Making Decisions under Model Misspecification*

Simone Cerreia–Vioglio[a], Lars Peter Hansen[b], Fabio Maccheroni[a]
and Massimo Marinacci[a]

[a]Università Bocconi and Igier, [b]University of Chicago

June 2022

Abstract

We use decision theory to confront uncertainty that is sufficiently broad to incorporate "models as approximations." We presume the existence of a featured collection of what we call "structured models" that have explicit substantive motivations. The decision maker confronts uncertainty through the lens of these models, but also views these models as simplifications, and hence, as misspecified. We extend the max-min analysis under model ambiguity to incorporate the uncertainty induced by acknowledging that the models used in decision-making are simplified approximations. Formally, we provide an axiomatic rationale for a decision criterion that incorporates model misspecification concerns.

JEL codes— C54, D81

**First Draft: December 2019.* We thank Attila Ambrus, Pierpaolo Battigalli, Benjamin Brooks, Tim Christensen, Roberto Corrao, Giacomo Lanzani, Marco Loseto, Philipp Sadowski, Todd Sarver, Jesse Shapiro as well as the audiences at Advances in Decision Analysis 2019, SAET 2019, Blue Collar Working Group 2.0, One World Mathematical Game Theory Seminar, RUD 2020, UCL-Osaka International Conference on the Mathematics for Risk and Decisions, MUSEES 2022, Confronting Uncertainty in Climate Change, Bicocca, Bilkent, Caltech, Duke, Glasgow, John Hopkins, LSE, Harvard, MIT, Parma, UAB, Warwick for their very useful comments. We thank for the financial support the Alfred P. Sloan Foundation (grant G-2018-11113), the European Research Council (grants SDDM-TEA and INDIMACRO) and the Ministero dell'università e della ricerca (grant 2017CY2NCA).

Come l'araba fenice:
che vi sia, ciascun lo dice;
dove sia, nessun lo sa.[1]

1 Introduction

The consequences of a decision may depend on exogenous contingencies and uncertain outcomes that are outside the control of a decision maker. This uncertainty takes on many forms. Economic applications typically feature *risk*, where the decision maker knows the correct probabilistic model governing the contingencies but not necessarily the decision outcomes. Yet, this is a demanding assumption. As a result, statisticians and econometricians have long wrestled with how to confront *ambiguity* over models or unknown parameters within a model. Each model is itself a simplification or an approximation designed to guide or enhance our understanding of some underlying phenomenon of interest. Thus, the model, by its very nature, is *misspecified*, but in typically uncertain ways. How should a decision maker acknowledge model misspecification in a way that guides the use of purposefully simplified models sensibly? This concern has certainly been on the radar screen of statisticians and control theorists, but it has been largely absent in formal approaches to decision theory.[2] Indeed, the statisticians Box and Cox have both stated the challenge succinctly in complementary ways:

> Since all models are wrong, the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad. Box (1976).
>
> ... it does not seem helpful just to say that all models are wrong. The very word "model" implies simplification and idealization. The idea that complex physical, biological or sociological systems can be exactly described by a few formulae is patently absurd. The construction of idealized representations that capture important stable aspects of such systems is, however, a vital part of general scientific analysis and statistical models, especially substantive ones ... Cox (1995).

While there are formulations of decision and control problems that intend to confront model misspecification, the aim of this paper is: (i) to develop an axiomatic approach that will provide a rigorous guide for applications and (ii) to enrich formal decision theory when applied to environments with uncertainty through the guise of models.

[1] "Like the Arabian phoenix: that it exists, everyone says; where it is, nobody knows." A passage from a libretto of Pietro Metastasio.

[2] In Hansen (2014) and Hansen and Marinacci (2016) three kinds of uncertainty are distinguished based on the knowledge of the decision maker, the most challenging being model misspecification viewed as uncertainty induced by the approximate nature of the models under consideration.

The protagonist of our analysis is a decision maker who is able to formulate models – for instance a policy maker having to decide a climate policy based on existing alternative climate models – but is concerned about their misspecification and wants to use a decision criterion which accounts for that. Our axiomatic analysis, which has a normative nature, aims to derive a criterion of this kind to help the decision maker to cope with model misspecification in a principled way. In this endeavour, we follow Hansen and Sargent (2022) by referring to the formulated models as "structured models." These structured models are ones that are explicitly motivated or featured, such as ones with substantive motivation or scientific underpinnings, consistent with the use of the term "models" by Box and Cox. They may be based on scientific knowledge relying on empirical evidence and theoretical arguments or on revealing parameterizations of probability models with parameters that are interpretable to the decision maker. In posing decision problems formally, it is often assumed, following Wald (1950), that the correct model belongs to the set of models that decision makers posit. The presumption that a decision maker identifies, among their hypotheses, the correct model is often questionable – recalling the initial quotation, the correct model is often a decision maker phoenix. We embrace, rather than push aside, the "models are approximations" perspective of many applied researchers, as articulated by Box, Cox and others. To explore misspecification formally, we introduce a potentially rich collection of probability distributions that depict possible representations of the data without formal substantive motivation. We refer to these as "unstructured models." We use such alternative models as a way to capture how models could be misspecified.[3]

This distinction between structured and unstructured is central to the analysis in this paper and is used to distinguish aversion to ambiguity over models and aversion to potential model misspecification. At a decision-theoretic level, a proper analysis of misspecification concerns has remained elusive so far. Indeed, many studies dealing with economic agents confronting model misspecification still assume that they are conventional expected utility decision makers who do not address formally potential model misspecification concerns in their preference ordering.[4] We extend the analysis of Hansen and Sargent (2022) by providing an axiomatic underpinning for a corresponding decision theory along with a representation of the implied preferences that can guide applications. In so doing, we show an important connection with the analysis of subjective and objective rationality of Gilboa et al. (2010).

Criterion This paper proposes a first decision-theoretic analysis of decision making under model misspecification. We consider a classic setup in the spirit of Wald (1950), but relative to his seminal work we explicitly remove the assumption that the correct model belongs to the set of posited models and we allow for nonneutrality toward this feature. More formally, in our purely normative approach we assume that decision makers posit a set Q of *structured*

[3]Such a distinction is also present in earlier work by Hansen and Sargent (2007) and Hansen and Miao (2018) but without specific reference to the terms "structured" and "unstructured."

[4]See, e.g., Esponda and Pouzo (2016) and Fudenberg et al. (2017).

(probabilistic) *models* q on states, motivated by their information, but they are afraid that none of them is correct and so face model misspecification. For this reason, decision makers contemplate what we call *unstructured models* in ranking acts f, according to a conservative decision criterion[5]

$$V(f) = \min_{p \in \Delta} \left\{ \int u(f)\, dp + \min_{q \in Q} c(p, q) \right\} \tag{1}$$

To interpret this criterion, let

$$C(p, Q) = \min_{q \in Q} c(p, q)$$

where we presume that $C(q, Q) = 0$ when $q \in Q$. In this construction, $C(p, Q)$ is what we call a Hausdorff statistical set distance between a model p and the posited set Q of structured models. This distance is nonzero if and only if p is unstructured, that is, $p \notin Q$. More generally, p's that are closer to the set of structured models Q have a less adverse impact on the preferences, as is evident by rewriting (1) as:

$$V(f) = \min_{p \in \Delta} \left\{ \int u(f)\, dp + C(p, Q) \right\}$$

This representation is a special case of the variational representation axiomatized by Maccheroni et al. (2006). The unstructured models are statistical artifacts that allow the decision maker to assess formally the potential consequences of misspecification as captured by the construction of $C(\cdot, Q)$. In this paper we provide a formal interpretation of $C(\cdot, Q)$ as an index of misspecification fear: the lower the index, the higher the fear.[6]

It is because of the ability to posit a set Q that the decision maker confronts uncertainty in the guise of models, so what we may call a decision problem under model uncertainty. In our normative approach, it is natural to enrich the standard decision-theoretic setting by taking Q as a given, a *datum* of the decision problem. For instance, in the climate policy problem, Q is the set of climate models that the policy maker considers. In this regard, observe that we are not after detecting which choice behavior of the decision maker may reveal model misspecification concerns, a different revealed preference exercise that would indeed require an endogenous Q.[7] In line with standard practice in applied economics, we imagine that the substantive modeling that underlies the construction of elements of Q is simplified with an explicit structure imposed to facilitate interpretation. Applied researchers commonly avoid reducing model building to the construction of the complex black boxes that a purely nonparametric exercise might well involve, especially in multivariate settings.

[5] Throughout the paper Δ denotes the set of all probabilities (Section 2.1).
[6] To ease terminology, we often refer to "misspecification" rather than "model misspecification."
[7] In this exercise, the findings of Denti and Pomatto (2022) may be useful.

A protective belt When c takes the entropic form $\lambda R(p||q)$, with $\lambda > 0$, criterion (1) takes the form

$$\min_{p\in\Delta}\left\{\int u\left(f\right)dp+\lambda\min_{q\in Q}R(p||q)\right\} \tag{2}$$

proposed by Hansen and Sargent (2022). It is the most tractable version of criterion (1), which for a singleton Q further reduces to a standard multiplier criterion a la Hansen and Sargent (2001, 2008). By exchanging orders of minimization, we preserve this tractability and provide a revealing link to this earlier research,

$$\min_{q\in Q}\left\{\min_{p\in\Delta}\left\{\int u\left(f\right)dp+\lambda R(p||q)\right\}\right\} \tag{3}$$

The inner minimization problem gives rise to the minimization problem featured by Hansen and Sargent (2001, 2008) to confront the potential misspecification of a given probability model q.[8] Unstructured models lack the substantive motivation of structured models, yet in (1) they act as a protective belt against model misspecification. The importance of their role is proportional – as quantified by λ in (2) – to their proximity to the set Q, a measure of their plausibility in view of the decision maker information. The outer minimization over structured models is the counterpart to the Wald (1950) and the more general Gilboa and Schmeidler (1989) max-min criterion.

Our analysis provides a decision-theoretic underpinning for incorporating misspecification concerns in a distinct way from ambiguity aversion. Observe that misspecification fear is absent when the index $\min_{q\in Q} c\left(p,q\right)$ equals the indicator function δ_Q of the set of structured models Q, that is,

$$\min_{q\in Q} c\left(p,q\right)=\begin{cases} 0 & \text{if } p\in Q \\ +\infty & \text{else}\end{cases}$$

In this case, which corresponds to $\lambda = +\infty$ in (2), criterion (1) takes a max-min form:

$$V\left(f\right)=\min_{q\in Q}\int u\left(f\right)dq$$

This max-min criterion thus characterizes decision makers who confront model misspecification but are not concerned by it, so are misspecification neutral (see Section 4.1). The criterion in (1) may thus be viewed as representing decision makers who use a more prudential variational criterion (1) than if they were to max-minimize over the set of structured models which they posited. In particular, the farther away an unstructured model is from the set Q (so the less plausible it is), the less it is weighted in the minimization.

[8]The Hansen and Sargent (2001, 2008) formulation of preferences builds on extensive literature in control theory starting with Jacobson (1973)'s deterministic robustness criterion and a stochastic extension given by Petersen et al. (2000), among several others.

Axiomatics We use the entropic case (2) to outline our axiomatic approach. Start with a singleton $Q = \{q\}$. Decision makers, being afraid that the reference model q might not be correct, contemplate also unstructured models $p \in \Delta$ and rank acts f according to the multiplier criterion

$$V_{\lambda,q}(f) = \min_{p\in\Delta} \left\{ \int u(f)\, dp + \lambda R(p||q) \right\} \tag{4}$$

Here the positive scalar λ is interpreted as an index of misspecification fear. When decision makers posit a nonsingleton set Q of structured models, but are concerned that none of them is correct, the multiplier criterion (4) then gives only an incomplete *dominance relation*:

$$f \succsim^* g \iff V_{\lambda,q}(f) \geq V_{\lambda,q}(g) \qquad \forall q \in Q \tag{5}$$

With (5), decision makers can safely regard f better than g. Through this ranking, the dominance relation provides a preferential account of the probabilistic information that Q represents. The dominance relation thus naturally arises when the set Q is posited.

Yet, the ranking (5) has little traction because of the incomplete nature of $\succsim^*$. Nonetheless, the burden of choice will have decision makers select between alternatives, be they rankable by $\succsim^*$ or not. A cautious way to complete the binary relation $\succsim^*$ is given by the preference $\succsim$ represented by (2), or equivalently by (3). This criterion thus emerges in our analysis as a cautious completion of a multiplier dominance relation $\succsim^*$. In this way, the probabilistic information gets embedded in the behavioral preference. Suitably extended to a general preference pair $(\succsim^*, \succsim)$, we support this approach by axiomatizing criterion (1) as the representation of the behavioral preference $\succsim$ and the unanimity criterion

$$f \succsim^* g \iff \min_{p\in\Delta} \left\{ \int u(f)\, dp + c(p,q) \right\} \geq \min_{p\in\Delta} \left\{ \int u(g)\, dp + c(p,q) \right\} \qquad \forall q \in Q$$

as the representation of the incomplete dominance relation $\succsim^*$.

To sum up, our two-preference approach is motivated by the natural way with which the dominance relation arises when the set Q is posited. In this approach, we then connect the dominance and behavioral preferences to derive their desired representations.

2 Preliminaries

2.1 Mathematics

Basic notions We consider a non-trivial *event σ-algebra* Σ in a *state space* S. We denote by Δ the set of finitely additive probabilities and endow Δ and any of its subsets with the weak* topology (see Appendix B for further details). In particular, Δ^σ denotes the subset of Δ formed by the countably additive probability measures. Given a subset Q in Δ, we denote by $\Delta^{\ll}(Q)$ the collection of all probabilities p which are absolutely continuous with respect to Q, that is, if $A \in \Sigma$ and $q(A) = 0$ for all $q \in Q$, then $p(A) = 0$. Moreover, $\Delta^{\sigma}(q)$ denotes the set of elements of Δ^{σ} which are absolutely continuous with respect to a single $q \in \Delta^{\sigma}$, i.e., $\Delta^{\sigma}(q) = \{p \in \Delta^{\sigma} : p \ll q\}$. Unless otherwise specified, all the subsets of Δ are to be intended non-empty.

The (convex analysis) indicator function $\delta_C : \Delta \to [0, \infty]$ of a convex subset C of Δ is defined by

$$\delta_C(p) = \begin{cases} 0 & \text{if } p \in C \\ +\infty & \text{else} \end{cases}$$

Throughout we adopt the convention $0 \cdot \pm\infty = 0$.

The *effective domain* of $f : C \to (-\infty, \infty]$, denoted by $\operatorname{dom} f$, is the set $\{p \in C : f(p) < \infty\}$ where f takes on a finite value. The function f is *grounded* if the infimum of its image is 0, i.e., $\inf_{p \in C} f(p) = 0$.

Statistical distances Consider a given collection $\mathcal{Q}$ of compact subsets Q of Δ^{σ} that contains, as singletons $\{q\}$, all the elements q of the sets Q. Denote by $\mathcal{S}$ the set of all these singletons.[9] For instance, we have $\mathcal{S} = \Delta^{\sigma}$ when $\mathcal{Q}$ covers the space Δ^{σ}.

We say that a function $C : \Delta \times \mathcal{Q} \to [0, \infty]$ is a *statistical set distance* if:

(C.i) for each $Q \in \mathcal{Q}$,

$$C(p, Q) = 0 \Longleftrightarrow p \in Q$$

(C.ii) for each $Q, Q' \in \mathcal{Q}$,

$$Q \supseteq Q' \Longrightarrow C(\cdot, Q) \leq C(\cdot, Q')$$

(C.iii) $C(\cdot, \{q\})$ is lower semicontinuous for all $q \in \mathcal{S}$.

The first two properties make possible to interpret the quantity $C(p, Q)$ as a distance between the probability p and the set Q of probabilities. The last property is a basic regularity condition whose usefulness will become clear momentarily.

A statistical set distance C induces a function $c : \Delta \times \mathcal{S} \to [0, \infty]$ defined by

$$c(p, q) = C(p, \{q\})$$

The value $c(p, q)$ of this function is a distance between two probabilities p and q.[10] The induced function c has the following properties:

[9] That is, $\mathcal{S} = \bigcup_{Q \in \mathcal{Q}} Q$. Different collections $\mathcal{Q}$ may share the same set $\mathcal{S}$. The smallest such collection is $\{\{q\} : q \in \mathcal{S}\}$.

[10] This distance is not a metric as, for instance, symmetry is not required.

(c.i) for each $q \in \mathcal{S}$,

$$c(p,q) = 0 \Longleftrightarrow p = q$$

(c.ii) $c(\cdot,q) : \Delta \to [0,\infty]$ is lower semicontinuous for all $q \in \mathcal{S}$.

Through the function c we can characterize an important class of statistical set distances. Specifically, we say that a statistical set distance $C : \Delta \times \mathcal{Q} \to [0,\infty]$ is *Hausdorff* if:

(C.iv) $C(\cdot,Q) = \min_{q\in Q} c(\cdot,q)$ for all $Q \in \mathcal{Q}$.

This property defines a Hausdorff-type distance between p and Q, in which the distance between points and sets is subsumed by that between points. This important class of statistical set distances will be the protagonist of our analysis. In this class, there is a duality between c and C. Indeed, we say that a function $c : \Delta \times \mathcal{S} \to [0,\infty]$ is a *statistical distance* if it satisfies (c.i) and (c.ii), now taken as defining properties, and if it induces a well-defined Hausdorff statistical set distance $C : \Delta \times \mathcal{Q} \to [0,\infty]$ given by

$$C(\cdot,Q) = \min_{q\in Q} c(\cdot,q)$$

This final property is automatically satisfied when c is jointly lower semicontinuous, which is an important case in our analysis.

Statistical distances and Hausdorff statistical set distances are thus dual notions that can be defined one in terms of the other. It is sometimes convenient to denote by c_Q the section $C(\cdot,Q) : \Delta \to [0,\infty]$ at Q of the Hausdorff statistical set distance C induced by c, that is,

$$c_Q(\cdot) = \min_{q\in Q} c(\cdot,q)$$

An important special case is when $\mathcal{Q}$ consists of a fixed set Q along with its elements (as singletons), so that $\mathcal{S} = Q$. In this case, to ease notation we just write

$$c : \Delta \times Q \to [0,\infty]$$

This function is, for instance, the protagonist of Theorem 1. In this result we also consider a *pseudo-statistical distance*, which is a function $c : \Delta \times Q \to [0,\infty]$ that satisfies all the properties of a statistical distance except (c.i), which is weakened to: for each $q \in \mathcal{S}$ there is $p \in \Delta$ such that $c(p,q) = 0$.[11]

Divergences We say that a statistical distance $c : \Delta \times \mathcal{S} \to [0,\infty]$ is a *divergence* if:

[11] In other words, (C.i) is satisifed for sets which are not singletons while, when Q is a singleton, (C.i) is weakened to groundedness.

(c.iii) for each $q \in \mathcal{S}$,

$$c(p,q) < \infty \Longrightarrow p \ll q$$

Divergences thus assign an infinite penalty when p is not absolutely continuous with respect to q. In the important "universal" case $\mathcal{S} = \Delta^{\sigma}$, there is a well-known class of divergences. To introduce it, given a continuous strictly convex function $\phi : [0,\infty) \to [0,\infty)$, with $\phi(1) = 0$ and $\lim_{t\to\infty} \phi(t)/t = +\infty$, define $D_{\phi} : \Delta \times \Delta^{\sigma} \to [0,\infty]$ by

$$D_{\phi}(p||q) = \begin{cases} \int \phi\left(\frac{dp}{dq}\right) dq & \text{if } p \in \Delta^{\sigma}(q) \\ +\infty & \text{otherwise} \end{cases} \tag{6}$$

under the conventions $0/0 = 0$ and $\ln 0 = -\infty$.[12] It can be proved that $D_{\phi} : \Delta \times \Delta^{\sigma} \to [0,\infty]$ is a convex divergence, called *ϕ-divergence*.[13] The most important example of ϕ-divergence is the *relative entropy* given by $\phi(t) = t\ln t - t + 1$ and denoted by $R(p||q)$.[14] Another important example is the *Gini relative index* given by the quadratic function $\phi(t) = (t-1)^2/2$ and denoted by $\chi^2(p||q)$.

Given a coefficient $\lambda \in (0,\infty]$, the function $\lambda D_{\phi} : \Delta \times \Delta^{\sigma} \to [0,\infty]$ is also a convex divergence. In particular, when $\lambda = \infty$ we have

$$(\infty) D_{\phi}(p||q) = \delta_{\{q\}}(p) = \begin{cases} 0 & \text{if } p = q \\ \infty & \text{else} \end{cases}$$

because of the convention $0 \cdot \infty = 0$.

Variational statistical distances We say that a statistical set distance $C : \Delta \times \mathcal{Q} \to [0,\infty]$ is *variational* if:

(C.v) $C(\cdot, Q)$ is lower semicontinuous and convex for all $Q \in \mathcal{Q}$.

This is a regularity condition that, when assumed, strengthens property (C.iii). We say that a (pseudo-)statistical distance c is *variational* when it induces a variational Hausdorff statistical set distance. For instance, when $\mathcal{Q}$ consists of compact and convex subsets of Δ^{σ}, a statistical distance is variational if it is convex and lower semicontinuous (see Lemma 12). Thus, ϕ-divergences are variational with such a $\mathcal{Q}$.

2.2 Decision theory

Setup We consider a generalized Anscombe and Aumann (1963) setup where a decision maker chooses among uncertain alternatives described by (simple) acts $f : S \to X$, which are Σ-

[12] The function dp/dq is any version of the Radon-Nikodym derivative of p with respect to q.

[13] See Chapter 1 of Liese and Vajda (1987). We refer to this book for properties of ϕ-divergences.

[14] Given the conventions $0/0 = 0 \cdot \pm\infty = 0$, it holds $\phi(0) = 1$.

measurable simple (i.e., finite-valued) functions from a *state space* S to a *consequence space* X. This latter set is assumed to be a non-empty convex subset of a vector space (for instance, X is the set of all simple lotteries defined on a prize space). The triple

$$(S, \Sigma, X) \tag{7}$$

forms an (Anscombe-Aumann) *decision framework*.

Let us denote by $\mathcal{F}$ the set of all acts. Given any consequence $x \in X$, we denote by $x \in \mathcal{F}$ also the constant act that takes value x. Thus, with a standard abuse of notation, we identify X with the subset of constant acts in $\mathcal{F}$. Given a function $u : X \to \mathbb{R}$, we denote by $\operatorname{Im} u$ its image. Observe that $u \circ f$ is a simple real-valued Σ-measurable function.

A preference $\succsim$ is a binary relation on $\mathcal{F}$ that satisfies the so-called *basic conditions* (cf. Gilboa et al., 2010), i.e., it is:

(i) *reflexive* and *transitive*;

(ii) *monotone*: for all $f, g \in \mathcal{F}$, if $f(s) \succsim g(s)$ for all $s \in S$, then $f \succsim g$;

(iii) *continuous*: for all $f, g, h \in \mathcal{F}$, the sets

$$\{\alpha \in [0,1] : \alpha f + (1-\alpha) g \succsim h\} \quad \text{and} \quad \{\alpha \in [0,1] : h \succsim \alpha f + (1-\alpha) g\}$$

are closed;

(iv) *non-trivial*: there exist $f, g \in \mathcal{F}$ such that $f \succ g$.

Moreover, a preference $\succsim$ is *unbounded* if, for each $x, y \in X$ with $x \succ y$, there exist $z, z' \in X$ such that

$$\frac{1}{2}z + \frac{1}{2}y \succsim x \succ y \succsim \frac{1}{2}x + \frac{1}{2}z'$$

Bets are binary acts that play a key role in decision theory. Formally, given any two prizes $x \succ y$, a bet on an event A is the act xAy defined by

$$xAy(s) = \begin{cases} x & \text{if } s \in A \\ y & \text{else} \end{cases}$$

In words, a bet on event A is a binary act that yields a more preferred consequence if A obtains.

Comparative uncertainty aversion As in Ghirardato and Marinacci (2002), given two preferences $\succsim_1$ and $\succsim_2$ on $\mathcal{F}$, we say that $\succsim_1$ is *more uncertainty averse than* $\succsim_2$ if, for each consequence $x \in X$ and act $f \in \mathcal{F}$,

$$f \succsim_1 x \implies f \succsim_2 x$$

In words, a preference is more uncertainty averse than another one if, whenever this preference is "bold enough" to prefer an uncertain alternative over a sure one, so does the other one.

Decision criteria A complete preference $\succsim$ on $\mathcal{F}$ is *variational* if it is represented by a decision criterion $V : \mathcal{F} \to \mathbb{R}$ given by

$$V(f) = \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p) \right\} \tag{8}$$

where the affine utility function u is non-constant and the index of uncertainty aversion $c : \Delta \to [0, \infty]$ is grounded, lower semicontinuous and convex. In particular, given two unbounded variational preferences $\succsim_1$ and $\succsim_2$ on $\mathcal{F}$ that share the same u, but different indexes c_1 and c_2, we have that $\succsim_1$ is more uncertainty averse than $\succsim_2$ if and only if $c_1 \leq c_2$ (see Maccheroni et al., 2006, Propositions 6 and 8).

When the function c has the entropic form $c(p, q) = \lambda R(p||q)$ with respect to a reference probability $q \in \Delta^\sigma$, criterion (8) takes the *multiplier* form

$$V_{\lambda,q}(f) = \min_{p \in \Delta} \left\{ \int u(f)\, dp + \lambda R(p||q) \right\}$$

analyzed by Hansen and Sargent (2001, 2008).[15] If, instead, the function c has the indicator form δ_C, with C compact and convex, criterion (8) takes the *max-min* form

$$V(f) = \min_{p \in C} \int u(f)\, dp$$

axiomatized by Gilboa and Schmeidler (1989).

All these criteria are here considered in their classical interpretation, so Waldean for the max-min criterion, in which the elements of Δ are interpreted as models.

3 Models and preferences

3.1 Models

The consequences of the acts among which decision makers have to choose depend on exogenous states that are outside their control. They know that states obtain according to a probabilistic model described by a probability measure in Δ, the so-called *true* or *correct model*. If decision makers knew the true model, they would confront only risk, which is the randomness inherent to the probabilistic nature of the model. Our decision makers, unfortunately, may not know the

[15]Strzalecki (2011) provides the behavioral assumptions that characterize multiplier preferences among variational preferences.

true model. Yet, they are able to posit a set of *structured* probabilistic models Q, based on their information (which might well include existing scientific theories, say economic or physical), that form a set of alternative hypotheses regarding the true model. It is a classical assumption, in the spirit of Wald (1950), in which Q is a set of posited hypotheses about the probabilistic behavior of a, natural or social, phenomenon of interest.

A *classical decision framework* is described by a quartet:

$$(S, \Sigma, X, Q) \tag{9}$$

in which a set Q of models is added to a standard decision framework (7), as discussed in the Introduction. The true model might not be in Q, that is, the decision makers information may be unable to pin it down. Throughout the paper we assume that decision makers know this limitation of their information and so confront model misspecification.[16] This is in contrast with Wald (1950) and most of the subsequent decision-theoretic literature, which assumes that decision makers either know the true model and so face risk or, at least, know that the true model belongs to Q and so face model ambiguity.[17]

In Theorem 1, but not in Theorem 2, we assume that Q is a convex subset of Δ^{σ}. As usual, convexity significantly simplifies the analysis. Yet, conceptually it is not an innocuous property: a hybrid model that mixes two structured models can only be less well motivated than either of them. Decision criterion (1), however, accounts for the lower appeal of hybrid models when $c(p, q)$ is also convex in q (as, for instance, when c is a ϕ-divergence). To see why, observe that $\min_{p \in \Delta} \left\{ \int u(f) \, dp + c(p, q) \right\}$ is, for each act f, convex in q. In turn, this implies that hybrid models negatively affect criterion (1). This negative impact of mixing thus features an "aversion to model hybridization" attitude, behaviorally captured by axiom A.9. Remarkably, the relative entropy criterion (2) turns out to be neutral to model hybridization. In this important special case, the assumption of convexity of Q is actually without any loss of generality (as Appendix B.2 clarifies).

Convexity of Q can be also justified in a robust Bayesian interpretation of our analysis that regards Q as the set of the so-called predictive distributions, which are combinations of "primitive" models (typically extreme points of Q) weighted according to alternative priors μ over them. For instance, if the primitive models describe states through i.i.d. processes, the elements of Q describe them via exchangeable processes that combine primitive models and priors (as in the Hewitt and Savage, 1955, version of the de Finetti Representation Theorem). Under this interpretation, the p's are introduced to provide a protective shield for each of the predictive distributions constructed from the alternative priors that are considered.

In this robust Bayesian vein, a second approach to constructing a convex Q is based on the

[16] Aydogan et al. (2018) propose an experimental setting that reveals the relevance of model misspecification for decision making.

[17] The model ambiguity (or uncertainty) literature is reviewed in Marinacci (2015).

potential misspecification of the likelihood along with uncertainty over the choice of a prior distribution. Let $Q^0 = \{q_\theta\}_{\theta\in\Theta}$ be a parameterized family where each θ denotes an alternative structured model. For simplicity, we consider the case of a finite collection of such models, that is, Θ has n elements. When needed, we identify Q^o with the vector $(q_\theta)_{\theta\in\Theta}$ of the Cartesian product Δ^n of simplexes. Each prior μ on Θ induces a predictive probability:

$$q = \sum_{\theta\in\Theta} q_\theta \mu_\theta \tag{10}$$

We consider a convex family of priors $\mu \in \Pi$. The set of predictive probabilities, Q, formed in this manner inherits the convexity of Π. One possible choice of Π is the set of all possible prior distributions over Θ giving rise to $Q = \operatorname{co} Q^o$ (we include our more general construction of $Q \subseteq \operatorname{co} Q^o$ to capture the perspective of a robust Bayesian with prior uncertainty). Assume that the reductive map $\mu \mapsto \sum_{\theta\in\Theta} q_\theta\mu_\theta$ is injective (so bijective).[18] Each predictive q is thus represented with a unique prior $\mu^q \in \Pi$ that quantifies a possible belief of the decision maker over the structured models of substantive interest.

The decision maker entertains misspecified likelihoods denoted p_θ over θ using, to fix ideas, a ϕ-divergence $D_\phi(p_\theta\|q_\theta)$. This calculation depends on θ. Alternative priors in Π provide alternative ways to average across the θ-specific divergences. For a given prior μ, we construct the composite divergence as:

$$\sum_{\theta\in\Theta} D_\phi(p_\theta\|q_\theta)\mu_\theta \tag{11}$$

This formula, considered in Hansen and Sargent (2022, 2022b), gives a measure of potential likelihood misspecification for a pre-specified predictive q given by (10). For implementing the decision formulation in practice, it would suffice to stop here while letting the decision maker investigating prior sensitivity by searching over misspecified likelihoods and priors in the set Π. Moreover, notice that for a given family of p_θ, minimizing (11) over all possible priors will result in a degenerate prior putting all the weight on the structured model θ that is least misspecified according to $D_\phi(p_\theta\|q_\theta)$. Only when multiple models have the same low measure of misspecification will the minimization include non-degenerate priors over models.

The weighted divergence (11) implies a divergence between q and a predictive p formed with the same prior μ^q as q:

$$p = \sum_{\theta\in\Theta} p_\theta \mu_\theta^q$$

There will be multiple ways to represent this predictive distribution. For instance, $p_\theta = p$ for all $\theta \in \Theta$ is one obvious choice. Thus, there is an induced distance between predictive distributions

[18] For instance, this is the case when the structured models q_θ are suitably orthogonal. Lemma 1 of Cerreia-Vioglio et al. (2013) characterizes the injectivity of the reductive map.

given, for each $(p, q) \in \Delta \times Q$, by

$$c(p, q) = \min_{(p_\theta)_{\theta\in\Theta}\in\Delta^n : p=\sum_{\theta\in\Theta} p_\theta \mu_\theta^q} \sum_{\theta\in\Theta} D_\phi(p_\theta \| q_\theta)\mu_\theta^q \tag{12}$$

This distance is a lower semicontinuous and convex variational divergence, as detailed in Lemma 14 of Appendix B.1 with a general statistical distance playing the role of D_ϕ here.

3.2 Preferences

We consider a two-preference setup, as in Gilboa et al. (2010), with a mental preference $\succsim^*$ and a behavioral preference $\succsim$.

Definition 1 *A preference* $\succsim$ *is (*subjectively*)* rational *if it is:*

a. complete;

b. risk independent: for all $x, y, z \in X$ *and* $\alpha \in (0,1)$*, if* $x \sim y$ *then* $\alpha x + (1-\alpha) z \sim \alpha y + (1-\alpha) z$*.*

The behavioral preference $\succsim$ governs the decision maker choice behavior and so it is natural to require it to be complete because, eventually, the decision maker has to choose between alternatives (burden of choice). It is subjectively rational because, in an "argumentative" perspective, the decision maker cannot be convinced that it leads to incorrect choices. Risk independence ensures that $\succsim$ is represented on the space of consequences X by an affine utility function $u : X \to \mathbb{R}$, for instance an expected utility functional when X is the set of simple lotteries. So, risk is addressed in a standard way and we abstract from non-expected utility issues.

The mental preference $\succsim^*$ on $\mathcal{F}$ represents the decision maker's "genuine" preference over acts, so it has the nature of a dominance relation for the decision maker. As such, it might well not be complete because of the decision maker inability to compare some pairs of acts.

Definition 2 *A preference* $\succsim^*$ *is a* dominance relation *(or is* objectively rational*) if it is:*

a. c-complete: for all $x, y \in X$*,* $x \succsim^* y$ *or* $y \succsim^* x$*;*

b. completeness: when Q *is a singleton, for all* $f, g \in \mathcal{F}$*,* $f \succsim^* g$ *or* $g \succsim^* f$*;*

c. weak c-independent: for all $f, g \in \mathcal{F}$*,* $x, y \in X$ *and* $\alpha \in (0,1)$*,*

$$\alpha f + (1-\alpha)x \succsim^* \alpha g + (1-\alpha)x \implies \alpha f + (1-\alpha)y \succsim^* \alpha g + (1-\alpha)y$$

d. convex: for all $f, g, h \in \mathcal{F}$ and $\alpha \in (0,1)$,

$$f \succsim^* h \text{ and } g \succsim^* h \implies \alpha f + (1-\alpha) g \succsim^* h$$

If $f \succsim^* g$ we say that f *dominates* g (*strictly* if $f \succ^* g$). It is objectively rational because the decision maker can convince others of its reasonableness, for instance through arguments based on scientific theories (a case especially relevant for our purposes). Momentarily, axiom A.3 will further clarify its nature. The dominance relation is, axiomatically, a variational preference which is not required to be complete, unless Q is a singleton.[19] When Q is a singleton, the dominance relation is complete and yet, because of model misspecification, satisfies only a weak form of independence. In other words, in our approach model misspecification may cause violations of the independence axiom for the dominance relation. Later in the paper, Proposition 6 will show that relaxing independence to weak c-independence is conceptually necessary as, otherwise, the behavioral preference would be misspecification neutral. This is a key observation for our analysis.

Along with the classical decision framework (9), the preferences $\succsim^*$ and $\succsim$ form a *two-preference classical decision environment*

$$(S, \Sigma, X, Q, \succsim^*, \succsim) \tag{13}$$

The next two assumptions, which we take from Gilboa et al. (2010), connect the two preferences $\succsim^*$ and $\succsim$.

A.1 *Consistency*: for all $f, g \in \mathcal{F}$,

$$f \succsim^* g \Longrightarrow f \succsim g$$

Consistency asserts that, whenever possible, the mental ranking informs the behavioral one. The next condition says that the decision maker opts, by default, for a sure alternative x over an uncertain one f, unless the dominance relation says otherwise.

A.2 *Caution*: for all $x \in X$ and $f \in \mathcal{F}$,

$$f \not\succsim^* x \Longrightarrow x \succsim f$$

Unlike the previous assumptions, the next two are peculiar to our analysis. They both link the posited set Q to the two preferences $\succsim^*$ and $\succsim$ of the decision maker. We begin with the

[19]Convexity is stronger than uncertainty aversion a la Schmeidler (1989), which merely requires that $f \sim^* g$ implies $\alpha f + (1-\alpha) g \succsim^* g$ for all $\alpha \in (0,1)$. Yet, convexity and uncertainty aversion coincide under completeness (see, e.g., Lemma 56 of Cerreia-Vioglio et al., 2011b). Nascimento and Riella (2011) study incomplete variational preferences, but their result is not applicable to our setting because their axioms are over lotteries of acts (and their state space is finite).

dominance relation $\succsim^*$. Here we write $f \overset{Q}{=} g$ when $q(f = g) = 1$ for all $q \in Q$, i.e., f and g are equal almost everywhere according to each structured model.

A.3 *Objective Q-coherence*: for all $f, g \in \mathcal{F}$,

$$f \overset{Q}{=} g \implies f \sim^* g$$

This axiom provides a preferential translation of the special status of structured models over unstructured ones: if they all regard two acts to be almost surely identical, the decision maker's "genuine" preference $\succsim^*$ follows suit and ranks them indifferent.

Previously, we noted that for some applications it may be important to allow the set of structured models, Q, not to be convex. Nevertheless, the closed convex hull, $\overline{\text{co}}\, Q$, of Q will play an important role in our next axiom.[20] Even when Q is not convex, we assign a special role to the probabilities in its convex hull relative to other unstructured models. Our rationale is that hybrid models retain an epistemic status and are more than just statistical artifacts used to assess model misspecification.[21]

To introduce our next axiom, recall that a rational preference $\succsim$ satisfies risk independence and thus admits an affine utility function $u : X \to \mathbb{R}$ that can be used to represent it over consequences as an expected utility.[22] Given a model $p \in \Delta$ and an act f, we define an indifference class $X_f^p \subseteq X$ of consequences x_f^p via the equality

$$u(x_f^p) = \int u(f)\, dp \tag{14}$$

We can interpret each x_f^p as a consequence that would be indifferent, so equivalent, to act f if p were the correct model. By constructing these equivalent consequences for alternative acts and models, our next axiom relates the posited set of models Q with the behavioral preference $\succsim$.

A.4 *Subjective Q-coherence*: for all $f \in \mathcal{F}$ and $x \in X$, we have

$$x \succ^* x_f^p \implies x \succ f$$

if and only if $p \in \overline{\text{co}}\, Q$.

In words, $p \in \Delta$ is a structured or hybrid model, so belongs to $\overline{\text{co}}\, Q$, if and only if decision makers take it seriously, that is, they never choose an act f that would be strictly dominated

[20]The need to consider the weak*-closure of the convex hull is a technical detail (with a finite set Q we can just consider convex hulls).

[21]In the robust Bayesian perspective previously discussed, the elements of $\overline{\text{co}}\, Q$ are the predictive distributions determined by alternative priors over Q.

[22]Under the usual identification of constant acts with consequences.

if p were the correct model. Such a salience of p for the decision makers' preference is the preferential footprint of a structured or hybrid model that decision makers take seriously under consideration because of its epistemic status – as opposed to a purely unstructured model, which they regard as a mere statistical artifact with no epistemic content.

More can be said in the original Anscombe-Aumann setting with lottery-valued acts. For a given model $p \in \Delta$ and act f, we construct the integral $\int f dp$, which is a lottery that describes the prize distribution induced by act f when states are generated by model $p \in \Delta$.[23] If $u : X \to \mathbb{R}$ is any affine utility function that represents $\succsim$ on X, then this integral obviously satisfies (14). This particular construction adds further clarity to axiom A.4 because it identifies one lottery in the indifference class X_f^p that depends directly on the model p. This axiom can now be written as

$$x \succ^* \int f dp \implies x \succ f$$

As an additional benefit, this formulation makes it clear that the definition of x_f^p is independent of the choice in (14) of the specific utility u that represents $\succsim$ on X.

To conclude, observe that in the traditional purely subjective axiomatizations, there is no way (actually, no language) to embed the probabilistic information that Q represents in the decision maker preference.[24] The last two axioms provide the needed embedding, as the representation theorems will show momentarily.

4 Representation with given structured information

We now show how the assumptions on the mental and behavioral preferences permit to characterize criterion (1) for a given set Q in Δ^σ, that is, for a DM's given structured information.

To this end, throughout this section we assume that Q is a compact and convex set and we say that a function $c : \Delta \times Q \to [0, \infty]$ is *uniquely null* if, for all $(p, q) \in \Delta \times Q$, the sets $c_p^{-1}(0)$ and $c_q^{-1}(0)$ are at most singletons. For instance, statistical distances are uniquely null because of the distance property (c.i).

We are now ready to state our first representation result.

Theorem 1 *Let $(S, \Sigma, X, Q, \succsim^*, \succsim)$ be a two-preference classical decision environment, where (S, Σ) is a standard Borel space. The following statements are equivalent:*

(i) $\succsim^*$ is an unbounded dominance relation and $\succsim$ is a rational preference that are both Q-coherent and jointly satisfy consistency and caution;

[23] For the simple act $f = \sum_i 1_{A_i} x_i$, by definition $\left(\int f dp\right)(z)$ is the probability $\sum_i p(A_i) x_i(z)$ of obtaining prize z by choosing f under p.

[24] For instance, in the Gilboa and Schmeidler (1989) seminal axiomatization the derived set of probabilities C is purely subjective. There is no formal connection with any underlying probabilistic information, something left to the decision maker personal, unmodelled, elaborations. A notable exception is Gajdos et al. (2008), which considers probabilistic information. Its analysis proceeds along lines very different from ours.

(ii) there exist an onto affine function $u : X \to \mathbb{R}$ *and a variational pseudo-statistical distance* $c : \Delta \times Q \to [0, \infty]$*, with* $\operatorname{dom} c_Q \subseteq \Delta^{\ll}(Q)$*, such that, for all acts* $f, g \in \mathcal{F}$*,*

$$f \succsim^* g \Longleftrightarrow \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p,q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + c(p,q) \right\} \qquad \forall q \in Q \quad (15)$$

and

$$f \succsim g \Longleftrightarrow \min_{p \in \Delta} \left\{ \int u(f)\, dp + \min_{q \in Q} c(p,q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + \min_{q \in Q} c(p,q) \right\} \quad (16)$$

If, in addition, c *is uniquely null, then it can be chosen to be a variational statistical distance.*

This result identifies, in particular, the main preferential assumptions underlying a representation of the type

$$V(f) = \min_{p \in \Delta} \left\{ \int u(f)\, dp + \min_{q \in Q} c(p,q) \right\} \quad (17)$$

for the preference $\succsim$ when a set Q of structured models is given. While this representation is of interest for a general variational pseudo-statistical distance c, it is of particular interest when c is a variational statistical distance. In this case, the partial ordering $\succsim^*$ is more easily interpreted. Though a technical condition of "unique nullity" is imposed to pin down statistical distances, our representation arguably has more general applicability and captures the preferential underpinning of criterion (17).

The Hausdorff statistical set distance $\min_{q \in Q} c(p,q)$ between p and Q is strictly positive if and only if p is an unstructured model, i.e., $p \notin Q$. In particular, the more distant from Q is an unstructured model, the more it is penalized as reflected in the minimization problem that criterion (17) features. In terms of uniqueness of the representation, the variational representation (u, c_Q) is unique, up to scaling, as in Maccheroni et al. (2006). As to the uniqueness of c, it will be established in the richer framework of Theorem 2.

A misspecification index A behavioral preference $\succsim$ represented by (17) is variational with index $\min_{q \in Q} c(p,q)$. So, if two unbounded preferences $\succsim_1$ and $\succsim_2$ represented by (17) share the same u but feature different statistical distances $\min_{q \in Q} c_1(p,q)$ and $\min_{q \in Q} c_2(p,q)$, then $\succsim_1$ is more uncertainty averse than $\succsim_2$ if and only if

$$\min_{q \in Q} c_1(p,q) \leq \min_{q \in Q} c_2(p,q)$$

In the present "classical" setting we interpret this comparative result as saying that the lower is $\min_{q \in Q} c(p,q)$, the higher is the fear of misspecification. Indeed, Q is fixed and the differences in behavior cannot be due to model ambiguity. We thus regard the section c_Q, i.e., the map

$$p \mapsto \min_{q\in Q} c\left(p,q\right) \tag{18}$$

as an index of aversion to model misspecification and we call it, for short, a *misspecification index*. The lower is this index, the higher is the fear of misspecification. The index is maximal when

$$c_Q\left(p\right) = \delta_Q\left(p\right) = \begin{cases} 0 & \text{if } p \in Q \\ +\infty & \text{else} \end{cases}$$

Later we will interpret this maximal case as representing a neutral attitude toward model misspecification (cf. Definition 4). In this case, the decision maker does not care about unstructured models and maximally penalizes them, so they play no role in the decision criterion. In contrast, unstructured models are penalized less, so play a bigger role in the criterion, when the decision maker wants to keep them on the table to express a concern about model misspecification. Comparing two indexes, when

$$c_{1,Q} \leq c_{2,Q}$$

we interpret the lower penalization of unstructured models in $c_{1,Q}$ as modelling a higher concern for model misspecification.

Specifications and computability Two specifications of our representation are noteworthy. First, when c is the entropic statistical distance $\lambda R(p||q)$, with $\lambda \in (0,\infty]$, we have the following important special case of our representation

$$V\left(f\right) = \min_{p\in\Delta}\left\{\int u\left(f\right)dp + \lambda \min_{q\in Q} R(p||q)\right\} \tag{19}$$

which gives tractability to our decision criterion under model misspecification. Specifically, for $\lambda \in (0,\infty)$,[25]

$$\min_{p\in\Delta}\left\{\int u\left(f\right)dp + \lambda \min_{q\in Q} R(p||q)\right\} = \min_{q\in Q} -\lambda \log \int e^{-\frac{u(f)}{\lambda}}dq \tag{20}$$

This result is well known when Q is a singleton, that is, when (19) is a standard multiplier criterion.

A second noteworthy special case of our representation is the Gini criterion

$$V\left(f\right) = \min_{p\in\Delta}\left\{\int u\left(f\right)dp + \lambda \min_{q\in Q} \chi^2(p||q)\right\} \tag{21}$$

[25] When $\lambda = \infty$, we have $\min_{p\in\Delta}\left\{\int u\left(f\right)dp + \lambda \min_{q\in Q} R(p||q)\right\} = \min_{q\in Q}\int u\left(f\right)dq$. See Appendix B.2 for the simple proof of (20).

Remarkably, we have

$$\min_{p\in\Delta}\left\{\int u(f)\,dp+\lambda\min_{q\in Q}\chi^2(p||q)\right\}=\min_{q\in Q}\left\{\int u(f)\,dq-\frac{1}{2\lambda}\mathrm{Var}_q(u(f))\right\}\tag{22}$$

for all acts f for which the *mean-variance* (in utils) criteria on the r.h.s. are monotone. So, the Gini criterion is a monotone version of the max-min mean-variance criterion.[26]

As to computability, in the important case when criterion (1) features a ϕ-divergence, like the specifications just discussed, we need only to know the set Q to compute it, no integral with respect to unstructured models is needed. This is proved in the next result which is a consequence of a duality formula of Ben-Tal and Teboulle (2007).[27]

Proposition 1 *Given $Q\subseteq\Delta^\sigma$ and $\lambda\in(0,\infty)$, for each act $f\in\mathcal{F}$ it holds*

$$V(f)=\min_{p\in\Delta}\left\{\int u(f)\,dp+\lambda\min_{q\in Q}D_\phi(p||q)\right\}=\lambda\min_{q\in Q}\sup_{\eta\in\mathbb{R}}\left\{\eta-\int\phi^*\left(\eta-\frac{u(f)}{\lambda}\right)dq\right\}$$

The r.h.s. formula computes criterion (1) for ϕ-divergences by using only integrals with respect to structured models. This formula substantially simplifies computations and thus confirms the analytical tractability of the previous specifications.

4.1 Interpretation of the decision criterion

In the Introduction we outlined a "protective belt" interpretation of decision criterion (17), i.e.,

$$V(f)=\min_{p\in\Delta}\left\{\int u(f)\,dp+\min_{q\in Q}c(p,q)\right\}$$

To elaborate, we begin by observing that the misspecification index (18) has the following bounds

$$0\leq\min_{q\in Q}c(p,q)\leq\delta_Q(p)\qquad\forall p\in\Delta\tag{23}$$

So, fear of misspecification is absent when the misspecification index is δ_Q – e.g., when $\lambda=+\infty$ in (19) – in which case criterion (17) takes a Wald (1950) max-min form

$$V(f)=\min_{q\in Q}\int u(f)\,dq\tag{24}$$

This max-min criterion characterizes a decision maker who confronts model misspecification, but is not concerned by it, and exhibits only aversion to model ambiguity. In other words, this Waldean decision maker is a natural candidate to be (model) misspecification neutral. The next

[26] At the end of Appendix B.2 we further discuss this point.

[27] Here ϕ^* denotes the convex Fenchel conjugate of ϕ, once extended to $\mathbb{R}$ by setting $\phi(t)=+\infty$ if $t<0$. In particular, ϕ^* is increasing.

limit result further corroborates this insight by showing that, when the fear of misspecification vanishes, the decision maker becomes Waldean.[28]

Proposition 2 *For each act $f \in \mathcal{F}$, we have*

$$\lim_{\lambda \uparrow \infty} \min_{p \in \Delta} \left\{ \int u(f)\, dp + \lambda \min_{q \in Q} R(p \| q) \right\} = \min_{q \in Q} \int u(f)\, dq$$

These observations, via bounds and limits, call for a proper decision-theoretic analysis of misspecification neutrality. To this end, note that structured models may be incorrect, yet useful as Box (1976) famously remarked. This motivates the next notion. Recall that act xAy, with $x \succ y$, represents a bet on event A.

Definition 3 *A preference $\succsim$ is* bet-consistent *if, given any $x \succ y$,*

$$q(A) \geq q(B) \quad \forall q \in Q \Longrightarrow xAy \succsim xBy$$

for all events $A, B \in \Sigma$.

Under bet-consistency, a decision maker may fear model misspecification yet regards structured models as good enough to choose to bet on events that they unanimously rank as more likely. Preferences that are bet-consistent can be classified as exhibiting a mild form of fear of model misspecification. The following result shows that an important class of preferences, which includes the ones represented by criterion (19), are bet-consistent.

Proposition 3 *If $\lambda \in (0, \infty]$ and $c = \lambda D_\phi$, then a preference $\succsim$ represented by (17) is bet-consistent.*

Next we substantially strengthen bet-consistency by considering all acts, not just bets.

Definition 4 *A preference $\succsim$ is (*model*)* misspecification neutral *if*

$$\int u(f)\, dq \geq \int u(g)\, dq \quad \forall q \in Q \Longrightarrow f \succsim g$$

for all acts $f, g \in \mathcal{F}$.

In this case, a decision maker trusts models enough so to follow them when they unanimously rank pairs of acts. Fear of misspecification thus plays no role in the decision maker preference, so it is decision-theoretically irrelevant. For this reason, the decision maker attitude toward model misspecification can be classified as neutral. The next result shows that this may happen if and only if the decision maker adopts the max-min criterion (24).

[28] To ease matters, we state the result in terms of criterion (19). A general version can be easily established via an increasing sequence of misspecification indexes.

Proposition 4 *A preference $\succsim$ represented by criterion (17) is misspecification neutral if and only if it is represented by the max-min criterion (24).*

This result provides the sought-after decision-theoretic argument for the interpretation of the max-min criterion as the special case of decision criterion (17) that corresponds to aversion to model ambiguity, with no fear of misspecification.[29] As remarked in the Introduction, it suggests that a decision maker using criterion (17) may be viewed as a decision maker who, under model ambiguity, would max-minimize over the set of structured models which she posited but that, for fear of misspecification, ends up using the more prudential variational criterion (17). Unstructured models lack the informational status of structured models, yet in the criterion (17) they act as a "protective belt" against model misspecification.

Under this interpretation of the criterion (17), the special multiplier case of a singleton $Q = \{q\}$ corresponds to a decision maker who, with no fear of misspecification, would adopt the expected utility criterion $\int u(f)\, dq$ to confront the risk inherent to q. In other words, a singleton Q in (17) corresponds to an expected utility decision maker who fears misspecification.

Summing up, in our analysis decision makers adopt the max-min criterion (24) if they either confront only model ambiguity (an information trait) or are averse to model ambiguity with no fear of model misspecification (a taste trait).

4.2 Interpretation of the dominance relation

As just argued, the singleton $Q = \{q\}$ special case

$$\min_{p\in\Delta}\left\{\int u(f)\, dp + c(p,q)\right\} \tag{25}$$

of decision criterion (17) is an expected utility criterion under fear of misspecification (of the unique posited q). Via the relation

$$f \succsim^* g \Longleftrightarrow \min_{p\in\Delta}\left\{\int u(f)\, dp + c(p,q)\right\} \geq \min_{p\in\Delta}\left\{\int u(g)\, dp + c(p,q)\right\} \qquad \forall q \in Q \tag{26}$$

the representation theorem thus clarifies the interpretation of $\succsim^*$ as a dominance relation under model misspecification by showing that it amounts to uniform dominance across all structured models with respect to criterion (25). The preference $\succsim^*$ thus arises naturally when a set Q is posited by providing a preferential account of the decision maker's probabilistic information that this set represents. In the two-preference setting that we adopted, the axiomatic connections between $\succsim^*$ and $\succsim$, via consistency and caution, then allow us to embed this information in the behavioral preference.

[29]This result actually holds without any convexity assumption on Q. The same applies to Propositions 1, 3 and 5 of this section.

It is easy to see that strict dominance amounts to (26), with strict inequality for some $q \in Q$. This observation raises a question: is there a notion of dominance that corresponds to strict inequality for all $q \in Q$? To address this question, we introduce a *strong dominance* relation by writing $f \succ\!\!\succ^* g$ if, for all acts $h, l \in \mathcal{F}$,

$$(1-\delta) f + \delta h \succ^* (1-\delta) g + \delta l$$

for all small enough $\delta \in [0,1]$.[30] By taking $h = f$ and $l = g$, we have the basic implication

$$f \succ\!\!\succ^* g \Longrightarrow f \succ^* g$$

Strong dominance is a strengthening of strict dominance in which the decision maker can convince others "beyond reasonable doubt." The next characterization corroborates this interpretation and, at the same time, answers the previous question in the positive.[31]

Proposition 5 *Let $c : \Delta \times Q \to [0, \infty]$ be a variational statistical distance, $u : X \to \mathbb{R}$ an onto and affine function and $\succsim^*$ an unbounded dominance relation represented by (26). For all acts $f, g \in \mathcal{F}$, we have $f \succ\!\!\succ^* g$ if and only if there exists $\varepsilon > 0$ such that*

$$\min_{p\in\Delta}\left\{\int u(f)\,dp + c(p,q)\right\} \geq \min_{p\in\Delta}\left\{\int u(g)\,dp + c(p,q)\right\} + \varepsilon \qquad \forall q \in Q \tag{27}$$

This characterization shows that $\succ^*$ and $\succ\!\!\succ^*$ agree on consequences and, more importantly, that

$$f \succ\!\!\succ^* g \Longrightarrow \min_{p\in\Delta}\left\{\int u(f)\,dp + c(p,q)\right\} > \min_{p\in\Delta}\left\{\int u(g)\,dp + c(p,q)\right\} \qquad \forall q \in Q$$

At the same time, (27) implies

$$f \succ\!\!\succ^* g \Longrightarrow f \succ g \tag{28}$$

We can diagram the relationships among the different dominance notions as follows:

$$\begin{array}{ccccc} \succ\!\!\succ^* & \Longrightarrow & \succ^* & \not\Longrightarrow & \succ \\ \Downarrow & & \Downarrow & & \\ \succ & \Longrightarrow & \succsim & & \end{array}$$

An instance when

$$f \succ^* g \;\Longrightarrow\; f \succ g \tag{29}$$

may fail is the max-min criterion (24).

[30]Strong dominance has been introduced by Cerreia-Vioglio et al. (2020).

[31]Up to an ε that ensures a needed uniformity of the strict inequality across structured models.

We close by discussing misspecification neutrality, which in view of Proposition 4 is characterized by the misspecification index $\min_{q\in Q} c(p,q) = \delta_Q(p)$.

Lemma 1 *Let c be a variational statistical distance $c : \Delta\times Q \to [0,\infty]$. We have $\min_{q\in Q} c(p,q) = \delta_Q(p)$ if and only if, for each $q \in Q$, $c(p,q) = \infty$ for all $p \notin Q$.*

Misspecification neutrality is thus characterized by a statistical distance that maximally penalizes unstructured models, which end up playing no role. From a statistical distance angle, this confirms that misspecification neutrality is the attitude of a decision maker who confronts model misspecification, but does not care about it (and so has no use for unstructured models).

This angle becomes relevant here because it shows that, under misspecification neutrality, the representation (26) of the dominance relation becomes

$$f \succsim^* g \Longleftrightarrow \min_{q'\in Q}\left\{\int u(f)\,dq' + c(q',q)\right\} \geq \min_{q'\in Q}\left\{\int u(g)\,dq' + c(q',q)\right\} \qquad \forall q\in Q \tag{30}$$

Unstructured models play no role here. This is shown by the next result which also demonstrates how relaxing independence to weak c-independence is conceptually necessary. For, if the dominance relation $\succsim^*$ satisfies the stronger assumption of c-independence, then the behavioral preference $\succsim$ is necessarily misspecification neutral.

A.5 *C-independence.* For all $f \in \mathcal{F}$, $x, y \in X$ and $\alpha \in (0,1]$,

$$f \succsim^* x \iff \alpha f + (1-\alpha)\,y \succsim^* \alpha x + (1-\alpha)\,y$$

When the dominance relation $\succsim^*$ is complete, our version is equivalent to the original version of Gilboa and Schmeidler (1989). Otherwise, ours is weaker.

Proposition 6 *Let $(S,\Sigma,X,Q,\succsim^*,\succsim)$ be a two-preference classical decision environment, where (S,Σ) is a standard Borel space. The following statements are equivalent:*

(i) $\succsim^*$ is an unbounded dominance relation that satisfies c-independence and $\succsim$ is a rational preference that are both Q-coherent and jointly satisfy consistency and caution;

(ii) there exist an onto affine function $u : X \to \mathbb{R}$ and a variational pseudo-statistical distance $c : \Delta\times Q \to [0,\infty]$, with $c_Q = \delta_Q$, such that, for all acts $f, g \in \mathcal{F}$, it holds (30) and

$$f \succsim g \iff \min_{q\in Q}\int u(f)\,dq \geq \min_{q\in Q}\int u(g)\,dq \tag{31}$$

Moreover, $\succsim^*$ satisfies independence if and only if $c : \Delta\times Q \to [0,\infty]$ can be chosen to be the variational statistical distance $c(p,q) = \delta_{\{q\}}(p)$ for all $(p,q) \in \Delta\times Q$.

To sum up, only a genuine variational dominance relation can accommodate fear of model misspecification and an approach where structured models always have a different and more relevant status than unstructured models.

The last part of the statement,[32] which is the version for our setting of the main result of Gilboa et al. (2010), shows that also statistical distances play no role, so (30) reduces to

$$f \succsim^* g \Longleftrightarrow \int u(f)\, dq \geq \int u(g)\, dq \qquad \forall q \in Q$$

when the dominance relation satisfies the independence axiom.

5 Representation with varying structured information

So far, we carried out our analysis for a given set Q of structured models. Indeed, a two-preference classical decision environment (13) should be more properly written as

$$\left(S, \Sigma, X, Q, \succsim_Q^*, \succsim_Q\right)$$

with the dependence of preferences on Q highlighted. Decision environments, however, may share common state and consequence spaces, but differ on the posited sets of structured models because of different information that decision makers may have. It then becomes important to ensure that decision makers use decision criteria that, across such environments, are consistent.

To address this issue, in this section we consider a family

$$\left\{\left(S, \Sigma, X, Q, \succsim_Q^*, \succsim_Q\right)\right\}_{Q \in \mathcal{Q}}$$

of classical decision environments that differ in the set Q of posited models and we introduce axioms on the family $\left\{\succsim_Q^*\right\}_{Q \in \mathcal{Q}}$ that connect these environments. We assume that $\mathcal{Q}$ is a collection of compact subsets of Δ^σ that contains all singletons and that covers all doubletons, that is, for each $q, q' \in \Delta^\sigma$ there exists some $Q \in \mathcal{Q}$ such that $\{q, q'\} \subseteq Q$. These assumptions are satisfied, for example, by the collection of finite sets of Δ^σ as well as by the collection $\mathcal{K}$ of its compact and convex sets.

A.6 *Monotonicity* (*in model ambiguity*): for all $f, g \in \mathcal{F}$, if $Q' \subseteq Q$ then

$$f \succsim_Q^* g \Longrightarrow f \succsim_{Q'}^* g$$

According to this axiom, if the "structured" information underlying a set Q is good enough for the decision maker to establish that an act dominates another one, a better information which

[32]In proving this last part, we can dispense with the assumption of (S, Σ) being a standard Borel space. Similarly, (i) would still imply (31), again without any assumption on (S, Σ).

decreases model ambiguity can only confirm such judgement. Its reversal would be, indeed, at odds with the objective rationality spirit of the dominance relation.

Next we consider a separability assumption.

A.7 *Q-separability*: for all $f, g \in \mathcal{F}$,

$$f \succsim_q^* g \quad \forall q \in Q \implies f \succsim_Q^* g$$

In words, an act dominates another one when it does, separately, through the lenses of each structured model. In this axiom the incompleteness of $\succsim_Q^*$ arises as that of a Paretian order over the, complete but possibly misspecification averse, preferences $\succsim_q^*$ determined by the elements of Q.

We close with a continuity axiom. To state it, we need a last piece of notation: we denote by $x_{f,q}$ the consequence indifferent to act f for preference $\succsim_q^*$.[33]

A.8 *Lower semicontinuity*: for all $x \in X$ and $f \in \mathcal{F}$, the set $\left\{q \in \Delta^\sigma : x \succsim_q^* x_{f,q}\right\}$ is closed.

The next class of two-preference families $P_{\mathcal{Q}} = \left\{\left(\succsim_Q^*, \succsim_Q\right)\right\}_{Q \in \mathcal{Q}}$ builds on the properties that we have introduced.

Definition 5 *A two-preference family $P_{\mathcal{Q}}$ is (*misspecification*)* robust *if:*

(i) $\left\{\succsim_Q^*\right\}_{Q \in \mathcal{Q}}$ *is monotone, separable, and lower semicontinuous;*

(ii) for each $Q \in \mathcal{Q}$, $\succsim_Q^*$ is an unbounded dominance relation, $\succsim_Q$ is a rational preference, both are Q-coherent and jointly satisfy caution and consistency.

We can now state our first representation result.

Theorem 2 *Let $P_{\mathcal{Q}}$ be a two-preference family. The following statements are equivalent:*

(i) $P_{\mathcal{Q}}$ is robust;

(ii) there exist an onto affine $u : X \to \mathbb{R}$ and a lower semicontinuous divergence $c : \Delta \times \Delta^\sigma \to [0, \infty]$, convex in p, such that, for each $Q \in \mathcal{Q}$,

$$f \succsim_Q^* g \Longleftrightarrow \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p, q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + c(p, q) \right\} \qquad \forall q \in Q$$

and

$$f \succsim_Q g \Longleftrightarrow \min_{p \in \Delta} \left\{ \int u(f)\, dp + \min_{q \in Q} c(p, q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + \min_{q \in Q} c(p, q) \right\}$$

for all acts $f, g \in \mathcal{F}$.

[33] In symbols, $f \sim_q^* x_{f,q}$. In particular, $x_{f,q}$ should not be confused with x_f^q as in (14).

Moreover, u is cardinal and, given u, c is unique.

A robust P_Q is thus characterized by a utility and divergence pair (u, c) that, consistently across decision environments, represents each $\succsim_Q^*$ via the unanimity rule (15) and each $\succsim_Q$ via the decision criterion

$$V_Q(f) = \min_{p \in \Delta} \left\{ \int u(f)\, dp + \min_{q \in Q} c(p, q) \right\} \tag{32}$$

An unstructured model p may play a role in this criterion when $c(p, q) < \infty$ for some structured model q, that is, when it has a finite distance from a structured model.

In this representation theorem we do not make any convexity assumption on the sets of structured models. Next we sharpen this result by assuming that they are compact and convex subsets of Δ^σ. We introduce a new axiom based on this added structure on sets of models. Under the hypotheses of Theorem 2, all dominance relations $\succsim_Q^*$ agree on X and so we can just write $\succsim^*$, dropping the subscript Q.

A.9 *Model hybridization aversion*: for all $q, q' \in \Delta^\sigma$, $\lambda \in (0, 1)$ and $f \in \mathcal{F}$,

$$\lambda x_{f,q} + (1 - \lambda)\, x_{f,q'} \succsim^* x_{f, \lambda q + (1-\lambda) q'}$$

According to this axiom, the decision maker dislikes, *ceteris paribus*, facing a hybrid structured model $\lambda q + (1 - \lambda)\, q'$ that, by mixing two structured models q and q', could only have a less substantive motivation (cf. Section 3.1).

The next result extends Theorem 1 to families of decision environments. It also sharpens Theorem 2 by dealing with sets of structured models that are also convex; in particular, here we get a variational divergence.

Proposition 7 *Let $P_\mathcal{K}$ be a two-preference family. The following statements are equivalent:*

(i) $P_\mathcal{K}$ is robust and model hybridization averse;

(ii) there exist an onto affine $u : X \to \mathbb{R}$ and a lower semicontinuous and convex variational divergence $c : \Delta \times \Delta^\sigma \to [0, \infty]$ such that, for each $Q \in \mathcal{K}$,

$$f \succsim_Q^* g \Longleftrightarrow \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p, q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + c(p, q) \right\} \qquad \forall q \in Q$$

and

$$f \succsim_Q g \Longleftrightarrow \min_{p \in \Delta} \left\{ \int u(f)\, dp + \min_{q \in Q} c(p, q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + \min_{q \in Q} c(p, q) \right\}$$

for all acts $f, g \in \mathcal{F}$.

Moreover, u is cardinal and, given u, c is unique.

This result ensures that the decision maker uses consistently criterion (1) across decision environments. In particular, the same statistical distance function is used (e.g., the relative entropy). Moreover, axioms A.6-A.9 further clarify the nature of structured models and their connection with the dominance relation.

Besides its broader scope, Proposition 7 improves Theorem 1 on two counts. First, it features a statistical distance without the need of a unique nullity condition. Second, it contains a sharp uniqueness part. The cost of these improvements is a less parsimonious setting in which the set Q is permitted to vary across the collection $\mathcal{K}$ of compact and convex subsets of Δ^{σ}.

6 Admissibility

A *two-preference classical decision problem* is a septet

$$\left(F, S, \Sigma, X, Q, \succsim_Q^*, \succsim_Q\right) \tag{33}$$

where $F \subseteq \mathcal{F}$ is a non-empty choice set formed by the acts among which a decision maker has actually to choose, and the preferences $\succsim_Q^*$ and $\succsim_Q$ are represented as in Theorem 2-(ii).

Given a set Q in $\mathcal{Q}$, the decision maker chooses the best act in F according to $\succsim_Q$. In particular, the *value function* $v : \mathcal{Q} \to (-\infty, \infty]$ is given by

$$v(Q) = \sup_{f \in F} \min_{p \in \Delta} \left\{ \int u(f)\, dp + \min_{q \in Q} c(p, q) \right\} \tag{34}$$

Yet, it is the dominance relation $\succsim_Q^*$ that permits to introduce admissibility.

Definition 6 *An act $f \in F$ is (*weakly*)* admissible *if there is no act $g \in F$ that (strongly) strictly dominates f.*

To relate this notion to the usual notion of admissibility,[34] observe that $g \succ_Q^* f$ amounts to

$$\min_{p \in \Delta} \left\{ \int u(g)\, dp + c(p, q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p, q) \right\} \qquad \forall q \in Q$$

with strict inequality for some $q \in Q$. We are thus purposefully defining admissibility in terms of the structured models Q, not the larger class of models Δ, with a model-by-model adjustment for misspecification that makes our notion different from the usual one.

The next result relates optimality and admissibility.

[34] See, e.g., Ferguson (1967) p. 54. Weak admissibility is, *mutatis mutandis*, related via formula (27) to the notion of extended admissibility studied in Blackwell and Girschick (1954), Heath and Sudderth (1978) and, more recently, in Duanmu and Roy (2021). This connection was pointed out to us by Jesse Shapiro. A statistical risk version of Proposition 5 provides a preferential foundation for extended admissibility.

Proposition 8 *Consider a decision problem (33).*

(i) Optimal acts are weakly admissible. They are admissible provided (29) holds.

(ii) Unique optimal acts are admissible.

Optimal acts (if exist) might not be admissible because the max-min nature of decision criterion (1) may lead to violations of (29). Yet, the last result ensures that they belong to the collection of weakly admissible acts

$$F_Q^* = \left\{ f \in F : \nexists g \in F, g \succ\!\!\succ_Q^* f \right\}$$

Next we build on this property to establish a comparative statics exercise across decision problems (33) that differ on the posited set Q of structured models.

Proposition 9 *We have*

$$Q \subseteq Q' \Longrightarrow v(Q) \geq v(Q')$$

and

$$v(Q) = \max_{f \in F_Q^*} \min_{p \in \Delta} \left\{ \int u(f)\, dp + \min_{q \in Q} c(p, q) \right\}$$

provided the sup *in (34) is achieved.*

Smaller sets of structured models are, thus, more valuable. Indeed, in decision problems that feature a larger set of structured models – so, a more discordant information – the decision maker exhibits, *ceteris paribus*, a higher uncertainty aversion due to a larger model ambiguity:

$$Q \subseteq Q' \Longrightarrow \min_{q \in Q} c(p, q) \geq \min_{q \in Q'} c(p, q) \tag{35}$$

In turn, this easily implies $v(Q) \geq v(Q')$, as the proof shows.

In the comparison (35), the divergence c is invariant as we change the set of structured models. For this reason, in Proposition 9 a larger set of structured models implies a higher uncertainty aversion due to model ambiguity and aversion to it (as is the case for max-min utility).[35] This invariance, however, is not an innocuous assumption as it rules out the possibility that the divergence becomes larger when an enlarged set of structured models reduces misspecification concerns.[36] For instance, the entropic divergence may feature a higher λ when Q gets larger, something that may reverse the inequality (35) by making more valuable larger sets of structured models. Nevertheless, with an invariant c any probability measure outside the set of structured models will necessarily be closer to a larger set of such models, as captured by the divergence. In this sense, increasing the set of structured models may diminish misspecification concerns even under the maintained invariance.

[35] See Ghirardato and Marinacci (2002).

[36] We thank Tim Christensen for having alerted us on this issue.

7 Beyond caution

Caution is the axiom behind the prudential nature of our representations results: Theorems 1 and 2. It is natural to wonder about what happens when we remove this assumption. We formally establish a representation result that extends Theorem 2. A similar version can be discussed within the setup of Section 4. To discuss our more general result, we introduce a new class of two-preference families.

Definition 7 *A two-preference family $P_{\mathcal{Q}}$ is (*misspecification*)* sensitive *if:*

(i) $\left\{\succsim_Q^*\right\}_{Q\in\mathcal{Q}}$ *is monotone, separable, and lower semicontinuous;*

(ii) for each $Q \in \mathcal{Q}$, $\succsim_Q^*$ is an unbounded dominance relation, $\succsim_Q$ is a monotone binary relation, both are Q-coherent when restricted to singletons and jointly satisfy consistency.

Compared to the notion of robust family (cf. Definition 5), we made three changes. The most important is that we removed caution. Moreover, we require $\succsim_Q$ to be only a monotone binary relation and Q-coherence to hold only if Q is a singleton.[37] These two latter changes are immaterial as we will later discuss, when caution is present. Thus, given a sensitive $P_{\mathcal{Q}}$ and Q, this implies that the dominance relation $\succsim_Q^*$ keeps on being represented as before, that is,

$$f \succsim_Q^* g \iff \min_{p\in\Delta}\left\{\int u(f)\,dp + c(p,q)\right\} \geq \min_{p\in\Delta}\left\{\int u(g)\,dp + c(p,q)\right\} \quad \forall q \in Q$$

In particular, given an act f, we have an evaluation map $q \mapsto \min_{p\in\Delta}\left\{\int u(f)\,dp + c(p,q)\right\}$ which belongs to $B(Q)$: the collection of all real-valued bounded functions on Q. Our criterion (32) emerges when these evaluations are aggregated via the minimum on Q. But, a priori, less extreme stances are conceivable. This would require dropping caution as the next two results shows.

Proposition 10 *Let $P_{\mathcal{Q}}$ be a two-preference family. The following statements are equivalent:*

(i) $P_{\mathcal{Q}}$ is sensitive;

(ii) there exist an onto affine $u: X \to \mathbb{R}$, a lower semicontinuous divergence $c: \Delta \times \Delta^\sigma \to [0,\infty]$, convex in p, and a normalized and monotone functional $J_Q : B(Q) \to \mathbb{R}$ such that for each $Q \in \mathcal{Q}$,

$$f \succsim_Q^* g \iff \min_{p\in\Delta}\left\{\int u(f)\,dp + c(p,q)\right\} \geq \min_{p\in\Delta}\left\{\int u(g)\,dp + c(p,q)\right\} \qquad \forall q \in Q \tag{36}$$

[37]A binary relation $\succsim$ over acts $\mathcal{F}$ is a monotone binary relation if it is a non-trivial complete preorder which satisfies monotonicity, continuity and independence over X, and is solvable, that is, for each $f \in \mathcal{F}$ there exists $x \in X$ such that $f \sim x$. In particular, a monotone binary relation is a rational preference if and only if it satisfies continuity over $\mathcal{F}$.

and

$$f \succsim_Q g \iff J_Q\left(\min_{p\in\Delta}\left\{\int u(f)\,dp + c(p,\cdot)\right\}\right) \geq J_Q\left(\min_{p\in\Delta}\left\{\int u(g)\,dp + c(p,\cdot)\right\}\right) \tag{37}$$

for all acts $f, g \in \mathcal{F}$*.*

Moreover, u *is cardinal and, given* u*,* c *is unique.*

Decision-theoretically, Theorem 2 is the special case of this result when $\succsim_Q^*$ and $\succsim_Q$ jointly satisfy caution for all $Q \in \mathcal{Q}$, as Corollary 1 below shows. Analytically, it corresponds to the special case where J_Q is the minimum over Q of the maps

$$q \longmapsto \min_{p\in\Delta}\left\{\int u(f)\,dp + c(p,q)\right\} \tag{38}$$

In this case, by exchanging the order of minima, (37) reduces to the decision criterion (32).

An altogether different case is when J_Q is a quasi-arithmetic mean over the maps (38), so that (37) now becomes

$$V_Q(f) = \phi_Q^{-1}\left(\int_Q \phi_Q\left(\min_{p\in\Delta}\left\{\int_S u(f(s))\,dp(s) + c(p,q)\right\}\right) d\mu_Q(q)\right) \tag{39}$$

where $\phi_Q : \mathbb{R} \to \mathbb{R}$ is strictly increasing and continuous and $\mu_Q \in \Delta(Q)$.[38] Momentarily, this criterion will be the protagonist of the next section. We conclude by observing that the leading assumption driving our representation results is indeed caution while continuity of $\succsim_Q$ and Q-coherence could have been dispensed with.

Corollary 1 *Let* $P_{\mathcal{Q}}$ *be a two-preference family. The following statements are equivalent:*

(i) $P_{\mathcal{Q}}$ *is robust;*

(ii) $P_{\mathcal{Q}}$ *is sensitive and* $\succsim_Q^*$ *and* $\succsim_Q$ *jointly satisfy caution for all* $Q \in \mathcal{Q}$*;*

(iii) there exist an onto affine $u : X \to \mathbb{R}$ *and a lower semicontinuous divergence* $c : \Delta \times \Delta^\sigma \to [0,\infty]$*, convex in* p*, such that, for each* $Q \in \mathcal{Q}$*,*

$$f \succsim_Q^* g \iff \min_{p\in\Delta}\left\{\int u(f)\,dp + c(p,q)\right\} \geq \min_{p\in\Delta}\left\{\int u(g)\,dp + c(p,q)\right\} \qquad \forall q \in Q$$

[38] Here, with a small abuse of notation, the set $\Delta(Q)$ denotes the set of all Borel probability *measures* over Q. In particular, in order to discuss this functional form, we need the maps defined as in (38), to be Borel measurable: a property which is guaranteed by the joint lower semicontinuity of c.

and

$$f \succsim_Q g \Longleftrightarrow \min_{p\in\Delta}\left\{\int u(f)\,dp + \min_{q\in Q} c(p,q)\right\} \geq \min_{p\in\Delta}\left\{\int u(g)\,dp + \min_{q\in Q} c(p,q)\right\}$$

for all acts $f, g \in \mathcal{F}$.

Moreover, u *is cardinal and, given* u, c *is unique.*

8 A Bayesian approach

With the exception of the robust Bayesian interpretation of models outlined to interpret convex sets of models (Section 3.1), so far we conducted our analysis in a classic Waldean setting where the DM's beliefs over the likelihood of models, in particular their quantification via prior probabilities, play no role. In contrast, in this section we outline a Bayesian approach based on them.

Under model ambiguity, the DM has a prior probability μ_Q over the set of structured models Q. In particular, the prior probability $\mu_Q(q)$ of a structured model $q \in Q$ quantifies the DM belief that q is the correct model. Under model misspecification, this interpretation is no longer possible because the DM no longer regards structured models as alternative correct models, one of them being correct. So, they no longer form an exhaustive collection of mutually exclusive uncertain alternatives – a logical partition – over which a meaningful belief can be expressed.

We thus face two possibilities. The first one is to content ourselves with the interpretation of the prior μ_Q as an averaging device which specifies the quasi-arithmetic aggregator in (39). The second, better, one is to find a meaningful logical partition that gives μ_Q a proper Bayesian interpretation. In both cases, one could argue that the prior μ_Q, compared to the standard case, might quantify a fragile belief which might need to be robustified (cf. Hansen and Sargent, 2007). We begin by introducing a Bayesian criterion under the average view of the prior μ_Q. We then discuss a possible interpretation of this prior that gives the criterion a genuine Bayesian flavor.

8.1 A Bayesian criterion

Consider the quasi-arithmetic specification (39), that is,

$$V_Q(f) = \phi_Q^{-1}\left(\int_Q \phi_Q\left(\min_{p\in\Delta}\left\{\int_S u(f(s))\,dp(s) + c(p,q)\right\}\right) d\mu_Q(q)\right) \tag{40}$$

where $\mu_Q \in \Delta(Q)$ and $\phi_Q : \mathbb{R} \to \mathbb{R}$ is strictly increasing and continuous. This is, formally, a Bayesian criterion with the prior probability μ_Q interpreted as an averaging device over the structured models. The variational criteria, indexed by Q,

$$\min_{p\in\Delta}\left\{\int_S u\left(f\left(s\right)\right)dq\left(s\right)+c\left(p,q\right)\right\}$$

and the corresponding dominance relation account for fear of model misspecification about the posited models q, while the function ϕ_Q addresses the fear of prior misspecification.

The Bayesian criterion (40) generalizes to model misspecification the smooth ambiguity criterion

$$V_Q\left(f\right)=\phi_Q^{-1}\left(\int_Q \phi_Q\left(\int_S u\left(f\left(s\right)\right)dq\left(s\right)\right)d\mu_Q\left(q\right)\right)$$

under model ambiguity, which is the special case $c\left(p,q\right)=\delta_{\{q\}}\left(p\right)$ for all $\left(p,q\right)\in\Delta\times\Delta^\sigma$ that corresponds to model misspecification neutrality. When each ϕ_Q is the identity, the criterion (40) further specializes to a standard subjective expected utility criterion

$$V_Q\left(f\right)=\int_Q\left(\int_S u\left(f\left(s\right)\right)dq\left(s\right)\right)d\mu_Q\left(q\right)$$

An important entropic specification of criterion (40) is

$$V_Q^{\lambda,\xi}\left(f\right)=\phi_\xi^{-1}\left(\int_Q \phi_\xi\left(\min_{p\in\Delta}\left\{\int_S u\left(f\left(s\right)\right)dp\left(s\right)+\lambda R\left(p||q\right)\right\}\right)d\mu_Q\left(q\right)\right) \tag{41}$$

where $\mu_Q\in\Delta\left(Q\right)$ and $\phi_\xi\left(t\right)=-e^{-\frac{1}{\xi}t}$ has an exponential form (common across the sets of models Q). The parameter $\xi>0$ captures fear of (reference) prior misspecification, while the parameter $\lambda>0$ is a fear of model misspecification index. The lower the parameter, the higher the fear. Next we show that, as fear of either model or prior misspecification vanishes or explodes, we get the criteria that one would expect. This provides an analytical consistency check for criterion (40). In deriving, this result we focus on the entropic formulation, but the result can be generalized in different directions, for example, by replacing the relative entropy with a general divergence as in (6) and by replacing the conditions on ξ with similar conditions on the Arrow-Pratt index of ϕ_Q.

Proposition 11 *If* $\operatorname{supp}\mu_Q=Q$ *and* $f\in\mathcal{F}$, *then*

$$\lim_{\xi\to 0^+}V_Q^{\lambda,\xi}\left(f\right)=\min_{p\in\Delta}\left\{\int_S u\left(f\left(s\right)\right)dp\left(s\right)+\lambda\min_{q\in Q}R\left(p||q\right)\right\}\quad\forall\lambda\in\left(0,\infty\right] \tag{42}$$

and

$$\lim_{\xi\to\infty}V_Q^{\lambda,\xi}\left(f\right)=\int_Q\left(\min_{p\in\Delta}\left\{\int_S u\left(f\left(s\right)\right)dp\left(s\right)+\lambda R\left(p||q\right)\right\}\right)d\mu_Q\left(q\right)\quad\forall\lambda\in\left(0,\infty\right] \tag{43}$$

Moreover,

$$\lim_{\xi \to \infty} \lim_{\lambda \to \infty} V_Q^{\lambda,\xi}(f) = \lim_{\lambda \to \infty} \lim_{\xi \to \infty} V_Q^{\lambda,\xi}(f) = \int_Q \left(\int_S u(f(s))\, dq(s) \right) d\mu_Q(q) \tag{44}$$

In words, the limit (42) shows that, as fear of prior misspecification explodes, criterion (40) gets closer and closer to our criterion (19). In contrast, the limit (43) shows that, when such fear vanishes, we end up with a criterion that averages, via the prior μ_Q, multiplier criteria (one per structured model q). Finally, the limit (44) shows that, when both fear vanish, at the limit we have a standard subjective expected utility criterion.

8.2 On the interpretation of priors

As we previously remarked, under model misspecification a set Q of structured models is no longer a set of exhaustive and mutually exclusive alternatives, so a logical partition upon which to define a prior probability. What might be a new partition of this kind?

To address this question, denote by $p^* \in \Delta$ the correct model. The agents do not know whether or not it belongs to Q. Let q^* be the structured model, assumed to uniquely exist, such that

$$c(p^*, q^*) = \min_{q \in Q} c(p^*, q)$$

Model q^* best approximates, or best fits, the correct model p^* according to the variational statistical distance c that decision makers adopt. If they know that p^* is in Q (model ambiguity), we have $p^* = q^*$ and so q^* itself is the correct model.

Decision makers are uncertain about q^*, that is, about which structured model $q \in Q$ best fits the correct model. But, they know that one of them is, indeed, the best fit. Under this interpretation of its elements, Q thus forms a collection of exhaustive and mutually exclusive alternatives. Decision makers now regard each element q of Q as a "candidate best fitting model": this is how they interpret q and what they are uncertain about. The meaning of prior $\mu_Q(q)$ is then clear: it quantifies the DM belief that q is the best fit of the correct model (see Walker, 2013, for an insightful discussion).

This interpretation of μ_Q reduces to the standard one under model ambiguity because, as previously remarked, in this case the best fit coincides with the correct model itself. In the rest of the section, we make more rigorous this discussion.

To this end, let $c : \Delta \times \Delta^\sigma \to [0, \infty]$ be a lower semicontinuous statistical distance. Consider a compact set of structured models Q. For each $q \in Q$, define the set

$$B_c(q, Q) = \left\{ p \in \Delta : c(p, q) = \min_{\tilde{q} \in Q} c(p, \tilde{q}) \right\}$$

If decision makers believe that a structured model $q \in Q$ best fits the correct model, then they consistently believe that the correct model belongs to a subset $B_c(q,Q)$ of Δ. So, for a structured model $q \in Q$, the set $B_c(q,Q)$ consists of all unstructured models that, if correct, make q their best fit.

We can thus regard $B_c(q,Q)$ as the partial identification set that corresponds to the agents belief that the structured model q best fits the correct one. They can construct, at least in principle, this set by solving the minimization problem that it features. Next we report some basic properties of these partial identification sets. Let

$$\Delta_{c,Q} = \{p \in \Delta : c(p,\cdot) \text{ is proper and strictly convex on } Q\}$$

For instance, by Lemma 13 we have $\Delta_{D_\phi,Q} \supseteq \{p \in \Delta^\sigma : p \sim Q\}$ for a ϕ-divergence D_ϕ when $Q \in \mathcal{K}$.

Lemma 2 *If $Q \in \mathcal{K}$, then*

(i) $\Delta = \bigcup_{q \in Q} B_c(q,Q)$;

(ii) $B_c(q,Q) \cap Q = \{q\}$ *for all* $q \in Q$;

(iii) $B_c(q,Q) \cap B_c(q',Q) \cap (Q \cup \Delta_{c,Q}) = \emptyset$ *for all distinct* $q, q' \in Q$.

Properties (i) and (iii) ensure that the family of the partial identification sets

$$\{B_c(q,Q)\}_{q \in Q}$$

forms a partition of $Q \cup \Delta_{c,Q}$. As long as the correct model belongs to $Q \cup \Delta_{c,Q}$, this permits to interpret $\mu_Q(q)$ as the probability that the structured model $q \in Q$ is the best fit of the correct model. In particular, we can interpret in this way the prior μ_Q that the Bayesian criterion (40) features, thus giving this criterion a genuine Bayesian status. Property (ii) ensures that under model ambiguity we go back to the traditional interpretation of priors.

9 Conclusion

Quantitative researchers use models to enhance their understanding of economic phenomena and to make policy assessments. In essence, each model tells its own quantitative story. We refer to such models as "structured models." Typically, there are more than just one such type of model, with each giving rise to a different quantitative story. Statistical and economic decision theories have addressed how best to confront the ambiguity among structured models. Such structured models are, by their very nature, misspecified. Nevertheless, the decision maker seeks to use such models in sensible ways. This problem is well recognized by applied researchers, but it is typically not part of formal decision theory. In this paper, we extend decision theory to confront model misspecification concerns. In so doing, we recover a variational representation of preferences that includes penalization based on discrepancy measures between "unstructured alternatives" and the set of structured probability models.

In terms of future research, a natural generalization of our criterion is

$$V(f) = \min_{p \in \Delta} \left\{ \int u(f)\, dp + C(p, Q) \right\}$$

where C is a general statistical set distance, not necessarily Hausdorff (so not necessarily characterized by an underlying statistical distance). This variational criterion still represents a preference that is more uncertainty averse than the corresponding max-min one. It may also easily accommodate reversals of the inequality (35), along the lines previously discussed. Though the analysis of this general criterion is beyond the scope of this paper and left for future research, we close our exposition with it as its form should help to put our exercise in a final perspective.

A Proofs and related analysis

In this appendix, we provide the proofs of our main results. We relegate to the Appendix B the proofs of most of our ancillary results (e.g., Propositions 1, 2, 8 and 9). In the same appendix, we also formally discuss few results about statistical distances and divergences (Lemmas 11–13). Appendix A.1 contains the proofs of our representation results (Theorems 1 and 2, and Proposition 7). Appendix A.2 contains the proofs of the remaining analysis. In both appendices, we denote by $B_0(\Sigma)$ the space of Σ-measurable simple functions $\varphi : S \to \mathbb{R}$, endowed with the supnorm $\| \|_\infty$. The dual of $B_0(\Sigma)$ can be identified with the space $ba(\Sigma)$ of all bounded finitely additive measures on (S, Σ).

A.1 Representation results

The proof of Theorem 1 is based on three key steps. We first provide two results regarding variational preferences which will help isolate the set of structured models Q in the main representation (their routine proof is confined to Appendix B). Second, we provide a representation for an unbounded and objectively Q-coherent dominance relation $\succsim^*$ (Appendix A.1.1). Third, we prove Theorem 1 (Appendix A.1.2). The proof of Theorem 2 and Proposition 7 instead is presented as one result (Appendix A.1.3). In what follows, given a function $c : \Delta \times Q \to [0, \infty]$, where Q is a compact and convex subset of Δ^σ, we say that c is *variational* if c_q is grounded, lower semicontinuous and convex and $c_Q (= \min_{q \in Q} c(\cdot, q))$ is well defined, grounded, lower semicontinuous and convex. The next two lemmas, proved in Appendix B.3, are key in characterizing subjective and objective Q-coherence.

Lemma 3 *Let $\succsim$ be a variational preference represented by $V : \mathcal{F} \to \mathbb{R}$ defined by*

$$V(f) = \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p) \right\} \qquad \forall f \in \mathcal{F}$$

and let $\bar{p} \in \Delta$. If $\succsim$ is unbounded, then the following conditions are equivalent:

(i) $c(\bar{p}) = 0$;

(ii) $x_f^{\bar{p}} \succsim f$ *for all* $f \in \mathcal{F}$;

(iii) for each $f \in \mathcal{F}$ and for each $x \in X$

$$x \succ x_f^{\bar{p}} \implies x \succ f$$

Lemma 4 *Let $\succsim$ be a variational preference represented by $V : \mathcal{F} \to \mathbb{R}$ defined by*

$$V(f) = \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p) \right\} \qquad \forall f \in \mathcal{F}$$

If $\succsim$ is unbounded, then the following conditions are equivalent:

(i) For each $f, g \in \mathcal{F}$

$$f \overset{Q}{=} g \implies f \sim g$$

(ii) $\operatorname{dom} c \subseteq \Delta^{\ll}(Q)$.

A.1.1 A Bewley-type representation

The next result is a multi-utility (variational) representation for unbounded dominance relations.

Lemma 5 *Let $\succsim^*$ be a binary relation on $\mathcal{F}$, where (S, Σ) is a standard Borel space. The following statements are equivalent:*

(i) $\succsim^*$ is an unbounded dominance relation which satisfies objective Q-coherence;

(ii) there exist an onto affine function $u : X \to \mathbb{R}$ and a variational $c : \Delta \times Q \to [0, \infty]$ such that $\operatorname{dom} c(\cdot, q) \subseteq \Delta^{\ll}(Q)$ for all $q \in Q$ and

$$f \succsim^* g \iff \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p, q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + c(p, q) \right\} \quad \forall q \in Q \quad (45)$$

To prove this result, we need to introduce one mathematical object. Let $\succeq^*$ be a binary relation on $B_0(\Sigma)$. We say that $\succeq^*$ is *convex niveloidal* if and only if $\succeq^*$ is a preorder that satisfies the following five properties:

1. For each $\varphi, \psi \in B_0(\Sigma)$ and for each $k \in \mathbb{R}$
$$\varphi \succeq^* \psi \implies \varphi + k \succeq^* \psi + k$$
2. If $\varphi, \psi \in B_0(\Sigma)$ and $\{k_n\}_{n\in\mathbb{N}} \subseteq \mathbb{R}$ are such that $k_n \uparrow k$ and $\varphi - k_n \succeq^* \psi$ for all $n \in \mathbb{N}$, then $\varphi - k \succeq^* \psi$;
3. For each $\varphi, \psi \in B_0(\Sigma)$
$$\varphi \geq \psi \implies \varphi \succeq^* \psi$$
4. For each $k, h \in \mathbb{R}$ and for each $\varphi \in B_0(\Sigma)$
$$k > h \implies \varphi + k \succ^* \varphi + h$$
5. For each $\varphi, \psi, \xi \in B_0(\Sigma)$ and for each $\lambda \in (0,1)$
$$\varphi \succeq^* \xi \text{ and } \psi \succeq^* \xi \implies \lambda\varphi + (1-\lambda)\psi \succeq^* \xi$$

Lemma 6 *If $\succsim^*$ is an unbounded dominance relation, then there exists an onto affine function $u : X \to \mathbb{R}$ such that*
$$x \succsim^* y \iff u(x) \geq u(y) \tag{46}$$

Proof Since $\succsim^*$ is a non-trivial preorder on $\mathcal{F}$ that satisfies c-completeness, continuity and weak c-independence, it is immediate to conclude that $\succsim^*$ restricted to X satisfies weak order, continuity and risk independence.[39] By Herstein and Milnor (1953), it follows that there exists an affine function $u : X \to \mathbb{R}$ that satisfies (46). Since $\succsim^*$ is a non-trivial c-complete preorder on $\mathcal{F}$ that satisfies monotonicity, we have that $\succsim^*$ is non-trivial on X. By Lemma 59 of Cerreia-Vioglio et al. (2011b) and since $\succsim^*$ is non-trivial on X and satisfies unboundedness, we can conclude that u is onto. ■

[39] To prove that $\succsim^*$ satisfies risk independence, it suffices to deploy the same technique of Lemma 28 of Maccheroni et al. (2006) and observe that $\succsim^*$ is a complete preorder on X. This yields that
$$x \sim^* y \implies \frac{1}{2}x + \frac{1}{2}z \sim^* \frac{1}{2}y + \frac{1}{2}z \quad \forall z \in X$$
By Theorem 2 of Herstein and Milnor (1953) and since $\succsim^*$ satisfies continuity, we can conclude that $\succsim^*$ satisfies risk independence.

Since u is affine and onto, note that $\{u(f) : f \in \mathcal{F}\} = B_0(\Sigma)$. In light of this observation, we can define a binary relation $\succeq^*$ on $B_0(\Sigma)$ by

$$\varphi \succeq^* \psi \iff f \succsim^* g \text{ where } u(f) = \varphi \text{ and } u(g) = \psi \tag{47}$$

Lemma 7 *If $\succsim^*$ is an unbounded dominance relation, then $\succeq^*$, defined as in (47), is a well defined convex niveloidal binary relation. Moreover, if $\succsim^*$ is objectively Q-coherent, then $\varphi \overset{Q}{=} \psi$ implies $\varphi \sim^* \psi$.*

We confine the routine proof to Appendix B. The next three results (Lemmas 8 and 9 as well as Proposition 12) will help us representing $\succeq^*$. This paired with Lemma 6 and Proposition 13 will yield the proof of Lemma 5.

Lemma 8 *Let $\succeq^*$ be a convex niveloidal binary relation. If $\psi \in B_0(\Sigma)$, then $U(\psi) = \{\varphi \in B_0(\Sigma) : \varphi \succeq^* \psi\}$ is a non-empty convex set such that:*

1. *$\psi \in U(\psi)$;*

2. *if $\varphi \in B_0(\Sigma)$ and $\{k_n\}_{n\in\mathbb{N}} \subseteq \mathbb{R}$ are such that $k_n \uparrow k$ and $\varphi - k_n \in U(\psi)$ for all $n \in \mathbb{N}$, then $\varphi - k \in U(\psi)$;*

3. *if $k > 0$, then $\psi - k \notin U(\psi)$;*

4. *if $\varphi_1 \geq \varphi_2$ and $\varphi_2 \in U(\psi)$, then $\varphi_1 \in U(\psi)$;*

5. *if $k \geq 0$ and $\varphi_2 \in U(\psi)$, then $\varphi_2 + k \in U(\psi)$.*

Proof Since $\succeq^*$ is reflexive, we have that $\psi \in U(\psi)$, proving that $U(\psi)$ is non-empty and point 1. Consider $\varphi_1, \varphi_2 \in U(\psi)$ and $\lambda \in (0,1)$. By definition, we have that $\varphi_1 \succeq^* \psi$ and $\varphi_2 \succeq^* \psi$. Since $\succeq^*$ satisfies convexity, we have that $\lambda\varphi_1 + (1-\lambda)\varphi_2 \succeq^* \psi$, proving convexity of $U(\psi)$. Consider $\varphi \in B_0(\Sigma)$ and $\{k_n\}_{n\in\mathbb{N}} \subseteq \mathbb{R}$ such that $k_n \uparrow k$ and $\varphi - k_n \in U(\psi)$ for all $n \in \mathbb{N}$. It follows that $\varphi - k_n \succeq^* \psi$ for all $n \in \mathbb{N}$, then $\varphi - k \succeq^* \psi$, that is, $\varphi - k \in U(\psi)$, proving point 2. If $k > 0$, then $0 > -k$ and $\psi = \psi + 0 \succ^* \psi - k$, that is, $\psi - k \notin U(\psi)$, proving point 3. Consider $\varphi_1 \geq \varphi_2$ such that $\varphi_2 \in U(\psi)$, then $\varphi_1 \succeq^* \varphi_2$ and $\varphi_2 \succeq^* \psi$, yielding that $\varphi_1 \succeq^* \psi$ and, in particular, $\varphi_1 \in U(\psi)$, proving point 4. Finally, to prove point 5, it is enough to set $\varphi_1 = \varphi_2 + k$ in point 4. ∎

Before stating the next result, we define few properties that will turn out to be useful later on. A functional $I : B_0(\Sigma) \to \mathbb{R}$ is:

1. a niveloid if $I(\varphi) - I(\psi) \leq \sup_{s\in S}(\varphi(s) - \psi(s))$ for all $\varphi, \psi \in B_0(\Sigma)$;

2. normalized if $I(k) = k$ for all $k \in \mathbb{R}$;[40]

[40] With the usual abuse of notation, we denote by k both the real number and the constant function taking value k.

3. monotone if for each $\varphi, \psi \in B_0(\Sigma)$

$$\varphi \geq \psi \implies I(\varphi) \geq I(\psi)$$

4. $\succeq^*$-consistent if for each $\varphi, \psi \in B_0(\Sigma)$

$$\varphi \succeq^* \psi \implies I(\varphi) \geq I(\psi)$$

5. concave if for each $\varphi, \psi \in B_0(\Sigma)$ and $\lambda \in (0,1)$

$$I(\lambda\varphi + (1-\lambda)\psi) \geq \lambda I(\varphi) + (1-\lambda) I(\psi)$$

6. translation invariant if for each $\varphi \in B_0(\Sigma)$ and $k \in \mathbb{R}$

$$I(\varphi + k) = I(\varphi) + k$$

Lemma 9 *Let $\succeq^*$ be a convex niveloidal binary relation. If $\psi \in B_0(\Sigma)$, then the functional $I_\psi : B_0(\Sigma) \to \mathbb{R}$, defined by*

$$I_\psi(\varphi) = \max\{k \in \mathbb{R} : \varphi - k \in U(\psi)\} \qquad \forall \varphi \in B_0(\Sigma)$$

is a concave niveloid which is $\succeq^*$-consistent and such that $I_\psi(\psi) = 0$. Moreover, we have that:

1. *The functional $\bar{I}_\psi = I_\psi - I_\psi(0)$ is a normalized concave niveloid which is $\succeq^*$-consistent.*
2. *If $\succeq^*$ satisfies*
$$\psi \overset{Q}{=} \psi' \implies \psi \sim^* \psi'$$
then
$$\psi \overset{Q}{=} \psi' \implies I_\psi = I_{\psi'} \text{ and } \bar{I}_\psi = \bar{I}_{\psi'}$$

We confine the routine proof of the previous lemma to Appendix B.

Proposition 12 *Let $\succeq^*$ be a binary relation on $B_0(\Sigma)$. The following statements are equivalent:*

(i) *$\succeq^*$ is convex niveloidal;*

(ii) *there exists a family of concave niveloids $\{I_\alpha\}_{\alpha \in A}$ on $B_0(\Sigma)$ such that*

$$\varphi \succeq^* \psi \iff I_\alpha(\varphi) \geq I_\alpha(\psi) \qquad \forall \alpha \in A \tag{48}$$

(iii) there exists a family of normalized concave niveloids $\left\{\bar{I}_{\alpha}\right\}_{\alpha\in A}$ *on* $B_0(\Sigma)$ *such that*

$$\varphi \succeq^{*} \psi \iff \bar{I}_{\alpha}(\varphi) \geq \bar{I}_{\alpha}(\psi) \qquad \forall \alpha \in A \tag{49}$$

Proof (iii) implies (i). It is trivial.

(i) implies (ii). Let $A = B_0(\Sigma)$. We next show that

$$\varphi_1 \succeq^{*} \varphi_2 \iff I_{\psi}(\varphi_1) \geq I_{\psi}(\varphi_2) \qquad \forall \psi \in B_0(\Sigma)$$

where I_{ψ} is defined as in Lemma 9 for all $\psi \in B_0(\Sigma)$. By Lemma 9, we have that I_{ψ} is $\succeq^{*}$-consistent for all $\psi \in B_0(\Sigma)$. This implies that

$$\varphi_1 \succeq^{*} \varphi_2 \implies I_{\psi}(\varphi_1) \geq I_{\psi}(\varphi_2) \qquad \forall \psi \in B_0(\Sigma)$$

Vice versa, consider $\varphi_1, \varphi_2 \in B_0(\Sigma)$. Assume that $I_{\psi}(\varphi_1) \geq I_{\psi}(\varphi_2)$ for all $\psi \in B_0(\Sigma)$. Let $\psi = \varphi_2$. By Lemma 9, we have that $I_{\varphi_2}(\varphi_1) \geq I_{\varphi_2}(\varphi_2) = 0$, yielding that $\varphi_1 \geq \varphi_1 - I_{\varphi_2}(\varphi_1) \in U(\varphi_2)$. By point 4 of Lemma 8, this implies that $\varphi_1 \in U(\varphi_2)$, that is, $\varphi_1 \succeq^{*} \varphi_2$.

(ii) implies (iii). Given a family of concave niveloids $\{I_{\alpha}\}_{\alpha\in A}$, define $\bar{I}_{\alpha} = I_{\alpha} - I_{\alpha}(0)$ for all $\alpha \in A$. It is immediate to verify that $\bar{I}_{\alpha}$ is a normalized concave niveloid for all $\alpha \in A$. It is also immediate to observe that

$$I_{\alpha}(\varphi_1) \geq I_{\alpha}(\varphi_2) \quad \forall \alpha \in A \iff \bar{I}_{\alpha}(\varphi_1) \geq \bar{I}_{\alpha}(\varphi_2) \quad \forall \alpha \in A$$

proving the implication. ■

Remark 1 Given a convex niveloidal binary relation $\succeq^{*}$ on $B_0(\Sigma)$, we call *canonical* (resp., *canonical normalized*) the representation $\{I_{\psi}\}_{\psi\in B_0(\Sigma)}$ (resp., $\left\{\bar{I}_{\psi}\right\}_{\psi\in B_0(\Sigma)}$) obtained from Lemma 9 and the proof of Proposition 12. By the previous proof, clearly, $\{I_{\psi}\}_{\psi\in B_0(\Sigma)}$ and $\left\{\bar{I}_{\psi}\right\}_{\psi\in B_0(\Sigma)}$ satisfy (48) and (49) respectively.

The next result clarifies what is the relation between any representation of $\succeq^{*}$ and the canonical ones. This will be useful in establishing an extra property of $\left\{\bar{I}_{\psi}\right\}_{\psi\in B_0(\Sigma)}$ in Corollary 2.

Lemma 10 *Let* $\succeq^{*}$ *be a convex niveloidal binary relation. If* B *is an index set and* $\{J_{\beta}\}_{\beta\in B}$ *is a family of normalized concave niveloids such that*

$$\varphi \succeq^{*} \psi \iff J_{\beta}(\varphi) \geq J_{\beta}(\psi) \qquad \forall \beta \in B$$

then for each $\psi \in B_0(\Sigma)$

$$I_{\psi}(\varphi) = \inf_{\beta\in B}\left(J_{\beta}(\varphi) - J_{\beta}(\psi)\right) \qquad \forall \varphi \in B_0(\Sigma) \tag{50}$$

and

$$\bar{I}_\psi(\varphi) = \inf_{\beta \in B} \left(J_\beta(\varphi) - J_\beta(\psi) \right) + \sup_{\beta \in B} J_\beta(\psi) \qquad \forall \varphi \in B_0(\Sigma) \tag{51}$$

Proof Fix $\varphi \in B_0(\Sigma)$ and $\psi \in B_0(\Sigma)$. By definition, we have that

$$I_\psi(\varphi) = \max \{k \in \mathbb{R} : \varphi - k \in U(\psi)\}$$

Since $\{J_\beta\}_{\beta \in B}$ represents $\succeq^*$ and each J_β is translation invariant, note that for each $k \in \mathbb{R}$

$$\begin{aligned} \varphi - k \in U(\psi) &\iff \varphi - k \succeq^* \psi \iff J_\beta(\varphi - k) \geq J_\beta(\psi) \quad \forall \beta \in B \\ &\iff J_\beta(\varphi) - k \geq J_\beta(\psi) \quad \forall \beta \in B \iff J_\beta(\varphi) - J_\beta(\psi) \geq k \quad \forall \beta \in B \\ &\iff \inf_{\beta \in B} \left(J_\beta(\varphi) - J_\beta(\psi) \right) \geq k \end{aligned}$$

By definition of I_ψ and since $\varphi - I_\psi(\varphi) \in U(\psi)$, this implies that $I_\psi(\varphi) = \inf_{\beta \in B} (J_\beta(\varphi) - J_\beta(\psi))$. Since φ and ψ were arbitrarily chosen, (50) follows. Since $\bar{I}_\psi = I_\psi - I_\psi(0)$, we only need to compute $-I_\psi(0)$. Since each J_β is normalized, we have that $-I_\psi(0) = -\inf_{\beta \in B} (J_\beta(0) - J_\beta(\psi)) = -\inf_{\beta \in B} (-J_\beta(\psi)) = \sup_{\beta \in B} J_\beta(\psi)$, proving (51). ■

Corollary 2 *If $\succeq^*$ is a convex niveloidal binary relation, then $\bar{I}_0 \leq \bar{I}_\psi$ for all $\psi \in B_0(\Sigma)$.*

Proof By Lemma 10 and Remark 1 and since each $\bar{I}_{\psi'}$ is a normalized concave niveloid, we have that

$$\bar{I}_0(\varphi) = \inf_{\psi' \in B_0(\Sigma)} \left(\bar{I}_{\psi'}(\varphi) - \bar{I}_{\psi'}(0) \right) + \sup_{\psi' \in B_0(\Sigma)} \bar{I}_{\psi'}(0) = \inf_{\psi' \in B_0(\Sigma)} \bar{I}_{\psi'}(\varphi) \leq \bar{I}_\psi(\varphi) \quad \forall \varphi \in B_0(\Sigma)$$

for all $\psi \in B_0(\Sigma)$, proving the statement. ■

The next result will be instrumental in providing a niveloidal multi-representation of $\succsim^*$ when $|Q| \geq 2$. In order to discuss it, we need a piece of terminology. We denote by V the quotient space $B_0(\Sigma)/M$ where M is the vector subspace $\left\{ \varphi \in B_0(\Sigma) : \varphi \overset{Q}{=} 0 \right\}$. Recall that the elements of V are equivalence classes $[\psi]$ with $\psi \in B_0(\Sigma)$ where $\psi', \psi'' \in [\psi]$ if and only if $\psi \overset{Q}{=} \psi' \overset{Q}{=} \psi''$. Recall that Q is convex.

Proposition 13 *If (S, Σ) is a standard Borel space and $|Q| \geq 2$, then there exists a bijection $f : V \to Q$.*

The routine proof of the previous result is relegated to Appendix B. We next prove our representation result for incomplete variational preferences.

Proof of Lemma 5 (ii) implies (i). It is trivial.

(i) implies (ii). Since $\succsim^*$ is a dominance relation, if $|Q| = 1$, that is $Q = \{\bar{q}\}$, then $\succsim^*$ is complete. By Maccheroni et al. (2006) and since $\succsim^*$ is unbounded, it follows that there exists an onto and affine $u : X \to \mathbb{R}$ and a grounded, lower semicontinuous and convex $c_{\bar{q}} : \Delta \to [0, \infty]$ such that $V : \mathcal{F} \to \mathbb{R}$ defined by

$$V(f) = \min_{p \in \Delta} \left\{ \int u(f) \, dp + c_{\bar{q}}(p) \right\} \qquad \forall f \in \mathcal{F}$$

represents $\succsim^*$. If we define $c : \Delta \times Q \to [0, \infty]$ by $c(p, q) = c_{\bar{q}}(p)$ for all $(p, q) \in \Delta \times Q$, then we have that c is variational. By Lemma 4 and since $\succsim^*$ is objectively Q-coherent, it follows that $\operatorname{dom} c(\cdot, q) \subseteq \Delta^{\ll}(Q)$ for all $q \in Q$, proving the implication. Assume $|Q| > 1$. By Lemma 6, there exists an onto affine function $u : X \to \mathbb{R}$ which represents $\succsim^*$ on X. By Lemma 7, this implies that we can consider the convex niveloidal binary relation $\succeq^*$ defined as in (47). By definition of $\succeq^*$ and Proposition 12 (and Remark 1), we have that

$$f \succsim^* g \iff u(f) \succeq^* u(g) \iff \bar{I}_\psi(u(f)) \geq \bar{I}_\psi(u(g)) \quad \forall \psi \in B_0(\Sigma)$$

where each $\bar{I}_\psi$ is a normalized concave niveloid. As before, consider $V = B_0(\Sigma)/M$ where M is the vector subspace $\left\{ \varphi \in B_0(\Sigma) : \varphi \overset{Q}{=} 0 \right\}$. For each equivalence class $[\psi]$, select exactly one $\psi' \in B_0(\Sigma)$ such that $\psi' \in [\psi]$. In particular, let $\psi' = 0$ when $[\psi] = [0]$. We denote this subset of $B_0(\Sigma)$ by $\tilde{V}$. Clearly, we have that

$$\bar{I}_\psi(u(f)) \geq \bar{I}_\psi(u(g)) \quad \forall \psi \in B_0(\Sigma) \implies \bar{I}_\psi(u(f)) \geq \bar{I}_\psi(u(g)) \quad \forall \psi \in \tilde{V}$$

Vice versa, assume that $\bar{I}_\psi(u(f)) \geq \bar{I}_\psi(u(g))$ for all $\psi \in \tilde{V}$. Consider $\hat{\psi} \in B_0(\Sigma)$. It follows that there exists $[\psi]$ in V such that $\hat{\psi} \in [\psi]$. Similarly, consider $\psi' \in \tilde{V}$ such that $\psi' \in [\psi]$. It follows that $\hat{\psi} \overset{Q}{=} \psi'$. By Lemmas 7 and 9 and since $\succsim^*$ is objectively Q-coherent, then $\bar{I}_{\hat{\psi}} = \bar{I}_{\psi'}$, yielding that $\bar{I}_{\hat{\psi}}(u(f)) \geq \bar{I}_{\hat{\psi}}(u(g))$. Since $\hat{\psi}$ was arbitrarily chosen $\bar{I}_\psi(u(f)) \geq \bar{I}_\psi(u(g))$ for all $\psi \in B_0(\Sigma)$. By construction, observe that there exists a bijection $\tilde{f} : \tilde{V} \to V$. By Proposition 13, we have that there exists a bijection $f : V \to Q$. Define $\bar{f} = f \circ \tilde{f}$. By Corollary 2, if we define $\hat{I}_q = \bar{I}_{\bar{f}^{-1}(q)}$ for all $q \in Q$, then we have that $\hat{I}_{\bar{f}(0)} \leq \hat{I}_q$ for all $q \in Q$ and

$$\begin{aligned} f \succsim^* g &\iff \bar{I}_\psi(u(f)) \geq \bar{I}_\psi(u(g)) \quad \forall \psi \in B_0(\Sigma) \iff \bar{I}_\psi(u(f)) \geq \bar{I}_\psi(u(g)) \quad \forall \psi \in \tilde{V} \\ &\iff \hat{I}_q(u(f)) \geq \hat{I}_q(u(g)) \quad \forall q \in Q \end{aligned}$$

Since each $\hat{I}_q$ is a normalized concave niveloid, we have that for each $q \in Q$ there exists a function $c_q : \Delta \to [0, \infty]$ which is grounded, lower semicontinuous, convex and such that

$$\hat{I}_q(\varphi) = \min_{p \in \Delta} \left\{ \int \varphi dp + c_q(p) \right\} \qquad \forall \varphi \in B_0(\Sigma)$$

Define $c : \Delta \times Q \to [0, \infty]$ by $c(p, q) = c_q(p)$ for all $(p, q) \in \Delta \times Q$. Clearly, the q-sections of c are grounded, lower semicontinuous and convex and (45) holds. By Lemma 4 and (45) and since $\succsim^*$ is objectively Q-coherent, it follows that $\operatorname{dom} c(\cdot, q) \subseteq \Delta^{\ll}(Q)$ for all $q \in Q$. Finally, recall that

$$c(p,q) = \sup_{\varphi \in B_0(\Sigma)} \left\{ \hat{I}_q(\varphi) - \int \varphi dp \right\} \qquad \forall p \in \Delta, \forall q \in Q$$

Since $\hat{I}_{\bar{f}(0)} \leq \hat{I}_q$ for all $q \in Q$, we have that for each $q \in Q$

$$c\left(p, \bar{f}(0)\right) = \sup_{\varphi \in B_0(\Sigma)} \left\{ \hat{I}_{\bar{f}(0)}(\varphi) - \int \varphi dp \right\} \leq \sup_{\varphi \in B_0(\Sigma)} \left\{ \hat{I}_q(\varphi) - \int \varphi dp \right\} = c(p,q) \quad \forall p \in \Delta$$

Since $c\left(\cdot, \bar{f}(0)\right)$ is grounded, lower semicontinuous and convex and $\bar{f}(0) \in Q$, this implies that $c_Q(\cdot) = \min_{q \in Q} c(\cdot, q) = c\left(\cdot, \bar{f}(0)\right)$ is well defined and shares the same properties, proving that c is variational. ■

A.1.2 Proof of Theorem 1

(i) implies (ii). We proceed by steps. Before starting, we make one observation. By Lemma 5 and since $\succsim^*$ is an unbounded dominance relation which is objectively Q-coherent there exist an onto affine function $u : X \to \mathbb{R}$ and a variational $c : \Delta \times Q \to [0, \infty]$ such that $\operatorname{dom} c(\cdot, q) \subseteq \Delta^{\ll}(Q)$ for all $q \in Q$ (in particular, $\operatorname{dom} c_Q(\cdot) \subseteq \cup_{q \in Q} \operatorname{dom} c(\cdot, q) \subseteq \Delta^{\ll}(Q)$) and

$$f \succsim^* g \iff \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p,q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + c(p,q) \right\} \quad \forall q \in Q$$

We are left to show that $c_Q : \Delta \to [0, \infty]$ is such that

$$f \succsim g \iff \min_{p \in \Delta} \left\{ \int u(f)\, dp + c_Q(p) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + c_Q(p) \right\} \tag{52}$$

and $c_Q^{-1}(0) = Q$. To prove this we consider c as in the proof of (i) implies (ii) of Lemma 5. This covers both cases $|Q| = 1$ and $|Q| > 1$. In particular, for each $q \in Q$ define $\hat{I}_q : B_0(\Sigma) \to \mathbb{R}$ by

$$\hat{I}_q(\varphi) = \min_{p \in \Delta} \left\{ \int \varphi dp + c(p,q) \right\} \qquad \forall \varphi \in B_0(\Sigma)$$

and recall that there exists $\hat{q}(= \bar{f}(0) \in Q$ when $|Q| > 1)$ such that $c(\cdot, \hat{q}) \leq c(\cdot, q)$, thus $\hat{I}_{\hat{q}} \leq \hat{I}_q$, for all $q \in Q$.

Step 1. $\succsim$ agrees with $\succsim^*$ on X. In particular, $u : X \to \mathbb{R}$ represents $\succsim^*$ and $\succsim$.

Proof of the Step Note that $\succsim^*$ and $\succsim$ restricted to X are continuous weak orders that satisfy risk independence. Moreover, by the observation above, $\succsim^*$ is represented by u. By Herstein and Milnor (1953) and since $\succsim$ is non-trivial, it follows that there exists a non-constant and affine function $v : X \to \mathbb{R}$ that represents $\succsim$ on X. Since $(\succsim^*, \succsim)$ jointly satisfy consistency, it follows that for each $x, y \in X$

$$u(x) \geq u(y) \Longrightarrow v(x) \geq v(y)$$

By Corollary B.3 of Ghirardato et al. (2004), u and v are equal up to an affine and positive transformation, hence the statement. We can set $v = u$. $\square$

Step 2. There exists a normalized, monotone and continuous functional $I : B_0(\Sigma) \to \mathbb{R}$ *such that*

$$f \succsim g \Longleftrightarrow I(u(f)) \geq I(u(g))$$

Proof of the Step By Cerreia-Vioglio et al. (2011a) and since $\succsim$ is a rational preference relation, the statement follows. $\square$

Step 3. $I(\varphi) \leq \inf_{q \in Q} \hat{I}_q(\varphi)$ *for all* $\varphi \in B_0(\Sigma)$.

Proof of the Step Consider $\varphi \in B_0(\Sigma)$. Since each $\hat{I}_q$ is normalized and monotone and u is onto, we have that $\hat{I}_q(\varphi) \in [\inf_{s \in S} \varphi(s), \sup_{s \in S} \varphi(s)] \subseteq \operatorname{Im} u$ for all $q \in Q$. Since $\varphi \in B_0(\Sigma)$, it follows that there exists $f \in \mathcal{F}$ such that $\varphi = u(f)$ and $x \in X$ such that $u(x) = \inf_{q \in Q} \hat{I}_q(\varphi)$. For each $\varepsilon > 0$ there exists $x_\varepsilon \in X$ such that $u(x_\varepsilon) = u(x) + \varepsilon$. Since $\inf_{q \in Q} \hat{I}_q(\varphi) = u(x)$, it follows that for each $\varepsilon > 0$ there exists $q \in Q$ such that $\hat{I}_q(u(f)) = \hat{I}_q(\varphi) < u(x_\varepsilon) = \hat{I}_q(u(x_\varepsilon))$, yielding that $f \not\succsim^* x_\varepsilon$. Since $(\succsim^*, \succsim)$ jointly satisfy caution, we have that $x_\varepsilon \succsim f$ for all $\varepsilon > 0$. By Step 2, this implies that

$$u(x) + \varepsilon = u(x_\varepsilon) = I(u(x_\varepsilon)) \geq I(u(f)) = I(\varphi) \quad \forall \varepsilon > 0$$

that is, $\inf_{q \in Q} \hat{I}_q(\varphi) = u(x) \geq I(\varphi)$, proving the step. $\square$

Step 4. $I(\varphi) \geq \inf_{q \in Q} \hat{I}_q(\varphi)$ *for all* $\varphi \in B_0(\Sigma)$.

Proof of the Step Consider $\varphi \in B_0(\Sigma)$. We use the same objects and notation of Step 3. Note that for each $q' \in Q$

$$\hat{I}_{q'}(u(f)) = \hat{I}_{q'}(\varphi) \geq \inf_{q \in Q} \hat{I}_q(\varphi) = u(x) = \hat{I}_{q'}(u(x))$$

that is, $f \succsim^* x$. Since $(\succsim^*, \succsim)$ jointly satisfy consistency, we have that $f \succsim x$. By Step 2, this implies that

$$I(\varphi) = I(u(f)) \geq I(u(x)) = u(x) = \inf_{q \in Q} \hat{I}_q(\varphi)$$

proving the step. $\square$

Step 5. $I(\varphi) = \min_{p \in \Delta} \left\{ \int \varphi dp + c_Q(p) \right\}$ *for all* $\varphi \in B_0(\Sigma)$.

Proof of the Step By Steps 3 and 4 and since $\hat{I}_{\hat{q}} \leq \hat{I}_q$ for all $q \in Q$, we have that

$$I(\varphi) = \min_{q \in Q} \hat{I}_q(\varphi) = \hat{I}_{\hat{q}}(\varphi) \qquad \forall \varphi \in B_0(\Sigma)$$

Since $c(\cdot,\hat{q}) = c_Q(\cdot)$, it follows that for each $\varphi \in B_0(\Sigma)$

$$I(\varphi) = \hat{I}_{\hat{q}}(\varphi) = \min_{p\in\Delta}\left\{\int \varphi dp + c(p,\hat{q})\right\} = \min_{p\in\Delta}\left\{\int \varphi dp + c_Q(p)\right\}$$

proving the step. □

Step 6. $c_Q^{-1}(0) = Q$.

Proof of the Step By Steps 2 and 5, we have that $V : \mathcal{F} \to \mathbb{R}$ defined by

$$V(f) = \min_{p\in\Delta}\left\{\int u(f)\,dp + c_Q(p)\right\}$$

represents $\succsim$. By Lemma 3 and since $\succsim$ is subjectively Q-coherent and c_Q is well defined, grounded, lower semicontinuous and convex, we can conclude that $c_Q^{-1}(0) = Q$. □

Thus, (52) follows from Steps 2 and 5 while, by Step 6, $c_Q^{-1}(0) = Q$. This completes the proof.

(ii) implies (i). It is routine.

Next, assume that c is uniquely null. Define the correspondence $\Gamma : Q \rightrightarrows Q$ by

$$\Gamma(q) = \{p \in \Delta : c(p,q) = 0\} = \arg\min c_q$$

Since $c_Q \leq c_q$ for all $q \in Q$ and $c_Q^{-1}(0) = Q$, we have that Γ is well defined. Since c_q is grounded, it follows that $\Gamma(q) \neq \emptyset$ for all $q \in Q$. Since c is uniquely null and c_q is grounded, we have that $c_q^{-1}(0)$ is a singleton, that is,

$$c(p,q) = c(p',q) = 0 \Longrightarrow p = p'$$

This implies that $\Gamma(q)$ is a singleton, therefore Γ is a function. Since $c_Q^{-1}(0) = Q$, observe that

$$\cup_{q\in Q}\Gamma(q) = \cup_{q\in Q}\arg\min c_q = \arg\min c_Q = Q$$

that is, Γ is surjective. Since c is uniquely null, we have that $c_p^{-1}(0)$ is at most a singleton, that is,

$$c(p,q) = c(p,q') = 0 \implies q = q'$$

yielding that Γ is injective. To sum up, Γ is a bijection. Define $\tilde{c} : \Delta \times Q \to [0,\infty]$ by $\tilde{c}(p,q) = c(p,\Gamma^{-1}(q))$ for all $(p,q) \in \Delta\times Q$. Note that $\tilde{c}(\cdot,q)$ is grounded, lower semicontinuous, convex and $\operatorname{dom}\tilde{c}(\cdot,q) \subseteq \Delta^{\ll}(Q)$ for all $q \in Q$ and $\operatorname{dom}\tilde{c}_Q(\cdot) \subseteq \Delta^{\ll}(Q)$. Next, we show that $\tilde{c}_Q = c_Q$. Since c_Q is well defined, for each $p \in \Delta$ there exists $q_p \in Q$ such that

$$\tilde{c}(p,\Gamma(q_p)) = c(p,q_p) = \min_{q\in Q} c(p,q) \leq c(p,q') = \tilde{c}(p,\Gamma(q')) \quad \forall q' \in Q$$

Since Γ is a bijection, we have that $\tilde{c}(p, \Gamma(q_p)) \leq \tilde{c}(p,q)$ for all $q \in Q$. Since p was arbitrarily chosen, it follows that

$$c_Q(p) = \min_{q \in Q} c(p,q) = \tilde{c}(p, \Gamma(q_p)) = \min_{q \in Q} \tilde{c}(p,q) = \tilde{c}_Q(p) \quad \forall p \in \Delta$$

To sum up, $\tilde{c}_Q = c_Q$ and $\tilde{c}_Q^{-1}(0) = c_Q^{-1}(0) = Q$. In turn, since c_Q is grounded, lower semicontinuous and convex, this implies that $\tilde{c}_Q$ is grounded, lower semicontinuous and convex. Since Γ is a bijection, we can conclude that (15) holds with $\tilde{c}$ in place of c and (16) holds with $\tilde{c}_Q$ in place of c_Q.

We are left to show that $\tilde{c}(p,q) = 0$ if and only if $p = q$. Since $c_q^{-1}(0)$ is a singleton for all $q \in Q$ and Γ is a bijection, if $\tilde{c}(p,q) = 0$, then $c(p, \Gamma^{-1}(q)) = 0$, yielding that $p = \Gamma(\Gamma^{-1}(q)) = q$. On the other hand, $\tilde{c}(q,q) = c(q, \Gamma^{-1}(q)) = 0$. We can conclude that $\tilde{c}(p,q) = 0$ if and only if $p = q$, proving that $\tilde{c}$ is a statistical distance. ■

A.1.3 Proof of Theorem 2, Propositions 7 and 10

Proof of Theorem 2 We only prove (i) implies (ii), the converse being routine.[41] We proceed by steps.

Step 1. $\succsim_Q^*$ *agrees with* $\succsim_{Q'}^*$ *on* X *for all* $Q, Q' \in \mathcal{Q}$. *In particular, there exists an affine and onto function* $u : X \to \mathbb{R}$ *representing* $\succsim_Q^*$ *for all* $Q \in \mathcal{Q}$.

Proof of the Step Let $Q, Q' \in \mathcal{Q}$ be such that $Q \supseteq Q'$. Note that $\succsim_Q^*$ and $\succsim_{Q'}^*$, restricted to X, satisfy weak order, continuity and risk independence. By Herstein and Milnor (1953) and since $\succsim_Q^*$ and $\succsim_{Q'}^*$ are non-trivial, there exist two non-constant affine functions $u_Q, u_{Q'} : X \to \mathbb{R}$ which represent $\succsim_Q^*$ and $\succsim_{Q'}^*$, respectively. Since $\{\succsim_Q^*\}_{Q \in \mathcal{Q}}$ is monotone in model ambiguity, we have that

$$u_Q(x) \geq u_Q(y) \Longrightarrow u_{Q'}(x) \geq u_{Q'}(y)$$

By Corollary B.3 of Ghirardato et al. (2004), u_Q and $u_{Q'}$ are equal up to an affine and positive transformation. Next, fix $\bar{q} \in \Delta^\sigma$. Set $u = u_{\bar{q}}$. Given any other $q \in \Delta^\sigma$, consider $\bar{Q} \in \mathcal{Q}$ such that $\bar{Q} \supseteq \{\bar{q}, q\}$. By the previous part, it follows that $u_{\bar{Q}}$, u_q and $u_{\bar{q}}$ are equal up to an affine and positive transformation. Given that q was arbitrarily chosen, we can set $u = u_q$ for all $q \in Q$. Similarly, given a generic $Q \in \mathcal{Q}$, select $q \in Q$. Since $Q \supseteq \{q\}$, it follows that we can set $u = u_Q$. Since each $\succsim_Q^*$ is unbounded for all $Q \in \mathcal{Q}$, we have that u is onto. □

Step 2. For each $q \in \Delta^\sigma$ *there exists a normalized, monotone, translation invariant and concave functional* $I_q : B_0(\Sigma) \to \mathbb{R}$ *such that*

$$f \succsim_q^* g \Longleftrightarrow I_q(u(f)) \geq I_q(u(g)) \tag{53}$$

[41] The only exception is the proof that the representation implies subjective Q-coherence. This is a consequence of Theorem 2.4.18 in Zalinescu (2002) paired with Lemma 32 of Maccheroni et al. (2006).

Moreover, there exists a unique grounded, lower semicontinuous and convex function $c_q : \Delta \to [0, \infty]$ *such that*

$$I_q(\varphi) = \min_{p \in \Delta} \left\{ \int \varphi dp + c_q(p) \right\} \qquad \forall \varphi \in B_0(\Sigma) \tag{54}$$

Proof of the Step Fix $q \in \Delta^\sigma$. Since $\succsim_q^*$ is an unbounded dominance relation which is complete, we have that $\succsim_q^*$ is a variational preference. By the proof of Theorem 3 and Proposition 6 of Maccheroni et al. (2006) and Step 1, there exists an onto and affine function $u_q : X \to \mathbb{R}$, which can be set to be equal to u, and, given u, a unique grounded, lower semicontinuous and convex function $c_q : \Delta \to [0, \infty]$ such that (54) and (53) hold. □

Define $c : \Delta \times \Delta^\sigma \to [0, \infty]$ by $c(p, q) = c_q(p)$ for all $(p, q) \in \Delta \times \Delta^\sigma$.

Step 3. For each $Q \in \mathcal{Q}$ *we have that* $f \succsim_Q^* g$ *if and only if* $f \succsim_q^* g$ *for all* $q \in Q$*. In particular, we have that*

$$f \succsim_Q^* g \iff \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p, q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + c(p, q) \right\} \quad \forall q \in Q \tag{55}$$

Proof of the Step Fix $Q \in \mathcal{Q}$. Since $\left\{\succsim_Q^*\right\}_{Q \in \mathcal{Q}}$ is monotone in model ambiguity, we have that

$$f \succsim_Q^* g \implies f \succsim_q^* g \qquad \forall q \in Q$$

Since $\left\{\succsim_Q^*\right\}_{Q \in \mathcal{Q}}$ is Q-separable, we can conclude that $f \succsim_Q^* g$ if and only if $f \succsim_q^* g$ for all $q \in Q$. By Step 2 and the definition of c, (55) follows. □

Step 4. $\succsim_Q^*$ *agrees with* $\succsim_Q$ *on* X *for all* $Q \in \mathcal{Q}$*. Moreover,* $\succsim_Q$ *is represented by the function* u *of Step 1.*

Proof of the Step Fix $Q \in \mathcal{Q}$. Note that $\succsim_Q^*$ and $\succsim_Q$, restricted to X, satisfy weak order, continuity and risk independence. By Herstein and Milnor (1953) and since $\succsim_Q$ is non-trivial, there exists a non-constant affine function v_Q which represents $\succsim_Q$. By Step 1, $\succsim_Q^*$ is represented by u. Since $\left(\succsim_Q^*, \succsim_Q\right)$ jointly satisfy consistency, it follows that for each $x, y \in X$

$$u(x) \geq u(y) \Longrightarrow v_Q(x) \geq v_Q(y)$$

By Corollary B.3 of Ghirardato et al. (2004), v_Q and u are equal up to an affine and positive transformation. So we can set $v_Q = u$, proving the statement. □

Step 5. For each $Q \in \mathcal{Q}$ *we have that*

$$f \succsim_Q g \iff \inf_{p \in \Delta} \left\{ \int u(f)\, dp + \inf_{q \in Q} c(p, q) \right\} \geq \inf_{p \in \Delta} \left\{ \int u(g)\, dp + \inf_{q \in Q} c(p, q) \right\} \tag{56}$$

Proof of the Step Fix $Q \in \mathcal{Q}$. By Cerreia-Vioglio et al. (2011a) and since $\succsim_Q$ is a rational preference relation, there exists a normalized, monotone and continuous functional $I_Q : B_0(\Sigma) \to \mathbb{R}$

such that

$$f \succsim_Q g \iff I_Q(u(f)) \geq I_Q(u(g)) \tag{57}$$

By the same arguments in Steps 3 and 4 of Theorem 1, we have that $I_Q = \inf_{q\in Q} I_q$, yielding that

$$\begin{aligned} I_Q(\varphi) &= \inf_{q\in Q}\min_{p\in\Delta}\left\{\int \varphi dp + c(p,q)\right\} = \inf_{q\in Q}\inf_{p\in\Delta}\left\{\int \varphi dp + c(p,q)\right\} \\ &= \inf_{p\in\Delta}\inf_{q\in Q}\left\{\int \varphi dp + c(p,q)\right\} = \inf_{p\in\Delta}\left\{\int \varphi dp + \inf_{q\in Q} c(p,q)\right\} \quad \forall \varphi \in B_0(\Sigma) \end{aligned}$$

By (57), this implies that (56) holds. □

Step 6. $c(p,q) = 0$ if and only if $p = q$.

Proof of the Step By Steps 2 and 5, we have that $\succsim_q^*$ coincides with $\succsim_q$ on $\mathcal{F}$ for all $q \in \Delta^\sigma$. By Lemma 3 and since $\succsim_q$ is subjectively $\{q\}$-coherent, we have that $\operatorname{argmin} c(\cdot,q) = \operatorname{argmin} c_q = \{q\}$. □

Step 7. $\operatorname{dom} c(\cdot,q) \subseteq \Delta^{\ll}(q) \subseteq \Delta^{\ll}(Q)$ *for all $q \in Q$ and for all $Q \in \mathcal{Q}$.*

Proof of the Step By the previous part of the proof, we have that $\succsim_q^*$ coincides with $\succsim_q$ on $\mathcal{F}$ for all $q \in \Delta^\sigma$. By Lemma 4 and since $\succsim_q^*$ is objectively $\{q\}$-coherent, we can conclude that $\operatorname{dom} c(\cdot,q) \subseteq \Delta^{\ll}(q) \subseteq \Delta^{\ll}(Q)$ for all $q \in Q$ and for all $Q \in \mathcal{Q}$. □

Step 8. c is jointly lower semicontinuous.

Proof of the Step Define the map $J : B_0(\Sigma) \times \Delta^\sigma \to \mathbb{R}$ by $J(\varphi,q) = I_q(\varphi)$ for all $q \in Q$. Observe that, for each $(p,q) \in \Delta \times \Delta^\sigma$,

$$c(p,q) = c_q(p) = \sup_{\varphi\in B_0(\Sigma)}\left\{I_q(\varphi) - \int \varphi dp\right\} = \sup_{\varphi\in B_0(\Sigma)}\left\{J(\varphi,q) - \int \varphi dp\right\} \tag{58}$$

We begin by observing that J is lower semicontinuous in the second argument. Note that for each $\varphi \in B_0(\Sigma)$ and for each $q \in \Delta^\sigma$

$$J(\varphi,q) = I_q(\varphi) = u(x_{f,q}) \qquad \text{where } f \in \mathcal{F} \text{ is s.t. } \varphi = u(f)$$

Fix $\varphi \in B_0(\Sigma)$ and $t \in \mathbb{R}$. By the axiom of lower semicontinuity, the set

$$\{q \in \Delta^\sigma : J(\varphi,q) \leq t\} = \{q \in \Delta^\sigma : u(x) \geq u(x_{f,q})\} = \left\{q \in \Delta^\sigma : x \succsim_q^* x_{f,q}\right\}$$

is closed where $x \in X$ and $f \in \mathcal{F}$ are such that $u(x) = t$ as well as $u(f) = \varphi$. Since φ and t were arbitrarily chosen, this yields that J is lower semicontinuous in the second argument. Since J is lower semicontinuous in the second argument, the map $(p,q) \mapsto J(\varphi,q) - \int \varphi dp$, defined over $\Delta \times \Delta^\sigma$, is jointly lower semicontinuous for all $\varphi \in B_0(\Sigma)$. By (58) and the definition of c, we conclude that c is jointly lower semicontinuous. □

Step 1 proves that u is affine and onto. Steps 2, 6, 7 and 8 prove that c is a jointly lower semicontinuous divergence which is convex in the first argument. Steps 1, 3, 5 and 8 yield the representation of $\succsim_Q^*$ and $\succsim_Q$ for all $Q \in \mathcal{Q}$. As for uniqueness, assume that the function $\tilde{c} : \Delta \times \Delta^\sigma \to [0, \infty]$ is a divergence which is jointly lower semicontinuous, convex in the first argument and that represents $\succsim_Q^*$ and $\succsim_Q$ for all $Q \in \mathcal{Q}$. By Proposition 6 of Maccheroni et al. (2006) and since $\operatorname{Im} u = \mathbb{R}$ and $\succsim_q^*$ is a variational preference for all $q \in \Delta^\sigma$, it follows that $\tilde{c}(\cdot, q) = c(\cdot, q)$ for all $q \in \Delta^\sigma$, yielding that $c = \tilde{c}$. ■

Proof of Proposition 7 We only prove (i) implies (ii), the converse being routine. We keep the notation of the previous proof. Compared to Theorem 2, we only need to prove that c is jointly convex. By Lemma 12 in Appendix B.1, this will yield that c is variational. Fix $\varphi \in B_0(\Sigma)$, $q, q' \in \Delta^\sigma$ and $\lambda \in (0, 1)$. By model hybridization aversion and since u is affine, we have that

$$\begin{aligned} J(\varphi, \lambda q + (1-\lambda) q') &= u\left(x_{f,\lambda q+(1-\lambda)q'}\right) \leq u\left(\lambda x_{f,q} + (1-\lambda) x_{f,q'}\right) \\ &= \lambda u(x_{f,q}) + (1-\lambda) u(x_{f,q'}) = \lambda J(\varphi, q) + (1-\lambda) J(\varphi, q') \end{aligned}$$

where $f \in \mathcal{F}$ is such that $u(f) = \varphi$. Since φ, q, q' and λ were arbitrarily chosen, this yields that J is convex in the second argument. Since J is convex in the second argument, the map $(p, q) \mapsto J(\varphi, q) - \int \varphi dp$, defined over $\Delta \times \Delta^\sigma$, is jointly convex for all $\varphi \in B_0(\Sigma)$. By (58) and the definition of c, we conclude that c is convex, proving the implication. ■

Proof of Proposition 10 We only prove (i) implies (ii), being (ii) implies (i) routine. We keep the same notation and terminology as in the proof and statement of Theorem 2. It is then immediate to notice that Steps 1–4 of that proof continue to hold here. In particular, there exist an onto and affine utility function and a function $c : \Delta \times \Delta^\sigma \to [0, \infty]$, which is grounded, convex and lower semicontinuous in the first argument, such that for each $Q \in \mathcal{Q}$ and for each $f, g \in \mathcal{F}$

$$f \succsim_Q^* g \iff \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p, q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + c(p, q) \right\} \quad \forall q \in Q \tag{59}$$

and for each $x, y \in X$

$$x \succsim_Q^* y \iff x \succsim_Q y \iff u(x) \geq u(y) \tag{60}$$

Fix $Q \in \mathcal{Q}$. Since $\succsim_Q$ is solvable, for each $f \in \mathcal{F}$ there exists $x_{f,Q} \in X$ such that $x_{f,Q} \sim_Q f$. Since $\operatorname{Im} u = \mathbb{R}$, define $I_Q : B_0(\Sigma) \to \mathbb{R}$ by $I_Q(\varphi) = u(x_{f,Q})$ where $f \in \mathcal{F}$ is such that $u(f) = \varphi$. By (60) and since $\succsim_Q$ is a complete, transitive and monotone binary relation, we have that I_Q is well defined and monotone. Moreover, by construction, we have that $I_Q(k 1_S) = k$ for all $k \in \mathbb{R}$. By (60) and construction, note that

$$I_Q(u(f)) \geq I_Q(u(g)) \iff u(x_{f,Q}) \geq u(x_{g,Q}) \iff x_{f,Q} \succsim_Q x_{g,Q} \iff f \succsim_Q g \tag{61}$$

Since $\succsim_Q^*$and $\succsim_Q$ jointly satisfy consistency for all $Q \in \mathcal{Q}$,[42] this implies that $\succsim_q^*$and $\succsim_q$ coincide on $\mathcal{F}$ for all $q \in \Delta^\sigma$ and

$$I_q(\varphi) = \min_{p \in \Delta} \left\{ \int \varphi dp + c(p,q) \right\} \qquad \forall \varphi \in B_0(\Sigma)$$

In particular, we have that Steps 6–8 of Theorem 2 hold also in this case, proving that c is a lower semicontinuous divergence, which is convex in the first argument. Given $\varphi \in B_0(\Sigma)$, note that the map $q \mapsto I_q(\varphi)$ is such that $\min_{s \in S} \varphi(s) \leq I_q(\varphi) \leq \max_{s \in S} \varphi(s)$ for all $q \in Q$, yielding that the map $q \mapsto I_q(\varphi)$ is an element of $B(Q)$. Consider the set

$$M = \{\tilde{\varphi} \in B(Q) : \exists \varphi \in B_0(\Sigma) \text{ s.t. } \forall q \in Q, \tilde{\varphi}(q) = I_q(\varphi)\}$$

Since $I_q(k1_S) = k$ for all $k \in \mathbb{R}$, we have that M contains all the constants $k1_Q$ where $k \in \mathbb{R}$. Define $\tilde{J}_Q : M \to \mathbb{R}$ by $\tilde{J}_Q(\tilde{\varphi}) = I_Q(\varphi)$ where $\varphi \in B_0(\Sigma)$ is such that $\tilde{\varphi}(q) = I_q(\varphi)$ for all $q \in Q$. Note that for each $\varphi \in B_0(\Sigma)$ there exists $f \in \mathcal{F}$ such that $u(f) = \varphi$. Assume that given $\tilde{\varphi} \in M$ there exist $\varphi, \psi \in B_0(\Sigma)$ such that $\tilde{\varphi}(q) = I_q(\varphi) = I_q(\psi)$ for all $q \in Q$. Consider $f, g \in \mathcal{F}$ such that $u(f) = \varphi$ and $u(g) = \psi$. It follows that $I_q(u(f)) = I_q(u(g))$ for all $q \in Q$. By (59) and consistency, this implies that $f \sim_Q^* g$ and $f \sim_Q g$. By (61), it follows that $I_Q(\varphi) = I_Q(u(f)) = I_Q(u(g)) = I_Q(\psi)$, proving that $\tilde{J}_Q$ is well defined. Next, assume that $\tilde{\varphi}, \tilde{\psi} \in M$ are such that $\tilde{\varphi} \geq \tilde{\psi}$. Let $\varphi, \psi \in B_0(\Sigma)$ be such that $\tilde{\varphi}(q) = I_q(\varphi)$ and $\tilde{\psi}(q) = I_q(\psi)$ for all $q \in Q$. Consider $f, g \in \mathcal{F}$ such that $u(f) = \varphi$ and $u(g) = \psi$. It follows that $I_q(u(f)) \geq I_q(u(g))$ for all $q \in Q$. By (59) and consistency, this implies that $f \succsim_Q^* g$ and $f \succsim_Q g$. By (61), it follows that

$$\tilde{J}_Q(\tilde{\varphi}) = I_Q(\varphi) = I_Q(u(f)) \geq I_Q(u(g)) = I_Q(\psi) = \tilde{J}_Q\left(\tilde{\psi}\right)$$

proving that $\tilde{J}_Q$ is monotone. Moreover, by construction, we have $\tilde{J}_Q(k1_Q) = I_Q(k1_S) = k$ for all $k \in \mathbb{R}$, proving that $\tilde{J}_Q$ is normalized. By (61) and definition of $\tilde{J}_Q$, we can conclude that

$$f \succsim_Q g \iff \tilde{J}_Q\left(\min_{p \in \Delta}\left\{\int u(f)\, dp + c(p,\cdot)\right\}\right) \geq \tilde{J}_Q\left(\min_{p \in \Delta}\left\{\int u(g)\, dp + c(p,\cdot)\right\}\right) \tag{62}$$

We next extend $\tilde{J}_Q$ to the entire set $B(Q)$. Define $J_Q : B(Q) \to \mathbb{R}$ by

$$J_Q(\tilde{\varphi}) = \sup\left\{\tilde{J}_Q\left(\tilde{\psi}\right) : M \ni \tilde{\psi} \leq \tilde{\varphi}\right\} \qquad \forall \tilde{\varphi} \in B(Q)$$

[42]By (59) and (61), we have that $f \mapsto \min_{p \in \Delta}\left\{\int u(f)\, dp + c(p,q)\right\}$ and $f \mapsto I_Q(u(f))$ represent, respectively, $\succsim_Q^*$ and $\succsim_Q$. Since $\succsim_Q^*$ and $\succsim_Q$ satisfy consistency, we can conclude that there exists a (not necessarily strictly) monotone function $h : \mathbb{R} \to \mathbb{R}$ such that $I_Q(u(f)) = h\left(\min_{p \in \Delta}\left\{\int u(f)\, dp + c(p,q)\right\}\right)$ for all $f \in \mathcal{F}$. Since I_Q is normalized and $\operatorname{Im} u = \mathbb{R}$, we have that $h(u(x)) = u(x)$ for all $x \in X$, proving that h is the identity.

It is routine to check that J_Q in both cases extends $\tilde{J}_Q$ and it is normalized and monotone. Moreover, by (62) it satisfies (37), proving the implication. Uniqueness follows from the same arguments of Theorem 2. ■

Proof of Corollary 1 We only prove (ii) implies (iii), being (i) implies (ii) obvious and (iii) implies (i) an immediate consequence of Theorem 2. We keep the same notation and terminology as in the proof and statement of Proposition 10. By Proposition 10, there exist an onto and affine utility function $u : X \to \mathbb{R}$ and a lower semicontinuous divergence $c : \Delta \times \Delta^\sigma \to [0, \infty]$, convex in p, such that for each $Q \in \mathcal{Q}$ and for each $f, g \in \mathcal{F}$

$$f \succsim_Q^* g \iff \min_{p\in\Delta} \left\{ \int u(f)\, dp + c(p, q) \right\} \geq \min_{p\in\Delta} \left\{ \int u(g)\, dp + c(p, q) \right\} \quad \forall q \in Q$$

and for each $x, y \in X$

$$x \succsim_Q^* y \iff x \succsim_Q y \iff u(x) \geq u(y)$$

Moreover, for each $Q \in \mathcal{Q}$ there exists a normalized and monotone functional $I_Q : B_0(\Sigma) \to \mathbb{R}$ such that $f \succsim_Q g$ if and only if $I_Q(u(f)) \geq I_Q(u(g))$. Fix $Q \in \mathcal{Q}$. By the same arguments in Steps 3 and 4 of Theorem 1, we have that

$$\begin{aligned} I_Q(\varphi) &= \inf_{q\in Q} \min_{p\in\Delta} \left\{ \int \varphi dp + c(p, q) \right\} = \inf_{q\in Q} \inf_{p\in\Delta} \left\{ \int \varphi dp + c(p, q) \right\} \\ &= \inf_{p\in\Delta} \inf_{q\in Q} \left\{ \int \varphi dp + c(p, q) \right\} = \inf_{p\in\Delta} \left\{ \int \varphi dp + \inf_{q\in Q} c(p, q) \right\} \quad \forall \varphi \in B_0(\Sigma) \end{aligned}$$

where the infima become minima since c is lower semicontinuous. We can conclude that

$$f \succsim_Q g \iff I_Q(u(f)) \geq I_Q(u(g)) \iff \min_{p\in\Delta} \left\{ \int u(f)\, dp + \min_{q\in Q} c(p, q) \right\} \geq \min_{p\in\Delta} \left\{ \int u(g)\, dp + \min_{q\in Q} c(p, q) \right\}$$

proving the implication. Uniqueness follows from Proposition 10. ■

A.2 Other proofs

Proof of Proposition 3 Consider first $\lambda \in (0, \infty)$. Note that $c(\cdot, q) = \lambda D_\phi(\cdot || q)$ is Shur convex (with respect to q) for all $q \in Q$. Consider $A, B \in \Sigma$. Assume that $q(A) \geq q(B)$ for all $q \in Q$. Let $q \in Q$. Consider $x, y \in X$ such that $x \succ y$. It follows that

$$\int v(u(xAy))\, dq \geq \int v(u(xBy))\, dq$$

for each $v : \mathbb{R} \to \mathbb{R}$ increasing and concave. By Theorem 2 of Cerreia-Vioglio et al. (2012) and since q was arbitrarily chosen, it follows that

$$\min_{p\in\Delta}\left\{\int u\left(xAy\right)dp + \lambda D_{\phi}\left(p||q\right)\right\} \geq \min_{p\in\Delta}\left\{\int u\left(xBy\right)dp + \lambda D_{\phi}\left(p||q\right)\right\} \quad \forall q \in Q$$

yielding that $xAy \succsim^* xBy$ and, in particular, $xAy \succsim xBy$. If $\lambda = \infty$ instead, as pointed out in Section 2.1, we have that $c\left(\cdot, q\right) = \lambda D_{\phi}\left(\cdot||q\right) = \delta_{\{q\}}\left(\cdot\right)$ for all $q \in Q$. This implies that (17) takes the max-min form over the set Q, which trivially implies bet-consistency.[43] ■

Proof of Proposition 4 We prove the "only if", the converse being obvious. Define $\succsim^*$ by $f \succsim^* g$ if and only if $\int u\left(f\right)dq \geq \int u\left(g\right)dq$ for all $q \in Q$. By hypothesis, the pair $\left(\succsim^*, \succsim\right)$ satisfies consistency. Let $f \not\succsim^* x$. Then, there exists $q \in Q$ such that $u(x_f^q) = \int u\left(f\right)dq < u\left(x\right)$. Hence, $x \succ x_f^q$. Since $c_Q^{-1}\left(0\right) = Q$, by Lemma 3 we have that $x \succ f$. So, the pair $\left(\succsim^*, \succsim\right)$ satisfies default to certainty. By Theorem 4 of Gilboa et al. (2010), this pair admits the representation

$$f \succsim^* g \Longleftrightarrow \int u\left(f\right)dq \geq \int u\left(g\right)dq \quad \forall q \in Q$$

and

$$f \succsim g \Longleftrightarrow \min_{q\in Q}\int u\left(f\right)dq \geq \min_{q\in Q}\int u\left(g\right)dq$$

Note that, in the notation of Gilboa et al. (2010), we have $C = Q$ because C is unique up to closure and convexity and Q is closed and convex. ■

Proof of Proposition 5 For each $q \in Q$ define $I_q : B_0\left(\Sigma\right) \to \mathbb{R}$ by

$$I_q\left(\varphi\right) = \min_{p\in\Delta}\left\{\int \varphi dp + c\left(p, q\right)\right\} \qquad \forall \varphi \in B_0\left(\Sigma\right)$$

Recall that $f \ggcurly^* g$ if and only if for each $h, l \in \mathcal{F}$ there exists $\varepsilon > 0$ such that

$$\left(1-\delta\right)f + \delta h \succ^* \left(1-\delta\right)g + \delta l \qquad \forall \delta \in \left[0, \varepsilon\right] \tag{63}$$

Moreover, given $h \in \mathcal{F}$, define $k_h = \inf_{s\in S} u\left(h\left(s\right)\right)$ and $k^h = \sup_{s\in S} u\left(h\left(s\right)\right)$.

"Only if." Assume that $f \ggcurly^* g$. Let $\hat{\varepsilon} > 0$. Consider $u\left(x\right) = k_f - \hat{\varepsilon}$ and $u\left(y\right) = k^g + \hat{\varepsilon}$. By definition, there exists $\varepsilon > 0$ such that $\left(1-\delta\right)f + \delta x \succ^* \left(1-\delta\right)g + \delta y$ for all $\delta \in \left[0, \varepsilon\right]$. Note that for each $q \in Q$ and for each $\delta \in \left[0, 1\right]$

$$\begin{aligned} I_q\left(u\left(\left(1-\delta\right)f + \delta x\right)\right) &= I_q\left(\left(1-\delta\right)u\left(f\right) + \delta u\left(x\right)\right) = I_q\left(u\left(f\right) - \delta u\left(f\right) + \delta u\left(x\right)\right) \\ &\leq I_q\left(u\left(f\right) - \delta k_f + \delta\left(k_f - \hat{\varepsilon}\right)\right) = I_q\left(u\left(f\right)\right) - \delta\hat{\varepsilon} \end{aligned}$$

[43]The next result will indeed prove a much more general fact.

and

$$
\begin{aligned}
I_q(u((1-\delta)g+\delta y)) &= I_q((1-\delta)u(g)+\delta u(y)) = I_q(u(g)-\delta u(g)+\delta u(y)) \\
&\geq I_q(u(g)-\delta k^g+\delta(k^g+\hat{\varepsilon})) = I_q(u(g))+\delta\hat{\varepsilon}
\end{aligned}
$$

It follows that for each $q \in Q$ and for each $\delta \in [0,\varepsilon]$

$$
I_q(u(f)) - I_q(u(g)) - 2\delta\hat{\varepsilon} \geq I_q(u((1-\delta)f+\delta x)) - I_q(u((1-\delta)g+\delta y)) \geq 0
$$

If we set $\delta = \varepsilon > 0$, then $I_q(u(f)) \geq I_q(u(g)) + 2\varepsilon\hat{\varepsilon}$ for all $q \in Q$, proving the statement.

"If." Let $f, g \in \mathcal{F}$. Assume there exists $\varepsilon > 0$ such that $I_q(u(f)) \geq I_q(u(g)) + \varepsilon$ for all $q \in Q$. Consider $h, l \in \mathcal{F}$. Note that for each $q \in Q$ and for each $\delta \in [0,1]$

$$
\begin{aligned}
I_q(u((1-\delta)f+\delta h)) &= I_q((1-\delta)u(f)+\delta u(h)) = I_q(u(f)-\delta u(f)+\delta u(h)) \\
&= I_q(u(f)+\delta(u(h)-u(f))) \\
&\geq I_q\left(u(f)+\delta\left(k_h-k^f\right)\right) = I_q(u(f))+\delta\left(k_h-k^f\right)
\end{aligned}
$$

and

$$
\begin{aligned}
I_q(u((1-\delta)g+\delta l)) &= I_q((1-\delta)u(g)+\delta u(l)) = I_q(u(g)-\delta u(g)+\delta u(l)) \\
&= I_q(u(g)+\delta(u(l)-u(g))) \\
&\leq I_q\left(u(g)+\delta\left(k^l-k_g\right)\right) = I_q(u(g))+\delta\left(k^l-k_g\right)
\end{aligned}
$$

It follows that for each $q \in Q$ and for each $\delta \in [0,1]$

$$
\begin{aligned}
I_q(u((1-\delta)f+\delta h)) - I_q(u((1-\delta)g+\delta l)) &\geq I_q(u(f))+\delta\left(k_h-k^f\right) - I_q(u(g)) - \delta\left(k^l-k_g\right) \\
&\geq \varepsilon + \delta\hat{\varepsilon}
\end{aligned}
$$

where $\hat{\varepsilon} = k_h - k^f - k^l + k_g$. We have two cases:

1. $\hat{\varepsilon} \geq 0$. In this case, $I_q(u((1-\delta)f+\delta h)) - I_q(u((1-\delta)g+\delta l)) > 0$ for all $\delta \in [0,1]$ and all $q \in Q$, proving (63).

2. $\hat{\varepsilon} < 0$. In this case, $I_q(u((1-\delta)f+\delta h)) - I_q(u((1-\delta)g+\delta l)) > 0$ for all $\delta \in [0, -\varepsilon/2\hat{\varepsilon}]$ and all $q \in Q$, proving (63).

This completes the proof of the result. ■

Proof of Lemma 1 "If." Given $q \in Q$, if $c(p,q) = \infty$ for all $p \notin Q$, then $c_Q(p) = \infty$ for all $p \notin Q$. Since $c_Q(p) = 0$ for all $p \in Q$, we conclude that $c_Q(p) = \delta_Q(p)$ for all $p \in \Delta$. "Only if." Conversely, for each $q \in Q$ we have that $c(p,q) \geq c_Q(p) = \delta_Q(p) = \infty$ for all $p \notin Q$. ■

Proof of Proposition 6 Before starting, we make two observations. First, observe that (i) and (ii) of Proposition 6 are particular cases of (i) and (ii) of Theorem 1. We thus adopt the same notation, terminology, and arguments contained in the proof of this latter theorem. Second, consider an unbounded dominance relation $\succsim^*$ and a rational preference $\succsim$ that jointly satisfy consistency and caution. By consistency, it follows that for each $f \in \mathcal{F}$ and for each $x \in X$

$$f \succsim^* x \implies f \succsim x$$

By caution, we have that for each $f \in \mathcal{F}$ and for each $x \in X$

$$f \succ x \implies f \succsim^* x$$

Moreover, by the same arguments of Step 1 of the proof of Theorem 1, $\succsim$ agrees with $\succsim^*$ on X and they are represented by an onto affine utility function $u : X \to \mathbb{R}$. This implies that if $f \sim x$, then there exists $y \in X$ such that $x \succ \alpha x + (1 - \alpha) y \succ y$ for all $\alpha \in (0, 1)$, yielding that $f \succ \alpha x + (1 - \alpha) y$ and $f \succsim^* \alpha x + (1 - \alpha) y$ for all $\alpha \in (0, 1)$. Since $\succsim^*$ satisfies continuity, we have that $f \succsim^* x$. Thus, we can conclude that for each $f \in \mathcal{F}$ and for each $x \in X$

$$f \succsim^* x \iff f \succsim x \tag{64}$$

(i) implies (ii). By Theorem 1, we have that there exist an onto affine function $u : X \to \mathbb{R}$ and a variational pseudo-statistical distance $c : \Delta \times Q \to [0, \infty]$ such that, for all acts $f, g \in \mathcal{F}$,

$$f \succsim^* g \iff \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p, q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + c(p, q) \right\} \qquad \forall q \in Q \tag{65}$$

and

$$f \succsim g \iff \min_{p \in \Delta} \left\{ \int u(f)\, dp + \min_{q \in Q} c(p, q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + \min_{q \in Q} c(p, q) \right\}$$

By (64) and since $\succsim^*$ satisfies c-independence, we have that if $f \in \mathcal{F}$, $x, y \in X$, and $\alpha \in (0, 1]$, then

$$f \succsim x \iff f \succsim^* x \iff \alpha f + (1 - \alpha) y \succsim^* \alpha x + (1 - \alpha) y \iff \alpha f + (1 - \alpha) y \succsim \alpha x + (1 - \alpha) y$$

proving that $\succsim$ satisfies c-independence. By Propositions 6 and 19 of Maccheroni et al. 2006 and since u is onto, $c_Q : \Delta \to [0, \infty]$ is grounded, lower semicontinuous and convex and $c_Q^{-1}(0) = Q$, we have that $c_Q = \delta_Q$, proving (31). Since $c(\cdot, q) \geq c_Q = \delta_Q$ for all $q \in Q$, we have that $c(p, q) \geq c_Q(p) = \delta_Q(p) = \infty$ for all $p \notin Q$, yielding that the min in (65) can be restricted to Q and proving (30).

(ii) implies (i). By the same arguments contained above and since $c_Q = \delta_Q$, we have that the min in (30) can be taken over Δ. By (ii) implies (i) of Theorem 1, the implication follows with the exception of proving that $\succsim^*$ satisfies c-independence. Since $\succsim$ is represented as in (31), $\succsim$ satisfies c-independence. Since $\succsim^*$ is an unbounded dominance relation and $\succsim$ is a rational preference and jointly they satisfy consistency and caution, (64) holds, yielding that if $f \in \mathcal{F}$, $x, y \in X$, and $\alpha \in (0,1]$, then

$$f \succsim^* x \iff f \succsim x \iff \alpha f + (1-\alpha) y \succsim \alpha x + (1-\alpha) y \iff \alpha f + (1-\alpha) y \succsim^* \alpha x + (1-\alpha) y$$

proving that $\succsim^*$ satisfies c-independence and the implication.

We prove the second part of the statement independently and with a different technique in order to dispense with the assumption of (S, Σ) being a standard Borel space. We only need to prove the "only if" part, the "if" being trivial.

By Proposition 2 of Cerreia-Vioglio (2016) and since $\succsim^*$ is unbounded, there exists a compact and convex set $C \subseteq \Delta$ and an affine and onto map $u : X \to \mathbb{R}$ such that

$$f \succsim^* g \Longleftrightarrow \int u(f)\, dq \geq \int u(g)\, dq \qquad \forall q \in C \tag{66}$$

and

$$f \succsim g \iff \min_{q \in C} \int u(f)\, dq \geq \min_{q \in C} \int u(g)\, dq \tag{67}$$

By Lemma 3 and since $\succsim$ is subjectively Q-coherent and $\succsim^*$ and $\succsim$ coincide on X, we can conclude that $C = Q$. If we set $c : \Delta \times Q \to [0, \infty]$ to be $c(p, q) = \delta_{\{q\}}(p)$ for all $(p, q) \in \Delta \times Q$, then it is immediate to see that c is a variational statistical distance. By (66) and (67) and since $C = Q$, the implication follows. ■

Proof of Proposition 11 We begin by making two observations. It is well known that, given a bounded and measurable $F : Q \to \mathbb{R}$,

$$\lim_{\xi \to 0^+} \phi_\xi^{-1} \left(\int_Q \phi_\xi (F(q))\, d\mu_Q \right) = \min_{q \in \operatorname{supp} \mu_Q} F(q) = \min_{q \in Q} F(q) \tag{68}$$

and

$$\phi_\xi^{-1} \left(\int_Q \phi_\xi (F(q))\, d\mu_Q \right) = \min_{\mu \ll \nu} \left\{ \int F d\nu + \xi R(\nu || \mu) \right\} \tag{69}$$

Fix $f \in \mathcal{F}$ and $\lambda \in (0, \infty]$. Since c is lower semicontinuous and each $f \in \mathcal{F}$ is finitely valued, if we set $F_\lambda(q) = \min_{p \in \Delta} \{u(f)\, dp + \lambda R(p||q)\}$ for all $q \in Q$, it is immediate to see that F is bounded and measurable.

By (68), (42) follows. By Proposition 12 of Maccheroni et al. (2006) and (69) and since $\lim_{\xi \to \infty} \xi R(\nu || \mu) = \infty$ if $\nu \neq \mu$ and $\lim_{\xi \to \infty} \xi R(\nu || \mu) = \infty$ if $\nu = \mu$, (43) follows. By (43), we have that

$$\lim_{\xi \to \infty} V_Q^{\lambda, \xi}(f) = \int_Q \left(\min_{p \in \Delta} \left\{ \int_S u(f(s))\, dp(s) + \lambda R(p||q) \right\} \right) d\mu_Q(q)$$

By Proposition 12 of Maccheroni et al. (2006) and since $\lim_{\lambda\to\infty} \lambda R(p||q) = \infty$ if $p \neq q$ and $\lim_{\lambda\to\infty} \lambda R(p||q) = \infty$ if $p = q$, we have that $\lim_{\lambda\to\infty} F_\lambda(q) = \int u(f)\,dq = F_\infty(q)$ for all $q \in Q$. By the Lebesgue Dominanted Convergence Theorem and since $\{F_\lambda\}_{\lambda\in(0,\infty)}$ are uniformly bounded, the second equality of (44) follows. The first has a similar proof and we omit it. ■

Proof of Lemma 2 (i) Let $p \in \Delta$. Since c is lower semicontinuous, there exists $q_p \in Q$ such that $c(p, q_p) = \min_{q\in Q} c(p, q)$, that is, $p \in B_c(q_p, Q)$. This proves that $\Delta \subseteq \bigcup_{q\in Q} B_c(q, Q)$, as desired (the other inclusion is trivial).

(ii) For each $q \in Q$ we have that $0 = c(q, q) \geq \min_{\tilde{q}\in Q} c(q, \tilde{q}) \geq 0$. Thus, $c(q, q) = \min_{\tilde{q}\in Q} c(q, \tilde{q})$ and so $q \in B_c(q, Q)$. It remains to show that $B_c(q, Q) \cap Q \subseteq \{q\}$. So, let $\bar{q} \in B_c(q, Q) \cap Q$. Then, $c(\bar{q}, q) = \min_{\tilde{q}\in Q} c(\bar{q}, \tilde{q})$. Since $\bar{q} \in Q$, we have $\min_{\tilde{q}\in Q} c(\bar{q}, \tilde{q}) = 0$ and so $c(\bar{q}, q) = 0$, which implies $\bar{q} = q$, as desired.

(iii) Let $q, q' \in Q$ with $q \neq q'$. In view of (ii), it is enough to consider $p \in B_c(q, Q) \cap B_c(q', Q) \cap \Delta_{c,Q}$. Since $p \in \Delta_{c,Q}$, the map $c(p, \cdot) : Q \to [0, \infty]$ is proper and strictly convex. Thus, $c(p, q) = c(p, q') = \min_{\tilde{q}\in Q} c(p, \tilde{q}) < \infty$, which leads to the contradiction $q = q'$. ■

References

[1] F. J. Anscombe and R. J. Aumann, A definition of subjective probability, *Annals of Mathematical Statistics*, 34, 199–205, 1963.

[2] I. Aydogan, L. Berger, V. Bosetti and N. Liu, Three layers of uncertainty and the role of model misspecification: an experiment, IGIER WP 623, 2018.

[3] A. Ben-Tal and M. Teboulle, An old-new concept of convex risk measures: The optimized certainty equivalent, *Mathematical Finance*, 17, 449–476, 2007.

[4] D. A. Blackwell and M. A. Girschick, *Theory of games and statistical decisions*, Wiley, New York, 1954.

[5] G. E. P. Box, Science and statistics, *Journal of the American Statistical Association*, 71, 791–799, 1976.

[6] S. Cerreia-Vioglio, Objective rationality and uncertainty averse preferences, *Theoretical Economics*, 11, 523–545, 2016.

[7] S. Cerreia-Vioglio, P. Ghirardato, F. Maccheroni, M. Marinacci and M. Siniscalchi, Rational preferences under ambiguity, *Economic Theory*, 48, 341–375, 2011a.

[8] S. Cerreia-Vioglio, A. Giarlotta, S. Greco, F. Maccheroni and M. Marinacci, Rational preference and rationalizable choice, *Economic Theory*, 69, 61–105, 2020.

[9] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci and L. Montrucchio, Uncertainty averse preferences, *Journal of Economic Theory*, 146, 1275–1330, 2011b.

[10] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci and L. Montrucchio, Probabilistic sophistication, second order stochastic dominance and uncertainty aversion, *Journal of Mathematical Economics*, 48, 271–283, 2012.

[11] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci and L. Montrucchio, Classical subjective expected utility, *Proceedings of the National Academy of Sciences*, 110, 6754–6759, 2013.

[12] D. R. Cox, Comment on "Model uncertainty, data mining and statistical inference", *Journal of the Royal Statistical Society*, Series A, 158, 455–456, 1995.

[13] T. Denti and L. Pomatto, Model and predictive uncertainty: A Foundation for smooth ambiguity preferences, *Econometrica*, 90, 551–584, 2022.

[14] H. Duanmu and D. M. Roy, On extended admissible procedures and their nonstandard Bayes risk, *Annals of Statistics*, 49, 2053–2078, 2021.

[15] I. Esponda and D. Pouzo, Berk-Nash equilibrium: A framework for modeling agents with misspecified models, *Econometrica*, 84, 1093–1130, 2016.

[16] T. S. Ferguson, *Mathematical statistics: A decision theoretic approach*, Academic Press, New York, 1967.

[17] D. Fudenberg, G. Romanyuk and P. Strack, Active learning with a misspecified prior, *Theoretical Economics*, 12, 1155–1189, 2017.

[18] T. Gajdos, T. Hayashi, J.-M. Tallon and J.-C. Vergnaud, Attitude toward imprecise information, *Journal of Economic Theory*, 140, 27–65, 2008.

[19] P. Ghirardato, F. Maccheroni and M. Marinacci, Differentiating ambiguity and ambiguity attitude, *Journal of Economic Theory*, 118, 133–173, 2004.

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

[21] I. Gilboa, F. Maccheroni, M. Marinacci and D. Schmeidler, Objective and subjective rationality in a multiple prior model, *Econometrica*, 78, 755–770, 2010.

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

[23] L. P. Hansen, Nobel lecture: Uncertainty outside and inside economic models, *Journal of Political Economy*, 122, 945–987, 2014.

[24] L. P. Hansen and M. Marinacci, Ambiguity aversion and model misspecification: An economic perspective, *Statistical Science*, 31, 511–515, 2016.

[25] L. P. Hansen and J. Miao, Aversion to ambiguity and model misspecification in dynamic stochastic environments, *Proceedings of the National Academy of Sciences*, 115, 9163–9168, 2018.

[26] L. P. Hansen and T. J. Sargent, Robust control and model uncertainty, *American Economic Review*, 91, 60–66, 2001.

[27] L. P. Hansen and T. J. Sargent, Recursive robust estimation and control without commitment, *Journal of Economic Theory*, 136, 1–27, 2007.

[28] L. P. Hansen and T. J. Sargent, *Robustness*, Princeton University Press, Princeton, 2008.

[29] L. P. Hansen and T. J. Sargent, Structured ambiguity and model misspecification, *Journal of Economic Theory*, 199, 2022.

[30] L. P. Hansen and T. J. Sargent, Risk, ambiguity, and misspecification: Decision theory, robust control and statistics, mimeo, 2022b.

[31] D. Heath and W. Sudderth, On finitely additive priors, coherence, and extended admissibility, *Annals of Statistics*, 333–345, 1978.

[32] I. N. Herstein and J. Milnor, An axiomatic approach to measurable utility, *Econometrica*, 21, 291–297, 1953.

[33] E. Hewitt and L. J. Savage, Symmetric measures on Cartesian products, *Transactions of the American Mathematical Society*, 80, 470–501, 1955.

[34] D. H. Jacobson, Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games, *IEEE Transactions on Automatic Control*, AC-18, 124–131, 1973.

[35] F. Liese and I. Vajda, *Convex statistical distances*, Teubner, Leipzig, 1987.

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

[37] M. Marinacci, Model uncertainty, *Journal of the European Economic Association*, 13, 1022–1100, 2015.

[38] L. Nascimento and G. Riella, A class of incomplete and ambiguity averse preferences, *Journal of Economic Theory*, 146, 728–750, 2011.

[39] I. R. Petersen, M. R. James and P. Dupuis, Minimax optimal control of stochastic uncertain systems with relative entropy constraints, *IEEE Transactions on Automatic Control*, 45, 398–412, 2000.

[40] D. Schmeidler, Subjective probability and expected utility without additivity, *Econometrica*, 57, 571–587, 1989.

[41] T. Strzalecki, Axiomatic foundations of multiplier preferences, *Econometrica*, 79, 47–73, 2011.

[42] A. Wald, *Statistical decision functions*, John Wiley & Sons, New York, 1950.

[43] S. G. Walker, Bayesian inference with misspecified models, *Journal of Statistical Planning and Inference*, 143, 1621–1633, 2013.

[44] C. Zalinescu, *Convex analysis in general vector spaces*, World Scientific, Singapore, 2002.

B Additional material

In this appendix, we begin by proving few relevant properties of statistical distances which we discussed in Section 2.1. We then discuss the irrelevance of the convexity of the set Q for the entropic model (cf. Section B.2). We conclude by providing the proofs of few ancillary facts useful in obtaining and discussing our decision criterion (cf. Sections B.3 and B.4).

B.1 Statistical distances and divergences

We here collect few properties of statistical distances. In order to characterize variational statistical distances, we substantially need to prove that the function $c_Q : \Delta \to [0, \infty]$, defined by $c_Q(p) = \min_{q \in Q} c(p, q)$, is well defined, grounded, lower semicontinuous and convex. This fact follows from the following version of a well-known result (see, e.g., Fiacco and Kyparisis, 1986).

Lemma 11 *Let Q be a compact and convex subset of Δ^σ. If $c : \Delta \times Q \to [0, \infty]$ is a lower semicontinuous and convex function such that there exist $\bar{p} \in \Delta$ and $\bar{q} \in Q$ such that $c(\bar{p}, \bar{q}) = 0$, then $c_Q : \Delta \to [0, \infty]$ defined by*

$$c_Q(p) = \min_{q \in Q} c(p, q) \qquad \forall p \in \Delta$$

is well defined, grounded, lower semicontinuous and convex.

Proof Since c is lower semicontinuous and Q is non-empty and compact, c_Q is well defined. Moreover, we have that $0 \geq c(\bar{p}, \bar{q}) \geq c_Q(\bar{p}) \geq 0$, proving that c_Q is grounded. Even though $c(p, q)$ might be ∞ for some $(p, q) \in \Delta \times Q$, by the same proof of the Maximum Theorem (see, e.g., Lemma 17.30 in Aliprantis and Border, 2006), it follows that c_Q is lower semicontinuous. If $p_1, p_2 \in \Delta$, then define $q_1, q_2 \in Q$ to be such that

$$c(p_1, q_1) = \min_{q \in Q} c(p_1, q) = c_Q(p_1) \text{ and } c(p_2, q_2) = \min_{q \in Q} c(p_2, q) = c_Q(p_2)$$

Consider $\lambda \in (0, 1)$. Define $p_\lambda = \lambda p_1 + (1 - \lambda) p_2$ and $q_\lambda = \lambda q_1 + (1 - \lambda) q_2 \in Q$. Since c is jointly convex, it follows that

$$\begin{aligned} c_Q(p_\lambda) &= \min_{q \in Q} c(p_\lambda, q) \leq c(p_\lambda, q_\lambda) \leq \lambda c(p_1, q_1) + (1 - \lambda) c(p_2, q_2) \\ &= \lambda c_Q(p_1) + (1 - \lambda) c_Q(p_2) \end{aligned}$$

proving convexity. ■

Lemma 12 *Let $\mathcal{Q}$ consist of compact and convex subsets of Δ^σ. A lower semicontinuous and convex function $c : \Delta \times \mathcal{S} \to [0, \infty]$ is a variational statistical distance if and only if it satisfies the distance property:*

$$c(p, q) = 0 \iff p = q \tag{70}$$

Proof We first prove the "If" part. By (70) and since c is lower semicontinuous, (c.i) and (c.ii) are satisfied. Fix $Q \in \mathcal{Q}$. By (c.i) and since c restricted to $\Delta \times Q$ is jointly lower semicontinuous and convex, then we have that $p \mapsto \min_{q \in Q} c(p, q)$ is well defined, grounded, lower semicontinuous and convex. By (c.i), it follows that (C.i) is satisfied. By construction and since $p \mapsto \min_{q \in Q} c(p, q)$ is lower semicontinuous and convex and Q was arbitrarily chosen, (C.ii), (C.iii), (C.iv) as well as (C.v) are satisfied, proving that c is a variational statistical distance. As for the "Only if" part, it is trivial since a statistical distance, by definition, satisfies (70). ■

The next result shows, inter alia, that restricted ϕ-divergences are variational divergences.[44] A piece of notation and one of terminology: 1) we write $p \sim Q$ if there exists a control measure $q \in Q$ such that $p \sim q$;[45] 2) given a function $f : \Delta \to [0, \infty]$ we say it is *strictly convex* if, given any distinct $p, q \in \Delta$, we have $f(\alpha p + (1 - \alpha) q) < \alpha f(p) + (1 - \alpha) f(q)$ for all $\alpha \in (0, 1)$ such that $\alpha p + (1 - \alpha) q \in \operatorname{dom} f$.

Lemma 13 *Let $\mathcal{Q}$ consist of compact and convex subsets of Δ^σ. A restricted ϕ-divergence $D_\phi : \Delta \times \mathcal{S} \to [0, \infty]$ is a variational divergence. Moreover, for each $Q \in \mathcal{Q}$*

(i) if $q \in Q$, then $D_\phi(\cdot || q) : \Delta \to [0, \infty]$ is strictly convex;

(ii) if $p \in \Delta^\sigma$ and $p \sim Q$, then $D_\phi(p || \cdot) : Q \to [0, \infty]$ is strictly convex.

Proof It is well known that on $\Delta \times \Delta^\sigma$ the function D_ϕ is jointly lower semicontinuous and convex and satisfies the property

$$D_\phi(p || q) = 0 \iff p = q$$

The same properties are preserved by D_ϕ restricted to $\Delta \times \mathcal{S}$. By Lemma 12, it follows that $D_\phi : \Delta \times \mathcal{S} \to [0, \infty]$ is a variational statistical distance. Finally, by definition, we have that $D_\phi(p || q) = \infty$ whenever $p \notin \Delta^\sigma(q)$, yielding that it is a variational divergence. We next prove points (i) and (ii). Fix $Q \in \mathcal{Q}$.

[44] Though a routine result, for the sake of completeness, we provide a proof since we did not find one allowing for S being infinite (see Topsoe, 2001, p. 178 for the finite case).

[45] A probability $q \in Q$ is a *control measure* of Q if $q' \ll q$ for all $q' \in Q$. When Q is a compact and convex subset of Δ^σ, Q has a control measure (see, e.g., Maccheroni and Marinacci, 2001). Such a measure might not be unique, yet any two control measures of Q are equivalent. So, the notion $p \sim Q$ is well defined and independent of the chosen control measure.

(i). Consider $q \in Q$. Let $p', p'' \in \Delta$ and $\alpha \in (0,1)$ be such that $p' \neq p''$ and $D_\phi(\alpha p' + (1-\alpha) p'' || q) < \infty$. If either $D_\phi (p'||q)$ or $D_\phi (p''||q)$ are not finite, we trivially conclude that $D_\phi(\alpha p' + (1-\alpha) p''||q) < \infty = \alpha D_\phi (p'||q) + (1-\alpha) D_\phi (p''||q)$. Let us then assume that both $D_\phi (p'||q)$ and $D_\phi (p''||q)$ are finite. By definition of D_ϕ and since $\Delta^\sigma (q)$ is convex, this implies that $p', p'' \in \Delta^\sigma (q)$ as well as $\alpha p' + (1-\alpha) p'' \in \Delta^\sigma (q)$. Since p' and p'' are distinct, we have that dp'/dq and dp''/dq take different values on a set of strictly positive q-measure: call it $\tilde{S}$. Since ϕ is strictly convex, it follows that

$$\phi \left(\alpha \frac{dp'}{dq} (s) + (1-\alpha) \frac{dp''}{dq} (s) \right) < \alpha \phi \left(\frac{dp'}{dq} (s) \right) + (1-\alpha) \phi \left(\frac{dp''}{dq} (s) \right) \quad \forall s \in \tilde{S}$$

By definition of D_ϕ, this implies that

$$\begin{aligned}
D_\phi (\alpha p' + (1-\alpha) p''||q) &= \int_S \phi \left(\frac{d\left[\alpha p' + (1-\alpha) p''\right]}{dq} (s) \right) dq \\
&= \int_S \phi \left(\alpha \frac{dp'}{dq} (s) + (1-\alpha) \frac{dp''}{dq} (s) \right) dq \\
&= \int_{\tilde{S}} \phi \left(\alpha \frac{dp'}{dq} (s) + (1-\alpha) \frac{dp''}{dq} (s) \right) dq \\
&+ \int_{S \setminus \tilde{S}} \phi \left(\alpha \frac{dp'}{dq} (s) + (1-\alpha) \frac{dp''}{dq} (s) \right) dq \\
&< \alpha \int_S \phi \left(\frac{dp'}{dq} (s) \right) dq + (1-\alpha) \int_S \phi \left(\frac{dp''}{dq} (s) \right) dq \\
&= \alpha D_\phi (p'||q) + (1-\alpha) D_\phi (p''||q)
\end{aligned}$$

We conclude that $D_\phi (\cdot||q) : \Delta \to [0, \infty]$ is strictly convex.

(ii). Before starting, we make three observations.

a. Since Q is a non-empty, compact and convex subset of Δ^σ, note that there exists $\bar{q} \in Q$ such that $q \ll \bar{q}$ for all $q \in Q$. Since $p \sim Q$, we have that $p \sim \bar{q}$. This implies also that $q \ll p$ for all $q \in Q$.

b. If $q \sim p$, then $(dp/dq)^{-1}$ is well defined almost everywhere (with respect to either p or q) and can be chosen (after defining arbitrarily the function over a set of zero measure) to be the Radon-Nikodym derivative dq/dp.

c. Since ϕ is strictly convex, if we define $\phi^\star : (0, \infty) \to [0, \infty)$ by $\phi^\star (x) = x\phi (1/x)$ for all $x > 0$, then also $\phi^\star$ is strictly convex. By point b, if $p \in \Delta^\sigma$ and $q \in Q$ are such that $p \sim q$ and we define $\dot{p} = dp/dq$, then $p(\{\dot{p}=0\}) = 0 = q(\{\dot{p}=0\})$ and

$$\begin{aligned} D_\phi(p||q) &= \int_S \phi\left(\frac{dp}{dq}\right) dq = \int_{\{\dot{p}=0\}} \phi\left(\frac{dp}{dq}\right) dq + \int_{\{\dot{p}>0\}} \phi\left(\frac{dp}{dq}\right) dq \\ &= \int_{\{\dot{p}>0\}} \phi\left(\frac{1}{\left(\frac{dp}{dq}\right)^{-1}}\right) dq = \int_{\{\dot{p}>0\}} \phi^\star\left(\frac{dq}{dp}\right)\frac{dp}{dq} dq \\ &= \int_{\{\dot{p}>0\}} \phi^\star\left(\frac{dq}{dp}\right) dp \end{aligned}$$

We can now prove the statement. Let $q', q'' \in Q$ and $\alpha \in (0,1)$ be such that $q' \neq q''$ and $D_\phi(p||\alpha q' + (1-\alpha)q'') < \infty$. If either $D_\phi(p||q')$ or $D_\phi(p||q'')$ are not finite, we trivially conclude that $D_\phi(p||\alpha q' + (1-\alpha)q'') < \infty = \alpha D_\phi(p||q') + (1-\alpha) D_\phi(p||q'')$. Let us then assume that both $D_\phi(p||q')$ and $D_\phi(p||q'')$ are finite. By definition of D_ϕ, we can conclude that $p \ll q'$ and $p \ll q''$. By point a, this yields that $q' \sim p \sim q''$ and $p \sim \alpha q' + (1-\alpha)q''$. Since q' and q'' are distinct, we have that dq'/dp and dq''/dp take different values on a set of strictly positive p-measure. By point c, we have that

$$p\left(\left\{\frac{dp}{d[\alpha q' + (1-\alpha)q'']} = 0\right\}\right) = p\left(\left\{\frac{dp}{dq'} = 0\right\}\right) = p\left(\left\{\frac{dp}{dq''} = 0\right\}\right) = 0$$

Thus, by point c and since dq'/dp and dq''/dp take different values on a set of strictly positive p-measure and $\phi^\star$ is strictly convex, there exists a p-measure 1 set $\tilde{S}$ such that

$$\begin{aligned} D_\phi(p||\alpha q' + (1-\alpha)q'') &= \int_{\tilde{S}} \phi^\star\left(\frac{d[\alpha q' + (1-\alpha)q'']}{dp}\right) dp \\ &< \alpha \int_{\tilde{S}} \phi^\star\left(\frac{dq'}{dp}\right) dp + (1-\alpha)\int_{\tilde{S}} \phi^\star\left(\frac{dq''}{dp}\right) dp \\ &= \alpha D_\phi(p||q') + (1-\alpha) D_\phi(p||q'') \end{aligned}$$

proving point (ii). ■

Consider a finite set $Q = \{q_i\}_{i=1}^n$. Assume that for each q in the convex hull of Q there exists a unique collection $\{\mu_i^q\}_{i=1}^n \subseteq \mathbb{R}_+^n$ such that $\sum_{i=1}^n \mu_i^q = 1$ and $q = \sum_{i=1}^n \mu_i^q q_i$. Consider a function $d : \Delta \times Q \to [0,\infty]$ which is lower semicontinuous and convex in the first argument and such that $d(p,q) = 0$ if and only if $p = q$. Define $c : \Delta \times \operatorname{co} Q \to [0,\infty]$ by

$$c(p,q) = \min_{(p_i)_{i=1}^n \in \Delta^n : p = \sum_{i=1}^n \mu_i^q p_i} \sum_{i=1}^n d(p_i, q_i)\mu_i^q \qquad \forall (p,q) \in \Delta \times \operatorname{co} Q \tag{71}$$

Lemma 14 *If c is defined as in (71), then c is a jointly lower semicontinuous and convex variational statistical distance. Moreover, if $d(p,q) < \infty$ implies $p \ll q$, then c is also a* *divergence.*

Proof We endow Δ^n with the product topology. Clearly, Δ^n is compact. For each $p \in \Delta$ and $q \in \operatorname{co} Q$ set $\Psi(p,q) = \{(p_i)_{i=1}^n \in \Delta^n : p = \sum_{i=1}^n \mu_i^q p_i\}$. Note that $\Psi(p,q)$ is non-empty, closed (hence, compact) and convex. Since $d : \Delta \times Q \to [0,\infty]$ is lower semicontinuous in p, given q, we have that $(p_i)_{i=1}^n \mapsto \sum_{i=1}^n d(p_i, q_i)\mu_i^q$ is lower semicontinuous. Since $\Psi(p,q)$ is compact, this implies that c is well defined. Next, observe that if $p = q$, then $(\bar{p}_i)_{i=1}^n \in \Delta^n$ such that $\bar{p}_i = q_i$ for all $i \in \{1,...,n\}$ satisfies $p = q = \sum_{i=1}^n \mu_i^q q_i = \sum_{i=1}^n \mu_i^q \bar{p}_i$, that is, $(\bar{p}_i)_{i=1}^n \in \Psi(p,q) = \Psi(q,q)$ and $0 \le c(p,q) \le \sum_{i=1}^n d(\bar{p}_i, q_i)\mu_i^q = 0$. Vice versa, since $d \ge 0$, we have that if $c(p,q) = 0$, then there exists $(\bar{p}_i)_{i=1}^n \in \Psi(p,q)$ such that $c(p,q) = \sum_{i=1}^n d(\bar{p}_i, q_i)\mu_i^q = 0$, yielding that $\bar{p}_i = q_i$ for all $i \in \{1,...,n\}$ such that $\mu_i^q > 0$. Since $p = \sum_{i=1}^n \mu_i^q \bar{p}_i = \sum_{i:\mu_i^q>0} \mu_i^q \bar{p}_i$ and $q = \sum_{i=1}^n \mu_i^q q_i = \sum_{i:\mu_i^q>0} \mu_i^q q_i$, we can conclude that $p = q$. Consider $p, r \in \Delta$ as well as $q, q' \in \operatorname{co} Q$ and $\lambda \in (0,1)$. Let $(\bar{p}_i)_{i=1}^n \in \Psi(p,q)$ and $(\bar{r}_i)_{i=1}^n \in \Psi(r,q')$ be such that $c(p,q) = \sum_{i=1}^n d(\bar{p}_i, q_i)\mu_i^q$ and $c(r,q') = \sum_{i=1}^n d(\bar{r}_i, q_i)\mu_i^{q'}$. For each $i \in \{1,...,n\}$ set

$$\alpha_i = \begin{cases} \frac{\lambda\mu_i^q}{\lambda\mu_i^q + (1-\lambda)\mu_i^{q'}} & \text{if } \lambda\mu_i^q + (1-\lambda)\mu_i^{q'} > 0 \\ \frac{1}{2} & \text{if } \lambda\mu_i^q + (1-\lambda)\mu_i^{q'} = 0 \end{cases}$$

Clearly, we have that $\alpha_i \in [0,1]$ and

$$1 - \alpha_i = \begin{cases} \frac{(1-\lambda)\mu_i^{q'}}{\lambda\mu_i^q + (1-\lambda)\mu_i^{q'}} & \text{if } \lambda\mu_i^q + (1-\lambda)\mu_i^{q'} > 0 \\ \frac{1}{2} & \text{if } \lambda\mu_i^q + (1-\lambda)\mu_i^{q'} = 0 \end{cases} \qquad \forall i \in \{1,...,n\}$$

Define $(\hat{p}_i)_{i=1}^n \in \Delta^n$ to be such that $\hat{p}_i = \alpha_i \bar{p}_i + (1-\alpha_i)\bar{r}_i$ for all $i \in \{1,...,n\}$. Note that

$$\lambda q + (1-\lambda) q' = \lambda \sum_{i=1}^n \mu_i^q q_i + (1-\lambda) \sum_{i=1}^n \mu_i^{q'} q_i = \sum_{i=1}^n \left[\lambda\mu_i^q + (1-\lambda)\mu_i^{q'}\right] q_i$$

yielding that $\mu^{\lambda q + (1-\lambda)q'} = \lambda\mu^q + (1-\lambda)\mu^{q'}$. Moreover, since $\lambda\mu_i^q + (1-\lambda)\mu_i^{q'} = 0$ if and only if $\mu_i^q = \mu_i^{q'} = 0$, we have that

$$\begin{aligned} \sum_{i=1}^n \mu_i^{\lambda q+(1-\lambda)q'} \hat{p}_i &= \sum_{i=1}^n \mu_i^{\lambda q+(1-\lambda)q'} (\alpha_i \bar{p}_i + (1-\alpha_i)\bar{r}_i) \\ &= \sum_{i=1}^n \mu_i^{\lambda q+(1-\lambda)q'} \alpha_i \bar{p}_i + \sum_{i=1}^n \mu_i^{\lambda q+(1-\lambda)q'} (1-\alpha_i)\bar{r}_i \\ &= \lambda \sum_{i=1}^n \mu_i^q \bar{p}_i + (1-\lambda) \sum_{i=1}^n \mu_i^{q'} \bar{r}_i = \lambda p + (1-\lambda) r \end{aligned}$$

proving that $(\hat{p}_i)_{i=1}^n \in \Psi(\lambda p + (1-\lambda) r, \lambda q + (1-\lambda) q')$. Since d is convex in p, this implies that

$$\begin{aligned}
\lambda c(p,q)+(1-\lambda)c(r,q') &= \lambda\sum_{i=1}^{n} d(\bar{p}_i,q_i)\,\mu_i^{q}+(1-\lambda)\sum_{i=1}^{n} d(\bar{r}_i,q_i)\,\mu_i^{q'}\\
&= \sum_{i=1}^{n}\alpha_i d(\bar{p}_i,q_i)\,\mu_i^{\lambda q+(1-\lambda)q'}+\sum_{i=1}^{n}(1-\alpha_i)\,d(\bar{r}_i,q_i)\,\mu_i^{\lambda q+(1-\lambda)q'}\\
&= \sum_{i=1}^{n}\left[\alpha_i d(\bar{p}_i,q_i)+(1-\alpha_i)\,d(\bar{r}_i,q_i)\right]\mu_i^{\lambda q+(1-\lambda)q'}\\
&\geq \sum_{i=1}^{n} d(\alpha_i\bar{p}_i+(1-\alpha_i)\,\bar{r}_i,q_i)\,\mu_i^{\lambda q+(1-\lambda)q'}=\sum_{i=1}^{n} d(\hat{p}_i,q_i)\,\mu_i^{\lambda q+(1-\lambda)q'}\\
&\geq c(\lambda p+(1-\lambda)\,r,\lambda q+(1-\lambda)\,q')
\end{aligned}$$

yielding that c is jointly convex. Next, consider the map $\Gamma:\Delta^n\times\operatorname{co}Q\to[0,\infty]$ defined by

$$\Gamma\left((p_i)_{i=1}^n,q\right)=\sum_{i=1}^{n} d(p_i,q_i)\,\mu_i^{q}\qquad\forall\left((p_i)_{i=1}^n,q\right)\in\Delta^n\times\operatorname{co}Q$$

We endow $\Delta^n\times\operatorname{co}Q$ with the product topology. Consider a net $\left\{\left((p_{i,\alpha})_{i=1}^n,q_\alpha\right)\right\}_{\alpha\in A}$ which converges to $\left((p_i)_{i=1}^n,q\right)$. Observe that $\{(\mu_i^{q_\alpha})_{i=1}^n\}_{\alpha\in A}$ converges pointwise to $(\mu_i^{q})_{i=1}^n$, otherwise there would exist a subnet $\left\{\left(\mu_i^{q_{\alpha_\beta}}\right)_{i=1}^n\right\}_{\beta\in B}$ which converges to $(\bar{\mu}_i)_{i=1}^n\neq(\mu_i^{q})_{i=1}^n$. Since $\{q_\alpha\}_{\alpha\in A}$ converges to q, this would yield that

$$\sum_{i=1}^{n}\mu_i^{q}q_i=q=\lim_{\beta}q_{\alpha_\beta}=\lim_{\beta}\sum_{i=1}^{n}\mu_i^{q_{\alpha_\beta}}q_i=\sum_{i=1}^{n}\bar{\mu}_iq_i$$

Since $\sum_{i=1}^{n}\bar{\mu}_i=1$, it follows that $\mu_i^{q}=\bar{\mu}_i$ for all $i\in\{1,...,n\}$, a contradiction. Since d is lower semicontinuous, we can conclude that

$$\begin{aligned}
\Gamma\left((p_i)_{i=1}^n,q\right) &= \sum_{i=1}^{n} d(p_i,q_i)\,\mu_i^{q}\leq\sum_{i=1}^{n}\liminf_{\alpha} d(p_{i,\alpha},q_i)\lim_{\alpha}\mu_i^{q_\alpha}\\
&= \sum_{i=1}^{n}\liminf_{\alpha} d(p_{i,\alpha},q_i)\,\mu_i^{q_\alpha}\leq\liminf_{\alpha}\sum_{i=1}^{n} d(p_{i,\alpha},q_i)\,\mu_i^{q_\alpha}\\
&= \liminf_{\alpha}\Gamma\left((p_{i,\alpha})_{i=1}^n,q_\alpha\right)
\end{aligned}$$

proving that Γ is lower semicontinuous. Let $t\in\mathbb{R}$. Consider a net $\{(p_\alpha,q_\alpha)\}_{\alpha\in A}\in\Delta\times\operatorname{co}Q$ that converges to (p,q) and such that $c(p_\alpha,q_\alpha)\leq t$ for all $\alpha\in A$. For each $\alpha\in A$, consider $(\bar{p}_{i,\alpha})_{i=1}^n\in\Psi(p_\alpha,q_\alpha)$ such that $c(p_\alpha,q_\alpha)=\sum_{i=1}^{n} d(\bar{p}_{i,\alpha},q_i)\,\mu_i^{q_\alpha}=\Gamma\left((\bar{p}_{i,\alpha})_{i=1}^n,q_\alpha\right)$. Since Δ^n is compact and the previous part of the proof, there exists a subnet $\left\{\left(\bar{p}_{i,\alpha_\beta}\right)_{i=1}^n\right\}_{\beta\in B}$ which converges to

$(\bar{p}_i)_{i=1}^n \in \Delta^n$ while $\left\{\left(\mu_i^{q_{\alpha_\beta}}\right)_{i=1}^n\right\}_{\beta\in B}$ converges to $(\mu_i^q)_{i=1}^n$. Since $\left\{p_{\alpha_\beta}\right\}_{\beta\in B}$ converges to p, it follows that $p = \lim_\beta p_{\alpha_\beta} = \lim_\beta \sum_{i=1}^n \mu_i^{q_{\alpha_\beta}} \bar{p}_{i,\alpha_\beta} = \sum_{i=1}^n \mu_i^q \bar{p}_i$, yielding that $(\bar{p}_i)_{i=1}^n \in \Psi(p,q)$. By definition of c and since Γ is lower semicontinuous, this implies that

$$
\begin{aligned}
c(p,q) &\leq \sum_{i=1}^n d(\bar{p}_i, q_i)\,\mu_i^q = \Gamma\left((\bar{p}_i)_{i=1}^n, q\right) \leq \liminf_\beta \Gamma\left(\left(\bar{p}_{i,\alpha_\beta}\right)_{i=1}^n, q_{\alpha_\beta}\right) \\
&= \liminf_\beta c\left(p_{\alpha_\beta}, q_{\alpha_\beta}\right) \leq t
\end{aligned}
$$

proving that c is jointly lower semicontinuous. By Lemma 12, if $\mathcal{Q}$ is a collection of compact and convex subsets of Δ^σ such that $\mathcal{S} = \operatorname{co} Q$, then we can conclude that c is a variational statistical distance. Finally, assume that $d(p,q) < \infty$ implies $p \ll q$. Consider $(p,q) \in \Delta \times \operatorname{co} Q$ and assume that $c(p,q) < \infty$. Let $(\bar{p}_i)_{i=1}^n \in \Psi(p,q)$ be such that $c(p,q) = \sum_{i=1}^n d(\bar{p}_i, q_i)\,\mu_i^q$. Since $c(p,q) < \infty$, we have that $d(\bar{p}_i, q_i) < \infty$ for all $i \in \{1,...,n\}$ such that $\mu_i^q > 0$, proving that $\bar{p}_i \ll q_i$ for all $i \in \{1,...,n\}$ such that $\mu_i^q > 0$. Next, consider $A \in \Sigma$ such that $q(A) = 0$. Since $q = \sum_{i=1}^n \mu_i^q q_i$, we have that $q_i(A) = 0$ for all $i \in \{1,...,n\}$ such that $\mu_i^q > 0$, yielding that $\bar{p}_i(A) = 0$ for all $i \in \{1,...,n\}$ such that $\mu_i^q > 0$. Since $p = \sum_{i=1}^n \mu_i^q \bar{p}_i = \sum_{i:\mu_i^q>0} \mu_i^q \bar{p}_i$, this implies that $p(A) = 0$, that is, $p \ll q$, yielding that c is a divergence. ■

B.2 Non-convex set of structured models

Let us consider two decision makers who adopt criterion (19), the first one posits a, possibly non-convex, set of structured models Q and the second one posits its closed convex hull $\overline{\operatorname{co}}\, Q$. So, the second decision maker considers also all the mixtures of structured models posited by the first decision maker. Next we show that their preferences over acts actually agree. We deal with the case $\lambda \in (0,\infty)$, being $\lambda = \infty$ trivial. It is thus without loss of generality to assume that the set of posited structured models is convex, as it was assumed in mostly of the main text. Before doing so we prove formula (20). Observe that given a compact subset $Q \subseteq \Delta^\sigma$, be that convex or not, we have

$$
\begin{aligned}
\min_{p\in\Delta}\left\{\int u(f)\,dp + \lambda \min_{q\in Q} R(p||q)\right\} &= \min_{p\in\Delta}\min_{q\in Q}\left\{\int u(f)\,dp + \lambda R(p||q)\right\} \\
&= \min_{q\in Q}\min_{p\in\Delta}\left\{\int u(f)\,dp + \lambda R(p||q)\right\} \\
&= \min_{q\in Q}\phi_\lambda^{-1}\left(\int \phi_\lambda(u(f))\,dq\right)
\end{aligned}
$$

where $\phi_\lambda(t) = -e^{-\frac{1}{\lambda}t}$ for all $t \in \mathbb{R}$ where $\lambda > 0$.

Proposition 14 *If $Q \subseteq \Delta^\sigma$ is compact, then for each $f \in \mathcal{F}$*

$$\min_{p\in\Delta}\left\{\int u(f)\,dp + \lambda \min_{q\in Q} R(p||q)\right\} = \min_{p\in\Delta}\left\{\int u(f)\,dp + \lambda \min_{q\in\overline{\mathrm{co}}\,Q} R(p||q)\right\}$$

Proof First observe that $\overline{\mathrm{co}}\,Q \subseteq \Delta^\sigma$. Indeed, since Q is a compact subset of Δ^σ, the set function $\nu : \Sigma \to [0,1]$, defined by $\nu(E) = \min_{q\in Q} q(E)$ for all $E \in \Sigma$ is an exact capacity which is continuous at S. This implies that $Q \subseteq \mathrm{core}\,\nu \subseteq \Delta^\sigma$, yielding that $\overline{\mathrm{co}}\,Q \subseteq \mathrm{core}\,\nu \subseteq \Delta^\sigma$. Given what we have shown before we can conclude that

$$\begin{aligned}
\min_{p\in\Delta}\left\{\int u(f)\,dp + \lambda \min_{q\in Q} R(p||q)\right\} &= \min_{q\in Q} \phi_\lambda^{-1}\left(\int \phi_\lambda(u(f))\,dq\right) \\
&= \phi_\lambda^{-1}\left(\min_{q\in Q}\left(\int \phi_\lambda(u(f))\,dq\right)\right) \\
&= \phi_\lambda^{-1}\left(\min_{q\in\overline{\mathrm{co}}\,Q}\left(\int \phi_\lambda(u(f))\,dq\right)\right) \\
&= \min_{q\in\overline{\mathrm{co}}\,Q} \phi_\lambda^{-1}\left(\int \phi_\lambda(u(f))\,dq\right) \\
&= \min_{p\in\Delta}\left\{\int u(f)\,dp + \lambda \min_{q\in\overline{\mathrm{co}}\,Q} R(p||q)\right\}
\end{aligned}$$

proving the statement. ■

After (22), we claimed that the Gini criterion is a monotone version of the max-min mean-variance criterion. To be more precise, given a probability $q \in \Delta^\sigma$ and a weight $1/2\lambda > 0$ for the variance, the mean-variance criterion is not monotone over its entire domain, but it is normalized, translation invariant, and monotone in an area containing the constant functions (see Theorem 24 of Maccheroni et al., 2006). At the same time, the variational preference with cost function the Gini index $\lambda\chi^2(\cdot||q)$ is monotone and coincides with the mean-variance criterion over such an area. A similar argument, *mutatis mutandis*, holds for the max-min mean-variance criterion and our formula (21). This allows us to see the corresponding variational criteria as a monotonization of the corresponding mean-variance ones.

B.3 Main theorems: ancillary results

We begin by proving the two ancillary variational lemmas.

Proof of Lemma 3 We actually prove that (i)$\Longrightarrow$(ii)$\Longleftrightarrow$(iii), with equivalence when $\succsim$ is unbounded.

(i) implies (ii). Let $f \in \mathcal{F}$. It is enough to observe that $c(\bar{p}) = 0$ implies

$$V\left(x_f^{\bar{p}}\right) = u\left(x_f^{\bar{p}}\right) = \int u(f)\,d\bar{p} + c(\bar{p}) \geq \min_{p\in\Delta}\left\{\int u(f)\,dp + c(p)\right\} = V(f)$$

yielding that $x_f^{\bar{p}} \succsim f$.

(ii) implies (iii). Assume that $x_f^{\bar{p}} \succsim f$ for all $f \in \mathcal{F}$. Since $\succsim$ is complete and transitive, it follows that if $x \succ x_f^{\bar{p}}$, then $x \succ f$.

(iii) implies (ii). By contradiction, suppose that there exists $f \in \mathcal{F}$ such that $f \succ x_f^{\bar{p}}$. Let $x_f \in X$ be such that $x_f \sim f$. This implies that $x_f \succ x_f^{\bar{p}}$ and so $x_f \succ f$, a contradiction.

(ii) implies (i). Let $\succsim$ be unbounded. Assume that $x_f^{\bar{p}} \succsim f$ for all $f \in \mathcal{F}$, i.e., $V(f) \leq \int u(f)\, d\bar{p}$ for all $f \in \mathcal{F}$. So, $\bar{p}$ corresponds to a SEU preference that is less ambiguity averse than $\succsim$. By Lemma 32 of Maccheroni et al. (2006), we can conclude that $c(\bar{p}) = 0$. ■

Proof of Lemma 4 We begin by observing that in proving the two implications, Q being either compact or convex plays no role.

(i) implies (ii). Let $p \in \Delta \backslash \Delta^{\ll}(Q)$. It follows that there exists $A \in \Sigma$ such that $q(A) = 0$ for all $q \in Q$ as well as $p(A) > 0$. Define $I : B_0(\Sigma) \to \mathbb{R}$ by $I(\varphi) = \min_{p \in \Delta} \left\{ \int \varphi dp + c(p) \right\}$ for all $\varphi \in B_0(\Sigma)$. Since u is unbounded, for each $\lambda \in \mathbb{R}$ there exists $x_\lambda \in X$ such that $u(x_\lambda) = \lambda$. Similarly, there exists $y \in X$ such that $u(y) = 0$. For each $\lambda \in \mathbb{R}$ define $f_\lambda = x_\lambda A y$. By construction, we have that $f_\lambda \overset{Q}{=} y$ for all $\lambda \in \mathbb{R}$. This implies that $I(\lambda 1_A) = V(f_\lambda) = V(y) = I(0) = 0$ for all $\lambda \in \mathbb{R}$. By Maccheroni et al. (2006) and since u is unbounded and $p(A) > 0$, we have that

$$c(p) = \sup_{\varphi \in B_0(\Sigma)} \left\{ I(\varphi) - \int \varphi dp \right\} \geq \sup_{\lambda \in \mathbb{R}} \left\{ I(\lambda 1_A) - \lambda p(A) \right\} = \infty$$

Since p was arbitrarily chosen, it follows that $\operatorname{dom} c \subseteq \Delta^{\ll}(Q)$.

(ii) implies (i). Assume that $\operatorname{dom} c \subseteq \Delta^{\ll}(Q)$. If $f \overset{Q}{=} g$, then $u(f) \overset{Q}{=} u(g)$. This implies that $u(f) \overset{p}{=} u(g)$ for all $p \in \Delta^{\ll}(Q)$ and, in particular,

$$\begin{aligned} V(f) &= \min_{p \in \Delta} \left\{ \int u(f)\, dp + c(p) \right\} = \min_{p \in \Delta^{\ll}(Q)} \left\{ \int u(f)\, dp + c(p) \right\} \\ &= \min_{p \in \Delta^{\ll}(Q)} \left\{ \int u(g)\, dp + c(p) \right\} = \min_{p \in \Delta} \left\{ \int u(g)\, dp + c(p) \right\} = V(g) \end{aligned}$$

proving that $f \sim g$. ■

Proof of Lemma 7 We begin by showing that $\succeq^*$ is well defined and does not depend on the representing elements of ψ and φ. Assume that $f_1, f_2, g_1, g_2 \in \mathcal{F}$ are such that $u(f_i) = \varphi$ and $u(g_i) = \psi$ for all $i \in \{1, 2\}$. It follows that $u(f_1(s)) = u(f_2(s))$ and $u(g_1(s)) = u(g_2(s))$ for all $s \in S$. By Lemma 6, this implies that $f_1(s) \sim^* f_2(s)$ and $g_1(s) \sim^* g_2(s)$ for all $s \in S$. Since $\succsim^*$ is a preorder that satisfies monotonicity, this implies that $f_1 \sim^* f_2$ and $g_1 \sim^* g_2$. Since $\succsim^*$ is a preorder, if $f_1 \succsim^* g_1$, then

$$f_2 \succsim^* f_1 \succsim^* g_1 \succsim^* g_2 \implies f_2 \succsim^* g_2$$

that is, $f_1 \succsim^* g_1$ implies $f_2 \succsim^* g_2$. Similarly, we can prove that $f_2 \succsim^* g_2$ implies $f_1 \succsim^* g_1$. In other words, $f_1 \succsim^* g_1$ if and only if $f_2 \succsim^* g_2$, proving that $\succeq^*$ is well defined and does not depend on the representing elements of ψ and φ. It is immediate to prove that $\succeq^*$ is a preorder. We next prove properties 1–5.

1. Consider $\varphi, \psi \in B_0(\Sigma)$ and $k \in \mathbb{R}$. Assume that $\varphi \succeq^* \psi$. Let $f, g \in \mathcal{F}$ and $x, y \in X$ be such that $u(f) = 2\varphi$, $u(g) = 2\psi$, $u(x) = 0$ and $u(y) = 2k$. Since u is affine, it follows that

$$\begin{aligned} u\left(\frac{1}{2}f + \frac{1}{2}x\right) &= \frac{1}{2}u(f) + \frac{1}{2}u(x) = \varphi \succeq^* \psi \\ &= \frac{1}{2}u(g) + \frac{1}{2}u(x) = u\left(\frac{1}{2}g + \frac{1}{2}x\right) \end{aligned}$$

proving that $\frac{1}{2}f + \frac{1}{2}x \succsim^* \frac{1}{2}g + \frac{1}{2}x$. Since $\succsim^*$ satisfies weak c-independence and u is affine, we have that $\frac{1}{2}f + \frac{1}{2}y \succsim^* \frac{1}{2}g + \frac{1}{2}y$, yielding that

$$\begin{aligned} \varphi + k &= \frac{1}{2}u(f) + \frac{1}{2}u(y) = u\left(\frac{1}{2}f + \frac{1}{2}y\right) \succeq^* u\left(\frac{1}{2}g + \frac{1}{2}y\right) \\ &= \frac{1}{2}u(g) + \frac{1}{2}u(y) = \psi + k \end{aligned}$$

2. Consider $\varphi, \psi \in B_0(\Sigma)$ and $\{k_n\}_{n\in\mathbb{N}} \subseteq \mathbb{R}$ such that $k_n \uparrow k$ and $\varphi - k_n \succeq^* \psi$ for all $n \in \mathbb{N}$. We have two cases:

   (a) $k > 0$. Consider $f, g, h \in \mathcal{F}$ such that

   $$u(f) = \varphi,\ u(g) = \varphi - k \text{ and } u(h) = \psi$$

   Since $k > 0$ and $k_n \uparrow k$, there exists $\bar{n} \in \mathbb{N}$ such that $k_n > 0$ for all $n \geq \bar{n}$. Define $\lambda_n = 1 - k_n/k$ for all $n \in \mathbb{N}$. It follows that $\lambda_n \in [0,1]$ for all $n \geq \bar{n}$. Since u is affine, for each $n \geq \bar{n}$

   $$u(\lambda_n f + (1-\lambda_n) g) = \lambda_n u(f) + (1-\lambda_n) u(g) = \varphi - k_n \succeq^* \psi = u(h)$$

   yielding that $\lambda_n f + (1-\lambda_n) g \succsim^* h$ for all $n \geq \bar{n}$. Since $\succsim^*$ satisfies continuity and $\lambda_n \to 0$, we have that $g \succsim^* h$, that is,

   $$\varphi - k = u(g) \succeq^* u(h) = \psi$$

   (b) $k \leq 0$. Since $\{k_n\}_{n\in\mathbb{N}}$ is convergent, $\{k_n\}_{n\in\mathbb{N}}$ is bounded. Thus, there exists $h > 0$ such that $k_n + h > 0$ for all $n \in \mathbb{N}$. Moreover, $k_n + h \uparrow k + h > 0$. By point 1, we also have that $\varphi - (k_n + h) = (\varphi - k_n) - h \succeq^* \psi - h$ for all $n \in \mathbb{N}$. By subpoint a, we can conclude that $(\varphi - k) - h = \varphi - (k + h) \succeq^* \psi - h$. By point 1, we obtain that $\varphi - k \succeq^* \psi$.

3. Consider $\varphi, \psi \in B_0(\Sigma)$ such that $\varphi \geq \psi$. Let $f, g \in \mathcal{F}$ be such that $u(f) = \varphi$ and $u(g) = \psi$. It follows that $u(f(s)) \geq u(g(s))$ for all $s \in S$. By Lemma 6, this implies that $f(s) \succsim^* g(s)$ for all $s \in S$. Since $\succsim^*$ satisfies monotonicity, this implies that $f \succsim^* g$, yielding that $\varphi = u(f) \succeq^* u(g) = \psi$.

4. Consider $k, h \in \mathbb{R}$ and $\varphi \in B_0(\Sigma)$. We first assume that $k > h$ and $k = 0$. By point 3, we have that $\varphi = \varphi + k \succeq^* \varphi + h$. By contradiction, assume that $\varphi \not\succ^* \varphi + h$. It follows that $\varphi \sim^* \varphi + h$, yielding that $I = \{w \in \mathbb{R} : \varphi \sim^* \varphi + w\}$ is a non-empty set which contains 0 and h. We next prove that I is an unbounded interval, that is, $I = \mathbb{R}$. First, consider $w_1, w_2 \in I$. Without loss of generality, assume that $w_1 \geq w_2$. By point 3 and since $w_1, w_2 \in I$, we have that for each $\lambda \in (0, 1)$

$$\varphi \succeq^* \varphi + w_1 \succeq^* \varphi + (\lambda w_1 + (1 - \lambda) w_2) \succeq^* \varphi + w_2 \succeq^* \varphi$$

proving that $\varphi \sim^* \varphi + (\lambda w_1 + (1 - \lambda) w_2)$, that is, $\lambda w_1 + (1 - \lambda) w_2 \in I$. Next, we observe that $I \cap (-\infty, 0) \neq \emptyset \neq I \cap (0, \infty)$. Since $h \in I$ and $h < 0$, we have that $I \cap (-\infty, 0) \neq \emptyset$. Since I is an interval and $0, h \in I$, we have that $h/2 \in I$. By point 1 and since $\varphi \sim^* \varphi + h/2$, we have that $\varphi - h/2 \sim^* (\varphi + h/2) - h/2 = \varphi$, proving that $0 < -h/2 \in I \cap (0, \infty)$. By definition of I, note that if $w \in I \setminus \{0\}$, then $\varphi + w \sim^* \varphi$. By point 1 and since $w/2 \in I$ and $\succeq^*$ is a preorder, we have that $(\varphi + w) + w/2 \sim^* \varphi + w/2 \sim^* \varphi$, that is, $\frac{3}{2}w, \frac{1}{2}w \in I$. Since I is an interval, we have that either $\left[\frac{3}{2}w, \frac{1}{2}w\right] \subseteq I$ if $w < 0$ or $\left[\frac{1}{2}w, \frac{3}{2}w\right] \subseteq I$ if $w > 0$. This will help us in proving that I is unbounded from below and above. By contradiction, assume that I is bounded from below and define $m = \inf I$. Since $I \cap (-\infty, 0) \neq \emptyset$, we have that $m < 0$. Consider $\{w_n\}_{n \in \mathbb{N}} \subseteq I \cap (-\infty, 0)$ such that $w_n \downarrow m$. Since $\left[\frac{3}{2}w_n, \frac{1}{2}w_n\right] \subseteq I$ for all $n \in \mathbb{N}$, it follows that $m \leq \frac{3}{2}w_n$ for all $n \in \mathbb{N}$. By passing to the limit, we obtain that $m \leq \frac{3}{2}m < 0$, a contradiction. By contradiction, assume that I is bounded from above and define $M = \sup I$. Since $I \cap (0, \infty) \neq \emptyset$, we have that $M > 0$. Consider $\{w_n\}_{n \in \mathbb{N}} \subseteq I \cap (0, \infty)$ such that $w_n \uparrow M$. Since $\left[\frac{1}{2}w_n, \frac{3}{2}w_n\right] \subseteq I$ for all $n \in \mathbb{N}$, it follows that $M \geq \frac{3}{2}w_n$ for all $n \in \mathbb{N}$. By passing to the limit, we obtain that $M \geq \frac{3}{2}M > 0$, a contradiction. To sum up, I is a non-empty unbounded interval, that is, $I = \mathbb{R}$. This implies that $\varphi \sim^* \varphi + w$ for all $w \in \mathbb{R}$. In particular, select $w_1 = \|\varphi\|_\infty + 1$ and $w_2 = -\|\varphi\|_\infty - 1$. Since $\succeq^*$ is a preorder, we have that $\varphi + w_1 \sim^* \varphi + w_2$. Moreover, $\varphi + w_1 \geq 1 > -1 \geq \varphi + w_2$. By point 3, this implies that $\varphi + w_1 \succeq^* 1 \succeq^* -1 \succeq^* \varphi + w_2$. Since $\succeq^*$ is a preorder and $\varphi + w_1 \sim^* \varphi + w_2$, we can conclude that $1 \sim^* -1$. Note also that there exist $x, y \in X$ such that $u(x) = 1$ and $u(y) = -1$. By Lemma 6, this implies that $x \succ^* y$. By definition of $\succeq^*$ and since $u(x) = 1 \sim^* -1 = u(y)$, we also have that

$y \succsim^* x$, a contradiction. Thus, we proved that if $k > h$ and $k = 0$, then $\varphi + k \succ^* \varphi + h$. Assume simply that $k > h$. This implies that $0 > h - k$ and $\varphi \succ^* \varphi + (h - k)$. By point 1, we can conclude that $\varphi + k \succ^* \varphi + (h - k) + k = \varphi + h$.

5. Consider $\varphi, \psi, \xi \in B_0(\Sigma)$ and $\lambda \in (0, 1)$. Assume that $\varphi \succeq^* \xi$ and $\psi \succeq^* \xi$. Let $f, g, h \in \mathcal{F}$ be such that $u(f) = \varphi$, $u(g) = \psi$ and $u(h) = \xi$. By assumption and definition of $\succeq^*$, we have that $f \succsim^* h$ and $g \succsim^* h$. Since $\succsim^*$ satisfies convexity and u is affine, this implies that $\lambda f + (1 - \lambda) g \succsim^* h$, yielding that $\lambda\varphi + (1 - \lambda)\psi = \lambda u(f) + (1 - \lambda) u(g) = u(\lambda f + (1 - \lambda) g) \succeq^* u(h) = \xi$.

Points 1–5 prove the first part of the statement. Finally, consider $\varphi, \psi \in B_0(\Sigma)$. Note that there exist a partition $\{A_i\}_{i=1}^n \subseteq \Sigma$ of S and $\{\alpha_i\}_{i=1}^n$ and $\{\beta_i\}_{i=1}^n$ in $\mathbb{R}$ such that

$$\varphi = \sum_{i=1}^n \alpha_i 1_{A_i} \text{ and } \psi = \sum_{i=1}^n \beta_i 1_{A_i}$$

Note that $\{s \in S : \varphi(s) \neq \psi(s)\} = \cup_{i \in \{1,\dots,n\}:\alpha_i \neq \beta_i} A_i$. Since $\varphi \overset{Q}{=} \psi$, we have that $q(A_i) = 0$ for all $q \in Q$ and for all $i \in \{1, ..., n\}$ such that $\alpha_i \neq \beta_i$. Since u is unbounded, define $\{x_i\}_{i=1}^n \subseteq X$ to be such that $u(x_i) = \alpha_i$ for all $i \in \{1, ..., n\}$. Since u is unbounded, define $\{y_i\}_{i=1}^n \subseteq X$ to be such that $y_i = x_i$ for all $i \in \{1, ..., n\}$ such that $\alpha_i = \beta_i$ and $u(y_i) = \beta_i$ otherwise. Define $f, g : S \to X$ by $f(s) = x_i$ and $g(s) = y_i$ for all $s \in A_i$ and for all $i \in \{1, ..., n\}$. It is immediate to see that $f \overset{Q}{=} g$ as well as $u(f) = \varphi$ and $u(g) = \psi$. Since $\succsim^*$ is objectively Q-coherent, we have that $f \sim^* g$, yielding that $\varphi \sim^* \psi$ and proving the second part of the statement. ■

Proof of Lemma 9 Consider $\varphi \in B_0(\Sigma)$. Define $C_\varphi = \{k \in \mathbb{R} : \varphi - k \in U(\psi)\}$. Note that C_φ is non-empty. Indeed, if we set $k = -\|\varphi\|_\infty - \|\psi\|_\infty$, then we obtain that $\varphi - k = \varphi + \|\varphi\|_\infty + \|\psi\|_\infty \geq 0 + \|\psi\|_\infty \geq \psi \in U(\psi)$. By property 4 of Lemma 8, we can conclude that $\varphi - k \in U(\psi)$, that is, $k \in C_\varphi$. Since $U(\psi)$ is convex, it follows that C_φ is an interval. Since $\varphi \in B_0(\Sigma)$, note that there exists $\hat{k} \in \mathbb{R}$ such that $\psi \geq \varphi - \hat{k}$. It follows that $\psi \succeq^* \varphi - \hat{k}$. In particular, we can conclude that $\psi \succ^* \varphi - \left(\hat{k} + \varepsilon\right)$ for all $\varepsilon > 0$. This yields that C_φ is bounded from above. Finally, assume that $\{k_n\}_{n\in\mathbb{N}} \subseteq C_\varphi$ and $k_n \uparrow k$. By property 2 of Lemma 8, we can conclude that $k \in C_\varphi$. To sum up, C_φ is a non-empty bounded from above interval of $\mathbb{R}$ that satisfies the property

$$\{k_n\}_{n\in\mathbb{N}} \subseteq C_\varphi \text{ and } k_n \uparrow k \implies k \in C_\varphi \tag{72}$$

The first part yields that $\sup\{k \in \mathbb{R} : \varphi - k \in U(\psi)\} = \sup C_\varphi \in \mathbb{R}$ is well defined. By (72), we also have that $\sup C_\varphi \in C_\varphi$, that is, $\sup C_\varphi = \max C_\varphi$, proving that I_ψ is well defined. Next, we prove that I_ψ is a concave niveloid. We first show that I_ψ is monotone and translation invariant. By Proposition 2 of Cerreia-Vioglio et al. (2014), this implies that I_ψ is a niveloid. Rather than proving monotonicity, we prove that I_ψ is $\succeq^*$-consistent.[46] Consider $\varphi_1, \varphi_2 \in B_0(\Sigma)$ such that $\varphi_1 \succeq^* \varphi_2$. By the properties of $\succeq^*$ and definition of I_ψ, we have that

$$\varphi_1 - I_\psi(\varphi_2) \succeq^* \varphi_2 - I_\psi(\varphi_2) \text{ and } \varphi_2 - I_\psi(\varphi_2) \in U(\psi)$$

and, in particular, $\varphi_2 - I_\psi(\varphi_2) \succeq^* \psi$. Since $\succeq^*$ is a preorder, this implies that $\varphi_1 - I_\psi(\varphi_2) \succeq^* \psi$, that is, $\varphi_1 - I_\psi(\varphi_2) \in U(\psi)$ and $I_\psi(\varphi_2) \in C_{\varphi_1}$, proving that $I_\psi(\varphi_1) \geq I_\psi(\varphi_2)$. We next prove translation invariance. Consider $\varphi \in B_0(\Sigma)$ and $k \in \mathbb{R}$. By definition of I_ψ, we can conclude that

$$(\varphi + k) - (I_\psi(\varphi) + k) = \varphi - I_\psi(\varphi) \in U(\psi)$$

This implies that $I_\psi(\varphi) + k \in C_{\varphi+k}$ and, in particular, $I_\psi(\varphi + k) \geq I_\psi(\varphi) + k$. Since k and φ were arbitrarily chosen, we have that

$$I_\psi(\varphi + k) \geq I_\psi(\varphi) + k \qquad \forall \varphi \in B_0(\Sigma), \forall k \in \mathbb{R}$$

This yields that $I_\psi(\varphi + k) = I_\psi(\varphi) + k$ for all $\varphi \in B_0(\Sigma)$ and for all $k \in \mathbb{R}$.[47]

We move to prove that I_ψ is concave. Consider $\varphi_1, \varphi_2 \in B_0(\Sigma)$ and $\lambda \in (0,1)$. By definition of I_ψ, we have that

$$\varphi_1 - I_\psi(\varphi_1) \in U(\psi) \text{ and } \varphi_2 - I_\psi(\varphi_2) \in U(\psi)$$

Since $U(\psi)$ is convex, we have that

$$\begin{aligned}&(\lambda\varphi_1 + (1-\lambda)\varphi_2) - (\lambda I_\psi(\varphi_1) + (1-\lambda) I_\psi(\varphi_2)) \\ &= \lambda(\varphi_1 - I_\psi(\varphi_1)) + (1-\lambda)(\varphi_2 - I_\psi(\varphi_2)) \in U(\psi)\end{aligned}$$

yielding that $\lambda I_\psi(\varphi_1) + (1-\lambda) I_\psi(\varphi_2) \in C_{\lambda\varphi_1 + (1-\lambda)\varphi_2}$ and, in particular, $I_\psi(\lambda\varphi_1 + (1-\lambda)\varphi_2) \geq \lambda I_\psi(\varphi_1) + (1-\lambda) I_\psi(\varphi_2)$.

Finally, since $\psi \in U(\psi)$, note that $0 \in C_\psi$ and $I_\psi(\psi) \geq 0$. By definition of I_ψ, if $I_\psi(\psi) > 0$, then $\psi - I_\psi(\psi) \in U(\psi)$, a contradiction with property 3 of Lemma 8.

1. It is routine to check that $\bar{I}_\psi$ is a normalized concave niveloid which is $\succeq^*$-consistent.

2. Clearly, we have that if $\psi \sim^* \psi'$, then $U(\psi) = U(\psi')$, yielding that $I_\psi = I_{\psi'}$ and, in particular, $I_\psi(0) = I_{\psi'}(0)$ as well as $\bar{I}_\psi = \bar{I}_{\psi'}$. The point trivially follows. ■

Proof of Proposition 13 We begin by observing that:

$$|ca(\Sigma)| \leq |ca_+(\Sigma) \times ca_+(\Sigma)| = |ca_+(\Sigma)| = |(0,\infty) \times \Delta^\sigma| = |\Delta^\sigma|$$

---

[46] Since if $\varphi_1 \geq \varphi_2$, then $\varphi_1 \succeq^* \varphi_2$, it follows that $\succeq^*$-consistency implies monotonicity.

[47] Observe that if $\varphi \in B_0(\Sigma)$ and $k \in \mathbb{R}$, then $-k \in \mathbb{R}$ and

$$I_\psi(\varphi) = I_\psi((\varphi + k) - k) \geq I_\psi(\varphi + k) - k$$

yielding that $I_\psi(\varphi + k) \leq I_\psi(\varphi) + k$.

The first inequality holds because the map $g : ca(\Sigma) \to ca_+(\Sigma) \times ca_+(\Sigma)$, defined by $\mu \mapsto (\mu^+, \mu^-)$, is injective. By Theorem 1.4.5 of Srivastava (1998) and since Σ is non-trivial, we have that $ca_+(\Sigma)$ is infinite, yielding that a bijection justifying the first equality exists. As to the second equality, the map $g : ca_+(\Sigma) \setminus \{0\} \to (0, \infty) \times \Delta^\sigma$, defined by $\mu \mapsto (\mu(S), \mu/\mu(S))$, is a bijection and so $|ca_+(\Sigma) \setminus \{0\}| = |(0,\infty) \times \Delta^\sigma|$. By Theorem 1.3.1 of Srivastava (1998), we can conclude that $|ca_+(\Sigma)| = |ca_+(\Sigma) \setminus \{0\}| = |(0,\infty) \times \Delta^\sigma|$. As to the last equality, by Theorem 1.4.5 and Exercise 1.5.1 of Srivastava (1998), being $|(0,\infty)| = |(0,1)| \leq |\Delta^\sigma|$, we have $|\Delta^\sigma| \leq |(0,\infty) \times \Delta^\sigma| = |(0,1) \times \Delta^\sigma| \leq |\Delta^\sigma \times \Delta^\sigma| = |\Delta^\sigma|$, yielding that $|(0,\infty) \times \Delta^\sigma| = |\Delta^\sigma|$.

We conclude that $|ca(\Sigma)| \leq |\Delta^\sigma|$, that is, there exists an injective map $g : ca(\Sigma) \to \Delta^\sigma$. Since Q is a compact and convex subset of Δ^σ, there exists $\bar{q} \in Q$ such that $q \ll \bar{q}$ for all $q \in Q$. We define $h : V \to ca(\Sigma)$ by

$$h([\psi])(A) = \int_A \psi d\bar{q} \qquad \forall A \in \Sigma$$

Note that h is well defined. For, if $\psi' \in [\psi]$, that is, $\psi \overset{Q}{=} \psi'$, then $\psi \overset{\bar{q}}{=} \psi'$, yielding that $\int_A \psi d\bar{q} = \int_A \psi' d\bar{q}$ for all $A \in \Sigma$. Similarly, $h([\psi]) = h([\psi'])$ implies that $\psi \overset{\bar{q}}{=} \psi'$. Since $q \ll \bar{q}$ for all $q \in Q$, this implies that $\psi \overset{Q}{=} \psi'$ and $[\psi] = [\psi']$, proving h is injective. This implies that $\tilde{f} = g \circ h$ is a well defined injective function from V to Δ^σ. Clearly, we have that $|\Delta^\sigma| \geq \left|\tilde{f}(V)\right| \geq |[0,1]|$. Since (S, Σ) is a standard Borel space and Q is convex and $|Q| \geq 2$, we also have that $|[0,1]| \geq |\Delta^\sigma| \geq |Q| \geq |[0,1]|$. This implies that $|V| = \left|\tilde{f}(V)\right| = |Q|$, proving the statement. ■

B.4 Analysis of the decision criterion: missing proofs

The proof of Proposition 1 follows from the following lemma. Here, as usual, ϕ is extended to $\mathbb{R}$ by setting $\phi(t) = +\infty$ if $t \notin [0, \infty)$. In particular, ϕ^* is non-decreasing.

Lemma 15 *For each $Q \subseteq \Delta^\sigma$ and each $\lambda \in (0, \infty)$,*

$$\inf_{p \in \Delta} \left\{ \int u(f)\, dp + \lambda \inf_{q \in Q} D_\phi(p||q) \right\} = \lambda \inf_{q \in Q} \sup_{\eta \in \mathbb{R}} \left\{ \eta - \int \phi^* \left(\eta - \frac{u(f)}{\lambda} \right) dq \right\}$$

for all $u : X \to \mathbb{R}$ and all $f : S \to X$ such that $u \circ f$ is bounded and Σ-measurable.

Proof By Theorem 4.2 of Ben-Tal and Teboulle (2007), for each $q \in \Delta^\sigma$ it holds

$$\inf_{p \in \Delta} \left\{ \int \xi dp + D_\phi(p||q) \right\} = \sup_{\eta \in \mathbb{R}} \left\{ \eta - \int \phi^*(\eta - \xi)\, dq \right\}$$

for all $\xi \in L^\infty(q)$. Then, if $u \circ f$ is bounded and measurable, from $u \circ f \in L^\infty(q)$ for all $q \in \Delta^\sigma$, it follows that

$$\inf_{p\in\Delta}\left\{\int u(f)\,dp+\lambda D_\phi(p||q)\right\}=\lambda\inf_{p\in\Delta}\left\{\int\frac{u(f)}{\lambda}dp+D_\phi(p||q)\right\}$$
$$=\lambda\sup_{\eta\in\mathbb{R}}\left\{\eta-\int\phi^*\left(\eta-\frac{u(f)}{\lambda}\right)dq\right\}$$

for all $\lambda>0$, as desired. By taking the inf over Q on both sides of the equation, the statement follows. ■

Proof of Proposition 1 In view of the last lemma, it is enough to observe that, if $f:S\to X$ is simple and measurable, then $u\circ f$ is simple and Σ-measurable for all $u:X\to\mathbb{R}$ and the infima are achieved. ■

Proof of Proposition 2 First, note that $\min_{q\in Q}R(p||q)=0$ if and only if $p\in Q$. Indeed, we have that

$$\min_{q\in Q}R(p||q)=0\iff\exists\bar{q}\in Q\text{ s.t. }R(p||\bar{q})=0\iff\exists\bar{q}\in Q\text{ s.t. }p=\bar{q}$$

Define $\lambda_n=n$ for all $n\in\mathbb{N}$. For each $n\in\mathbb{N}$, we have $\lambda_n\min_{q\in Q}R(p||q)=0$ if and only if $p\in Q$. So, for each $p\in\Delta$,

$$\lim_n\lambda_n\min_{q\in Q}R(p||q)=\begin{cases}0 & \text{if } p\in Q\\ +\infty & \text{if } p\notin Q\end{cases}$$

Since $\lambda_n\min_{q\in Q}R(p||q)=0$ for each $n\in\mathbb{N}$ if and only if $p\in Q$, by Proposition 12 of Maccheroni et al. (2006) we have

$$\lim_n\min_{p\in\Delta}\left\{\int u(f)\,dp+\lambda_n\min_{q\in Q}R(p||q)\right\}=\min_{q\in Q}\int u(f)\,dq\quad\forall f\in\mathcal{F}$$

Finally, by (23), we have that for each $f\in\mathcal{F}$

$$\min_{q\in Q}\int u(f)\,dq\le\lim_n\min_{p\in\Delta}\left\{\int u(f)\,dp+\lambda_n\min_{q\in Q}R(p||q)\right\}$$
$$\le\lim_{\lambda\uparrow\infty}\min_{p\in\Delta}\left\{\int u(f)\,dp+\lambda\min_{q\in Q}R(p||q)\right\}\le\min_{q\in Q}\int u(f)\,dq$$

yielding the statement. ■

Proof of Proposition 8 (i) Let $\hat{f}\in F$ be optimal. By (28), if there is $g\in F$ such that $g\succ\!\!\succ^*_Q\hat{f}$, then $g\succ_Q\hat{f}$, a contradiction with $\hat{f}$ being optimal. We conclude that $\hat{f}$ is weakly admissible. A similar argument proves that there is no $g\in F$ such that $g\succ^*_Q\hat{f}$ when (29) holds.

(ii) Suppose $\hat{f} \in F$ is the unique optimal act, that is, $\hat{f} \succ_Q f$ for all $f \in F \setminus \left\{ \hat{f} \right\}$. If $g \in F$ is such that $g \succ_Q^* \hat{f}$, then $g \neq \hat{f}$ and $g \succsim_Q \hat{f}$. In turn, this implies $g \succsim_Q \hat{f} \succ_Q g$, a contradiction. We conclude that $\hat{f}$ is admissible. ■

Proof of Proposition 9 Since $Q \subseteq Q'$, it follows that $\min_{q \in Q} c(p,q) \geq \min_{q \in Q'} c(p,q)$ for all $p \in \Delta$. We thus have

$$\min_{p \in \Delta} \left\{ \int u(f)\, dp + \min_{q \in Q} c(p,q) \right\} \geq \min_{p \in \Delta} \left\{ \int u(f)\, dp + \min_{q \in Q'} c(p,q) \right\} \qquad \forall f \in F$$

yielding that $v(Q) \geq v(Q')$. Next, fix Q and assume that the sup in (34) is achieved. Let $\bar{f} \in F$ be such that

$$\min_{p \in \Delta} \left\{ \int u\left(\bar{f}\right) dp + \min_{q \in Q} c(p,q) \right\} = v(Q)$$

By contradiction, assume that $\bar{f} \in F / F_Q^*$. By Proposition 5 and since $\bar{f} \notin F_Q^*$ and $\bar{f} \in F$, there exists $g \in F$ such that $g \ggcurly_Q^* \bar{f}$, that is, there exists $\varepsilon > 0$ such that

$$\min_{p \in \Delta} \left\{ \int u(g)\, dp + c(p,q) \right\} \geq \min_{p \in \Delta} \left\{ \int u\left(\bar{f}\right) dp + c(p,q) \right\} + \varepsilon \qquad \forall q \in Q$$

Since g is finite-valued, this implies that $v(Q) < \infty$ and

$$\begin{aligned} v(Q) &\geq \min_{p \in \Delta} \left\{ \int u(g)\, dp + \min_{q \in Q} c(p,q) \right\} = \min_{p \in \Delta} \min_{q \in Q} \left\{ \int u(g)\, dp + c(p,q) \right\} \\ &\geq \inf_{q \in Q} \min_{p \in \Delta} \left\{ \int u(g)\, dp + c(p,q) \right\} \geq \inf_{q \in Q} \min_{p \in \Delta} \left\{ \int u\left(\bar{f}\right) dp + c(p,q) \right\} + \varepsilon \\ &\geq \min_{p \in \Delta} \min_{q \in Q} \left\{ \int u\left(\bar{f}\right) dp + c(p,q) \right\} + \varepsilon = \min_{p \in \Delta} \left\{ \int u\left(\bar{f}\right) dp + \min_{q \in Q} c(p,q) \right\} + \varepsilon \\ &= v(Q) + \varepsilon \end{aligned}$$

a contradiction. ■

References

[1] C. D. Aliprantis and K. C. Border, *Infinite dimensional analysis*, 3rd ed., Springer, New York, 2006.

[2] A. Ben-Tal and M. Teboulle, An old-new concept of convex risk measures: The optimized certainty equivalent, *Mathematical Finance*, 17, 449–476, 2007.

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

[4] A. V. Fiacco and J. Kyparisis, Convexity and concavity properties of the optimal value function in parametric nonlinear programming, *Journal of Optimization Theory and Applications*, 48, 95–126, 1986.

[5] F. Maccheroni and M. Marinacci, A Heine-Borel Theorem for $ba\,(\Sigma)$, *RISEC*, 48, 353–362, 2001.

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

[7] S. M. Srivastava, *A course on Borel sets*, Springer, New York, 1998.

[8] F. Topsoe, Basic concepts, identities and inequalities – the toolkit of information theory, *Entropy*, 3, 162–190, 2001.