subseteq \Theta$ and $\rho, \gamma, \gamma' \in \mathcal{D}$,

$$\rho_T \gamma_{\{\theta\}} \gamma' = \begin{cases} \rho_{\{\theta\}}\gamma', & \text{if } \theta \in T \\ \gamma_{\{\theta\}}\gamma', & \text{if } \theta \in \Theta \setminus T. \end{cases} \tag{16}$$

   In particular, since $\succsim$ satisfies IIP, for any $\gamma \in \mathcal{D}$ and $\theta \in \Theta$, $(\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*)$ implies $(\rho_{\{\theta\}}\gamma_{\{\theta'\}}\gamma', P^*) \succsim (\tau_{\{\theta\}}\gamma_{\{\theta'\}}\gamma', P^*)$ for all $\gamma' \in \mathcal{D}$ and $\theta' \in \Theta$, and thus $(\rho_{\{\theta\}}\gamma, P^*) \mathrel{\hat{\succsim}} (\tau_{\{\theta\}}\gamma, P^*)$. Since $\hat{\succsim} \subseteq \succsim$, we have that for all $\theta \in \Theta$ and $\rho, \tau, \gamma \in \mathcal{D}$, $(\rho_{\{\theta\}}\gamma, P^*) \mathrel{\hat{\succsim}} (\tau_{\{\theta\}}\gamma, P^*)$ if, and only if, $(\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*)$. By Consequentialism, we also have that $(\rho, P^*) \mathrel{\hat{\succsim}} (\tau, P^*)$ if, and only if, $(\rho_{\{\theta\}}\gamma, P^*) \succsim (\tau_{\{\theta\}}\gamma, P^*)$ for all $\theta \in \Theta$ and $\gamma \in \mathcal{D}$. Taken together, these two facts facts imply axioms 4 and 7.

Recalling the definition in (14) and since $\Delta(A)$ is a metric space, we have that $\{(f(\theta), g(\theta)) \in \Delta(A)^2 : f(\theta) \succeq_\theta g(\theta)\}$ is closed if, and only if, it is sequentially closed. Also recall from Appendix A.2 that $(\rho, P) \hat{\succsim} (\tau, Q) \iff f(\theta) \succeq_\theta g(\theta)$ for all $f = \rho P$, $g = \tau Q$ and $\theta \in \Theta$. Therefore, $\{(\rho P, \tau Q) \in \mathcal{F}^2 : (\rho, P) \hat{\succsim} (\tau, Q)\}$ is closed if, and only if, $\bigcap_{\theta\in\Theta}\{(f, g) \in \mathcal{F}^2 : f(\theta) \succeq_\theta g(\theta)\}$ is closed. So consider sequences $\{f^n\}_{n\geq 1}$ and $\{g^n\}_{n\geq 1} \in \mathcal{F}$ such that $f^n \to f$, $g^n \to g$, and $f^n(\theta) \succeq_\theta g^n(\theta)$ for all $\theta \in \Theta$, $n \geq 1$. Since $\succeq_\theta$ is continuous by Appendix A.2, we have that $\{(f, g) \in \mathcal{F}^2 : f(\theta) \succeq_\theta g(\theta)\}$ is closed for every $\theta \in \Theta$. Therefore, the intersection is also closed, which then implies continuity of $\hat{\succsim}$.

3. To prove that $\hat{\succsim}$ satisfies the Sure Thing Principle, first consider $\rho, \tau, \gamma \in \mathcal{D}$ such that $(\rho_T\gamma, P^*) \hat{\succsim} (\tau_T\gamma, P^*)$. Then, by eq. (16), $(\rho_{\{\theta\}}\kappa, P^*) \succsim (\tau_{\{\theta\}}\kappa, P^*)$ for all $\theta \in T$ and $\kappa \in \mathcal{D}$, which by Monotonicity of $\succsim$, implies $(\rho_T\gamma', P^*) \succsim (\tau_T\gamma', P^*)$ for any $\gamma' \in \mathcal{D}$. Again using eq. (16) and the fact that $(\rho_{\{\theta\}}\kappa, P^*) \succsim (\tau_{\{\theta\}}\kappa, P^*)$ for all $\theta \in T$ and $\kappa \in \mathcal{D}$, gives us $(\rho_T\gamma', P^*) \hat{\succsim} (\tau_T\gamma', P^*)$.

4. Take any $P \in \mathcal{P}$ and $\rho, \tau \in \mathcal{D}$ with $(\rho, P) \hat{\succsim} (\tau, P)$. Then, by definition, there exists $\rho', \tau' \in \mathcal{D}$ such that $\rho' P^* = \rho P$, $\tau' P^* = \tau P$ and $(\rho'_{\{\theta\}}\gamma, P^*) \succsim (\tau'_{\{\theta\}}\gamma, P^*)$ for all $\theta \in \Theta$ and $\gamma \in \mathcal{D}$. Now CMI implies that $(\alpha\rho'_{\{\theta\}}\gamma + (1-\alpha)\kappa'_{\{\theta\}}\gamma, P^*) \succsim (\alpha\tau'_{\{\theta\}}\gamma + (1-\alpha)\kappa'_{\{\theta\}}\gamma, P^*)$ for all $\theta \in \Theta$, $\alpha \in (0, 1]$ and $\kappa', \gamma \in \mathcal{D}$. By Lemma 1, for any $\kappa P \in \mathcal{F}$, there exists $\kappa' \in \mathcal{D}$ such that $\kappa' P^* = \kappa P$. Therefore, for all $\kappa \in \mathcal{D}$ and $\alpha \in (0, 1]$, there exists $\kappa' \in \mathcal{D}$ such that $(\alpha\rho + (1-\alpha)\kappa)P = (\alpha\rho' + (1-\alpha)\kappa')P^*$, $(\alpha\tau + (1-\alpha)\kappa)P = (\alpha\tau' + (1-\alpha)\kappa')P^*$, and $((\alpha\rho' + (1-\alpha)\kappa')_{\{\theta\}}\gamma, P^*) \succsim ((\alpha\tau' + (1-\alpha)\kappa')_{\{\theta\}}\gamma, P^*)$ for all $\theta \in \Theta$ and $\gamma \in \mathcal{D}$. We conclude that $\hat{\succsim}$ satisfies Mixture Independence.

5. Suppose $\succsim$ is non-trivial. Then, by completeness, there exists $(\rho, P) \succ (\tau, Q)$. Lemma 1 now guarantees that there exists $\rho', \tau' \in \mathcal{D}$ with $(\rho', P^*) \succ (\tau', P^*)$. Let $T = \{\theta \in \Theta : (\rho'_{\{\theta\}}\gamma, P^*) \succsim (\tau'_{\{\theta\}}\gamma, P^*)\ \forall\gamma \in \mathcal{D}\}$, and note that by IIP and Monotonicity, $T \neq \emptyset$. Monotonicity now implies that $(\rho'_T\tau', P^*) \succsim (\rho', P^*) \succ (\tau', P^*)$, and by transitivity, $(\rho'_T\tau', P^*) \succ (\tau', P^*)$. By construction of T and eq. (16), $(\rho'_T\tau', P^*) \hat{\succsim} (\tau', P^*)$. Again by Monotonicity, there exists $\theta \in T$ and $\gamma \in \mathcal{D}$ such that $(\tau'_{\{\theta\}}\gamma, P^*) \not\succsim (\rho'_T\tau'_{\{\theta\}}\gamma, P^*)$. Therefore $(\tau', P^*) \hat{\not\succsim} (\rho'_T\tau', P^*)$, which concludes the proof. $\square$

Now, from Proposition 1, we know that $\hat{\succsim}$ is a dominance relation. By Theorem 1, it has a dominance representation by a parameter-wise cardinally unique utility function $u : A \times \Theta \to \mathbb{R}$.

A.4 Proof of Theorem 2

If $\succsim$ is trivial, it can clearly be represented by any utility function u paired with a constant aggregator functional I. So assume that $\succsim$ is not trivial. Let $\succsim$ satisfy axioms 1 to 6 and fix a utility function $u : A \times \Theta \to \mathbb{R}$ that represents its unanimously preferred sub-relation $\hat{\succsim}$. Consider the risk mapping $r_u : \mathcal{S} \to \mathbb{R}^\Theta$ under u defined by (2), and set $\mathcal{R}_u = \{r_u(\sigma) : \sigma \in \mathcal{S}\}$. Note that, as a product of (possibly degenerate) intervals, $\mathcal{R}_u$ is connected, and as a consequence of the Hewitt-Marczewski-Pondiczery theorem (Hewitt, 1946), it is also separable.

Consider the preference relation $\succeq^u$ on $\mathcal{R}_u$ defined by (6). It is easy to see that $\succeq^u$ is transitive, since $r_u(\sigma) \succeq^u r_u(\sigma')$ and $r_u(\sigma') \succeq^u r_u(\tilde{\sigma})$ imply that $\sigma \succsim \sigma'$ and $\sigma' \succsim \tilde{\sigma}$, hence $\sigma \succsim \tilde{\sigma}$ by transitivity of $\succsim$, which in turn implies $r_u(\sigma) \succeq^u r_u(\tilde{\sigma})$. By definition, $\succeq^u$ is also complete on $\mathcal{R}_u$.

Now we prove that $\succeq^u$ is continuous, i.e., that $\{(r, r') \in \mathcal{R}_u^2 : r \succeq^u r'\}$ is closed in the product topology. First note that for every $\theta \in \Theta$, the functional $\Delta(A) \ni f(\theta) \mapsto \int_A u(a, \theta)\mathrm{d}f(\theta)$ is continuous in the topology of weak convergence. Thus, $\tilde{r}_u : \mathcal{F} \to \mathbb{R}^{\Theta}$, defined as

$$\tilde{r}_u(f)(\theta) = \int_A u(a, \theta)\mathrm{d}f(\theta) \quad \forall f \in \mathcal{F},\ \theta \in \Theta, \tag{17}$$

is continuous in the product topology on $\mathbb{R}^{\Theta}$. Now note that $\{(r_u(\sigma), r_u(\sigma')) \in \mathcal{R}_u^2 : r_u(\sigma) \succeq^u r_u(\sigma')\} = \{(\tilde{r}_u(\rho P), \tilde{r}_u(\tau Q)) \in \mathcal{R}_u^2 : (\rho, P) \succsim (\tau, Q)\}$. Let the map $\tilde{R}_u : \mathcal{F}^2 \to \mathcal{R}_u^2$ be defined by $\tilde{R}_u(f, g) = (\tilde{r}_u(f), \tilde{r}_u(g))$. Then $\tilde{R}_u$ is continuous, as a product of continuous functions, and

$$\tilde{R}_u(\{(\rho P, \tau Q) \in \mathcal{F}^2 : (\rho, P) \succsim (\tau, Q)\}) = \{(\tilde{r}_u(\rho P), \tilde{r}_u(\tau Q)) \in \mathcal{R}_u^2 : (\rho, P) \succsim (\tau, Q)\}.$$

Since A is compact, so is $\Delta(A)$. Thus, by Tychonoff's theorem, $\mathcal{F}$ is a compact Hausdorff space, and so is $\mathcal{F}^2$. Then, by Continuity of $\succsim$, $\{(\rho P, \tau Q) \in \mathcal{F}^2 : (\rho, P) \succsim (\tau, Q)\}$ is also compact, since it is a closed subset of a compact space. This makes $\tilde{R}_u(\{(\rho P, \tau Q) \in \mathcal{F}^2 : (\rho, P) \succsim (\tau, Q)\})$ compact as well – because $\tilde{R}_u$ is continuous – hence closed, since $\mathcal{R}_u^2$ is Hausdorff. We conclude that $\succeq^u$ is continuous.

Therefore, $\succeq^u$ is a continuous weak order on a connected and separable topological space. By Herden (1989, Corollary 3.2), there exists a continuous utility function $I : \mathcal{R}_u \to \mathbb{R}$ such that, for all $r_u(\sigma), r_u(\sigma') \in \mathcal{R}_u$, $r_u(\sigma) \succeq^u r_u(\sigma') \iff I(r_u(\sigma)) \geq I(r_u(\sigma'))$. By definition of $\succeq^u$, we obtain, for all $\sigma, \sigma' \in \mathcal{S}$,

$$\sigma \succsim \sigma' \iff r_u(\sigma) \succeq^u r_u(\sigma') \iff I(r_u(\sigma)) \geq I(r_u(\sigma')).$$

Now we prove that $\succsim$ has a representation with an ex-post utility function $u : A \times \Theta \to \mathbb{R}$ such that $u(\cdot, \theta)$ is non-constant for all $\theta \in \Theta$. From Appendix A.2, for any dominance representation $\hat{u}$ of $\hat{\succsim}$, $\hat{u}(\cdot, \theta)$ is constant for all $\theta \in \Theta^I$. Let $(\hat{u}, \hat{I})$ be an MRA representation of $\succsim$ such that $\hat{u}(a, \theta) = k \in \mathbb{R}$ for all $a \in A$ and $\theta \in \Theta^I$. Such a representation exists by the uniqueness properties of Theorem 1. Set $u(\cdot, \theta) = \hat{u}(\cdot, \theta)$ for $\theta \in \Theta \setminus \Theta^I$, and let $u(\cdot, \theta)$ be any non-constant function with $k \in co(u(A, \theta))$ for all $\theta \in \Theta^I$, where $co(\cdot)$ denotes the convex hull of a set. Note that $\mathcal{R}_{\hat{u}} \subseteq \mathcal{R}_u$, and let $I : \mathcal{R}_u \to \mathbb{R}$ be given by $I(r) = \hat{I}(\hat{r})$ if $r(\theta) = \hat{r}(\theta)$ for all $\theta \in \Theta \setminus \Theta^I$, where $r \in \mathcal{R}_u$ and $\hat{r} \in \mathcal{R}_{\hat{u}}$. Clearly $I(r_u(\sigma)) \geq I(r_u(\sigma')) \iff \hat{I}(r_u(\sigma)_{(\Theta^I)}k) \geq \hat{I}(r_u(\sigma')_{(\Theta^I)}k)$, therefore (u, I) also represents $\succsim$.

A.5 Proofs of results in Section 5

Let $\succsim$ be a preference relation on $\mathcal{S}$ satisfying axioms 1 to 6, and take a utility function u that represents $\hat{\succsim}$. Consider the preference relation $\succeq^u$ on $\mathcal{R}_u$ defined by (6). We first formally state and prove a result that was only alluded to in the main text.

Lemma 6. The preference $\succeq^u$ on $\mathcal{R}_u$ satisfies the *basic A-A axioms*

- *Weak Order*: the preference $\succeq^u$ is complete and transitive.
- *A-A Monotonicity*: for all $r, r' \in \mathcal{R}_u$, if $r \geq r'$, then $r \succeq^u r'$.
- *Risk Independence:* for all constant risk functions $r, r', q \in \mathcal{R}_u$ and $\alpha \in [0, 1]$, $r \succeq^u r'$ implies $\alpha r + (1-\alpha)q \succeq^u \alpha r' + (1-\alpha)q$.
- *Mixture Continuity:* the sets $\{\alpha \in [0, 1] : \alpha r + (1-\alpha)r' \succeq^u q\}$ and $\{\alpha \in [0, 1] : q \succeq^u \alpha r + (1-\alpha)r'\}$ are closed for all $r, r', q \in \mathcal{R}_u$.

Proof: We showed in Appendix A.4 that $\succeq^u$ satisfies Weak Order and that $\{(r, r') \in \mathcal{R}_u^2 : r \succeq^u r'\}$ is closed. Take any $r, r', q \in \mathcal{R}_u$ and a sequence $(\alpha_n)_{n\geq 1} \in [0, 1]$, with $\lim_{n\to\infty} \alpha_n = \alpha$, such that $\alpha_n r + (1-\alpha_n)r' \succeq^u q$ for all $n \geq 1$. Since $\{(r, r') \in \mathcal{R}_u^2 : r \succeq^u r'\}$ is closed, we have that $\lim_n(\alpha_n r + (1-\alpha_n)r') = \alpha r + (1-\alpha)r' \succeq^u q$. Hence $\{\alpha \in [0, 1] : \alpha r + (1-\alpha)r' \succeq^u q\}$ is closed. An analogous argument proves that $\{\alpha \in [0,1] : q \succeq^u \alpha r + (1-\alpha) r'\}$ is closed, thus $\succeq^u$ satisfies Mixture Continuity.

A-A Monotonicity follows directly from Theorem 2, since I is monotone. To prove Risk Independence, note that if $r, r' \in \mathcal{R}_u$ are constant, then $r \succeq^u r'$ if and only if $r \geq r'$, by A-A Monotonicity. The result follows immediately. $\square$

Now we turn to Lemma 3.

Proof of Lemma 3: The proof of statement 1 consists of applying the definitions and doing simple algebra, thus will be omitted. To prove statement 2, recall that for all $\rho, \gamma \in \mathcal{D}$, $T \in \Sigma$ and $\theta \in \Theta$,

$$r_u(\rho_T \gamma, P^*)(\theta) = \int_X u(\rho_T \gamma, \theta) \mathrm{d}P^*_\theta = \int_A \int_X u(a,\theta) \rho_T \gamma(x, \mathrm{d}a) P^*_\theta(\mathrm{d}x) = \begin{cases} \int_A u(a,\theta) \mathrm{d}\rho P^*_\theta(a), & \text{if } \theta \in T \\ \int_A u(a,\theta) \mathrm{d}\gamma P^*_\theta(a), & \text{if } \theta \notin T. \end{cases}$$

Therefore, $r_u(\rho_T \gamma, P^*)(\theta) = r_u(\rho, P^*)(\theta)$ for $\theta \in T$, and $r_u(\rho_T \gamma, P^*)(\theta) = r_u(\gamma, P^*)(\theta)$ for $\theta \in \Theta \setminus T$, which proves the statement.

Now take another MRA representation $(\tilde{u}, \tilde{I})$ of $\succsim$, normalized in such a way that $\mathcal{R}_{\tilde{u}} = [0,1]^\Theta$. It is possible to do this because, by Theorem 2, $\succsim$ has an MRA representation with parameter-wise non-constant utility. Then, parameter-wise cardinal uniqueness of u allows us normalize this utility function on each parameter. Note that $r_{\tilde{u}}(\overline{\rho}, P^*)(\theta) = 1$ and $r_{\tilde{u}}(\underline{\rho}, P^*)(\theta) = 0$ for all $\theta \in \Theta$.

If $\sigma \in \mathcal{K}(\hat{\succsim})$, Lemma 1 implies that there exists $\alpha \in [0,1]$ such that $\sigma \in [(\alpha\overline{\rho} + (1-\alpha)\underline{\rho}, P^*)]$, thus by statement 1, $r_{\tilde{u}}(\sigma)(\theta) = \alpha$, for all $\theta \in \Theta$. Conversely, take $r_{\tilde{u}}(\tau, P) \in \mathcal{R}_{\tilde{u}}$ such that $r_{\tilde{u}}(\tau, P)(\theta) = \alpha$ for all $\theta \in \Theta$, and take $\rho P^* = \tau P$. Since $r_{\tilde{u}}(\rho, P^*)_{(\{\theta\})} r_{\tilde{u}}(\gamma, P^*) = r_{\tilde{u}}(\alpha\overline{\rho} + (1-\alpha)\underline{\rho}, P^*)_{(\{\theta\})} r_{\tilde{u}}(\gamma, P^*)$ for any $\gamma \in \mathcal{D}$ and $\theta \in \Theta$, we may apply statement 2 and the definition of $\succeq^u$ to obtain $r_{\tilde{u}}(\rho_{\{\theta\}}\gamma, P^*) \simeq^u r_{\tilde{u}}((\alpha\overline{\rho} + (1-\alpha)\underline{\rho})_{\{\theta\}}\gamma, P^*)$ for all $\theta \in \Theta$. By definition of $\simeq^u$, $(\rho_{\{\theta\}}\gamma, P^*) \sim ((\alpha\overline{\rho} + (1-\alpha)\underline{\rho})_{\{\theta\}}\gamma, P^*)$ for all $\theta \in \Theta$, and by IIP, $(\rho, P^*) \hat{\sim} (\alpha\overline{\rho} + (1-\alpha)\underline{\rho}, P^*)$. Since $\tau P_\theta = \rho P^*_\theta$ for all $\theta \in \Theta$, we have $(\tau, P) \in \mathcal{K}(\hat{\succsim})$, which finishes the proof. $\square$

In view statement 3 of Lemma 3, for the remainder of this section we work with a representation (u, I) of $\succsim$ such that $\mathcal{R}_u = [0,1]^\Theta$, so that $r_u(\sigma)$ being constant implies $\sigma \in \mathcal{K}(\hat{\succsim})$. We start by proving the lemmas which translate SDT axioms on $\succsim$ to the corresponding Anscombe-Aumann axioms on $\succeq^u$.

Lemma 7. If $\succsim$ satisfies Mixture Independence, then $\succeq^u$ satisfies Independence.

Proof: Take $r, r', q \in \mathcal{R}_u$ and assume $\succsim$ satisfies Axiom 9. From Lemma 1 and the definition of $\succeq^u$, there exists $\rho, \tau, \gamma \in \mathcal{D}$ and $P \in \mathcal{P}$ such that, $r = r_u(\rho, P)$, $r' = r_u(\tau, P)$, $q = r_u(\gamma, P)$ and $(\rho, P) \succsim (\tau, P)$. By Mixture Independence, we have $(\alpha\rho + (1-\alpha)\gamma, P) \succsim (\alpha\tau + (1-\alpha)\gamma, P)$ for all $\alpha \in (0,1]$. Therefore, from statement 1 of Lemma 3, we obtain $\alpha r + (1-\alpha) q \succeq^u \alpha r' + (1-\alpha) q$ for all $\alpha \in (0,1]$. $\square$

Lemma 8. If $\succsim$ satisfies Monotone Continuity*, then $\succeq^u$ satisfies Monotone Continuity.

Proof: Take any $r, r', q \in \mathcal{R}_u$ with q constant, and let $(T_n)_{n \geq 1} \in \Sigma$ be such that $T_1 \supseteq T_2 \supseteq \cdots$ and $\bigcap_{n \geq 1} T_n = \emptyset$. Suppose $r \succ^u r'$. By Lemma 1, there exist $\rho, \tau, \kappa \in \mathcal{D}$, with $(\kappa, P^*) \in \mathcal{K}(\hat{\succsim})$, such that $r = r_u(\rho, P^*)$, $r' = r_u(\tau, P^*)$ and $q = r_u(\gamma, P^*)$. From the definition of $\succeq^u$, $(\rho, P^*) \succ (\tau, P^*)$. Then Monotone Continuity* implies that there exists $m \geq 1$ such that $(\kappa_{T_m}\rho, P^*) \succ (\tau, P^*)$. The result then follows from statement 2 of Lemma 3. $\square$

Lemma 9. If $\succsim$ satisfies Hedging, then $\succeq^u$ satisfies Uncertainty Aversion.

Proof: Let $r, r' \in \mathcal{R}_u$ be such that $r \simeq^u r'$. By Lemma 1, there exists $P \in \mathcal{P}$ and $\rho, \tau \in \mathcal{D}$ such that $r = r_u(\rho, P)$, $r' = r_u(\tau, P)$ and $(\rho, P) \sim (\tau, P)$. Then, by Hedging, $(\alpha\rho + (1-\alpha)\tau, P) \succsim (\rho, P)$ for all $\alpha \in (0,1)$. This implies that $r_u(\alpha\rho + (1-\alpha)\tau, P) \succeq^u r_u(\rho, P)$, thus from statement 1 of Lemma 3, we obtain $\alpha r + (1-\alpha) r' \succeq^u r$. $\square$

Lemma 10. If $\succsim$ satisfies CRE-Independence, then $\succeq^u$ satisfies Certainty Independence.

Proof: Take any $r, r', q \in \mathcal{R}_u$ with constant q, and assume $r \succeq^u r'$. A now familiar argument implies that there exists $P \in \mathcal{P}$ and $\rho, \tau, \kappa \in \mathcal{D}$ such that $r = r_u(\rho, P)$, $r' = r_u(\tau, P)$, $q = r_u(\kappa, P)$ and $(\rho, P) \succsim (\tau, P)$. Since q is constant, Lemma 3 guarantees that $(\tau, P) \in \mathcal{K}(\hat{\succsim})$. By CRE-Independence, $(\alpha\rho + (1-\alpha)\kappa, P) \succsim (\alpha\tau + (1-\alpha)\kappa, P)$ for all $\alpha \in (0, 1]$. Applying Lemma 3 once again, we obtain $\alpha r + (1-\alpha)q \succeq^u \alpha r' + (1-\alpha)q$. □

Lemma 11. If $\succsim$ satisfies symmetry, then $\succeq^u$ satisfies:

- *Risk Symmetry*: for all $r, r', q, k \in \mathcal{R}_u$, with q, k constant, and any $T, F \in \Sigma$ such that $T \cap F = \emptyset$, $r_{(T)}q_{(F)}k \succeq^u r'_{(T)}q_{(F)}k$ implies $r_{(T)}k_{(F)}q \succeq^u r'_{(T)}k_{(F)}q$.

Proof: Assume that $r_{(T)}q_{(F)}k \succeq^u r'_{(T)}q_{(F)}k$, where $r, r', q, k \in \mathcal{R}_u$, with q, k constant, and $T \cap F = \emptyset$. There exist $\rho, \tau \in \mathcal{D}$ and $(\kappa, P^*), (\gamma, P^*) \in \mathcal{K}(\hat{\succsim})$ such that $r = r_u(\rho, P^*)$, $r' = r_u(\tau, P^*)$, $q = r_u(\kappa, P^*)$ and $k = r_u(\gamma, P^*)$. Then, by Lemma 3, $(\rho_T\kappa_F\gamma, P^*) \succsim (\tau_T\kappa_F\gamma, P^*)$. From Axiom 13, we get $(\rho_T\gamma_F\kappa, P^*) \succsim (\tau_T\gamma_F\kappa, P^*)$. Applying Lemma 3 once more, we obtain $r_{(T)}k_{(F)}q \succeq^u r'_{(T)}k_{(F)}q$. □

Lemma 12. If $\succsim$ satisfies Weak CRE-Independence, then $\succeq^u$ satisfies:

- *Weak Certainty Independence:* if $r, r', q \in \mathcal{R}_u$ and q is constant, then $\alpha r + (1-\alpha)q \succeq^u \alpha r' + (1-\alpha)q$ implies $\alpha r + (1-\alpha)q' \succeq^u \alpha r' + (1-\alpha)q'$ for any $\alpha \in (0, 1)$ and constant $q' \in \mathcal{R}_u$.

Proof: Suppose $\alpha r + (1-\alpha)q \succeq^u \alpha r' + (1-\alpha)q$, where q is constant and $\alpha \in (0, 1)$. By a familiar argument, there exist $\rho, \tau, \kappa \in \mathcal{D}$ such that $r = r_u(\rho, P^*)$, $r' = r_u(\tau, P^*)$, $q = r_u(\kappa, P^*)$ and $\alpha(\rho, P^*) + (1-\alpha)(\kappa, P^*) \succsim \alpha(\tau, P^*) + (1-\alpha)(\kappa, P^*)$. Applying Weak CRE-Independence, we get $\alpha(\rho, P^*) + (1-\alpha)(\kappa', P^*) \succsim \alpha(\tau, P^*) + (1-\alpha)(\kappa', P^*)$ for all $(\kappa', P^*) \in \mathcal{K}(\hat{\succsim})$. Since any constant $q' \in \mathcal{R}_u$ correspond to $r_u(\kappa', P^*)$ for some κ' (by Lemma 1), we obtain $\alpha r + (1-\alpha)q' \succeq^u \alpha r' + (1-\alpha)q'$ for any $\alpha \in (0, 1)$ and constant $q' \in \mathcal{R}_u$. □

Lemma 13. If $\succsim$ satisfies the Sure Thing Principle, then $\succeq^u$ satisfies:

- *Savage's STP:* for all $r, r', q, q' \in \mathcal{R}_u$ and $T \in \Sigma$, if $r_{(T)}q \succeq^u r'_{(T)}q$ then $r_{(T)}q' \succeq^u r'_{(T)}q'$.

Proof: Take $r, r', q, q' \in \mathcal{R}_u$ and $T \in \Sigma$ with $r_{(T)}q \succeq^u r'_{(T)}q$. As usual, there exists $\rho, \tau, \gamma, \gamma' \in \mathcal{D}$ such that $r = r_u(\rho, P^*)$, $r' = r_u(\tau, P^*)$, $q = r_u(\gamma, P^*)$, $q' = r_u(\gamma', P^*)$ and, by Lemma 3, $(\rho_T\gamma, P^*) \succsim (\tau_T\gamma, P^*)$. The Sure Thing Principle then implies that $(\rho_T\gamma', P^*) \succsim (\tau_T\gamma', P^*)$, and applying Lemma 3 once again we obtain the result. □

Now we can proceed to the proofs of the propositions in Section 5. Again, I will show only sufficiency of the axioms, since necessity is easily verifiable.

Proof of Proposition 2: Assume that $\succsim$ satisfies axioms 1 to 4 and 6, Mixture Independence and Monotone Continuity*. Then $\succeq^u$ satisfies the basic A-A axioms, Independence and Monotone Continuity. Applying the classic SEU representation theorem (see, e.g., Fishburn (1970, Theorem 13.3)),[5] the axioms on $\succeq^u$ are equivalent to the existence a cardinally unique linear function $U : [0, 1] \to \mathbb{R}$ and a finitely additive probability distribution $\pi \in \Delta(\Theta)$ such that, for all $r, r' \in \mathcal{R}_u$,

$$r \succeq^u r' \iff \int_\Theta U \circ r \mathrm{d}\pi \geq \int_\Theta U \circ r' \mathrm{d}\pi.$$

[5]Although SEU theorems in the Anscombe-Aumann framework are usually stated in terms of acts mapping Θ to some set Δ of probability distributions, the same proofs also work when acts take values in Y, where Y is any convex set.

Monotone Continuity then guarantees that $\pi \in \Delta(\Theta)$ is countably additive (Arrow, 1971). From the uniqueness properties of the representation and linearity, we can take U to be the identity function, thus the result obtains. $\square$

Denote by $B(\Sigma)$ the set of bounded real-valued Σ-measurable functions on Θ. Let $B(\Sigma)^*$ be its continuous dual space, i.e., the space of all linear functionals $\phi : B(\Sigma) \to \mathbb{R}$ that are continuous with respect to the topology induced by the uniform norm $\|r\|_\infty = \sup_{\theta\in\Theta} |r(\theta)|$. A well known fact is that $B(\Sigma)^*$ is isomorphic to the set $ba(\Sigma)$ of finitely additive real-valued set functions on Σ which are bounded in the total variation norm.

Recall that we assumed $\mathcal{R}_u$ was endowed with the product topology, rather than the uniform topology. So denote by $B(\Sigma)^\times$ the continuous dual space of $B(\Sigma)$ in the product topology, and note that, since this topology is coarser than the uniform topology, we have that $B(\Sigma)^\times \subseteq B(\Sigma)^*$ (because any functional that is continuous in the former topology is also in the latter). For the same reason, any weak* compact subset of $B(\Sigma)^*$ remains compact as a subset of $B(\Sigma)^\times$.

Proof of Proposition 3: In view of Lemmas 9 and 10, if $\succsim$ satisfies axioms 1 to 4 and 6, Hedging and CRE-Independence, then $\succeq^u$ satisfies the basic A-A axioms, Uncertainty Aversion and Certainty Independence. Following the proof of the MPEU representation theorem from Gilboa and Schmeidler (1989, Theorem 1 and Proposition 4.1), $\succeq^u$ satisfies these axioms if, and only if, there exists a convex set $\Phi \subseteq B(\Sigma)^\times$ of monotone continuous linear functionals on $B(\Sigma)$ such that, for all $r, r' \in \mathcal{R}_u$,

$$r \succeq^u r' \iff \inf_{\phi\in\Phi} \phi(r) \geq \inf_{\phi\in\Phi} \phi(r').$$

Since $B(\Sigma)^*$ is isomorphic to $ba(\Sigma)$ and $B(\Sigma)^\times \subseteq B(\Sigma)^*$, there also exists an isomorphism between Φ and $\Pi \subseteq ba(\Sigma)$. Monotonicity of each $\phi \in \Phi$ implies that every $\pi \in \Pi$ is a (finitely additive) measure. Therefore, for all $r, r' \in \mathcal{R}_u$,

$$r \succeq^u r' \iff \inf_{\pi\in\Pi} \int_\Theta r \mathrm{d}\pi \geq \inf_{\pi\in\Pi} \int_\Theta r' \mathrm{d}\pi.$$

Lemma 3.3 in Gilboa and Schmeidler (1989) then guarantees that $\phi(\mathbf{1}_\Theta) = 1$ for all $\phi \in \Phi$, thus $\pi(\Theta) = 1$ for every $\pi \in \Pi$. Finally, since $\succsim$ satisfies Monotone Continuity*, then Monotone Continuity of $\succeq^u$ guarantees that $\Pi \subseteq \Delta(\Theta)$ (Chateauneuf et al., 2005, Theorem 1). $\square$

Many models in the Anscombe-Aumann setting are characterized only for simple acts, i.e., acts that have a finite range. To make use of such results, we first prove that the propositions in Section 5 hold when restricted to the set $\mathcal{R}_u^0 \equiv \{r \in \mathcal{R}_u : r(\Theta) \subseteq F,\ F \text{ finite}\}$ of all simple risk functions. Then we use standard results and continuity of I to approximate each $r \in \mathcal{R}_u$ by a sequence of functions in $(r_n^0)_{n\geq1} \in \mathcal{R}_u^0$. This leads to the lemma below.

Lemma 14. Suppose $\succsim$ has an MRA representation with utility function u, and $I_0 : \mathcal{R}_u^0 \to \mathbb{R}$ represents the restriction of $\succeq^u$ to $\mathcal{R}_u^0$. Then I_0 has a unique continuous extension $I : \mathcal{R}_u \to \mathbb{R}$ and (u, I) represents $\succsim$.

Proof: Since $\succsim$ has an MRA representation, Theorem 2 applied to $\mathcal{R}_u^0$ implies that I_0 is monotone and continuous. Because I_0 represents the restriction of $\succeq^u$ to $\mathcal{R}_u^0$, there exists (u, I) representing $\succsim$ such that I_0 is a restriction of I to $\mathcal{R}_u^0$. A well known result in functional analysis is that for any measurable function $r \in \mathcal{R}_u$, there exists a sequence $(r_n)_{n\geq1} \in \mathcal{R}_u^0$ such that $r_n \leq r$ for all $n \geq 1$ and $r_n \to r$. Therefore, $\mathcal{R}_u^0$ is dense in $\mathcal{R}_u$.

Take any two continuous extensions $\hat{I}$ and $\tilde{I}$ of I_0. Fix an arbitrary $r \in \mathcal{R}_u \setminus \mathcal{R}_u^0$ and suppose that $\hat{I}(r) \neq \tilde{I}(r)$. Since $\mathbb{R}$ is Hausdorff, there exist open neighborhoods $\hat{N}$ of $\hat{I}(r)$ and $\tilde{N}$ of $\tilde{I}(r)$ such that

$\hat{N} \cap \tilde{N} = \emptyset$. Since $\hat{I}$ and $\tilde{I}$ are continuous, $\hat{I}^{-1}(\hat{N})$ and $\tilde{I}^{-1}(\tilde{N})$ are open neighborhoods of r. However, $\hat{I}^{-1}(\hat{N}) \cap \tilde{I}^{-1}(\tilde{N}) \cap \mathcal{R}_u^0 = I_0^{-1}(\hat{N}) \cap I_0^{-1}(\tilde{N}) \cap \mathcal{R}_u^0 = I_0^{-1}(\hat{N} \cap \tilde{N}) \cap \mathcal{R}_u^0 = \emptyset$, which is absurd. We conclude that $\hat{I}(r) = \tilde{I}(r)$ for all $r \in \mathcal{R}_u$. Therefore, I is the unique continuous extension of I_0 and (u, I) represents $\succsim$. $\square$

Proof of Proposition 4: Since $\succsim$ has a MPEU representation, $\succeq^u$ satisfies the basic A-A axioms and Uncertainty Aversion. By Stoye (2011, Theorem 1(iii)), the restriction of $\succeq^u$ to $\mathcal{R}_u^0$ satisfies Risk Symmetry if, and only if, it has a MEU representation. Using Lemma 14, we can then extend the representation to $\mathcal{R}_u$. $\square$

Proof of Proposition 5: Suppose $\succsim$ satisfies axioms 1 to 4, 6 and 10, Hedging and Weak CRE-Independence. Then $\succeq^u$ satisfies the basic A-A axioms, Uncertainty Aversion and Weak Certainty Independence, and so does the restriction of $\succeq^u$ to $\mathcal{R}_u^0$. According to Maccheroni et al. (2006, Appendix B), this holds if, and only if, there exists a monotone functional $I_0 : \mathcal{R}_u^0 \to \mathbb{R}$ that represents $\succeq^u$ on $\mathcal{R}_u^0$ and satisfies, for all $r, q \in \mathcal{R}_u^0$ and $\alpha, k \in [0, 1]$,

(a) *Translation Invariance:* $I_0(\alpha r + (1 - \alpha)k\,\mathbf{1}_\Theta) = I_0(\alpha r) + (1 - \alpha)k$;

(b) *Normalization:* $I_0(k\,\mathbf{1}_\Theta) = k$;

(c) *Concavity:* $I_0(\alpha r + (1 - \alpha)q) \geq \alpha I_0(r) + (1 - \alpha)I_0(q)$.

Moreover, by Theorem 2 and the fact that, for a given u, $I_0 : \mathcal{R}_u^0 \to \mathbb{R}$ is unique up to normalization, we have that I_0 is continuous in the topology of point-wise convergence.

Define $I : \mathcal{R}_u \to \mathbb{R}$ as the unique continuous extension of I_0. By Lemma 14, (u, I) represents $\succsim$. Clearly, I satisfies (b). To see that it satisfies (a), let $r \in \mathcal{R}_u$ and take any $(r_n) \in \mathcal{R}_u^0$ such that $r_n \to r$. Then $I(\alpha r + (1 - \alpha)k\,\mathbf{1}_\Theta) = I(\lim_n \alpha r_n + (1 - \alpha)k\,\mathbf{1}_\Theta) = \lim_n I_0(\alpha r_n + (1 - \alpha)k\,\mathbf{1}_\Theta) = \lim_n I_0(\alpha r_n) + (1 - \alpha)k = I(\alpha r) + (1 - \alpha)k$, which implies that I is translation invariant. A similar argument implies that I is concave.

Applying the Fenchel-Moreau theorem as in Cerreia-Vioglio et al. (2014, Section 5), we can identify $I : \mathcal{R}_u \to \mathbb{R}$ with its convex biconjugate $I(r) = \inf_{\phi \in B(\Sigma)^\times} \{\phi(r) - I^*(\phi)\}$, where $\phi \mapsto I^*(\phi) = \inf_{r \in \mathcal{R}_u} \{\phi(r) - I(r)\}$ is concave, upper semicontinuous and $\sup_{\phi \in B(\Sigma)^\times} I^*(\phi) = 0$. Finally, recall that $B(\Sigma)^\times \subseteq B(\Sigma)^*$, thus $B(\Sigma)^\times$ is isomorphic to a subset $\Pi_0 \subseteq ba(\Sigma)$. Defining $c^* : \Pi_0 \to [0, \infty]$ by $c^*(\pi) = -\inf_{r \in \mathcal{R}_u} \{\int r\mathrm{d}\pi - I(r)\}$, we can thus write

$$I(r) = \inf \left\{ \int r\mathrm{d}\pi + c^*(\pi) : \pi \in \Pi_0 \right\} \quad \forall r \in \mathcal{R}_u.$$

From Monotone Continuity and Normalization, we have $c^*(\pi) = \infty$ for all $\pi \in \Pi_0 \setminus \Delta(\Theta)$ (see Maccheroni et al. (2006, Lemma 30)). Letting $\Pi = \{\pi \in \Pi_0 : c^*(\pi) < \infty\}$ and defining $c : \Pi \to [0, \infty)$ by the restriction of c^* to Π, we obtain the result. $\square$

Proof of Proposition 6: Since $\succsim$ has a variational representation and satisfies Axiom 15, then $\succeq^u$ satisfies the basic A-A axioms, Uncertainty Aversion, Weak Certainty Independence and Savage's STP. Strzalecki (2011, Theorem 1) implies that the restriction of $\succeq^u$ to $\mathcal{R}_u^0$ satisfies the aforementioned axioms if, and only if, it has a multiplier representation (u, I_0), where $I_0(r) = \min_{\pi \in \Delta(\Theta)} \left\{ \int_\Theta r\mathrm{d}\pi + kD(\pi \| \mu) \right\}$. From the SEOU representation of I_0 found in Strzalecki (2011), i.e., $I_0(r) = \int_\Theta -\exp(-k^{-1}r)\mathrm{d}\mu$ for $k < \infty$ and $\int_\Theta r\mathrm{d}\mu$ for $k = \infty$, it is easily checked that I_0 is continuous in the topology of point-wise convergence. We can thus extend I_0 to $I : \mathcal{R}_u \to \mathbb{R}$ using the same technique as in the proof of Proposition 5. $\square$

A.6 Proofs of results in Section 6

A.6.1 Proof of Theorem 3

If $\succsim \subseteq \mathcal{S}^2$ satisfies axioms 1 to 6, then Lemma 5 implies that the relation $\succeq$ defined by (8) is a weak order that satisfies axioms A3 to A6. Then it is a straightforward exercise to check that the state dependent stochastic choice function $c : \Phi \to \mathcal{F} \setminus \{\emptyset\}$ given by $c(F) = \{f \in F : f \succeq g \ \forall g \in F\}$ for all $F \in \Phi$, satisfies axioms C1 to C4.

Now assume $c : \Phi \to \mathcal{F} \setminus \{\emptyset\}$ satisfies axioms C1 to C4. We want to construct a complete extension $\succeq$ of $\trianglerighteq$ that satisfies axioms A3 to A6, which will allow us to define a preference $\succsim$ on $\mathcal{S}$ that rationalizes c and satisfies axioms 1 to 6.

Because $\trianglerighteq_\theta$ satisfies C-LARP and Conditional Continuity, we can apply Theorem 4 from Clark (1993) and Theorem 3 from Clark (2000) to obtain a continuous total order (a complete, transitive and antisymmetric preference) $\succeq_\theta$ on $\Delta(A)$ that extends $\trianglerighteq_\theta$ and satisfies Independence. That is, if $p \succeq_\theta q$, then $\lambda p + (1-\lambda)r \succeq_\theta \lambda q + (1-\lambda)r$ for all $\lambda \in [0,1]$ and $r \in \Delta(A)$. By RP-Monotonicity, $f = c(F)$ implies that for all $g \in F$ with $g \neq f$, there exists $\theta \in \Theta$ such that $p \not\trianglerighteq_\theta f(\theta)$ for all $p \in \Delta(A)$ such that $g(\theta) \trianglerighteq_\theta p$. Therefore, a well-known argument (see, e.g., Ok and Riella (2021), Theorem 9.3) guarantees that $\{\succeq_\theta : \theta \in \Theta\}$ can be constructed so that, if $f = c(F)$, $g \in F$ and $g \neq f$, there is a $\theta \in \Theta$ such that $f(\theta) \succeq_\theta g(\theta)$.

Define, for all $f, g \in \mathcal{F}$, $f \mathrel{\hat{\succeq}} g$ if, and only if, $f(\theta) \succeq_\theta g(\theta)$ for every $\theta \in \Theta$. It is easily checked that, by construction, $\hat{\succeq}$ is a partial order that satisfies axioms A3 to A5, A7 and A8. Suppose Θ is countable, so that $\mathcal{F}$ is metrizable. If $(f^m), (g^m) \in \mathcal{F}$ are such that $f^m \to f$, $g^m \to g$ and $f^m \mathrel{\hat{\succeq}} g^m$ for all $m \in \mathbb{N}$, then $f^m(\theta) \to f(\theta)$ and $g^m(\theta) \to g(\theta)$ for all $\theta \in \Theta$, which by definition of $\hat{\succeq}$ and the fact that $\succeq_\theta$ is closed for all $\theta \in \Theta$, implies that $f \mathrel{\hat{\succeq}} g$. Therefore, $\hat{\succeq}$ satisfies axiom A6.

From GARP and the fact that we are working with choice correspondences, if $f^1 \in c(F_1) \cap F_2, f^2 \in c(F_2) \cap F_3, \ldots, f^k \in c(F_k) \cap F_1$, then $f^1 = \cdots = f^k = f$. Moreover, by construction, $g(\theta) \mathrel{\hat{\not\succeq}} f(\theta)$ for any $g \in \bigcup_{i=1}^k F_k \setminus \{f\}$. In the language of Nishimura et al. (2017), $((\mathcal{F}, \hat{\succeq}), \Phi)$ is a continuous choice environment and c satisfies cyclical $\hat{\succeq}$-consistency. Therefore, we can apply their Theorem 1 to obtain a weak order $\succeq$ that extends $\hat{\succeq}$ and, for all $F \in \Phi$,

$$c(F) = \{f \in F : f \succeq g \ \forall g \in F\}.$$

Furthermore, since $\succeq$ extends $\hat{\succeq}$, it is easy to see that it satisfies axioms A3 to A5.

Finally, define $\succsim$ on $\mathcal{S}$ by

$$\forall (\rho, P), (\tau, Q) \in \mathcal{S}, \ (\rho, P) \succsim (\tau, Q) \iff \rho P \succeq \tau Q.$$

Obviously, $\succsim$ satisfies Consequentialism. It is easy to check, using arguments similar to the proof of Lemma 5, that $\succsim$ also satisfies axioms 1 and 3 to 6. Therefore, by Theorem 2, it has an MRA representation (u, I), which completes the proof.

A.6.2 Proof of Theorem 4

Suppose $\succcurlyeq_.$ satisfies Axiom D1. Note that item 1 of Axiom D1 implies that each $\succcurlyeq_P$ satisfies a version of Consequentialism. Indeed, taking $P = P'$ in the statement, we have that $\rho P = \tau P$ implies $\rho \sim_P \tau$.

Consider the following relation on $\mathcal{D}$:

$$\forall \rho, \tau \in \mathcal{D} : \rho \mathrel{\hat{\succcurlyeq}} \tau \iff \rho \succcurlyeq_\theta \tau \text{ for all } \theta \in \Theta.$$

Since $\succcurlyeq_{P^*}$ is reflexive, $\hat{\succcurlyeq}$ is non-empty. Also define a preference relation $\hat{\succeq}$ on $\mathcal{F}$ by

$$f \mathrel{\hat{\succeq}} g \iff \exists \rho, \tau \in \mathcal{D} \text{ such that } \rho \mathrel{\hat{\succcurlyeq}} \tau,\ \rho P^* = f \text{ and } \tau P^* = g.$$

Substituting $\hat{\succeq}$ for $\succeq$ in the proof of Lemma 5, it can be seen that $\hat{\succeq}$ is reflexive, transitive, and satisfies axioms A3 to A8. Therefore, substituting $\hat{\succeq}$ for $\succeq$ in the Proof of Theorem 1, we obtain a dominance representation with utility index $v : A \times \Theta \to \mathbb{R}$, i.e., for all $\rho, \tau \in \mathcal{D}$,

$$\rho \mathrel{\hat{\succcurlyeq}} \tau \iff \int_X v(\rho, \theta) \mathrm{d}P^*_\theta \geq \int_X v(\tau, \theta) \mathrm{d}P^*_\theta \quad \forall \theta \in \Theta.$$

Note that, due to Lemma 1, we have $\mathcal{R}_v = \{r_v(\delta, P^*) : \delta \in \mathcal{D}\}$. Define a preference $\succeq_*$ on $\mathcal{R}_v$ by

$$\forall r, r' \in \mathcal{R}_v :\ r \succeq_* r' \iff \exists \rho \succcurlyeq_{P^*} \tau \text{ such that } r = r_v(\rho, P^*) \text{ and } r' = r_v(\tau, P^*).$$

An argument entirely analogous to the Proof of Theorem 2 guarantees that $\succeq_*$ is a continuous and monotone weak order. Therefore, there exists a monotone and continuous utility function $J : \mathcal{R}_v \to \mathbb{R}$ such that, for all $\rho, \tau \in \mathcal{D}$,

$$r_v(\rho, P^*) \succeq_* r_v(\tau, P^*) \iff J(r_v(\rho, P^*)) \geq J(r_v(\tau, P^*)) \iff \rho \succcurlyeq_{P^*} \tau.$$

We now extend the representation to the collection $\succcurlyeq_{\boldsymbol{\cdot}} = \{\succcurlyeq_P : P \in \mathcal{P}\}$. Take any $P \in \mathcal{P}$ and $\rho, \tau \in \mathcal{D}$. By Lemma 1, there exists $\rho', \tau' \in \mathcal{D}$ such that $\rho P = \rho' P^*$ and $\tau P = \tau' P^*$. From item 1 of Axiom D1, we obtain $\rho \succcurlyeq_P \tau \iff \rho' \succcurlyeq_{P^*} \tau'$. Therefore, $\rho \succcurlyeq_P \tau \iff J(r_v(\rho', P^*)) \geq J(r_v(\tau', P^*))$. Since $r_v(\rho', P^*) = r_v(\rho, P)$ and $r_v(\tau', P^*) = r_v(\tau, P)$, we have

$$\rho \succcurlyeq_P \tau \iff J(r_v(\rho, P)) \geq J(r_v(\tau, P)),$$

for all $\rho, \tau \in \mathcal{D}$ and $P \in \mathcal{P}$.

A.6.3 Proof of Theorem 5

Assume $\succcurlyeq^{\boldsymbol{\cdot}}$ satisfies axioms E1 to E3. By Lemma 4, there exists $\delta^* \in \overline{\mathcal{D}}$ such that for any $f \in \mathcal{F}$, there is a $P^f \in \Delta(X)^\Theta$ such that $f = \delta^* P^f$.

For any given $f, h \in \mathcal{F}$, take $P^f, P^h \in \mathcal{P}$ such that $f = \delta^* P^f$ and $g = \delta^* P^h$. Since δ^* is invariant, we have:

1. For all $\alpha \in [0, 1]$, $\delta^*(\alpha P^f + (1 - \alpha) P^h) = \alpha \delta^* P^f + (1 - \alpha) \delta^* P^h = \alpha f + (1 - \alpha) h$;
2. For all $\theta \in \Theta$, $\delta^*(P^f{}_{(\theta)} P^h) = \delta^* P^f{}_{(\theta)} \delta^* P^h = f_{(\theta)} h$.

Define the preference $\succeq^*$ on $\mathcal{F}$ by

$$f \succeq^* h \iff \exists P^f, P^h \in \mathcal{P} \text{ such that } P^f \succcurlyeq^{\{\delta^*\}} P^h,\ \delta^* P^f = f \text{ and } \delta^* P^h = h.$$

Consistency guarantees that $\succeq^*$ is well-defined and reflexive, since by taking $M = N = \{\delta^*\}$, $Q = P'$ and $P = Q'$ in the statement of the axiom, we get that $\delta^* P = \delta^* P'$ implies $P \sim^{\{\delta^*\}} P'$. Item 1 of Axiom E1 and Lemma 4 immediately imply that $\succeq^*$ is complete and transitive, whereas item 5 readily implies that $\succeq^*$ satisfies Axiom A6. We now show that $\succeq^*$ satisfies axioms A3 to A6.

Indeed, take any $f, g, h, h' \in \mathcal{F}$ and $\theta \in \Theta$ such that $f_{(\theta)} h \succeq^* g_{(\theta)} h$. Then, $P^f{}_{(\theta)} P^h \succcurlyeq^{\{\delta^*\}} P^g{}_{(\theta)} P^h$ for some $\delta^* P^f = f$, $\delta^* P^h = h$ and $\delta^* P^g = g$. Item 2 of Axiom E1 then implies that $P^f{}_{(\theta)} R \succcurlyeq^{\{\delta^*\}} P^g{}_{(\theta)} R$ for all $R \in \mathcal{P}$, and thus $f_{(\theta)} h' \succeq^* g_{(\theta)} h'$, by Lemma 4. Thus $\succeq^*$ satisfies Axiom A3.

Now fix any $f, g, h \in \mathcal{F}$ such that $f_{(\theta)}h \succeq^* g_{(\theta)}h$ for all $\theta \in \Theta$. This implies that there exist $P^f, P^g, P^h \in \mathcal{P}$, where $\delta^* P^f = f$, $\delta^* P^g = g$ and $\delta^* P^h = h$, such that $P^f{}_{(\theta)}P^h \succcurlyeq^{\{\delta^*\}} P^g{}_{(\theta)}P^h$ for all $\theta \in \Theta$. By item 2 in Axiom E1, we have that $P^f{}_{(\theta)}R \succcurlyeq^{\{\delta^*\}} P^g{}_{(\theta)}R$ for all $\theta \in \Theta$ and $R \in \mathcal{P}$. From item 3, we obtain $P^f \succcurlyeq^{\{\delta^*\}} P^g$, which implies $f \succeq^* g$. Therefore, $\succeq^*$ satisfies Axiom A4.

Next fix any $\theta \in \Theta$ and take any $f, g, h \in \mathcal{F}$ such that $f(\theta') = g(\theta') = h(\theta')$ for all $\theta' \neq \theta$. Then there exist $P^f, P^g, P^h \in \mathcal{P}$ such that $P^f_{\theta'} = P^g_{\theta'} = P^h_{\theta'}$ for all $\theta' \neq \theta$, $f = \delta^* P^f$, $g = \delta^* P^g$ and $h = \delta^* P^h$. Suppose $f \succeq^* g$, implying $P^f \succcurlyeq^{\{\delta^*\}} P^g$. By item 4 of Axiom E1, we have $\alpha P^f + (1-\alpha)P^h \succcurlyeq^{\{\delta^*\}} \alpha P^g + (1-\alpha)P^h$ for all $\alpha \in (0,1]$, and thus $\alpha f + (1-\alpha)h \succeq^* \alpha g + (1-\alpha)h$. We conclude that $\succeq^*$ satisfies Axiom A5.

For each $\theta \in \Theta$, define the conditional preferences $\succeq_\theta$ as in (14). Since $\succeq^*$ is complete, so are $\succeq_\theta$. Appendix A.2 then shows that each $\succeq_\theta$ has a representation given by (15). Fix one such representation $w : A \times \Theta \to \mathbb{R}$ and define its associated risk functional $\tilde{r}_w : \mathcal{F} \to \mathbb{R}^\Theta$ as in eq. (17). Consider the set of all such risk functions $\mathcal{R}_w = \{\tilde{r}_w(f) : f \in \mathcal{F}\}$ and define a preference relation $\succeq^w$ on $\mathcal{R}_w$ by $r \succeq^w r' \iff \exists f, g \in \mathcal{F}$ such that $f \succeq^* g$, $r = \tilde{r}_w(f)$ and $r' = \tilde{r}_w(g)$. Note that if $f = \delta^* P$, for any $f \in \mathcal{F}$ and $P \in \mathcal{P}$, then $\tilde{r}_w(f) = r_w(\delta^*, P)$. A straightforward adaptation of the argument in Appendix A.4 then implies that there exist $w : A \times \Theta \to \mathbb{R}$, continuous in the first argument, and a monotone and continuous $H : \mathcal{R}_w \to \mathbb{R}$, such that

$$P \succcurlyeq^{\{\delta^*\}} Q \iff H(r_w(\delta^*, P)) \geq H(r_w(\delta^*, Q)).$$

Moreover, we clearly have $C_P(\{\delta^*\}) = \{\delta^*\}$ for all $P \in \mathcal{P}$, hence $\succcurlyeq^{\{\delta^*\}}$ has the desired representation.

It remains to show that $\succcurlyeq^M$ has such a representation for each $M \in \mathcal{M}$, and that these representations coincide. To that end, first consider any $M \in \mathcal{M}$ such that $C_P(M)$ is a singleton for all $P \in \mathcal{P}$. Take any $P, Q \in \mathcal{P}$ and let $\{\rho\} = C_P(M)$ and $\{\tau\} = C_Q(M)$. By Lemma 4, there exist $P', Q' \in \mathcal{P}$ such that $\rho P = \delta^* P'$ and $\tau Q = \delta^* Q'$. Consistency then implies that $P \succcurlyeq^M Q \iff P' \succcurlyeq^{\{\delta^*\}} Q'$, and thus

$$P \succcurlyeq^M Q \iff H(r_w(\delta^*, P')) \geq H(r_w(\delta^*, Q')) \iff H(r_w(\rho, P)) \geq H(r_w(\tau, Q)).$$

We now extend the representation to menus M with $|C_P(M)| > 1$ for some $P \in \mathcal{P}$. Take any $M \in \mathcal{M}$ and $P \in \mathcal{P}$. By Lemma 4, for all $\rho \in C_P(M)$ there exists $P^\rho \in \mathcal{P}$ such that $\rho P = \delta^* P^\rho$. Since $\succcurlyeq^{\{\delta^*\}}$ is complete and transitive, there exists $\rho^*(P) \in C_P(M)$ such that $P^{\rho^*(P)} \succcurlyeq^{\{\delta^*\}} P^\rho$ for all $\rho \in C_P(M)$. Define a decision rule $\bar{\delta}$ by $\bar{\delta}_P = \rho^*(P)_P$ for every $P \in \mathcal{P}$ and note that $\bar{\delta}P = \rho^*(P)P$. Thus by Axiom E2, $P \succcurlyeq^{\{\bar{\delta}\}} Q \iff P \succcurlyeq^{\tilde{M}(P,Q)} Q$ for all $P, Q \in \mathcal{P}$, where $\tilde{M}(P,Q) = \{\delta \in \mathcal{D} : \delta_P = \rho^*(P)_P,\ \delta_Q = \rho^*(Q)_Q$ and $\forall R \neq P,\ \exists \gamma \in M$ s.t. $\delta_R = \gamma_R\}$. Applying Axiom E3 twice, we obtain the following equivalences:

$$P \succcurlyeq^{\{\bar{\delta}\}} Q \iff P \succcurlyeq^{\tilde{M}(P,Q)} Q \iff P \succcurlyeq^M Q.$$

Finally, note that since $r_w(\rho^*(P), P) = r_w(\bar{\delta}, P)$ for all $P \in \mathcal{P}$ and $\succcurlyeq^{\{\bar{\delta}\}}$ can be represented by (w, H) we get that for all $P, Q \in \mathcal{P}$, $P \succcurlyeq^M Q \iff H(r_w(\rho^*(P), P)) \geq H(r_w(\rho^*(Q), Q))$ subject to $\rho^*(P) \in C_P(M)$, $\rho^*(Q) \in C_Q(M)$.

References

Adams, C. P. (2016). Finite mixture models with one exclusion restriction. *Econometrics Journal 19*(2), 150–165.

Al-Najjar, N. I. and L. De Castro (2014). Parametric representation of preferences. *Journal of Economic Theory 150*(1), 642–667.

Amarante, M. (2009). Foundations of neo-Bayesian statistics. *Journal of Economic Theory 144*(5), 2146–2173.

Anscombe, F. J. and R. J. Aumann (1963). A Definition of Subjective Probability. *The Annals of Mathematical Statistics 34*(1), 199–205.

Arrow, K. J. (1971). *Essays in the theory of risk-bearing*. Markham Pub. Co.

Blackwell, D. (1951). Comparison of Experiments. *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability*, 93–102.

Blackwell, D. (1953, 6). Equivalent Comparisons of Experiments. *Annals of Mathematical Statistics 24*(2), 265–272.

Caplin, A. and M. Dean (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review 105*(7), 2183–2203.

Caplin, A. and D. Martin (2015). A testable theory of imperfect perception. *Economic Journal 125*(582), 184–202.

Cerreia-Vioglio, S., P. Ghirardato, F. Maccheroni, M. Marinacci, and M. Siniscalchi (2011). Rational preferences under ambiguity. *Economic Theory 48*(2), 341–375.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011). Uncertainty averse preferences. *Journal of Economic Theory 146*(4), 1275–1330.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2013). Ambiguity and robust statistics. *Journal of Economic Theory 148*(3), 974–1049.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and A. Rustichini (2014). Niveloids and their extensions: Risk measures on small domains. *Journal of Mathematical Analysis and Applications 413*(1), 343–360.

Cerreia-Vioglio, S., L. P. Hansen, F. Maccheroni, and M. Marinacci (2020). Making Decisions under Model Misspecification. *SSRN Electronic Journal*.

Chateauneuf, A., F. Maccheroni, M. Marinacci, and J. M. Tallon (2005). Monotone continuous multiple priors. *Economic Theory 26*(4), 973–982.

Clark, S. A. (1993). Revealed preference and linear utility. *Theory and Decision 34*(1), 21–45.

Clark, S. A. (2000). Revealed preference and expected utility. *Theory and Decision 49*(2), 159–174.

Crémer, J. (1982). A simple proof of Blackwell's "comparison of experiments" theorem. *Journal of Economic Theory 27*(2), 439–443.

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review 67*(3), 297–308.

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics 75*(4), 643–669.

Epstein, L. G. and K. Seo (2010). Symmetry of evidence without evidence of symmetry. *Theoretical Economics 5*(3), 313–368.

Fishburn, P. C. (1970). *Utility Theory for Decision Making* (1st ed.). New York: Research Analysis Corporation.

Furtado, B. (2022). Identifiability of Finite Mixture Models via Nonnegative Matrix Factorization.

Gentzkow, M. and E. Kamenica (2011). Bayesian Persuasion. *American Economic Review 101*(6), 2590–2615.

Ghirardato, P. and M. Marinacci (2002). Ambiguity made precise: A comparative foundation. *Journal of Economic Theory 102*(2), 251–289.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*(2), 141–153.

Hansen, L. P. and T. J. Sargent (2001). Robust Control and Model Uncertainty. *American Economic Review 91*(2), 60–66.

Henry, M., Y. Kitamura, and B. Salanié (2014). Partial identification of finite mixtures in econometric models. *Quantitative Economics 5*(1), 123–144.

Herden, G. (1989). On the existence of utility functions. *Mathematical Social Sciences 17*(3), 297–313.

Hewitt, E. (1946, 8). A remark on density characters. *Bulletin of the American Mathematical Society 52*(8), 641–643.

Inoue, L. Y. T. (2009). *Decision theory: principles and approaches* (1st ed.), Volume 47. Chichester: John Wiley & Sons, Inc.

Jakobsen, A. M. (2021). An Axiomatic Model of Persuasion. *Econometrica 89*(5), 2081–2116.

Kreps, D. (2018). *Notes On The Theory Of Choice*. New York: Routledge.

Kuzmics, C. (2017). Abraham Wald's complete class theorem and Knightian uncertainty. *Games and Economic Behavior 104*, 666–673.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica 74*(6), 1447–1498.

Manski, C. F. (2021). Econometrics for Decision Making: Building Foundations Sketched by Haavelmo and Wald. *Econometrica 89*(6), 2827–2853.

Manski, C. F. and A. Tetenov (2014). The Quantile Performance of Statistical Treatment Rules using Hypothesis Tests to Allocate a Population to Two Treatments. *Cemmap Working Paper CWP44/14*(January).

Marschak, J. and K. Miyasawa (1968). Economic Comparability of Information Systems. *International Economic Review 9*(2), 137.

Matêjka, F. and A. McKay (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review 105*(1), 272–298.

Neilson, W. S. (2010). A simplified axiomatic approach to ambiguity aversion. *Journal of Risk and Uncertainty 41*(2), 113–124.

Nishimura, H., E. A. Ok, and J. K. Quah (2017). A comprehensive approach to revealed preference theory. *American Economic Review 107*(4), 1239–1263.

Ok, E. A. and G. Riella (2021). Fully Preorderable Groups. *Order 38*(1), 127–142.

Savage, L. J. (1954). *Foundations of Statistics* (1st ed.). New York: John Wiley & Sons, Inc.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Stoye, J. (2011). Statistical decisions under ambiguity. *Theory and Decision 70*(2), 129–148.

Stoye, J. (2012). New perspectives on statistical decisions under ambiguity. *Annual Review of Economics 4*, 257–282.

Strzalecki, T. (2011). Axiomatic Foundations of Multiplier Preferences. *Econometrica 79*(1), 47–73.

Wald, A. (1939). Contributions to the Theory of Statistical Estimation and Testing Hypotheses. *The Annals of Mathematical Statistics 10*(4), 299–326.

Wald, A. (1947a). An Essentially Complete Class of Admissible Decision Functions. *The Annals of Mathematical Statistics 18*(4), 549–555.

Wald, A. (1947b). Foundations of a General Theory of Sequential Decision Functions. *Econometrica 15*(4), 279.

Wald, A. (1949). Statistical Decision Functions. *The Annals of Mathematical Statistics 20*(2), 165–205.