A more robust definition of multiple priors*

Paolo Ghirardato§ Marciano Siniscalchi¶

June 2009
Revised: June 2010

Abstract

This paper provides a multiple-priors representation of ambiguous beliefs à la Ghirardato, Maccheroni, and Marinacci (2004) and Nehring (2002) for any preference that is (i) *monotonic*, (ii) *Bernoullian*, i.e. admits an affine utility representation when restricted to constant acts, and (iii) suitably *continuous*. Monotonicity is the main substantive assumption: we do not require either Certainty Independence or Uncertainty Aversion. We characterize the set of ambiguous beliefs in terms of Clarke-Rockafellar differentials. This allows us to provide an explicit calculation of the set of priors for several recent decision models: multiplier preferences, the smooth ambiguity model, the vector expected utility model, as well as confidence function, variational, general "uncertainty-averse" preferences, and mean-dispersion preferences. With an eye towards applications, our results indicate how assumptions about the individual's perception of ambiguity translate to parametric restrictions in these functional representations.

*We thank Simone Cerreia-Vioglio, Theo Diasakos, Fabio Maccheroni, Mark Machina, Pietro Ortoleva and Tomasz Strzalecki, as well as audiences at the 2009 RUD and SAET conferences, the Workshop in honor of Daniel Ellsberg (Vienna, May 2010), and at seminars in Rice, Northwestern, Penn, Princeton and Montreal for helpful comments. The usual disclaimer applies. Ghirardato is also grateful to the Italian MIUR for financial support.

§DSMA and Collegio Carlo Alberto, Università di Torino; ghirardato@econ.unito.it

¶Northwestern University; marciano@northwestern.edu

1 Introduction

Sets of probabilities, and the related notions of "upper and lower probabilities" employed by statisticians (Smith, 1961; Dempster, 1967), have long been a key ingredient of *intuitions* about ambiguity in both theoretical and applied work. For example, to rationalize the modal choices in his now-famous examples, Ellsberg (1961, p. 61) suggested that the thought process of his subjects may have unfolded as follows:

> [o]ut of the set Y of all possible distributions there remains a set Y^0 of distributions that still seem 'reasonable,' ... that his information — perceived as scanty, unreliable, *ambiguous* — does not permit him confidently to rule out."

Similar sentiments are echoed, for instance, by Gilboa and Schmeidler (1989, p. 142) in their seminal contribution. More recently, Klibanoff, Marinacci, and Mukerji (2005, p. 1850) focus on "behavior in instances where the [decision-maker]'s information is *explicitly* consistent with multiple probabilities on the state space relevant to the decision at hand." The theory of "model uncertainty" (e.g. Hansen and Sargent, 2001) has a very similar intuitive motivation: "[t]he decision maker believes that data will come from an unknown member of a *set* of unspecified models near his approximating model." (Hansen and Sargent, 2008, p. 5). For analogous intepretations in applied work, see e.g. Epstein and Wang (1994, p. 283), Mukerji (1998, pp. 1207–1208), Billot, Chateauneuf, Gilboa, and Tallon (2000, pp. 685–6), or Ju and Miao (2009, p. 5).

At the same time, sets of probabilities, or "multiple priors," also play a prominent *formal* role in several popular models of choice in the face of ambiguity. For instance, according to the maxmin-expected utility (MEU) model of Gilboa and Schmeidler (1989), the individual evaluates an uncertain prospect ("act") h, mapping states in a set S to outcomes in a set X, according to their minimum expected utility over a weak*-closed, convex set D of probabilities: formally, $V(h) = \min_{Q \in D} \mathrm{E}_Q[u \circ h]$. In the multiplier model of Hansen and Sargent (2001), the evaluating functional is instead $V(h) = \min_{Q \in \Delta(S)} \mathrm{E}_Q[u \circ h] + \theta R(Q||P)$, where $\Delta(S)$ is the set of all measures on S, $\theta \geq 0$ is a parameter, and R denotes the relative entropy of Q with respect to the "approximating model" P. In the smooth ambiguity model of Klibanoff et al. (2005), the evaluating functional is a double expectation: $V(h) = \int_{\Delta(S)} \phi \left(\mathrm{E}_Q[u \circ h] \right) d\mu(Q)$, where ϕ and μ are the "second-order" utility function and probability measure respectively.

The aim of the present paper is to provide a robust, preference-based link between the intuitive interpretation of sets of priors in both theory and applications, and their formal role in functional representations of preferences. To this end, we adopt a formal definition of *relevant priors*, which relates the above intuition to preferences, and show that it identifies a unique set of probabilities under mild regularity conditions. This provides a robust subjective foundation of the notion of "multiple priors," analogous to Machina and Schmeidler (1992)'s foundation of "single-prior" non-expected utility. The main result of this paper shows that the definition we adopt is operational; specifically, we provide a general "differential" characterization, which we then specialize to compute relevant priors for virtually all models of choice under ambiguity, including the most recent ones.

Linking the intuitive and formal roles of priors has significant practical consequences. Applications of ambiguity-sensitive decision models typically rely on substantive assumptions about agents' beliefs and perceived ambiguity. Our results characterize the preferences consistent with such assumptions; in particular, given a specific functional representation of preferences, *they identify the parametric restrictions that correctly represent the intended assumptions on perceived ambiguity*. Similarly, our results clarify the scope and interpretation of comparative-statics assertions. This is all the more important in light of the fact that the link between parameters and relevant priors is *not* always obvious. We provide an extended example in Sec. 2.

Our "robust" derivation of ambiguous beliefs also has potential direct applications to information economics. For instance, sets of priors representing ambiguous beliefs play a crucial role in several recent results on the absence of trade and efficient risk-sharing (e.g. Billot et al., 2000). These results are obtained under the assumption of Ambiguity Hedging (see below), so a question arises as to their robustness to less extreme assumptions on ambiguity attitudes. Our findings may provide the tools to investigate this and other issues of economic significance.

This paper is organized as follows. Section 2 provides a heuristic discussion of our results. Section 3 discusses the related literature, with special emphasis on Ghirardato et al. (2004, GMM henceforth) and Nehring (2002, 2007). Section 4 introduces notation and other preliminaries. Section 5 describes the preference axioms we adopt and provides a functional representation of preferences. Section 6 provides the main results: it defines relevant priors, states our differential characterization, and then calculates relevant priors for a variety of decision models.

Section 7 discusses crisp acts and unambiguous events for MBC preferences. All proofs are in the Appendix. A companion paper, Ghirardato and Siniscalchi (2010), analyzes the general class of preferences we study in this paper; it examines their representation, attitudes toward ambiguity, and probabilistic sophistication, and also features some applications.

2 Heuristic treatment

Relevant priors

The notion of relevant priors that we adopt is essentially due to GMM and Nehring (2002); in particular, our definition is a slight restatement of Nehring's notion of "Bernoulli priors" (see Nehring, 2002, p.27). For brevity, we henceforth refer to these contributions as GMM/N. Suppose that the individual's preferences over the (convex) set X of consequences are represented by a Bernoulli utility function u; notice that this is the case for all the models we discussed above. We deem a weak*-closed, convex set D of probabilities, or priors, over a state space S *relevant* if the following two conditions hold:

(i) For every pair of acts f and g, if $\mathrm{E}_Q[u \circ f] \geq \mathrm{E}_Q[u \circ g]$ for all probabilities $Q \in D$, then f is weakly preferred to g; and

(ii) for every weak*-closed, convex, proper subset set D' of D, there exist acts f and g such that $\mathrm{E}_Q[u \circ f] \geq \mathrm{E}_Q[u \circ g]$ for all probabilities $Q \in D'$, but g is strictly preferred to f.

To interpret, note first that, for a fixed Bernoulli utility u, we can identify probabilities in the candidate set D with alternative *expected-utility decision models* that the individual deems sensible.[1] If every such sensible model ranks f above g, the individual should arguably prefer f to g: this is precisely what Condition (i) requires.

A moment's reflection shows that this requirement is too weak in and of itself: any "sufficiently large" set of priors (for instance, the set $\Delta(S)$ of all priors) satisfies it. We then require

[1] This is consistent with the way probabilities are treated in all the functional specifications discussed above (e.g. MEU), and indeed with most models of ambiguity-sensitive preferences in the literature. However, one advantage of our approach is that it is "modular": although we do not do so here, one replace expected utility with another preference functional over lotteries which involves a cardinal utility function.

that, loosely speaking, the *entire* candidate set of priors should "matter": if only *some* of the expected-utility models the individual deems sensible rank one act above another, the individual's preferences could still rank them the opposite way. Condition (ii) embodies this requirement: more precisely, it asserts that, for any proper (weak*-closed, convex) subset D' of the candidate set D, we can find acts that are ranked one way by priors in D', and the opposite way by the individual.

We suggest that the definition of relevant priors captures the spirit of the quotations in the Introduction. The question arises whether a set of relevant priors exists for a given preference relation, and whether, if one exists, it is unique. Our first result, contained in Propositions 3 and 4, delivers existence *and* uniqueness of relevant priors for a broad class of preferences, discussed below. Furthermore, it does so by drawing a tight connection with the work of GMM/N; we discuss this in more detail in Sec. 3.

Our second and main contribution is to make this definition of relevant priors *operational*. Except in trivial or very special cases, Conditions (i) and (ii) are cumbersome to verify directly. Our Theorem 5 provides a solution. The preferences we consider (details will be provided momentarily) admit a representation via a pair (I, u), where u is a Bernoulli utility function and I is a suitable functional over bounded, measurable real functions: an act f is preferred to another act g if and only if $I(u \circ f) \geq I(u \circ g)$. Theorem 5 states that for such preferences the set of relevant priors is, up to convex closure, the *union of all normalized Clarke-Rockafellar differentials of* I (Rockafellar, 1979), evaluated at all points in the interior of its domain.

We then leverage this general result to explicitly compute the set of relevant priors for a variety of recent decision models: we consider multiplier preferences (Hansen and Sargent, 2001) in §6.2.1; the smooth ambiguity model of Klibanoff et al. (2005) in §6.2.2; the vector expected utility model (Siniscalchi, 2009) in §6.2.3; confidence function (Chateauneuf and Faro, 2009), variational (Maccheroni, Marinacci, and Rustichini, 2006), and general "uncertainty-averse" preferences (Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio, 2008) in §6.2.5; and monotone mean-dispersion preferences (Grant and Polak, 2007) in §6.2.4. These results complement analogous calculations carried out in GMM/N for earlier decision models, such as MEU, Choquet Expected Utility (Schmeidler, 1989), and of course subjective expected utility. Their calculations deliver intuitive answers for such earlier models; this lends support to the notion of

relevant priors that we, too adopt. Our results concerning more recent, richer decision models can yield new insight into the properties and distinguishing features of such models; moreover, they provide further validation for the GMM/N approach that we generalize.

It is important to emphasize that, when we deem a set of priors "relevant," we mean that it is relevant for the individual's *behavior*, in the sense of Conditions (i) and (ii) above. For instance, the individual could contemplate a large set of priors, but if her behavior is consistent with a single subjective expected utility model, then we deem the corresponding prior to be her sole relevant prior. This is a consequence of our revealed-preference approach: see GMM, p. 137, for further discussion.

An example

To illustrate our results, consider a simple specification of smooth ambiguity preferences. The state space is $S = \{s_1, s_2\}$; outcomes are monetary, with linear utility for simplicity. Preferences over acts (i.e., vectors in $\mathbb{R}^2_+$) are represented by $V(f) = \mu(\{Q_1\})\phi\left(\mathrm{E}_{Q_1}[f]\right) + \mu(\{Q_2\})\phi\left(\mathrm{E}_{Q_2}[f]\right)$, where $Q_1, Q_2 \in \Delta(S)$ are (first-order) probability distributions on S and $\mathrm{E}_{Q_1}, \mathrm{E}_{Q_2}$ are the corresponding expectation operators, $\mu \in \Delta(\Delta(S))$ is the second-order probability, with support $\{Q_1, Q_2\}$, and ϕ is the second-order utility function. Henceforth, let $\mu \equiv \mu(\{Q_1\}) = 1 - \mu(\{Q_2\})$.

If ϕ is continuously differentiable, so is V; in this case, the Clarke-Rockafeller differential coincides with the gradient of V. Our main result, Theorem 5, then provides an expression for the set C in terms of this gradient: see Eq. (6) in §6.2.2. This expression can be further simplified for common specifications of the second-order utility ϕ. To build intuition, we consider two simple examples (see Appendix C.9.2 for detailed calculations). First, if ϕ is a CARA utility function, then, regardless of the absolute risk-aversion parameter in ϕ,

$$C = \{\alpha Q_1 + (1-\alpha)Q_2 : \alpha \in [0,1]\}, \tag{1}$$

i.e., the convex hull of the support of the second-order probability. Second, if ϕ is a CRRA function with relative risk aversion parameter $\gamma > 0$, $\mu = \frac{1}{2}$, and $Q_1(\{s_1\}) = Q_2(\{s_2\}) = \frac{3}{4}$, then

$$C = \left\{\alpha Q_1 + (1-\alpha)Q_2 : \alpha \in \left[\frac{1}{1+3^\gamma}, \frac{3^\gamma}{1+3^\gamma}\right]\right\}, \tag{2}$$

Hence, in the CARA case, C only depends upon the support of the second-order prior, while in the CRRA case, both μ and ϕ determine the set of relevant priors. In particular, higher values of

γ yield larger sets of relevant priors. (Recall that the curvature of ϕ characterizes second-order risk aversion, which Klibanoff et al. (2005) interpret as a dislike for ambiguity.) Thus, in the CRRA case, one cannot model the individual's attitudes toward ambiguity separately from her set of behaviorally relevant priors. This is instead possible in the CARA case.

These results can assist in the interpretation of comparative-statics results. To illustrate, consider a simple, two-period consumption-savings problem. Three individuals, Ann, Bob and Chloe, share the same initial wealth $w_0 > 0$, and have smooth ambiguity preferences with the same second-order beliefs, characterized by $\mu = \frac{1}{2}$ as above. The remaining parameters of their preferences are given in the second and third columns of Table 1; the fourth column indicates the corresponding set C (up to five significant digits), calculated as per Appendix C.9.2. (Note that Ann and Bob's set C is that following from Eq. (2).)

	$Q_1(\{s_1\}) = Q_2(\{s_2\})$	γ	C	α^*
Ann	0.75	2	$\{Q \in \Delta(S) : Q(\{s_1\}) \in [0.30000, 0.70000]\}$	0.77882
Bob	0.75	3	$\{Q \in \Delta(S) : Q(\{s_1\}) \in [0.26786, 0.73214]\}$	0.51844
Chloe	0.72361	3	$\{Q \in \Delta(S) : Q(\{s_1\}) \in [0.30000, 0.70000]\}$	0.64930

Table 1: A consumption-savings example

Each individual must allocate their initial wealth at time 0 between consumption c_0 and savings $w_0 - c_0$; the latter are invested in an asset $\tilde{r}$ whose uncertain payoffs per unit of investment are $\tilde{r}(s_1) = 1.2$ and $\tilde{r}(s_2) = 0.9$ in states s_1 and s_2 respectively. Consumption at time 1 is then equal to $\tilde{r} \cdot (w_0 - c_0)$. For simplicity, assume that the utility from the (non-random) consumption stream (c_0, c_1) is $u(c_0, c_1) = c_0 + \delta c_1$. Each individual then solves

$$\max_{c_0 \in [0, w_0]} \left(\frac{1}{2(1-\gamma)} \left(\mathrm{E}_{Q_1}[c_0 + \delta \tilde{r} \cdot (w_0 - c_0)] \right)^{1-\gamma} + \frac{1}{2(1-\gamma)} \left(\mathrm{E}_{Q_2}[(c_0 + \delta \tilde{r} \cdot (w_0 - c_0)] \right)^{1-\gamma} \right).$$

Straightforward calculations show that, under suitable assumptions on the parameters (which are satisfied in this example), optimal savings are given by $s_0 = \alpha^* w_0$, where

$$a^* = \frac{1 - \left(\frac{1 - \delta \mathrm{E}_{Q_2}[\tilde{r}]}{-1 + \delta \mathrm{E}_{Q_1}[\tilde{r}]} \right)^{-\frac{1}{\gamma}}}{(1 - \delta \mathrm{E}_{Q_1}[\tilde{r}]) + \left(\frac{1 - \delta \mathrm{E}_{Q_2}[\tilde{r}]}{-1 + \delta \mathrm{E}_{Q_1}[\tilde{r}]} \right)^{-\frac{1}{\gamma}} (-1 + \delta \mathrm{E}_{Q_2}[\tilde{r}])}.$$

The fourth column of Table 1 calculates the optimal savings rate α^* for Ann, Bob and Chloe, assuming that $\delta = 0.96$. As may be seen, Bob and Chloe, who have a higher second-order risk-aversion parameter than Ann, save less. However, unlike Chloe, Bob also has a *larger set of relevant priors* than Ann; that is, he perceives more ambiguity than Ann and Chloe.

Consequently, the fact that Bob saves less than Ann cannot be wholly ascribed to the fact that he is more "ambiguity-averse": the difference in the set of relevant priors is a confounding factor. On the other hand, Chloe perceives the same set of relevant priors as Ann (the value of $Q_1(\{s_1\}) = Q_2(\{s_2\})$ in Table 1 is chosen so as to ensure this), so the difference in their savings rate can be safely ascribed to differences in their ambiguity attitudes.

Robustness: MBC preferences

To ensure that our identification of relevant priors apply to a diverse collection of decision models, our results require only a minimal common structure on preferences. We adopt a setting à la Anscombe and Aumann (1963), and allow for any preference relation that is (i) *monotonic*, (ii) *Bernoullian*, i.e. admits an expected utility representation when restricted to constant acts, and (iii) suitably *continuous*. We refer to such preferences as "MBC" for short. Virtually all preference specifications employed in applications, including all of our initial examples, are members of the MBC class. Thus, while assumptions (i)-(iii) are not innocuous—in particular, we do assume EU risk preferences (but see footnote 1)— they are satisfied for a broad and diverse family of preference models of theoretical and applied interest.

In particular, our preference assumptions *do not restrict ambiguity attitudes* in any way. To elaborate, as is well-known, MEU preferences embody a specific, strong form of aversion to ambiguity, which we call Ambiguity Hedging.[2] The same axiom appears in the characterization of variational preferences (Maccheroni et al., 2006), of which multiplier preferences are a special case; for a general analysis of "uncertainty-averse preferences," see Cerreia-Vioglio et al. (2008). Smooth ambiguity preferences featuring a concave second-order utility function display a (stronger) form of Ambiguity Hedging.

Ambiguity Hedging is a mathematically convenient assumption, naturally related to con-

[2]Following Schmeidler (1989), this is usually named "Uncertainty Aversion." See Ghirardato and Marinacci (2002) for an explanation of our choice of terminology.

vexity in the realms of traditional consumer and firm theory. However, beginning with Epstein (1999), concerns have been voiced regarding the extent to which it satisfactorily formalizes the behavioral phenomenon of aversion to ambiguity; alternatives have been proposed (Epstein, 1999; Chateauneuf and Tallon, 2002, and GMM). More recently, Baillon, L'Haridon, and Placido (forthcoming) employ two intriguing examples due to Machina (2009) to show that (i) decision models that incorporate the Ambiguity Hedging axiom, as well as smooth ambiguity preferences with a concave second-order utility, preclude intuitive patterns of behavior that are clearly indicative of dislike for ambiguity; and (ii) such patterns are actually observed in experimental settings. Furthermore, Ambiguity Hedging and related axioms impose a *uniform* dislike for ambiguity: loosely speaking, they require that, for instance, the individual avoid bets on *any* event she perceives as being ambiguous. By way of contrast, several influential contributions have demonstrated experimentally that the same individual may exhibit different ambiguity attitudes, depending upon the choice problem she faces (see, e.g., Heath and Tversky, 1991), the stakes involved (Hogarth and Einhorn, 1990), etc.

Again with an eye toward robustness, we *do not impose any form of "independence" on non-constant acts*. To elaborate, the classical Anscombe and Aumann (1963) Independence axiom requires that the preference ranking between two acts f and g be preserved under (state-wise) mixtures with *any* third act h. MEU preferences satisfy (see Gilboa and Schmeidler, 1989) a relaxed form of this axiom, called Certainty Independence, which only requires that preferences be preserved under mixtures with *constant acts*. This assumption is mathematically convenient, as we illustrate below in our discussion of the results in GMM. However, Certainty Independence also significantly restricts ambiguity attitudes: for instance, a linear-utility decision maker who satisfies this axiom displays the same attitudes toward ambiguity regardless of her initial wealth, as well as the magnitude of the payoffs.[3] Recent theoretical contributions by Klibanoff et al. (2005) and Grant and Polak (2007) underscore the importance of relaxing or removing these restrictions; this is also in accordance with economic intuition in risk settings.

[3] In the case of non-neutral risk attitudes, ambiguity attitudes are independent of initial *utility*, etc.

3 Related literature

As noted above, our definition of relevant priors builds upon ideas in GMM/N.[4] Nehring (2002) (see also Nehring, 2007) deems a DM "utility sophisticated" with respect to some set D of priors if her preferences satisfy Condition (i) above; he calls D the set of "Bernoulli priors" if it is the minimal set (with respect to set inclusion) for which this is the case. Nehring argues that, under the behavioral assumptions considered by GMM, the set C of priors that GMM identify (see below) is the set of Bernoulli priors. Formally, our Condition (ii) makes Nehring's minimality requirement a bit more explicit.

The interpretation and use of relevant priors in the papers just cited is slightly different from ours. Nehring (2002) suggests that every DM comes equipped with *two* primitive, consistent binary relations: her preferences over acts, and an incomplete likelihood ordering over events. Under a non-atomicity condition (which we do not require), the likelihood ordering admits a Bewley-style representation, via a convex-ranged set Π of priors. The notion of "Bernoulli priors" is then mainly intended as a way to relate what is behaviorally identified (GMM's set C) and the DM's "true" beliefs (the set Π): cf. Theorem 3 in Nehring (2002). Nehring (2007) instead emphasizes utility sophistication as a property of independent interest.

We do not distinguish between a "true but unobservable" set of priors and the relevant priors we identify from behavior. Our objective is to identify the collection of alternative models of the world that have behavioral, rather than epistemic significance. Moreover, our main contribution is the differential characterization of relevant priors, which Nehring does not provide.

The paper closest to ours is GMM. These authors derive a set C of priors as representation *à la* Bewley of the DM's *unambiguous preference relation*, which is constructed from her primitive preferences. The analysis in GMM is carried out under the assumption that the primitive preference satisfies Certainty Independence; our Proposition 3 shows that their construction extends to our environment. More importantly, our Proposition 4 shows that the set C they obtain is, in fact, the *unique* set of relevant priors for any primitive MBC preference relation.

In contrast, our notion of relevant priors is (i) stated in terms of the individual's primitive preference relation, rather than the unambiguous preference ordering; and (ii) explicitly moti-

[4] For the precise connection between the work of GMM and Nehring, see GMM and Nehring (2002).

vated by the interpretation of priors as alternative models that the DM considers possible.

GMM also provide a "differential" characterization of the set C as the *Clarke differential* (Clarke, 1983) of the representing functional I (see above), evaluated at the origin. They then use this characterization to compute relevant priors for models that fit their assumptions.

Our differential characterization of relevant priors (Theorem 5) is intended to deal with a broader, more diverse class of preferences; for this reason, the analytical tools we use to establish it are different from GMM's. The main reason is that, under Certainty Independence, the Clarke derivative and differential of I enjoy especially convenient properties that are not satisfied in our environment (see the discussion following Def. 3). We obtain the GMM result as a special case in Corollary 8, and also consider partial weakenings of Certainty Independence—see Corollaries 6 and 7.

Other related literature

Gilboa, Maccheroni, Marinacci, and Schmeidler (2010) consider a DM who is endowed with a possibly incomplete preference over acts reflecting "objective" information, and a complete preference reflecting her actual behavior. These preferences are related via two requirements: "Consistency" states that, if f is objectively superior to g, then it should also be behaviorally preferred; "Caution" essentially states that the DM is pessimistic when comparing acts that are not objectively ranked. The objective preference has a Bewley-style representation via a set C of priors; under the stated requirements, the behavioral preference is MEU, with priors C.

While there are natural similarities, our objectives are clearly different. We do not posit the existence of objective information, and, as noted above, do not restrict ambiguity attitudes. Again, our main contribution is the operational characterization of relevant priors.

Finally, Siniscalchi (2006) proposes a related notion of "plausible priors." Loosely speaking, a prior is deemed plausible if it is the *unique* probability that provides an expected-utility representation of the individual's preferences over a subset of acts. The cited paper provides axioms (which include Certainty Independence) under which plausible priors exist.

The main difference with the present paper, and with the GMM/N approach, is the fact that plausible priors are identified individually, rather than as elements of a set. This requires restrictions on preferences that we do not need (in addition to Certainty Independence).

4 Notation and preliminaries

We consider a state space S, endowed with a sigma-algebra Σ. The notation $B_0(\Sigma,\Gamma)$ indicates the set of simple Σ–measurable real functions on S with values in the interval[5] $\Gamma \subset \mathbb{R}$, endowed with the topology induced by the supremum norm; for simplicity, write $B_0(\Sigma,\mathbb{R})$ as $B_0(\Sigma)$.

Similarly, we denote by $B_b(\Sigma,\Gamma)$ the set of Σ–measurable real functions $a : S \to \mathbb{R}$ for which there exist $\alpha,\beta \in \Gamma$ such that $\alpha \geq a(s) \geq \beta$ for all $s \in S$; note that, since Γ is assumed to be an interval, $a(s) \in \Gamma$ for all $s \in S$. We equip the set $B_b(\Sigma,\Gamma)$ with the supremum norm, and let $B(\Sigma) \equiv B_b(\Sigma,\mathbb{R})$. Recall that, since Σ is a sigma-algebra, $B(\Sigma)$ is the closure of $B_0(\Sigma)$, and it is a Banach space.

The set of finitely additive probabilities on Σ is denoted $ba_1(\Sigma)$. The (relative) weak* topology on $ba_1(\Sigma)$ is the topology induced by $B_0(\Sigma)$ or, equivalently, by $B(\Sigma)$.

If B is one of the spaces $B_0(\Sigma,\Gamma)$ or $B_b(\Sigma,\Gamma)$ for some interval Γ, a functional $I : B \to \mathbb{R}$ is: **monotonic** if $I(a) \geq I(b)$ for all $a \geq b$; **continuous** if it is sup-norm continuous; **normalized** if $I(\alpha 1_S) = \alpha$ for all $\alpha \in \Gamma$; **isotone** if, for all $\alpha,\beta \in \Gamma$, $I(\alpha 1_S) \geq I(\beta 1_S)$ if and only if $\alpha \geq \beta$; **translation-invariant** if $I(a+\alpha 1_S) = I(a)+\alpha$ for all $a \in B$ and $\alpha \in \mathbb{R}$ such that $a+\alpha 1_S \in B$; **positively homogeneous** if $I(\alpha a) = \alpha I(a)$ for all $a \in B$ and $\alpha \in \mathbb{R}_+$ such that $\alpha a \in B$; and **constant-linear** if it is translation-invariant and positively homogeneous.

Finally, fix a convex subset X of a vector space. (Simple) acts are Σ-measurable functions $f : S \to X$ such that $f(S) = \{f(s) : s \in S\}$ is finite; the set of all (simple) acts is denoted by $\mathscr{F}$. We define mixtures of acts pointwise: for any $\alpha \in [0,1]$, $\alpha f + (1-\alpha)g$ is the act that delivers the prize $\alpha f(s) + (1-\alpha)g(s)$ in state s. Given $f, g \in \mathscr{F}$ and $A \in \Sigma$, we denote by $f\,A\,g$ the act in $\mathscr{F}$ which yields $f(s)$ for $s \in A$ and $g(s)$ for $s \in A^c \equiv S \setminus A$.

5 MBC preferences

5.1 Axioms

The main object of interest is a preference relation $\succcurlyeq$ on $\mathscr{F}$. The first axiom is standard.

[5] By interval, we mean a convex subset of the real line, which may be open or closed on the left or right, and may also be unbounded on one or both sides.

Axiom 1 (Weak Order) $\succcurlyeq$ *is complete and transitive on* $\mathscr{F}$.

The next two axioms, which employ the convex structure of X, are familiar from expected utility. As usual, $\succ$ (resp. $\sim$) denotes the asymmetric (resp. symmetric) component of $\succcurlyeq$, and we abuse notation by identifying the prize x with the constant act that delivers x for every s.

Axiom 2 (Constant Independence) $\forall x, y, z \in X$, $\forall \lambda \in (0,1]$*:* $x \succ y$ *implies* $\lambda x + (1-\lambda)z \succ \lambda y + (1-\lambda)z$.

Axiom 3 (Constant Archimedean) $\forall x, y \in X$ *and* $f \in \mathscr{F}$*: if* $x \succ f \succ y$ *then there are* $\alpha, \beta \in (0,1)$ *such that* $\alpha x + (1-\alpha)y \succ f \succ \beta x + (1-\beta)y$.

As it is well-known, the above three axioms (restricted to X) imply the existence of a *Bernoulli utility index* on X; that is, $u : X \to \mathbb{R}$ which is affine and represents the restriction of $\succcurlyeq$ to X. Thus, they embody what we called the *Bernoullian* structure of our preference. Furthermore,these axioms also imply the existence of *certainty equivalents* for all acts $f \in \mathscr{F}$.

The next axiom is key, and it applies to all of $\mathscr{F}$.

Axiom 4 (Monotonicity) $\forall f, g \in \mathscr{F}$*: if* $f(s) \succcurlyeq g(s)$ *for all* $s \in S$ *then* $f \succcurlyeq g$.

A first departure from GMM in required at this point. Alongside Monotonicity, the Certainty Independence axiom implies that preferences are "uniformly continuous."[6] Since we drop Certainty Independence, here we explicitly assume a form of continuity.

Furthermore, functional representations of preferences via uniformly continuous functionals defined on some subset B_0 of $B_0(\Sigma)$ can be extended by standard arguments to representations on the closure of B_0 in $B(\Sigma)$, with non-empty interior. This is important for our purposes, because the techniques we use apply to functionals defined on Banach, rather than normed spaces. Norm continuity alone is not sufficient to ensure the existence of such an extension, and uniform continuity is more than is actually needed. Instead, we propose an axiom that is both necessary and sufficient for the existence of a suitable, unique continuous extension. Our axiom, *Cauchy Continuity*, adapts a well-known condition (Bourbaki, 1998, §8.5, Theorem 1) to

[6]More generally, Maccheroni et al. (2006) obtain uniform continuity as a consequence of Monotonicity and Weak Certainty Independence (Axiom 9 below).

the present decision-theoretic setting; Observation 1, preceding the proof of Proposition 1 in the Appendix, provides additional discussion, and motivates the name of our axiom.

We require some notation. Following Gilboa and Schmeidler (1989), let $\mathscr{F}_b$ denote the collection of functions $f : S \to X$ that satisfy the following two properties:

(i) for all $x \in X$, $\{s \in S : f(s) \succ x\} \in \Sigma$; and

(ii) there exist $x, y \in X$ such that $x \succcurlyeq f(s) \succcurlyeq y$ for all $s \in S$.

Informally, $\mathscr{F}_b$ is the set of acts that are both measurable and bounded with respect to the restriction of the preference $\succcurlyeq$ to X.

Now fix a sequence of acts $(f^k) \subset \mathscr{F}$ and an act $f \in \mathscr{F}_b$. Say that (f^k) **converges to** f, written $f^k \to f$, iff, for all prizes $x, y \in X$ with $x \succ y$, there exists K such that $k \geq K$ implies for all $s \in S$

$$\frac{1}{2}f(s)+\frac{1}{2}y \prec \frac{1}{2}f^k(s)+\frac{1}{2}x \quad \text{and} \quad \frac{1}{2}f^k(s)+\frac{1}{2}y \prec \frac{1}{2}f(s)+\frac{1}{2}x.$$

The proof of Proposition 1 shows that this definition corresponds to sup-norm convergence in the space of utility profiles. Note that, like the definition of $\mathscr{F}_b$, a statement such as $f^k \to f$ involves preferences over X exclusively.

We can now state our continuity assumption.

Axiom 5 (Cauchy Continuity) *Consider sequences $(f^k) \subset \mathscr{F}$, $(x^k) \subset X$ such that $f^k \to f \in \mathscr{F}_b$. If $f^k \sim x^k$ for all k, then there exists $x \in X$ such that $x^k \to x$.*

Three observations are in order. First of all, Axiom 5 only involves preferences over simple acts in $\mathscr{F}$; after all, that is the domain of the preference relation $\succcurlyeq$ we assume given. To elaborate, the assumption $f^k \to f$ concerns preferences over X, even if f is not a simple act; furthermore, each act f^k is simple, so $f^k \sim x^k$ is a well-defined statement about preferences over $\mathscr{F}$. In particular, the axiom does not require that "$f \sim x$": indeed, at this stage such a statement would be meaningless in case $f \notin \mathscr{F}$, because $\succcurlyeq$ is *not* defined on $\mathscr{F}_b$.

Second, if the limit act f *is*, in fact, simple, then the proof of Proposition 1 below shows that in this case $f \sim x$ must follow. Furthermore, the same result implies that it *is* possible to extend $\succcurlyeq$ from $\mathscr{F}$ to $\mathscr{F}_b$ uniquely, and in this case the extended preference relation *will* necessarily deem f and x indifferent.

Finally, Proposition 1 also implies that, if the DM were to have suitably continuous preferences over all of $\mathscr{F}_b$, then the restriction of her preferences to $\mathscr{F}$ would satisfy Axiom 5. In other

words, Cauchy Continuity is *necessary and sufficient* for the existence of a continuous extension of preferences from $\mathscr{F}$ to $\mathscr{F}_b$.

Our final axiom is standard.

Axiom 6 (Non-triviality) *There exist $x, y \in X$ such that $x \succ y$.*

5.2 Basic functional representation

Proposition 1 *A preference relation $\succcurlyeq$ satisfies Axioms 1–6 if and only if there exists a non-constant, affine function $u : X \to \mathbb{R}$ and a monotonic, normalized, continuous functional $I : B_b(\Sigma, u(X)) \to \mathbb{R}$ such that for all $f, g \in \mathscr{F}$*

$$f \succcurlyeq g \iff I(u \circ f) \geq I(u \circ g). \tag{3}$$

Moreover, if (I_v, v) also satisfies Eq. (3), and I_v is normalized, then there exist $\lambda, \mu \in \mathbb{R}$ with $\lambda > 0$ such that $v(x) = \lambda u(x) + \mu$ for all $x \in X$, and $I_v(b) = \lambda I(\lambda^{-1}[b - \mu]) + \mu$ for all $b \in B_b(\Sigma, v(X))$.

We emphasize that Proposition 1 establishes the equivalence between axioms on a preference relation $\succcurlyeq$ on *simple* acts, and the existence of a monotonic, normalized and continuous functional I on a set of *bounded* functions. This allows for a succinct statement of the result that involves standard properties of functionals, such as continuity; we could of course provide an (equivalent) characterization in terms of the restriction of I to $B_0(\Sigma, u(X))$, but this would require adding to continuity a functional counterpart (Cauchy continuity) to Axiom 5. As noted above, our analysis requires a functional defined on a suitable subset of the Banach space $B(\Sigma)$, which is why we chose to employ the current formulation.

Notice that differently from Lemma 1 in GMM, the functional I depends upon the normalization chosen for the utility function; the *pair* (I, u) is unique.

Henceforth, in light of Proposition 1, a binary relation $\succcurlyeq$ on $\mathscr{F}$ that satisfies Axioms 1–6 will be called an **MBC preference** (for **M**onotonic, **B**ernoullian and **C**ontinuous).

Finally, it is convenient to allow for representation of preferences by means of functionals that are *not* normalized. The following straightforward result guarantees that, as long as such functionals are isotone, they can be re-normalized while preserving the essential properties of monotonicity and, especially, continuity.

Lemma 2 *Let Γ be an interval with non-empty interior, and consider an isotone, monotonic, continuous functional $I : B_b(\Sigma,\Gamma) \to \mathbb{R}$. Then there exists a unique monotonic, continuous and normalized functional $\hat{I} : B_b(\Sigma,\Gamma) \to \mathbb{R}$ such that $I(a) \geq I(b)$ if and only if $\hat{I}(a) \geq \hat{I}(b)$ for all $a, b \in B_b(\Sigma,\Gamma)$.*

5.3 Additional axioms and properties of I

We conclude by listing five axioms that are employed in specific models, and imply that the preference can be represented per Proposition 1 by a pair (I,u) in which the functional I enjoys additional properties. The first is the well-known Ambiguity Hedging axiom due to Schmeidler (1989). It is equivalent to the existence of a representation with I quasiconcave.

Axiom 7 (Ambiguity Hedging) *For all $f, g \in \mathscr{F}$ and $\lambda \in [0,1]$: $f \sim g$ implies $\lambda f + (1-\lambda)g \succcurlyeq g$.*

Next, we recall three weakenings of the Anscombe and Aumann (1963) Independence axiom. The first is due to Gilboa and Schmeidler (1989), and is equivalent to (the existence of a representation with) a constant-linear I.

Axiom 8 (Certainty Independence) *For all $f, g \in \mathscr{F}$, $x \in X$, and $\lambda \in (0,1]$: $f \succ g$ if and only if $\lambda f + (1-\lambda)x \succ \lambda g + (1-\lambda)x$.*

The second is due to Maccheroni et al. (2006); it is equivalent to translation invariance of I.

Axiom 9 (Weak Certainty Independence) *For all $f, g \in \mathscr{F}$, $x, y \in X$, and $\lambda \in (0,1)$: $\lambda f + (1-\lambda)x \succcurlyeq \lambda g + (1-\lambda)x$ implies $\lambda f + (1-\lambda)y \succcurlyeq \lambda g + (1-\lambda)y$.*

For the third axiom, fix a distinguished prize $x_0 \in X$; given a representation (I,u) with $u(x_0) = 0$, it is equivalent to positive homogeneity of I (Cerreia-Vioglio et al., 2008; Chateauneuf and Faro, 2009).

Axiom 10 (Homotheticity) *For all $f, g \in \mathscr{F}$ and $\lambda, \mu \in (0,1]$: $\lambda f + (1-\lambda)x_0 \succcurlyeq \lambda g + (1-\lambda)x_0$ implies $\mu f + (1-\mu)x_0 \succcurlyeq \mu g + (1-\mu)x_0$.*

6 Relevant priors: definition and characterization

We now formalize the central notion of this paper (cf. Sec. 2):

Definition 1 Let $\succcurlyeq$ be an MBC preference with representation (I,u). A weak*-closed, convex set $D \subset ba_1(\Sigma)$ is a set of *relevant priors for* $\succcurlyeq$ iff it satisfies the following two properties:

(i) for all $f,g \in \mathscr{F}$, $Q(u\circ f) \geq Q(u\circ g)$ for all $Q \in D$ implies $f \succcurlyeq g$; and

(ii) if $D' \subsetneq D$ is non-empty, weak*-closed and convex, there exist $f,g \in \mathscr{F}$ such that $Q(u\circ f) \geq Q(u\circ g)$ for all $Q \in D'$, but $f \prec g$.

As promised in the Introduction, we establish existence and uniqueness of relevant priors by relating this notion to the definition of unambiguous preference due to GMM/N:

Definition 2 Let $f,g \in \mathscr{F}$. We say that f is **unambiguously preferred to** g, denoted $f \succcurlyeq^* g$, if for every $h \in \mathscr{F}$ and all $\lambda \in (0,1]$, $\lambda f + (1-\lambda)h \succcurlyeq \lambda g + (1-\lambda)h$.

Relative to GMM/N, our basic axioms on the primitive preference $\succcurlyeq$, are weaker: we do not impose Certainty Independence. Nevertheless, the unambiguous preference $\succcurlyeq^*$ still admits a Bewley-style representation as in GMM/N:

Proposition 3 *Consider an MBC preference $\succcurlyeq$, and let $\succcurlyeq^*$ be as in Definition 2. Then there exists a non-empty, unique, convex and weak*-closed set $C \subset ba_1(\Sigma)$ such that, for all $f,g \in \mathscr{F}$,*

$$f \succcurlyeq^* g \quad \Longleftrightarrow \quad \int u\circ f \, dP \geq \int u\circ g \, dP \quad \text{for all } P \in C,$$

where u is the function obtained in Proposition 1. Moreover, C is independent of the choice of normalization of u.

The last sentence —which follows from the structure of the Bewley-style representation and the uniqueness of C given u— shows that C is cardinally invariant, even though I is not.

We can finally relate the set C to the set of relevant priors:

Proposition 4 *For any MBC preference $\succcurlyeq$, the set C in Proposition 3 is the unique set of relevant priors.*

6.1 Clarke-Rockafellar differential(s) and C

We now come to the main contribution of the present paper: we establish a connection between relevant priors and *Clarke-Rockafellar differentials* of the functional I.

The following definitions are based on Clarke (1983, Def. 2.4.10 and Corollary to Theorem 2.9.1); we specialize them to the case of a continuous functional I, as delivered by Proposition 1. We note that given an MBR representation (I,u), the set $B_b(\Sigma,u(X))$ has non-empty interior (see Lemma 17 in the Appendix).

Definition 3 Consider a continuous functional $I:U\to\mathbb{R}$, where $U\subset B(\Sigma)$ is open. For every $e\in U$ and $a\in B(\Sigma)$, the **Clarke-Rockafellar (upper) derivative** of I in e in the direction a is

$$I^{CR}(e;a)=\lim_{\epsilon\downarrow 0}\limsup_{d\to e,t\downarrow 0}\inf_{b:\|b-a\|<\epsilon}\frac{I(d+tb)-I(d)}{t}.$$

The **Clarke-Rockafellar differential** of I at e is the set

$$\partial I(e)=\{Q\in ba(\Sigma):Q(a)\leq I^{CR}(e;a),\ \forall a\in B(\Sigma)\}.$$

Henceforth, we shall abbreviate the expression "Clarke-Rockafellar" by CR.

In the setting of GMM, the functional I is constant-linear, so that: 1) it extends uniquely to all of $B(\Sigma)$; 2) it is (globally hence) locally Lipschitz. For locally Lipschitz functionals, the CR derivative and differential correspond to Clarke's notions, which have a simpler expression (see Appedix B). Moreover, the Clarke upper derivative at 0 in the direction $a\in B(\Sigma)$, denoted $I^c(0;a)$, when I is constant-linear takes the particularly simple form

$$I^c(0;a)=\sup_{d\in B(\Sigma)}I(d+a)-I(d)$$

(cf. GMM, Prop. A.3). We cannot take advantage of this simpler form, and will work directly with Def. 3; furthermore, unlike GMM, we shall need to compute CR differentials at all points in the interior of the domain of I. On the positive side, the CR differential retains most of the useful features of the Clarke differential, and it does not require Lipschitz behavior even locally.

It is also important to point out that, not unlike the usual notion of gradient, the definition of CR differential is seldom used directly (although we do so in proving Theorem 5). It is useful chiefly because of its convenient calculus properties.

We are ready to state our main result. We consider a representation (I,u) wherein the functional I is isotone, but not necessarily normalized (cf. Lemma 2); this simplifies the calculation of the CR differential for some decision models (in particular, smooth ambiguity preferences).

Theorem 5 *Consider an MBC preference $\succcurlyeq$ represented as in Eq. (3) by a non-constant, affine function $u : X \to \mathbb{R}$ and a monotonic, continuous and isotone functional $I : B_b(\Sigma, u(X)) \to \mathbb{R}$ Then, the set C in Proposition 3 can be computed as follows:*

$$C = \overline{\text{co}} \left(\bigcup_{e \in \text{int } B_b(\Sigma, u(X))} \left\{ \frac{Q}{Q(S)} : Q \in \partial I(e), Q(S) > 0 \right\} \right).$$

Thus, the set C of relevant priors is equal to the convex closure of the *union* of all the CR differentials, once the latter are properly renormalized so as to contain only probability measures.

We now consider interesting special cases, corresponding to different independence properties of the preference $\succcurlyeq$ that have been analyzed in the literature. This will also make apparent the extent to which our result generalizes its counterpart in GMM. First, if preferences satisfy Axiom 9, as is the case for variational and vector expected utility preferences, the functional I is also translation-invariant; in this case, all elements $Q \in \partial I(e)$ automatically satisfy the normalization $Q(S) = 1$ (see statement 2 of Prop. A.3 in GMM). Moreover, a monotonic, translation-invariant functional is globally Lipschitz with Lipschitz constant 1 (see, e.g., Maccheroni et al., 2006). For such preferences, Theorem 5 can be strengthened as follows:

Corollary 6 *If I is normalized and translation-invariant, then $C = \overline{\text{co}} \left(\bigcup_{e \in \text{int } B_b(\Sigma, u(X))} \partial I(e) \right)$.*

If instead the functional I is positively homogeneous (cf. Axiom 10, which is satisfied by confidence-function preferences), as well as locally Lipschitz, and furthermore if $0 \in \text{int } u(X)$, then $\partial I(e) \subset \partial I(0_S)$ for all interior points e, where 0_S denotes the constantly 0 function (cf. statement 1 of Prop. A.3 in GMM). By positive homogeneity, I can be extended to all of $B(\Sigma)$ by assuming that $u(X) \supset [-1,1]$. Therefore, Theorem 5 specializes as follows:

Corollary 7 *If I is normalized, positively homogeneous and locally Lipschitz, and $0 \in \text{int } u(X)$, then $C = \overline{\text{co}} \left\{ \frac{Q}{Q(S)} : Q \in \partial I(0), Q(S) > 0 \right\}$.*

Third, and finally, GMM consider preferences that satisfy Axiom 8 (Certainty Independence), and hence admit a representation with I constant-linear; i.e., both positively homogeneous and translation-invariant. For such preferences, we combine the preceding corollaries to obtain

Corollary 8 (GMM, Theorem 14) *If I is constant-linear, then $C = \partial I(0)$.*

6.2 Calculation for specific decision models

The differential characterization in Theorem 5 is mainly useful because the properties of CR differentials may be used to compute the set C of ambiguous beliefs for specific decision models. This section carries out these calculations for a variety of recent preference representations.

We begin with multiplier preferences, mainly because, in case the state space is finite, a direct computation based on the formula provided by Theorem 5 is straightforward. We find this useful to build intuition. We then consider three decision models, namely smooth ambiguity, VEU preferences, and mean-variance preferences, for which the computations leverage the basic calculus of CR differentials; in particular, they employ the so-called "fuzzy sum rule," the chain rule, and a convenient formula for the Clarke differential of integral functionals. Finally, we consider preferences which satisfy the Ambiguity Hedging axiom, such as the variational and Chateauneuf-Faro preferences. Here, the calculations rely mainly on the relationship between CR differentials and normal cones of sublevel sets.

6.2.1 Multiplier preferences

Hansen, Sargent, and Tallarini (1999); Hansen and Sargent (2001) introduced *multiplier preferences*, a specification that has gained considerable popularity due to its tractability. Each act $f \in \mathscr{F}$ is evaluated according to the functional $V(f) = I(u \circ f)$, where u is a Bernoulli utility function,

$$I(a) = \min_{Q \in ca_1(\Sigma)} \left(Q(a) + \theta R(Q\|P) \right),$$

$P \in ca_1(\Sigma)$ is a (countably additive) reference prior, and $R(Q\|P)$ denotes the relative entropy of $Q \in ca_1(\Sigma)$ with respect to P. The functional I is well-defined, monotonic and continuous on $B(\Sigma)$, so multiplier preferences are MBC preferences; furthermore, I is translation-invariant, so

Corollary 6 can be applied. The cited authors note that the functional I can be represented as $I(a) = -\theta \log\left(P\left(e^{-\frac{a}{\theta}}\right)\right)$, which is convenient for the purposes of computing the set C.

To gain intuition, suppose that S is finite. Then I is a continuously differentiable function on $\mathbb{R}^{|S|}$; the Corollary to Proposition 2.2.1 and Proposition 2.2.4 in Clarke (1983) imply that $\partial I(e)$ is the gradient of I at e for every $e \in \text{int } B_0(\Sigma, u(X)) = \text{int } u(X)^{|S|}$. Therefore, denoting by $\Delta(S) \equiv ba_1(2^S)$ the probability simplex and invoking Corollary 6,

$$C = \overline{\text{co}}\left\{P_{a,\theta} \in \Delta(S) : a \in \mathbb{R}^{|S|}\right\}, \quad \text{where} \quad P_{a,\theta} \equiv \left(\frac{e^{-\frac{a(s)}{\theta}} P(\{s\})}{P(e^{-\frac{a}{\theta}})}\right)_{s \in S}. \tag{4}$$

Now fix $s \in S$ and consider the function $a_{K,s} : \mathbb{R}^{|S|} \to \mathbb{R}$ such that $a_{K,s}(s) = 0$ and $a_{K,s}(s') = K\theta$. Then $P_{a_{K,s},\theta}(\{s\}) = \frac{P(\{s\})}{P(\{s\}) + e^{-K} P(S \setminus \{s\})} \to 1$ as $K \to \infty$. This implies that

$$C = \Delta(S).$$

For the general case of S infinite, a formula similar to Eq. (4) can be obtained by adapting the techniques in the proof of Proposition 9 below. The details are omitted.

6.2.2 Smooth ambiguity preferences

Klibanoff et al. (2005, KMM henceforth) propose and axiomatize the *smooth ambiguity* model, according to which acts $f \in \mathscr{F}$ are evaluated by means of the functional $V : \mathscr{F} \to \mathbb{R}$ defined by

$$V(f) = \int_{ba_1(\Sigma)} \phi\left(Q(u \circ f)\right) d\mu(Q). \tag{5}$$

Here, u is a Bernoulli (first-order) utility function, $\phi : u(X) \to \mathbb{R}$ is a continuous and strictly increasing *second-order utility*, and similarly μ is a *second-order probability* defined over $ba_1(\Sigma)$ (the reader is referred to KMM for details).

To see that smooth ambiguity preferences are MBC preferences as well, write $V(f) = I(u \circ f)$, where the functional $I : B_b(\Sigma, u(X)) \to \mathbb{R}$ is defined by $I(a) = \int \phi(Q(a))\, d\mu(Q)$ for every $a \in B_b(\Sigma, u(X))$. Note that, for any sequence $(a^k) \subset B_b(\Sigma, u(X))$ that converges to $a \in B_b(\Sigma, u(X))$ in the supremum norm, the integrals $Q(a^k) = \int a^k\, dQ$ converge to $Q(a)$ uniformly in $Q \in ba_1(\Sigma)$; thus, $\phi(Q(a^k)) \to \phi(Q(a))$ uniformly in Q,[7] which guarantees that $I(a^k) \to I(a)$: that is, the

[7] Since $a^k \to a$ uniformly and $a, (a^k) \in B_b(\Sigma, u(X))$, there exists an interval $[\alpha, \beta] \subset u(X)$ such that $a^k(s) \in [\alpha, \beta]$

functional I is continuous on $B_b(\Sigma, u(X))$. Monotonicity is immediate, as is the fact that I is isotone (but not normalized in general); thus, smooth ambiguity preferences are members of the MBC class.

Before we provide a general formula for the set C of ambiguous beliefs, consider the simple case of a finite state space S and continuously differentiable second-order utility ϕ. Then I is a continuously differentiable function on a subset of $\mathbb{R}^{|S|}$, and we immediately get

$$C = \overline{\text{co}} \left\{ \left(\frac{\int_{\Delta(S)} \phi'(Q(e)) Q(\{s\})\, d\mu(Q)}{\int_{\Delta(S)} \phi'(Q(e))\, d\mu(Q)} \right)_{s \in S} : e \in \text{int } u(X)^{|S|} \right\}. \tag{6}$$

For general state spaces, we can leverage results by Clarke (1983) on the differential of integral functionals to provide a formula that covers most cases of thoretical and applied interest. We require a notion of regularity for functions and functionals. As it will play a role in other results, we provide a general definition.

Definition 4 (Clarke, 1983, p. 39) A function $f : B \to \mathbb{R}$ on a Banach space B is **regular** at $b \in B$ if, for all $v \in B$, the directional derivative

$$f'(b; v) \equiv \lim_{t \downarrow 0} \frac{f(b + t v) - f(b)}{t}$$

is well-defined and coincides with $f^{\text{CR}}(b; v)$.

Note that f is regular if it is convex (hence $-f$ is regular if f is concave), or if it is continuously differentiable (cf. Clarke, 1983, Corollary to Prop. 2.2.4 and Prop. 2.3.6).

Proposition 9 *Suppose that S is a Polish space and Σ is its Borel sigma-algebra. Consider a smooth ambiguity preference represented by (u, μ, ϕ), where ϕ is locally Lipschitz, either ϕ or $-\phi$ is regular, and μ is a Borel measure on the space $ca_1(\Sigma)$ of countably additive probabilities, endowed with the $\sigma(ca(\Sigma), C(S))$ topology (i.e., the usual weak-convergence topology).*

---

for all s and k, and similarly $a(s) \in [\alpha, \beta]$ for all s. Hence, for all $Q \in ba_1(\Sigma)$, $Q(a^k) \in [\alpha, \beta]$ for all k, and also $Q(a) \in [\alpha, \beta]$. Now ϕ is continuous, and hence uniformly continuous on the compact interval $[\alpha, \beta]$; hence, for every $\epsilon > 0$, there is $\delta > 0$ such that $|\gamma - \gamma'| < \delta$ implies $|\phi(\gamma) - \phi(\gamma')| < \epsilon$. In particular, if k is large enough so that $\|a^k - a\| < \delta$, we have $|Q(a^k) - Q(a)| < \delta$ for all Q, and hence $|\phi(Q(a^k)) - \phi(Q(a))| < \epsilon$ for all Q, as required.

Then

$$C = \overline{\text{co}}\left(\frac{\int_{ca_1(\Sigma)} \alpha_{e,Q}\, Q\, d\mu(Q)}{\int_{ca_1(\Sigma)} \alpha_{e,Q}\, d\mu(Q)} : \begin{array}{l} e \in \text{int } B_b(\Sigma, u(X)),\ \int_{ca_1(\Sigma)} \alpha_{e,Q}\, d\mu(Q) > 0, \\ Q \mapsto \alpha_{e,Q} \text{ a measurable selection from } Q \mapsto \partial\phi(Q(e)) \end{array} \right). \quad (7)$$

To interpret the notation employed above, each $\alpha_{e,Q}$ is a real number, and $L = \int_{ca_1(\Sigma)} \alpha_{e,Q} Q\, d\mu(Q)$ is the linear functional on $B(\Sigma)$ such that $L(a) = \int_{ca_1(\Sigma)} \alpha_{e,Q} Q(a)\, d\mu(Q)$ for all a. In case ϕ is continuously differentiable, $\partial\phi(Q(e)) = \{\phi'(Q(e))\}$, and one obtains a simpler formula similar to the one provided above for the case of S finite. The existence of suitable measurable selections from the correspondence $Q \mapsto \partial\phi(Q(e))$ is established in the proof of Proposition 9.

A function is locally Lipschitz if it is concave or convex, or if it is continuously differentiable; also, as noted above, any one of these conditions implies regularity of either ϕ or $-\phi$. Hence, the above result covers virtually all specifications of smooth ambiguity preferences used in applications, as well as specifications that are neither ambiguity-averse nor ambiguity-loving.

6.2.3 Vector expected utility

Siniscalchi (2009) introduces and axiomatizes the *vector expected utility* (VEU) model, according to which acts $f \in \mathscr{F}$ are evaluated via the index

$$V(f) = I(u \circ f), \quad \text{where} \quad I(a) = P(a) + A\left((P(\zeta_i a))_{0 \le i < n}\right) \quad \forall a \in B_0(\Sigma, u(X));$$

here, u is a Bernoulli utility function, P is a countably additive *baseline probability*, $0 \le n \le \infty$, $\zeta_i : S \to \mathbb{R}$ is a bounded *adjustment factor* that satisfies $P(\zeta_i) = 0$ for $0 \le i < n$, and the *adjustment function* $A : \mathbb{R}^n \to \mathbb{R}$ satisfies $A(0) = 0$ and $A(\varphi) = A(-\varphi)$ for all $\varphi \in \mathbb{R}^n$. Suitable assumptions ensure that I is monotonic; also, the VEU functional I admits a monotonic, continuous extension to $B_b(\Sigma, u(X))$. Hence VEU preferences also belong to the MBC class.

VEU preferences are translation-invariant, but not positively homogeneous; furthermore, they need not be ambiguity-averse (or -loving). For instance, Siniscalchi (2008) shows by example that VEU preferences can accommodate the "reflection" example of Machina (2009), which requires relaxing Ambiguity Hedging. Indeed, VEU preferences can exhibit non-uniform ambiguity attitudes (cf. Siniscalchi, 2007). Thus, they provide an ideal testing ground for the tools developed in this paper.

We now explicitly compute the set C as a function of the elements of the representation for a rich class of VEU preferences. Specifically, we assume that $n < \infty$ (which, again, can be characterized behaviorally) and that the adjustment function A, or its negative $-A$, is regular in the sense of Definition 4. The latter is true if, say, A is continuously differentiable (taking advantage of the fact that in VEU A is defined on some Euclidean space $\mathbb{R}^n$), or if it is concave or convex.[8]

To state the result, it is convenient to define, for every $0 \leq i < n$, the functional $M_i : B_0(\Sigma) \to \mathbb{R}$ by $M_i(a) = P(\zeta_i a) = \int \zeta_i a \, dP$ for all $a \in B_0(\Sigma)$. Also note that, since A is a function on $\mathbb{R}^n$, every $L \in \partial A$ can be identified with a vector $(\alpha_0)_{0 \leq i < n}$ in $\mathbb{R}^n$. Observe that no restriction is imposed on the cardinality of the state space S.

Proposition 10 *Consider a VEU preference with representation* $(u, P, n, (\zeta_i)_{0 \leq i < n}, A)$. *If* $u(X)$ *has non-empty interior,* $n < \infty$ *and either* A *or* $-A$ *is regular, then*

$$C = \overline{\text{co}} \left\{ P + \sum_{0 \leq i < n} \alpha_i M_i \, : \, (\alpha_i)_{0 \leq i < n} \in \partial A(M_0(a), \ldots, M_{n-1}(a)) \text{ for some } a \in \text{int } B_b(\Sigma, u(X)) \right\}.$$

6.2.4 Monotonic mean-dispersion preferences

Grant and Polak (2007) propose a model of "mean-dispersion" preferences that generalizes variational preferences (see below) for the case of a finite state space and unbounded utility. Formally, consider: a Bernoulli utility function $u : X \to \mathbb{R}$ with $u(X) = \mathbb{R}$; a vector $\pi \in \mathbb{R}^S$ with $\sum_s \pi_s = 1$; a continuous function $\varphi : \mathbb{R}^2 \to \mathbb{R}$, increasing in the first argument and decreasing in the second, and such that $\varphi(\gamma, 0) = \gamma$; a function $\rho : \{v \in \mathbb{R}^n : v \cdot \pi = 0\} \to \mathbb{R}_+$ with $\rho(0) = 0$. Preferences are represented by the functional

$$V(f) = \varphi(\mu, \rho(d)) \qquad \text{where} \quad \mu = \pi \cdot u \circ f \quad \text{and} \quad d = u \circ f - \mu 1_S. \tag{8}$$

Mean-dispersion preferences allow for a weakening of monotonicity. However, when monotonicity holds, these preferences are a member of the MBC family, and π is a probability distribution (other cross-restrictions on the elements of the representation hold as well). When $\varphi(\mu, \rho) = \mu + \rho$, the mean-dispersion functional is similar to the VEU functional (cf. Sec. 6.2.3).

[8] Indeed, a direct calculation of the set C is often feasible when the functional A is parametrically specified.

(See Siniscalchi (2009), Sec. 5.1 for discussion of the relationship between mean-dispersion and VEU preferences.)

We compute the set C for monotone mean-dispersion preferences under regularity assumptions. We let $I(a) = \varphi(a \cdot \pi, \rho(a - (a \cdot \pi))1_S)$.

Proposition 11 *Assume that S is finite, I is monotonic, and that maps $(\mu, \rho) \mapsto \varphi(\mu, -\rho)$ and $d \mapsto -\rho(d)$ are locally Lipschitz and regular. Then*

$$C = \overline{\text{co}} \left\{ \pi + \frac{\beta_2}{\beta_1} \sum_s \alpha_s (1_{\{s\}} - \pi) : \begin{array}{c} e \in \text{int}\, \mathbb{R}^S,\ (\beta_1, \beta_2) \in \partial \varphi(e \cdot \pi, \rho(e - e \cdot \pi 1_S)), \\ \beta_1 > 0,\ \alpha \in \partial \rho(e - e \cdot \pi 1_S) \end{array} \right\}.$$

An important special case is that of an additive aggregator: $\varphi(\mu, \rho) = \mu - \rho$. In this case, the only requirement is that the function $-\rho(\cdot)$ be locally Lipschitz and regular, which is the case, for instance, if $\rho(\cdot)$ is concave. With an additive aggregator, $\beta_1 = 1$ and $\beta_2 = -1$.

6.2.5 Preferences that satisfy Axiom 7 (Ambiguity Hedging)

A variety of recent models generalize the MEU preferences of Gilboa and Schmeidler (1989) in several ways, but retain their central Ambiguity Hedging axiom. Maccheroni et al. (2006, MMR henceforth) define *variational preferences* as a generalization of both MEU and multiplier preferences. Acts $f \in \mathscr{F}$ are evaluated via the functional

$$V(f) = \min_{Q \in ba_1(\Sigma)} \left(Q(u \circ f) + c(q) \right),$$

where u is a Bernoulli utility function, and $c : ba_1(\Sigma) \to \mathbb{R}_+ \cup \{\infty\}$ is a "cost function." For MEU preferences, $c(\cdot)$ can be taken to be the indicator function (in the sense of convex analysis) of the relevant set of priors; for multiplier preferences, $c(Q) = \theta R(Q \| P)$. As is apparent, these preferences are translation-invariant, but not necessarily positively homogeneous

Chateauneuf and Faro (2009) propose a complementary model, which we call *confidence-function preferences* (CF for short): in their representation,

$$V(f) = \min_{q \in ba_1(\Sigma) : \varphi(Q) \geq \alpha_0} \frac{Q(u \circ f)}{\varphi(Q)},$$

where $\alpha_0 \in (0, 1]$ is a minimal confidence level and $\varphi : ba_1(\Sigma) \to (0, 1]$ is a confidence function. These preferences may violate translation invariance, but satisfy positive homogeneity.

Cerreia-Vioglio et al. (2008) provide a representation of preferences that satisfy our Axioms 1–4 and 6 plus Ambiguity Hedging; their *uncertainty averse preferences* (UA henceforth) thus encompass both variational and CF preferences.[9] Smooth ambiguity preferences with $u(X)=\mathbb{R}$, μ countably additive, and concave second-order utility ϕ [henceforth, "smooth UA preferences"] are also elements of this family. The UA representation evaluates acts $f \in \mathscr{F}$ according to

$$V(f)=\inf_{Q\in ba_1(\Sigma)} G\big(Q(u\circ f),Q\big). \tag{9}$$

The function $G : u(X)\times ba_1(\Sigma)\to \mathbb{R}\cup\{+\infty\}$ is quasiconvex, non-decreasing in its first argument, and such that $\inf_{Q\in ba_1(\Sigma)} G(\gamma,Q)=\gamma$ for all $\gamma\in u(X)$. For variational preferences, $G(t,Q)=t+c(Q)$; for CF preferences, $G(t,Q)=\frac{t}{\varphi(Q)}$ if $\varphi(Q)\geq\alpha_0$, and $G(t,Q)=+\infty$ otherwise. For the purposes of our analysis, we let $I(a)=\inf_{Q\in ba_1(\Sigma)} G(Q(a),Q)$ for all $a\in B_0(\Sigma,u(X))$; notice that I is normalized, due to the properties of the UA representation.

We can leverage our Theorem 5 and the general functional characterization of UA preferences provided by Cerreia-Vioglio et al. (2008) to compute the set C for all the decision models considered in this section in a single result. Relative to the most general UA preference specification, we require mild regularity conditions that are satisfied by *all* the notable special cases alluded to above; i.e., variational (and hence also multiplier), CF, and smooth UA preferences.[10]

The key requirement is that the CR differential of the functional I not contain the zero measure (or functional); loosely speaking, this is equivalent to the assumption that, for every point a in the interior of $B_b(\Sigma,u(X))$, there be a neighborhood of a in which the functional representation I decreases approximately linearly for sufficiently small displacements in *some* direction. Rockafellar (1979) discusses this condition, which may be restated as saying that I admits no "substationary points" in the interior of is domain, and is thus related to monotonicity.

[9] In our language, such preferences would naturally be called "ambiguity-hedging preferences." But we prefer to adopt Cerreia-Vioglio et al. (2008)'s terminology. Also, strictly speaking, Cerreia-Vioglio et al. (2008) show that a variant of CF preferences, for which $u(X)=\mathbb{R}$ and two confidence functions are used, are UA preferences; in Chateauneuf and Faro (2009), $u(X)=\mathbb{R}_+$. However, the proof of our Proposition 12 below applies *verbatim* to the original CF model.

[10] The quasi-arithmetic preferences considered by Cerreia-Vioglio et al. (2008) also satisfy these conditions.

Definition 5 Denote by $Q_0 \in ba(\Sigma)$ the zero measure: $Q(E) = 0$ for all $E \in \Sigma$. An MBC representation (I, u) is **nice** if, for all $a \in \text{int } B_b(\Sigma, u(X))$, $Q_0 \notin \partial I(a)$.

As noted above, all commonly used UA preference models satisfy the above requirement; furthermore, while the UA functional is defined and characterized by Cerreia-Vioglio et al. (2008) on $B_0(\Sigma, u(X))$, these common models admit a continuous extension to all of $B_b(\Sigma, u(X))$, and hence belong to the MBC family as well:

Remark 6.1 Variational, CF and smooth UA preferences admit a nice MBC representation (I, u).

Our characterization result follows.

Proposition 12 *Consider a UA representation (u, G) that admits a nice MBC representation, and is such that the infimum in Eq. (9) is attained for every $f \in \mathscr{F}$. Then*

$$C = \overline{\text{co}} \left(\bigcup_{e \in \text{int } B_b(\Sigma, u(X))} \arg \min_{Q \in ba_1(\Sigma)} G(Q(e), e) \right).$$

For S finite (or, we conjecture, compact metric), the above result can be established without assuming that UA preferences are nice, and instead requiring that the functional I be locally Lipschitz. Finally, we observe that Cerreia-Vioglio et al. (2008) have independently obtained an analogous result for UA preferences. Their analysis does not employ Clarke-Rockafellar differentials, but highlights an interesting connection with the Greenberg and Pierskalla (1973) differential employed in quasiconvex optimization.

7 Consequences

We conclude by providing a behavioral test for membership in C and definitions and characterizations of "crisp acts" (cf. GMM, Def. 9) and unambiguous events for MBC preferences.

7.1 Certainty equivalents and a test for C

We first introduce convenient notation. For any measure $Q \in ba_1(\Sigma)$ and function $a \in B(\Sigma)$, let $Q(a) = \int a \, dQ$. Also, given a weak* closed set $D \subset ba_1(\Sigma)$ and function $a \in B(\Sigma)$, let $\underline{D}(a) = \min_{Q \in D} Q(a)$ and $\overline{D}(a) = \max_{Q \in D} Q(a)$.

For $f \in \mathscr{F}$ and $D = C$, the quantities $\underline{C}(u \circ f)$ and $\overline{C}(u \circ f)$ have a behavioral characterization. Recall that a prize $x_f \in X$ is a certainty equivalent of an act f if $f \sim x_f$. We want to define a corresponding notion for the unambiguous preference $\succcurlyeq^*$; however, since the latter is not complete in general, we identify lower and upper certainty equivalents. This is related to the definition of the set $C^*(f)$ in GMM, p. 158.

Definition 6 For any act $f \in \mathscr{F}$, a **lower certainty equivalent** of f is a prize $\underline{x}_f \in X$ such that (i) $f \succcurlyeq^* \underline{x}_f$ and (ii) for all $y \in X$ such that $f \succcurlyeq^* y$, $\underline{x}_f \succcurlyeq y$. Similarly, an **upper certainty equivalent** of f is a prize $\overline{x}_f \in X$ such that (i) $f \preccurlyeq^* \overline{x}_f$ and (ii) for all $y \in X$ such that $f \preccurlyeq^* y$, $\overline{x}_f \preccurlyeq y$.

We then have (the proof is a straightforward adaptation of GMM's argument):

Corollary 13 (GMM, Proposition 18) *Every act $f \in \mathscr{F}$ admits lower and upper certainty equivalents. Furthermore, for every lower and upper certainty equivalent $\underline{x}_f, \overline{x}_f$,*

$$u(\underline{x}_f) = \underline{C}(u \circ f) \quad \textit{and} \quad u(\overline{x}_f) = \overline{C}(u \circ f).$$

The notion of lower certainty equivalent can be used to verify whether a given probability $P \in ba_1(\Sigma)$ belongs to the set C —i.e., it is relevant— without invoking Proposition 3, and indeed using only behavioral data (i.e. the DM's preferences).

Further notation is needed. Fix $P \in ba_1(\Sigma)$ and $f \in \mathscr{F}$, and suppose that $f = \sum_{i=1}^n x_i 1_{E_i}$ for a collection of distinct prizes $x_1, \ldots, x_n$ and a measurable partition $E_1, \ldots, E_n$ of S. Then, define

$$x_{P,f} \equiv P(E_1)x_1 + \ldots + P(E_n)x_n.$$

That is, $x_{P,f} \in X$ is a mixture of the prizes $x_1, \ldots, x_n$ delivered by f, with weights given by the probabilities that P assigns to each event $E_1, \ldots, E_n$. Recall that, since u is affine with respect to the vector-space structure on X, $u(x_{P,f}) = P(u \circ f)$.

The sought test is obtained as a straightforward consequence of our results and well-known properties of support functionals. The details are as follows.

Corollary 14 *For a probability charge $P \in ba_1(\Sigma)$, the following statements are equivalent:*

(i) $P \in C$

(ii) for all $f \in \mathscr{F}$, $\underline{x}_f \preccurlyeq x_{P,f}$

7.2 Crisp acts and unambiguous events

GMM deem an act crisp if it cannot be used to hedge the ambiguity of any other act. In our setting without Certainty Independence, the behavioral condition has to be strengthened somehow: We say that an act is crisp if it is unambiguously indifferent to a constant.[11] More precisely, denote by $\sim^*$ the symmetric component of $\succsim^*$. Then, say that an act $f \in \mathscr{F}$ is **crisp** if $f \sim^* x$ for some $x \in X$; equivalently, f is crisp if, for all $g \in \mathscr{F}$ and $\lambda \in [0,1]$, $\lambda f + (1-\lambda)g \sim \lambda x + (1-\lambda)g$.

Corollary 15 *An act $f \in \mathscr{F}$ is crisp if and only if $\underline{C}(u \circ f) = \overline{C}(u \circ f)$.*

Crisp acts are clearly related to unambiguous acts. Ghirardato, Maccheroni, and Marinacci (2005, henceforth GMMu) characterize unambiguous acts and events for GMM (i.e. constant-linear) preferences. We now generalize part of their discussion of unambiguous events to the broader class of MBC preferences. We refer the reader to GMMu for the discussion of unambiguous acts (which also generalizes) and for more detailed interpretation.

An event is unambiguous if if any *bet on* such event is crisp. Precisely:

Definition 7 An event $A \in \Sigma$ is **unambiguous** if and only if for any $x, y \in X$ with $x \succ y$, the act $x\,A\,y$ is crisp. The set of unambiguous events is denoted Λ.

(It can be shown that the definition could be equivalently stated as "*for some $x \succ y$...*".) Conversely, A is ambiguous if $x\,A\,y \not\sim^* z$ for any $z \in X$; that is, if $x\,A\,y \sim z$, then there exist $\lambda \in (0,1]$, $g \in \mathscr{F}$ such that $\lambda x\,A\,y + (1-\lambda)g \not\sim \lambda z + (1-\lambda)g$.

Unambiguous events have a natural characterization in terms of the set of priors C. (Again, the proof is a straightforward extension of GMMu's argument.)

Proposition 16 (GMMu, Proposition 8) *For any MBC preference with set of priors C and any event $A \in \Sigma$, $A \in \Lambda$ if and only if $P(A) = Q(A)$ for all $P, Q \in C$.*

That is, an event is unambiguous when all the probabilities in C agree on it. (Notice that this is independent of the normalization chosen for u.) In light of this characterization, it is immediate to verify that Λ is a **(finite) λ-system** (GMMu, Corollary 9). That is: 1) $S \in \Lambda$; 2) if $A \in \Lambda$ then $A^c \in \Lambda$; 3) if $A, B \in \Lambda$ and $A \cap B = \emptyset$ then $A \cup B \in \Lambda$. (Cf. Zhang (2002) and Nehring (1999).)

[11] For GMM's preferences the two conditions are equivalent. See Corollary 15 and their Prop. 10.

A Preliminaries on subsets of $B(\Sigma)$

The following result characterizes properties of the subsets of $B(\Sigma)$ of interest in this paper. To fix ideas, it is also useful to compare $B_b(\Sigma,\Gamma)$ with $B(\Sigma,\Gamma)$, the set of bounded, Σ-measurable functions with values in Γ: in particular, if $S=\Gamma=(0,1)$ and Σ is the Borel sigma-algebra, the identity function is in $B(\Sigma,\Gamma)$ but not in $B_b(\Sigma,\Gamma)$.

Also recall that, given an MBC representation (I,u), non-triviality implies that u is non-constant and affine, so that $u(X)$ is an interval with non-empty interior.

Lemma 17 *Let Γ be an interval with non-empty interior. Then:*

1. $B_b(\Sigma,\Gamma)=\{a\in B(\Sigma) : \inf_s a(s), \sup_s a(s)\in\Gamma\}$;

2. $\text{int } B_b(\Sigma,\Gamma)=\{a\in B(\Sigma) : \inf_s a(s), \sup_s a(s)\in \text{int }\Gamma\}$;

3. $\text{cl } B_b(\Sigma,\Gamma)=B_b(\Sigma,\text{cl }\Gamma)=B(\Sigma,\text{cl }\Gamma)$.

Proof: For 1, fix $a\in B_b(\Sigma,\Gamma)$. By assumption there are $\alpha,\beta\in\Gamma$ such that $\alpha\geq a(s)\geq\beta$ for all s: hence $\alpha\geq\sup_s a\geq\inf_s a\geq\beta$, so $\inf_s a,\sup_s a\in\Gamma$, as required. The converse is immediate.

For 2, fix $a\in\text{int } B_b(\Sigma,\Gamma)$. Then there exists $\epsilon>0$ such that, for all $b\in B(\Sigma)$, $\|b-a\|<\epsilon$ implies $b\in B_b(\Sigma,\Gamma)$. Then in particular $a-1_S\frac{\epsilon}{2}, a+1_S\frac{\epsilon}{2}\in B_b(\Sigma,\Gamma)$, which implies that $\inf_s a-\frac{\epsilon}{2},\sup_s a+\frac{\epsilon}{2}\in\Gamma$ by part 1: that is, $\inf_s a,\sup_s a\in\text{int }\Gamma$. Conversely, consider $a\in B(\Sigma)$ such that $\inf_s a,\sup_s a\in\text{int }\Gamma$; choose $\epsilon>0$ so that $\inf_s a-\epsilon,\sup_s a+\epsilon\in\Gamma$, and consider $b\in B(\Sigma)$ such that $\|b-a\|<\epsilon$: then $\inf_s b\geq\inf_s a+\inf_s(b-a)=\inf_s a-\sup_s(a-b)\geq\inf_s a-\|a-b\|>\inf_s a-\epsilon$ and $\sup_s b\leq\sup_s a+\sup_s(b-a)\leq\sup_s a+\|b-a\|<\sup_s a+\epsilon$. Hence, $\inf_s b,\sup_s b\in\Gamma$, and part 1 implies that $b\in B_b(\Sigma,\Gamma)$. Thus, $a\in\text{int } B_b(\Sigma,\Gamma)$.

For 3, consider $a\in\text{cl } B_b(\Sigma,\Gamma)$, so there is a sequence $(a^k)\subset B_b(\Sigma,\Gamma)$ such that $a^k\to a$. By part 1, $\inf_s a^k,\sup_s a^k\in\Gamma$, and by continuity of inf and sup, $\inf_s a^k\to\inf_s a$ and $\sup_s a^k\to\sup_s a$; since the numerical sequences $(\inf_s a^k),(\sup_s a^k)$ lie in Γ, their limits by definition lie in cl Γ; furthermore, for every $s\in S$, the fact that $\sup_s a^k\geq a^k(s)\geq\inf_s a^k$ for all k implies $\sup_s a\geq a(s)\geq\inf_s a$ by continuity of the infimum and supremum on $B(\Sigma)$: hence, $a\in B_b(\Sigma,\text{cl }\Gamma)$. Next, if $a\in B_b(\Sigma,\text{cl }\Gamma)$, then $a(S)\subset\text{cl }\Gamma$ and furthermore there are $\alpha,\beta\in\text{cl }\Gamma$ with $+\infty>\alpha\geq a(s)\geq\beta>-\infty$ for all s, so a is bounded and thus $a\in B(\Sigma,\text{cl }\Gamma)$. Finally, if $a\in B(\Sigma,\text{cl }\Gamma)$, then $\inf_s a,\sup_s a\in$

cl Γ. If $\inf_s a = \sup_s b$, then $a = 1_S\gamma$ for some $\gamma \in$ cl Γ and it is clear that $a \in$ cl $B_b(\Sigma,\Gamma)$. Otherwise, let $\gamma = \frac{1}{2}\inf_s a + \frac{1}{2}\sup_s a \in$ int Γ, and for all $k \geq 1$ define $a^k = \frac{1}{k}\gamma + \frac{k-1}{k}a$. Then, for all k and all $s \in S$, $a^k(s) \geq \frac{1}{k}\gamma + \frac{k-1}{k}\inf_s a \in \Gamma$, and similarly $a^k(s) \leq \frac{1}{k}\gamma + \frac{k-1}{k}\sup_s a \in \Gamma$; hence, for all k, $\inf_s a^k, \sup_s a^k \in \Gamma$, so by part 1 $a^k \in B_b(\Sigma,\Gamma)$ (indeed, a^k is in the interior) and therefore $a \in$ cl $B_b(\Sigma,\Gamma)$, as claimed. ■

B Clarke-Rockafellar and Clarke derivatives and differentials

Note first that, relative to our Def. 3, Rockafellar's original definition (e.g., Rockafellar, 1979) allows for I discontinuous, in which case the limsup must be taken over sequences $b \to a$ such that $I(b) \to I(a)$. His definition also allows for I extended real-valued and defined on all of $B(\Sigma)$.

The following lemma clarifies that the definition of CR derivative is local: only behavior in a neighborhood of the point of interest matters. This is important for our purposes, as our functional I is in general defined only on a subset of $B(\Sigma)$. Note that continuity is assumed only in order to apply our slightly simplified definition of the CR derivative (i.e. Def. 3).

Lemma 18 *Let I, J be real-valued functionals on $B(\Sigma)$, and suppose that, for some open subset U of $B(\Sigma)$, both I and J are continuous on U, and $I(e) = J(e)$ for all $e \in U$. Then, for all $e \in U$ and $a \in B(\Sigma)$, $I^{CR}(e;a) = J^{CR}(e;a)$ and $\partial I(e) = \partial J(e)$. The same conclusion holds if I is defined and continuous only on U.*

Proof: It is of course enough to prove the assertion regarding CR derivatives; morevoer, by inspecting Def. 3, we only need to show that, for every $e \in U$, $a \in B(\Sigma)$, and $\epsilon > 0$,

$$\limsup_{d \to e, t \downarrow 0} \inf_{b:\|b-a\|<\epsilon} \frac{I(d+tb)-I(d)}{t} = \limsup_{d \to e, t \downarrow 0} \inf_{b:\|b-a\|<\epsilon} \frac{J(d+tb)-J(d)}{t}.$$

Thus, consider sequences $(d^k),(t^k)$ such that $d^k \to e$, $t^k \downarrow 0$, and $\inf_{b:\|b-a\|<\epsilon} \frac{I(d^k+t^k b)-I(d^k)}{t^k} \to \gamma \in \mathbb{R} \cup \{\pm\infty\}$ as $k \to \infty$. For the given a and ϵ, we can find $\eta, \lambda > 0$ such that $\|d-e\| < \eta$ and $t \in (0,\lambda)$ imply $d \in U$ and, importantly, also $d + tb \in U$ for all $b \in B(\Sigma)$ with $\|b-a\| < \epsilon$.[12] Therefore,

[12] Since U is open, there is $\delta > 0$ such that $\|d'-e\| < \delta$ implies $d' \in U$: thus, if $0 < t < \frac{\delta}{2(\|a\|+\epsilon)} \equiv \lambda$, then for all b with $\|b-a\| < \epsilon$ we have $\|(e+tb)-e\| = t\|b\| \leq t(\|a\|+\|b-a\|) < t(\|a\|+\epsilon) < \frac{1}{2}\delta$, and if $\|d-e\| < \frac{1}{2}\delta \equiv \eta$, then $\|(d+tb)-e\| \leq \|d-e\| + t\|b\| < \delta$ for all such b, as required.

there exists K such that $k \geq K$ implies $d^k, d^k + t^k b \in U$ for all b with $\|b-a\| < \epsilon$; hence, for such k, $\inf_{b:\|b-a\|<\epsilon} \frac{I(d^k+t^k b)-I(d^k)}{t^k} = \inf_{b:\|b-a\|<\epsilon} \frac{J(d^k+t^k b)-J(d^k)}{t^k}$, and so $\inf_{b:\|b-a\|<\epsilon} \frac{J(d^k+t^k b)-J(d^k)}{t^k} \to \gamma$. Switching the role of I and J shows that, for any sequence $d \to e$ and $t \downarrow 0$, $\inf_{b:\|b-a\|<\epsilon} \frac{I(d+tb)-I(d)}{t}$ has a limit λ (possibly infinite) if and only if $\inf_{b:\|b-a\|<\epsilon} \frac{J(d+tb)-J(d)}{t}$ has the same limit. The above equality then follows.

If I is defined and continuous on U, it can be extended arbitrarily to a functional on $B(\Sigma)$, which will then satisfy the stated conditions. ■

The main properties of interest of I^{CR} and ∂I are listed in the following proposition, which is mostly based on Aussel, Corvellec, and Lassonde (1995, ACL henceforth), specialized to our setting. Here and in the following, for $a, b \in B(\Sigma)$, we let $[a,b] = \{\lambda a + (1-\lambda)b : \lambda \in [0,1]\}$. The "intervals" (a,b), $[a,b)$ and $(a,b]$ are defined analogously.

Proposition 19 *Let $I : B_b(\Sigma,\Gamma) \to \mathbb{R}$ be continuous, with $\operatorname{int}\Gamma \neq \emptyset$. Then:*

1. *The set $\operatorname{dom}\partial I \equiv \{e \in \operatorname{int} B_b(\Sigma,\Gamma) : \partial I(e) \neq \emptyset\}$ is dense in $B_b(\Sigma,\Gamma)$.*

2. *For every $e \in \operatorname{int} B_b(\Sigma,\Gamma)$, either $I^{CR}(e;0) = -\infty$ or $e \in \operatorname{dom}\partial I$ and $I^{CR}(e;a) = \sup_{Q\in\partial I(e)} Q(a)$ for all $a \in B(\Sigma)$.*

3. *(Approximate Mean-Value Theorem) for all $a, b \in \operatorname{int} B_b(\Sigma,\Gamma)$, there exist a function $c \in [a,b)$ and sequences $(c_n) \in \operatorname{int} B_b(\Sigma,\Gamma)$ and $(Q_n) \in ba(\Sigma)$ such that $c_n \to c$, $Q_n \in \partial I(c_n)$ for all n, and*

$$\liminf_n Q_n(c_n - c) \geq 0, \qquad \liminf_n Q_n(b-a) \geq f(b) - f(a).$$

4. *I is monotonic if and only if, for all $e \in \operatorname{int} B_b(\Sigma,\Gamma)$ and $Q \in \partial I(e)$, $Q \geq 0$.*

5. *If I is convex, then CR derivatives are the usual directional derivatives and CR differentials are subdifferentials. That is, $I^{CR}(e;a) = \lim_{t\downarrow 0} t^{-1}(I(e+ta) - I(e))$ for all $e \in \operatorname{int} B_b(\Sigma,\Gamma)$ and $a \in B_b(\Sigma)$, and $\partial I(e) = \{Q \in ba(\Sigma) : Q(a) - Q(e) \leq I(a) - I(e)\ \forall a \in B_b(\Sigma,\Gamma)\}$ for all $e \in B_b(\Sigma,\Gamma)$.*

Proof: ACL consider l.s.c. functions defined on a Banach space with values in $\mathbb{R} \cup \{+\infty\}$. Thus, to establish the above claims, our general strategy is to use Lemma 18 and standard results to

extend I to a suitable continuous functional J on $B(\Sigma)$ that has the same CR derivatives and differential as I on $B_b(\Sigma,\Gamma)$. It is convenient to denote by $O_\epsilon(a)$ the open ball of radius $\epsilon>0$ around $a\in B(\Sigma)$, and by $\bar{O}_\epsilon(a)$ its closure.

(1) For every $e\in \text{int } B_b(\Sigma,\Gamma)$, there is $\epsilon>0$ such that $O_\epsilon(e)\subset \text{int } B_b(\Sigma,\Gamma)$; then $\bar{O}_{\frac{\epsilon}{2}}(e)$ is a closed subset of int $B_b(\Sigma,\Gamma)$, and since I is continuous on this set, by Tietze's extension theorem it admits a continuous extension J_e to all of $B(\Sigma)$. Corollary 3.2 in ACL then yields a dense subset $D_e\subset \text{dom } J_e$ such that $d\in D_e$ implies that $\partial J_e(d)\neq\emptyset$, i.e. $D_e\subset \text{dom } \partial J_e$; in particular, since $\bar{O}_{\frac{\epsilon}{2}}(e)\subset \text{dom } J_e$, D_e contains a sequence $(d^k)\to e$ that also lies in $O_{\frac{\epsilon}{2}}(e)\subset \text{int } B_b(\Sigma,\Gamma)$, so $\partial I(d^k)=\partial J_e(d^k)\neq\emptyset$ for each k; let $\tilde{D}_e=\{d^k\}$ be the collection of elements of this sequence. Then the set $\bigcup_{e\in\text{int } B_b(\Sigma,\Gamma)}\tilde{D}_e$ is included in dom ∂I and is dense in int $B_b(\Sigma,\Gamma)$.

(2) is the Corollary to Theorem 2.9.1 in Clarke (1983): define I arbitrarily outside $B_b(\Sigma,\Gamma)$, as no restrictions are imposed on I except that it be finite wherever the differential is calculated.

(3) Let $\epsilon>0$ be such that $O_\epsilon(a),O_\epsilon(b)\subset \text{int } B_b(\Sigma,\Gamma)$, and consider the set $F=\{d\in B(\Sigma):\exists\lambda\in[0,1] \text{ s.t. } \|d-\lambda a-(1-\lambda)b\|\leq\frac{\epsilon}{2}\}$. Then F is closed, has non-empty interior, and lies in the interior of $B_b(\Sigma,\Gamma)$,[13] where I is continuous and finite; hence, there is a continuous extension J_F of I to $B(\Sigma)$. Theorem 4.2 in ACL now yields a point $c\in[a,b]$ and sequences $(c^n)\in B(\Sigma)$, (Q^n) such that $c^n\to c$ and $Q^n\in\partial J_F(c^n)$ for all n, and the noted inequalities hold. Since F contains an open neighborhood of c, for n large enough $c^n\in F\subset \text{int } B_b(\Sigma,\Gamma)$, and therefore $Q^n\in\partial I(c^n)=\partial J_F(c^n)$, as required.

(4) follows from (2) and (3). Specifically, if I is monotonic, then for all $e\in\text{int } B_b(\Sigma,\Gamma)$, and for all $k\geq 1$, $I^{CR}(e;-1_S\frac{1}{k}-1_E)\leq 0$ from the definition; hence, from (2), $\sup_{Q\in\partial I(e)}Q(-1_S\frac{1}{k}-1_E)\leq 0$, or $Q(1_S\frac{1}{k}+1_E)\geq 0$. Since $\frac{1}{k}1_S+1_E\to 1_E$ uniformly as $k\to\infty$, $Q(1_E)\geq 0$, i.e $Q\geq 0$ for all $Q\in\partial I(e)$.

Conversely, suppose $Q\geq 0$ for all $e\in\text{int } B_b(\Sigma,\Gamma)$ and $Q\in\partial I(e)$, and consider first $a,b\in\text{int } B_b(\Sigma,\Gamma)$ with $a\geq b$. If $a=b$, there is nothing to prove; otherwise, (3) implies that, for

[13] Suppose $(d^k)\subset F$ and $d^k\to F\in B(\Sigma)$; for each k, there is $\lambda^k\in[0,1]$ such that $\|d^k-\lambda^k a-(1-\lambda^k)b\|\leq\frac{\epsilon}{2}$. Then there is a subsequence of (λ^k) that converges to some $\lambda\in[0,1]$, and by continuity of the norm also $\|d-\lambda a-(1-\lambda)b\|\leq\frac{\epsilon}{2}$: thus, F is closed. Each $c\in[a,b]$ is in the interior of F, because by definititon $\|d-c\|<\frac{\epsilon}{2}$ implies $d\in F$. Finally, for each $d\in F$ there is $c=\lambda a+(1-\lambda)b\in[a,b]$ such that $\|d-c\|\leq\frac{\epsilon}{2}$; let $a'=a+(d-c)$, $b'=b+(d-c)$ and note that $\lambda a'+(1-\lambda)b'=d$: furthermore, $\|a'-a\|=\|d-c\|\leq\frac{\epsilon}{2}<\epsilon$ and similarly $\|b'-b\|<\epsilon$, so a',b' lie in int $B_b(\Sigma,\Gamma)$, and hence so does d.

some $c \in \text{int } B_b(\Sigma,\Gamma)$ and $Q \in \partial I(c)$, $Q(b-a) \geq I(b) - I(a)$: but if $Q \geq 0$, then $Q(b-a) \leq 0$, so $I(a) \geq I(b)$ follows. For arbitrary $a, b \in B_b(\Sigma,\Gamma)$, consider $a^k = \frac{k-1}{k}a + \frac{1}{k}\gamma$ and $b^k = \frac{k-1}{k}b + \frac{1}{k}\gamma$ for some $\gamma \in \text{int } \Gamma$; then $a \geq b$ implies $a^k \geq b^k$ for all k, and $a^k, b^k \in \text{int } B_b(\Sigma,\Gamma)$ for all k. Thus, $I(a^k) \geq I(b^k)$, and the result follows from continuity.

(5) is Theorem 5 in Rockafellar (1980); again define I arbitrarily outside of $B_b(\Sigma,\Gamma)$. ■

Say that I is **locally Lipschitz** if, for every $e \in \text{int } B_b(\Sigma,\Gamma)$ there is a neighborhood $U \subset B_b(\Sigma,\Gamma)$ of e such that the restriction of I to U is Lipschitz. If I is locally Lipschitz, then $I^{CR}(e,0) \neq -\infty$, and $I^{CR}(e;\cdot)$ reduces to the Clarke derivative:

Definition 8 For $e \in \text{int } B_b(\Sigma,\Gamma)$ and $a \in B(\Sigma)$, the **Clarke (upper) derivative** of I in e in the direction a is

$$I^c(e;a) = \limsup_{\substack{c \to e \\ t \downarrow 0}} \frac{I(c+ta) - I(c)}{t}.$$

The **Clarke differential** of I in e is the set $\partial I(e) = \{Q \in ba(\Sigma) : Q(a) \leq I^c(e;a),\ \forall a \in B_b(\Sigma)\}$.

Section A.2 in GMM summarizes the key properties of interest of the Clarke differential. It is worth emphasizing that the Clarke differential and derivative are well-defined, and enjoy attractive properties, in any normed space. Thus, we could restrict attention to functionals defined on $B_0(\Sigma,\Gamma)$ rather than $B_b(\Sigma,\Gamma)$, and GMM do so (for the most part). By way of contrast, the results in Prop. 19 require that I be defined on a subset of a complete metric space.

C Proofs

C.1 Proof of Proposition 1

Observation 1: Bourbaki (1998, §8.5, Theorem 1) shows that a function $f : D \to Y$ defined on a dense subset D of a topological space E and taking values in a regular space Y admits a unique continuous extension to E if and only if, for every sequence (d^k) in D that converges to some $e \in E$, the sequence $f(d^k)$ converges to some $y \in Y$. If E and Y are complete metric spaces, then this condition requires that the image of any Cauchy sequence in D be a Cauchy sequence in Y; our Axiom 5, and its name, are inspired by this condition. However, note that $B_b(\Sigma,\Gamma)$ is

not equal to the closure of $B_0(\Sigma,\Gamma)$, unless Γ is closed. Hence, we provide a direct proof of the existence and uniqueness of the extension. Also see Observation 2 at the end of this proof.

Sufficiency: as observed, under axioms 1–3 the von Neumann-Morgenstern theorem delivers an affine function $u : X \to \mathbb{R}$ that represents the restriction of $\succcurlyeq$ to X. Such function is non-constant if axiom 6 also holds.

Next, we claim that, under axioms 1–3 and 6, plus 4, for every $f \in \mathscr{F}$, there exists $x_f \in X$ such that $f \sim x_f$. Axiom 4 implies that there exist $x, y \in X$ such that $x \succcurlyeq f \succcurlyeq y$. If $x \sim f$ or $f \sim y$, we are done. Otherwise, let $U = \{\lambda \in [0,1] : \lambda x + (1-\lambda)y \succcurlyeq f\}$ and $L = \{\lambda \in [0,1] : f \succcurlyeq \lambda x + (1-\lambda)y\}$. Since $x \succ f \succ y$, U and L are non-empty. Moreover, if $\lambda \in U$ and $\lambda' \in (\lambda, 1]$, then (using the representation provided by u) $\lambda' \in U$. Furthermore, suppose (λ^k) is a sequence in U, so $\lambda^k x + (1-\lambda^k)y \succcurlyeq f$, and $\lambda^k \to \lambda$; if $\lambda x + (1-\lambda)y \prec f$, then Axiom 3 yields $\beta \in (0,1)$ such that $[\beta \cdot 1 + (1-\beta)\lambda]x + (1-\beta)(1-\lambda)y = \beta x + (1-\beta)[\lambda x + (1-\lambda)y] \prec f$: but $\beta \cdot 1 + (1-\beta)\lambda > \lambda$, so there exists n such that $\beta \cdot 1 + (1-\beta)\lambda \geq \lambda^k$, so $\beta \cdot 1 + (1-\beta)\lambda \in U$ and therefore $[\beta \cdot 1 + (1-\beta)\lambda]x + (1-\beta)(1-\lambda)y = \beta x + (1-\beta)[\lambda x + (1-\lambda)y] \succcurlyeq f$: contradiction. Hence, $\lambda \in U$, so U is closed. Similarly the set L is closed. Since $U \cup L = [0,1]$, by connectedness there must be $\lambda \in U \cap L$, which implies the claim.

Notice now that, given a sequence $(f^k) \subset \mathscr{F}$ and a bounded act $f \in \mathscr{F}_b$ we have $f^k \to f$ as defined before axiom 5 iff for every $x, y \in X$ such that $u(x) > u(y)$, there exists N such that for $n \geq N$ and all $s \in S$,

$$\frac{1}{2}u(f(s)) + \frac{1}{2}u(y) < \frac{1}{2}u(f^k(s)) + \frac{1}{2}u(x) \quad \text{and} \quad \frac{1}{2}u(f^k(s)) + \frac{1}{2}u(y) < \frac{1}{2}u(f(s)) + \frac{1}{2}u(x).$$

This is equivalent to the requirement that $|u(f^k(s)) - u(f(s))| < u(x) - u(y)$ for all s. In other words, $f^k \to f$ iff $u \circ f^k \to u \circ f$ in the sup-norm topology, i.e. in $B(\Sigma)$.

Therefore, under Axioms 1–4 and 6, Axiom 5 is equivalent to the following statement: given $(f^k) \subset \mathscr{F}$, $(x^k) \subset X$ and $f \in \mathscr{F}_b$, if $f^k \sim x^k$ for all k and $u \circ f^k \to u \circ f$ in the sup-norm topology, then $u(x^k) \to u(x)$ [in the usual convergence in $\mathbb{R}$] for some $x \in X$. Furthermore, we claim that the axiom also implies two key properties:

Continuity on $\mathscr{F}$: given $(f^k) \subset \mathscr{F}$, $(x^k) \subset X$ and $f \in \mathscr{F}$, if $f^k \sim x^k$ for all k and $u \circ f^k \to u \circ f$, then there exists $x \in X$ such that $u(x^k) \to u(x)$ and $f \sim x$.

Proof: To see this, consider the sequence $(h^k) \subset \mathscr{F}$ constructed by letting $h^{2k-1} = f^k$ and

$h^{2k} = f$ for $k \geq 1$, and the sequence $(z^k) \subset X$ such that $x^{2k-1} = x^k$ and $x^{2k} = y$, where $y \in X$ satisfies $f \sim y$. Then $u \circ h^k \to u \circ f$, so Axiom 5 implies that $u(z^k) \to u(z)$ for some $z \in X$. Now, considering k even, it must be the case that $u(y) = u(z)$, i.e. $x_f \sim z$; and considering k odd, we must have $u(x) = \lim_k u(x^k) = u(z)$, i.e. $x \sim z$: thus, $f \sim y \sim z \sim x$, as claimed. *QED*

Unique limits on $\mathscr{F}_b$: if $f \in \mathscr{F}_b$, $(f^k),(g^k) \subset \mathscr{F}$, $(x^k),(y^k) \subset X$, $f^k \sim x^k$ and $g^k \sim y^k$ for all k, and $f^k \to f$, $g^k \to f$, $x^k \to x$ and $y^k \to y$ for $x, y \in X$, then $x \sim y$.

Proof: Define $h^{2k-1} = f^k$, $h^{2k} = g^k$, $z^{2k-1} = x^k$, $z^{2k} = y^k$. Then $h^{2k} \to f$: for any $\epsilon > 0$ there is K_1 such that $k \geq K_1$ implies $\|u \circ f^k - u \circ f\| < \epsilon$, and K_2 such that $k \geq K_2$ implies $\|u \circ g^k - u \circ f\| < \epsilon$, so $K = \max(2K_1 - 1, 2K_2)$ is such that, for all $k \geq K$, $\|u \circ h^k - u \circ f\| < \epsilon$. Hence, there is $z \in X$ such that $u(z^k) \to u(z)$. Now considering odd indices, $u(x) = \lim_k u(x^k) = u(z)$, and considering even indices, $u(y) = \lim_k u(y^k) = u(z)$, so $x \sim y$ as claimed. *QED*

We then define $I : B_b(\Sigma, u(X)) \to \mathbb{R}$ by letting $I(a) = u(x)$, where $a = u \circ f$ for some $f \in \mathscr{F}_b$, and there are sequences $(f^k) \subset \mathscr{F}$ and $(x^k) \subset X$ such that $f^k \sim x^k$ for all k, $f^k \to f$, and $x^k \to x \in X$. By the unique limits property just established, this is well-posed, and by considering the trivial sequences given by $f^k = f$ and $x^k = x \sim f$, it is clear that (I, u) represent $\succcurlyeq$ on $\mathscr{F}$ as stated in the Proposition.

The functional I is normalized: if $\alpha \in u(X)$, then there is $x \in X$ with $u(x) = \alpha$; let $f_\alpha \in \mathscr{F}$ be such that $f(s) = x$ for all $s \in S$; Axiom 4 implies that $f \sim x$, and so $I(\alpha 1_S) = u(x) = \alpha$.

The functional I is monotonic on $B_b(\Sigma, u(X))$: if $a, b \in B_b(\Sigma, u(X))$, then by standard arguments there are sequences $(a^k),(b^k) \in B_0(\Sigma, u(X))$ such that $a^k \geq a$ for all k and $a^k \to a$ in the sup-norm, and similarly $b^k \leq b$ for all k and $b^k \to b$. Find $(f^k),(g^k) \subset \mathscr{F}$, $f, g \in \mathscr{F}_b$, and $(x^k),(y^k) \subset X$ such that $a = u \circ f$, $b = u \circ g$, $a^k = u \circ f^k$, $b^k = u \circ g^k$ and $f^k \sim x^k$, $g^k \sim y^k$ for all k; then $f^k \to f$ and $g^k \to g$, so Axiom 5 yields $x, y \in X$ with $x^k \to x$ and $y^k \to y$, and therefore $I(a) = u(x)$, $I(b) = u(y)$. If now $a(s) \geq b(s)$ for all s, then $a^k(s) \geq a(s) \geq b(s) \geq b^k(s)$ as well, so $f^k(s) \succcurlyeq g^k(s)$ for all s, and therefore $x^k \sim f^k \succcurlyeq g^k \sim y^k$ by Axiom 4. But then $I(a) = u(x) = \lim_k u(x^k) \geq \lim_k u(y^k) = u(y) = I(b)$, as required.

Finally, the functional I is continuous on $B_b(\Sigma, u(X))$. Consider a sequence $(a^k) \subset B_b(\Sigma, u(X))$ such that $a^k \to a \in B_b(\Sigma, u(X))$, and find $(f^k) \subset \mathscr{F}_b$ and $f \in \mathscr{F}_b$ such that $a = u \circ f$ and $a^k = u \circ f^k$ for all k. For every k, consider a sequence $(a^{k,\ell}) \subset B_0(\Sigma, u(X))$ such that $a^{k\ell} \to_{\ell \to \infty} a^k$, and find acts $(f^{k,\ell}) \subset \mathscr{F}$ such that $u \circ f^{k,\ell} = a^{k,\ell}$ for every ℓ; correspondingly, find $(x^{k,\ell}) \subset X$ such that

$f^{k,\ell} \sim x^{k,\ell}$ for every ℓ: then Axiom 5 yields $x^k \in X$ such that $x^{k,\ell} \to_{\ell\to\infty} x^k$. Note that, by definition, this means that $I(a^k) = u(x^k)$ for every k. Now, for every k, there is $\ell(k)$ such that $\|a^{k,\ell(k)} - a^k\| < \frac{1}{k}$ and $|u(x^{k,\ell(k)}) - u(x^k)| < \frac{1}{k}$. Then $\|a^{k,\ell(k)} - a\| \leq \|a^{k,\ell(k)} - a^k\| + \|a^k - a\| \to 0$ as $k \to \infty$, which implies that the sequences $(f^{k,\ell(k)}) \subset \mathscr{F}$ and $(x^{k,\ell(k)}) \subset X$ satisfy $f^{k,\ell(k)} \sim x^{k,\ell(k)}$ for each k and $f^{k,\ell(k)} \to f$. Axiom 5 then yields $y \in X$ such that $u(x^{k,\ell(k)}) \to u(y)$; again, note that this implies that $I(a) = u(y)$. But then $|u(x^k) - u(y)| \leq |u(x^k) - u(x^{k,\ell(k)})| + |u(x^{k,\ell(k)}) - u(y)| \to 0$, i.e. $I(a^k) = u(x^k) \to u(y) = I(a)$, as required.

Necessity: we only show that Axiom 5 holds. Consider $(f^k) \subset \mathscr{F}$, $(x^k) \subset X$, $f \in \mathscr{F}_b$ such that $f^k \sim x^k$ for all k and $f^k \to f$. Then $I(u \circ f^k) \to I(u \circ f)$ because I is continuous on $B_b(\Sigma, u(X))$; furthermore, since $f \in \mathscr{F}_b$, there are $x, y \in X$ such that $x \succcurlyeq f(s) \succcurlyeq y$ for all s, and so $u(x) \geq I(u \circ f) \geq u(y)$ because I is monotonic on $B_b(\Sigma, u(X))$. Since $u(X)$ is an interval, there exists $z \in X$ such that $u(z) = I(u \circ f)$, and since $u(x^k) = I(u \circ f^k)$ because (I, u) represent preferences on $\mathscr{F}$, we conclude that $u(x^k) \to u(z)$, i.e. $x^k \to z$, as required by the Axiom.

Uniqueness: suppose that (I_u, u) and (I_v, v) are normalized representations of the same preference $\succcurlyeq$. Then, by standard arguments, $v = \lambda u + \mu$ for some $\lambda, \mu \in \mathbb{R}$ with $\lambda > 0$. Furthermore, fix $f \in \mathscr{F}$ and let $x \in X$ be such that $f \sim x$. Then

$$I_v(v \circ f) = I_v(v(x)) = v(x) = \lambda u(x) + \mu = \lambda I_u(u \circ f) + \mu,$$

which implies that the required invariance property holds on $B_0(\Sigma, u(X))$. By continuity of I on $B_b(\Sigma, u(X))$, it must hold on $B_b(\Sigma, u(X))$ as well.

Observation 2: Suppose a preference $\succcurlyeq$ is represented on $\mathscr{F}$ by a non-constant, affine utility u and a normalized, monotonic, *uniformly* continuous functional $I : B_0(\Sigma, u(X)) \to \mathbb{R}$. Then $\succcurlyeq$ satisfies Axiom 5. To see this, note that, if $(f^k) \subset \mathscr{F}$, $f \in \mathscr{F}_b$, $(x^k) \subset X$, $f^k \sim x^k$ and $f^k \to f$, then $(u \circ f^k)$ is a Cauchy sequence in $B_0(\Sigma, u(X))$ because it converges in $B(\Sigma)$. By assumption, for every $\epsilon > 0$ there is $\delta > 0$ such that $a, b \in B_0(\Sigma, u(X))$ and $\|a - b\| < \delta$ imply $|I(a) - I(b)| < \epsilon$. Now fix $\epsilon > 0$ and choose K such that $k, \ell \geq K$ imply $\|u \circ f^k - u \circ f^\ell\| < \delta$: then $|I(u \circ f^k) - I(u \circ f^\ell)| < \epsilon$, i.e. $(I(u \circ f^k))|$ is a Cauchy sequence in $u(X)$, and it converges in cl $u(X)$. Furthermore, by continuity of infimum and supremum, and normalization and monotonicity of I on $B_0(\Sigma, u(X))$, $\inf_s u(f(s)) = \lim_k \inf_s u(f^k(s)) \leq \lim_k I(u \circ f^k) \leq \lim_k \sup_s u(f^k(s)) = \sup_s u(f(s))$: but since

$u \circ f \in B_b(\Sigma, u(X))$ by assumption, by Lemma 17 $\inf_s u \circ f, \sup_s u \circ f \in u(X)$ and therefore also $\lim_k I(u \circ f^k) \in u(X)$, which implies the claim.

C.2 Proof of Lemma 2

Define $\phi : \Gamma \to \mathbb{R}$ by $\phi(\gamma) = I(\gamma 1_S)$. Then ϕ is continuous and strictly increasing, and its inverse ϕ^{-1} is also strictly increasing. We claim ϕ^{-1} is continuous. To see this, suppose $I(\gamma_n 1_S) \uparrow I(\gamma 1_S)$ for $\gamma, (\gamma_n) \in \Gamma$; then by isotony (γ_n) is an increasing sequence bounded above by γ, so it has a limit $\bar{\gamma}$. Since ϕ is continuous, $\phi(\bar{\gamma}) = \lim_n \phi(\gamma_n) = \lim_n I(\gamma_n 1_S) = I(\gamma 1_S) = \phi(\gamma)$, i.e. $\bar{\gamma} = \gamma$. A similar argument holds for $I(\gamma_n 1_S) \downarrow I(\gamma 1_S)$, and the claim follows.

The functional $\hat{I} : B_b(\Sigma, \Gamma) \to \mathbb{R}$ defined by $\hat{I}(a) = \phi^{-1}(I(a))$ is normalized by construction: $\hat{I}(\gamma 1_S) = \phi^{-1}(I(\gamma 1_S)) = \phi^{-1}(\phi(\gamma)) = \gamma$. If $a, b \in B_b(\Sigma, \Gamma)$, then as required $I(a) \geq I(b)$ holds if and only if $\hat{I}(a) = \phi^{-1}(I(a)) \geq \phi^{-1}(I(b)) = \hat{I}(b)$ because ϕ^{-1} is strictly increasing; in particular, if $a \geq b$ then $I(a) \geq I(b)$ by monotonicity of I, and thus $\hat{I}(a) \geq \hat{I}(b)$ as required. Finally, since I and ϕ^{-1} are both continuous on their respective domains, $\hat{I} = \phi^{-1} \circ I$ is continuous as well.

C.3 Proof of Proposition 3

For any MBC preference $\succcurlyeq$, the relation $\succcurlyeq^*$ satisfies: (i) for all $f, g \in \mathscr{F}$, $f \succcurlyeq^* g$ implies $f \succcurlyeq g$; (ii) for all $x, y \in X$, $x \succcurlyeq^* y$ if and only if $x \succcurlyeq y$; (iii) $\succcurlyeq^*$ is reflexive, transitive, monotonic (cf. Axiom 4) and continuous: if $(f^k), (g^k) \subset \mathscr{F}$ and $f, g \in \mathscr{F}$ satisfy $f^k \succcurlyeq g^k$ for all k and $f^k \to f$, $g^k \to g$, then $f \succcurlyeq g$; and (iv) $\succcurlyeq^*$ is independent on $\mathscr{F}$: for every $f, g, h \in \mathscr{F}$ and $\lambda \in [0,1]$, $f \succcurlyeq^* g$ implies $\lambda f + (1-\lambda)h \succcurlyeq^* \lambda g + (1-\lambda)h$. To see this, we argue as in the proof of Prop. 4 in GMM, with one exception. To show that $\succcurlyeq^*$ is monotonic, GMM invoke Certainty Independence, but only *state-by-state*. Hence, Constant Independence (Axiom 2) suffices.

The result then follows from Proposition A.2 in GMM.

C.4 Unambiguous preferences on bounded acts

The Bewley-style representation in Prop. 3 is fully determined by $\succcurlyeq$, which is defined on simple acts alone. However, just like $\succcurlyeq$ has a unique extension $\succcurlyeq_b$ to $\mathscr{F}_b$ defined via Prop. 1 by $f \succcurlyeq_b g$ iff $I(u \circ f) \geq I(u \circ g)$ for all $f, g \in \mathscr{F}_b$, so does $\succcurlyeq^*$: one can let $f \succcurlyeq^*_b g$ iff $P(u \circ f) \geq P(u \circ g)$ for

all $P \in C$. The following result show that $\succcurlyeq_b^*$ thus defined is the *only* possible extension of $\succcurlyeq^*$; in particular, if one were instead to define a preference $\hat{\succcurlyeq}_b^*$ using $\succcurlyeq_b$ as in Def. 2, one would obtain the same binary relation.

This is important for our analysis, because our definition of relevant priors deals with simple acts only, but we need to relate the set C to properties of the functional I on all of $B_b(\Sigma, u(X))$.

Corollary 20 *Consider an MBC preference, with (I,u) and C obtained in Propositions 1 and 3 respectively. For any two functions $a, b \in B_b(\Sigma, u(X))$, the following statements are equivalent:*

(i) for all $c \in B_b(\Sigma, u(X))$ and $\lambda \in (0,1]$, $I(\lambda a + (1-\lambda)c) \geq I(\lambda b + (1-\lambda)c)$;

(ii) for all $c \in B_0(\Sigma, u(X))$ and $\lambda \in (0,1]$, $I(\lambda a + (1-\lambda)c) \geq I(\lambda b + (1-\lambda)c)$;

(iii) for all $P \in C$, $\int a \, dP \geq \int b \, dP$.

Furthermore, C is the only weak closed, convex subset of $ba_1(\Sigma)$ for which (i) is equivalent to (iii) for all $a, b \in B_b(\Sigma, u(X))$.*

Proof: It is clear that (i) implies (ii). Now assume (ii), and by standard arguments construct sequences $(a^k), (b^k) \subset B_0(\Sigma, u(X))$ such that $a^k \downarrow a$ and $b^k \uparrow b$ (in the supremum norm). Then, for all $\lambda \in (0,1]$ and $c \in B_0(\Sigma, u(X))$, we have

$$I(\lambda a^k + (1-\lambda)c) \geq I(\lambda a + (1-\lambda)c) \geq I(\lambda b + (1-\lambda)c) \geq I(\lambda b^k + (1-\lambda)c)$$

by monotonicity of I on $B_b(\Sigma, u(X))$. By Proposition 3, for each $P \in C$, $P(a^k) \geq P(b^k)$ for all k, and thus (since the sequences converge uniformly) $P(a) \geq P(b)$: thus, (iii) holds.

Now assume that (iii) holds. Then, for every $P \in C$, $\lambda \in (0,1]$ and $c \in B_b(\Sigma, u(X))$, $P(\lambda a + (1-\lambda)c) \geq P(\lambda b + (1-\lambda)c)$; furthermore, if $a', b' \in B_0(\Sigma, u(X))$ are such that $a' \geq \lambda a + (1-\lambda)c$ and $b' \leq \lambda b + (1-\lambda)c$, then also $P(a') \geq P(b')$ for all $P \in C$, so by Proposition 3 and the fact that $f \succcurlyeq^* g$ implies $f \succcurlyeq g$, $I(a') \geq I(b')$. Thus, for given $\lambda \in (0,1]$ and $c \in B_b(\Sigma, u(X))$ construct sequences $(a^k), (b^k) \subset B_0(\Sigma, u(X))$ with $a^k \downarrow \lambda a + (1-\lambda)c$ and $b^k \uparrow \lambda b + (1-\lambda)c$: we have $I(a^k) \geq I(b^k)$ by the argument just given, so (i) must hold by continuity of I on $B_b(\Sigma, u(X))$.

Finally, suppose there exists another set C' for which (i) and (iii) are equivalent for all $a, b \in B_b(\Sigma, u(X))$. The argument just given shows that (i) and (ii) are equivalent, so we can just as well assume that C' is another set for which (ii) and (iii) are equivalent for all $a, b \in B_b(\Sigma, u(X))$. Define another preference $\succcurlyeq^\dagger$ on $\mathscr{F}$ by letting $f \succcurlyeq^\dagger g$ iff $P(u \circ f) \geq P(u \circ g)$ for all $P \in C'$. But

since (ii) and (iii) are in particular equivalent for $u \circ f$ and $u \circ g$, we must have $f \succcurlyeq^{\dagger} g$ iff $f \succcurlyeq^{*} g$. It follows that u and C' also provide a representation of $\succcurlyeq^{*}$ on $\mathscr{F}$: but Proposition 3 asserts the uniqueness of C, so $C' = C$. ■

C.5 Proof of Corollary 14

By Corollary 13, (ii) holds iff for all $a \in B_0(\Sigma, u(X))$, $\underline{C}(a) \leq P(a)$. The equivalence of (i) and (ii) then follows, e.g., from Clarke (1983), Proposition 2.1.4 (b).

C.6 Proof of Corollary 15

We first show the following preliminary result: for $f \in \mathscr{F}$, $\underline{C}(u \circ f) \leq I(u \circ f) \leq \overline{C}(u \circ f)$. To see this, let $\underline{x} \in X$ be such that $u(\underline{x}) = \underline{C}(u \circ f)$; such a prize exists because X is convex, u is affine, and there are $x', x'' \in X$ such that $x' \succcurlyeq f(s) \succcurlyeq x''$, and hence $u(x') \geq P(u \circ f) \geq u(x'')$ for every $P \in ba_1(\Sigma)$. Hence by Prop. 3 $f \succcurlyeq^{*} \underline{x}$, which, as argued in the proof of Prop. 3, implies that $f \succcurlyeq \underline{x}$, i.e. $I(u \circ f) \geq u(\underline{x}) = \underline{C}(u \circ f)$. The argument for $\overline{C}(u \circ f)$ is symmetric.

Now suppose f is crisp and let x_f be its certainty equivalent. Then by definition $f \sim^{*} x_f$, i.e. $P(u \circ f) = u(x_f)$ for every $P \in C$. Conversely, if $\underline{C}(u \circ f) = \overline{C}(u \circ f)$, then $u(x_f) = I(u \circ f) = P(u \circ f)$ for all $P \in C$ by the preceding claim, and therefore $f \sim^{*} x_f$.

C.7 Proof of Proposition 4

We show that C is the smallest set (by inclusion) that satisfies Property (i). That C satisfies that property is clear, because $Q(u \circ f) \geq Q(u \circ g)$ for all $Q \in C$ implies $f \succcurlyeq^{*} g$ by Proposition 3, and hence in particular $f \succcurlyeq g$. Now suppose another set $D \subset ba_1(\Sigma)$ also satisfies Property (i). If $Q(u \circ f) \geq Q(u \circ g)$ for all $Q \in D$, then, for all $\lambda \in (0,1]$ and $h \in \mathscr{F}$, also $Q(u \circ [\lambda f + (1-\lambda)h]) \geq Q(u \circ [\lambda g + (1-\lambda)h])$ for all $Q \in D$. Since D satisfies (i), this implies $\lambda f + (1-\lambda)h \succcurlyeq \lambda g + (1-\lambda)h$ for all $\lambda \in (0,1]$ and $h \in \mathscr{F}$: that is, $f \succcurlyeq^{*} g$. But by Prop. 3, this implies that $Q(a) \geq Q(b)$ for all $Q \in C$. By Proposition A.1 in GMM, this in turn implies that $C \subset \overline{\text{co}}\, D$, as claimed.

Next, suppose D is a weak* closed, convex subset of $ba_1(\Sigma)$ that satisfies Property (i), and $D \neq C$; we show that D cannot also satisfy Property (ii). As was just argued, it must be the case

that $C \subsetneq D$. Hence, for all $f, g \in \mathscr{F}$ such that $Q(u \circ f) \geq Q(u \circ g)$, we have $f \succcurlyeq g$ because C satisfies Property (i); therefore, D fails Property (ii).

Finally, we show that C satisfies Property (ii). Fix a weak* closed, convex set $D \subsetneq C$ and let $P \in C \setminus D$. Then, by a separation theorem (Megginson, 1998, Theorem 2.2.8 and Proposition 2.6.4), there exists $d' \in B_0(\Sigma)$ such that $\inf_{Q \in D} Q(d') > P(d')$. Let $d = d' - \inf_{Q \in D} Q(d')$, so $Q(d) \geq 0$ for all $Q \in D$ and $P(d) < 0$. Now choose $\gamma \in \operatorname{int} u(X)$; there is $\epsilon > 0$ such that $\gamma + \epsilon d \in B_0(\Sigma, u(X))$. For all $Q \in D$, $Q(\gamma + \epsilon d) = \gamma + \epsilon Q(d) \geq \gamma = Q(1_S \gamma)$; on the other hand, $P(\gamma + \epsilon d) = \gamma + \epsilon P(d) < \gamma = P(1_S \gamma)$. Take $a = \gamma + \epsilon d$ and $b = 1_S \gamma$: then $Q(a) \geq Q(b)$ for all $Q \in D$, but it is *not* the case that $Q(a) \geq Q(b)$ for all $Q \in C \supset D$, and hence, by Prop. 3, it is *not* the case that $f' \succcurlyeq^* g'$, where $f', g' \in \mathscr{F}$ are acts such that $u \circ f' = a$ and $u \circ g' = b$. By definition, there exist $\lambda \in (0,1]$ and $h \in \mathscr{F}$ such that $\lambda f' + (1-\lambda)h \prec \lambda g' + (1-\lambda)h$. Now define $f = \lambda f' + (1-\lambda)h$ and $g = \lambda g' + (1-\lambda)h$: then $f \prec g$, and for all $Q \in D$, $Q(u \circ f) = \lambda Q(u \circ f') + (1-\lambda) Q(u \circ h) = \lambda Q(a) + (1-\lambda) Q(u \circ h) \geq \lambda Q(b) + (1-\lambda) Q(u \circ h) = \lambda Q(u \circ g') + (1-\lambda) Q(u \circ h) = Q(u \circ g)$, as required.

C.8 Proof of Theorem 5

We first construct a pair of functionals that correspond intuitively to the "lower and upper envelope" preferences of GMM. In GMM, these functionals coincide with the upper and lower Clarke derivatives of I at 0 (i.e. at certainty). In our more general setting, a more delicate construction is required, but the functionals in the following definition retain a central role.

Definition 9 Let Γ be an interval with $\operatorname{int} \Gamma \neq \emptyset$, and consider an isotone, monotonic and continuous functional $I : B_b(\Sigma, \Gamma) \to \mathbb{R}$. Define the functionals $I_0, I^0 : B_b(\Sigma, \Gamma) \to \mathbb{R}$ as follows

$$I_0(a) = \sup \left\{ \xi \in \Gamma : \forall \lambda \in (0,1], c \in B_b(\Sigma, \Gamma),\ I(\lambda a + (1-\lambda)c) \geq I(\lambda \xi + (1-\lambda)c) \right\} \tag{10}$$

$$I^0(a) = \inf \left\{ \xi \in \Gamma : \forall \lambda \in (0,1], c \in B_b(\Sigma, \Gamma),\ I(\lambda a + (1-\lambda)c) \leq I(\lambda \xi + (1-\lambda)c) \right\} \tag{11}$$

These definitions are well-posed ($a \in B_b(\Sigma, \Gamma)$ implies that there are $\alpha, \beta \in \Gamma$ such that $\alpha \geq a(s) \geq \beta$; now invoke the monotonicity of I). Furthermore:

Remark C.1 If I is as in Def. 9, then $I^0(a) \in \Gamma$ for all $a \in B_b(\Sigma, \Gamma)$; furthermore, for all $\lambda \in (0,1]$ and $c \in B_b(\Sigma, u(X))$, $I(\lambda a + (1-\lambda)c) \leq I(\lambda I^0(a) + (1-\lambda)c)$. Similar statements hold for I_0.

Proof: By definition, there is a sequence $\xi^k \in \Gamma$ such that $\xi^k \downarrow I^0(a)$, and for each such k we have $I(\lambda a+(1-\lambda)c) \leq I(\lambda\xi^k+(1-\lambda)c)$ for all $\lambda \in (0,1]$ and $c \in B_b(\Sigma,\Gamma)$. In particular, this holds for $\lambda = 1$; since $a \in B_b(\Sigma,\Gamma)$, there is $\alpha \in \Gamma$ such that $\alpha \leq a(s)$ for all s, and this implies that $\xi^k \geq \alpha$ for all k: if $\xi^k < \alpha$ for some k, then by isotony and monotonicity of I we would conclude that $I(\xi^k) < I(\frac{1}{2}\xi^k + \frac{1}{2}\alpha) \leq I(a)$, contradiction. Therefore, $I^0(a) \geq \alpha$ and so $I^0(a) \in \Gamma$. Now continuity of I yields the required conclusion for any λ and c. ■

Corollary 21 *Let (I,u) and C be as in Prop. 3. For all $a \in B_b(\Sigma,u(X))$, $I_0(a) = \underline{C}(a)$ and $I^0(a) = \overline{C}(a)$.*

Proof: Choose $a \in B_b(\Sigma,u(X))$. Comparing the definition of I^0 to (i) in Cor. 20 shows that

$$I^0(a) = \inf\{\xi \in u(X) : \forall P \in C,\, P(a) \leq \xi\} = \inf\{\xi \in u(X) : \overline{C}(a) \leq \xi\} = \overline{C}(a).$$

The second equality follows by noting that, if $P(a) \leq \xi$ for all $P \in C$, then also $\max_{P\in C} P(a) \leq \xi$; the converse is obvious. The third follows by noting that, once again, $a \in B_b(\Sigma,u(X))$ implies the existence of $\alpha,\beta \in u(X)$ such that $\alpha \geq a(s) \geq \beta$ for all s, so that $\alpha \geq \overline{C}(a) \geq \beta$: thus, $\overline{C}(a)$ is not only the smallest real number not smaller than $\overline{C}(a)$, but also the smallest number in $u(X)$ having this property. The argument for I_0 is symmetric. ■

Proof of Theorem 5: Recall that, by Lemma 17, $e \in \text{int } B_b(\Sigma,u(X))$ if and only if $e \in B(\Sigma)$ and $\inf_s e, \sup_s e \in \text{int } u(X)$. Also, per 1 in Prop. 19, $\partial I(e) \neq \emptyset$ for a dense subset of $B_0(\Sigma,u(X))$.

Since I is monotonic, by Statement 4 in Prop. 19, $Q \in \partial I(e)$ implies that $Q(S) \geq 0$. Moreover, we claim that there exist $e \in \text{int } B_b(\Sigma,u(X))$ and $Q \in \partial I(e)$ such that $Q(S) > 0$. Suppose not: then $\partial I(e) = \{Q_0\}$ for all $e \in \text{int } B_b(\Sigma,u(X))$ such that $\partial I(e) \neq \emptyset$, where $Q_0(E) = 0$ for all $E \in \Sigma$. In this case, choose $\vartheta,\vartheta' \in \text{int } u(X)$ and $\vartheta > \vartheta'$: by the approximate Mean Value Theorem (Statement 3 in Prop. 19) applied to the constant functions $1_S\vartheta, 1_S\vartheta' \in \text{int } B_b(\Sigma,u(X))$, there are sequences $(c_n) \in \text{int } B_b(\Sigma,u(X))$ and $(Q_n) \in ba(\Sigma)$ with $c_n \in \text{int } B_b(\Sigma,u(X))$ and $Q_n \in \partial I(c_n)$ for each n such that $0 = \liminf_n Q_n(\vartheta 1_S - \vartheta' 1_S) \geq I(\vartheta 1_S) - I(\vartheta' 1_S)$. By isotony, this implies $\vartheta' \geq \vartheta$, a contradiction.

Now observe that, by Proposition 1, $\succcurlyeq$ satisfies Axioms 1–6, and so the set C is well-defined and has the properties in Proposition 3 and Corollary 20. Let $D = \bigcup_{e\in \text{int } B_b(\Sigma,u(X))} \partial I(e)$; consider

arbitrary acts $f, g \in \mathscr{F}$, and define $a = u \circ f - u \circ g$. We claim that $Q(a) \geq 0$ for all $Q \in D$ if and only if $P(a) \geq 0$ for all $P \in C$. To prove this claim, fix $\vartheta \in \text{int } u(X)$ and let $\delta > 0$ be such that $\vartheta \pm \delta a \in \text{int } B_b(\Sigma, u(X))$.

Suppose that $Q(a) \geq 0$ for all $Q \in D$. Fix an arbitrary $\lambda \in (0,1]$ and $d \in B_b(\Sigma, u(X))$, and let $\Delta \equiv I(\lambda(\vartheta + \delta a) + (1-\lambda)d) - I(\lambda\vartheta + (1-\lambda)d)$. (Notice that if $\lambda = 0$, $\Delta = 0$ holds trivially.) By the approximate Mean Value Theorem,[14] there are (c_n), (Q_n) such that $c_n \in \text{int } B_b(\Sigma, u(X))$ and $Q_n \in \partial I(c_n)$ for each n and $\liminf_n Q_n(\lambda\vartheta + (1-\lambda)d - \lambda(\vartheta + \delta a) - (1-\lambda)d) \geq -\Delta$. But

$$Q_n(\lambda\vartheta + (1-\lambda)d - \lambda(\vartheta + \delta a) - (1-\lambda)d) = Q_n(-\lambda\delta a) = -\lambda\delta Q_n(a) \leq 0$$

by assumption, and so $\Delta \geq 0$. Since λ and d were arbitrary, this implies that $I_0(\vartheta + \delta a) \geq \vartheta$, so it follows by Corollary 21 that $\underline{C}(\vartheta + \delta a) = \vartheta + \delta \underline{C}(a) \geq \vartheta$, implying $P(a) \geq 0$ for all $P \in C$.

Conversely, suppose that $P(a) \geq 0$ for all $P \in C$, so $I^0(\vartheta - \delta a) = \overline{C}(\vartheta - \delta a) = \vartheta + \delta \overline{C}(-a) \leq \vartheta$. Fix $e \in \text{int } B_b(\Sigma, u(X))$ such that $\partial I(e) \neq \emptyset$. By the definition of I^{CR},

$$\begin{aligned}
I^{CR}(e; -\delta a) &= \lim_{\epsilon \downarrow 0} \limsup_{d \to e, t \downarrow 0} \inf_{b : \|b - (-\delta a)\| < \epsilon} \frac{I(d + tb) - I(d)}{t} = \\
&= \lim_{\epsilon \downarrow 0} \limsup_{d \to e, t \downarrow 0} \inf_{b : \|b + \delta a\| < \epsilon} \frac{I((1-t)d + tb) - I((1-t)d)}{t} = \\
&= \lim_{\epsilon \downarrow 0} \limsup_{d \to e, t \downarrow 0} \inf_{b : \|b + \delta a\| < \epsilon} \frac{I\big((1-t)d + t(b + \delta a - \vartheta) + t(\vartheta - \delta a)\big) - I((1-t)d)}{t} = \\
&= \lim_{\epsilon \downarrow 0} \limsup_{d \to e, t \downarrow 0} \inf_{b : \|b + \delta a\| < \epsilon} \frac{I\left((1-t)[d + \frac{t}{1-t}(b + \delta a - \vartheta)] + t(\vartheta - \delta a)\right) - I((1-t)d)}{t} \leq \\
&\leq \lim_{\epsilon \downarrow 0} \limsup_{d \to e, t \downarrow 0} \inf_{b : \|b + \delta a\| < \epsilon} \frac{I\left((1-t)[d + \frac{t}{1-t}(b + \delta a - \vartheta)] + t\, I^0(\vartheta - \delta a)\right) - I((1-t)d)}{t} \leq \\
&\leq \lim_{\epsilon \downarrow 0} \limsup_{d \to e, t \downarrow 0} \inf_{b : \|b + \delta a\| < \epsilon} \frac{I\left((1-t)[d + \frac{t}{1-t}(b + \delta a - \vartheta)] + t\vartheta\right) - I((1-t)d)}{t} = \\
&= \lim_{\epsilon \downarrow 0} \limsup_{d \to e, t \downarrow 0} \inf_{b : \|b + \delta a\| < \epsilon} \frac{I((1-t)d + t(b + \delta a)) - I((1-t)d)}{t} = \\
&= \lim_{\epsilon \downarrow 0} \limsup_{d \to e, t \downarrow 0} \inf_{b' : \|b'\| < \epsilon} \frac{I(d + tb') - I(d)}{t} = I^{CR}(e; 0).
\end{aligned}$$

[14] In the notation of Statement 3 of Prop. 19, take $a = \lambda(\vartheta + \delta a) + (1-\lambda)d$ and $b = \lambda\vartheta + (1-\lambda)d$. We must verify that these points lie in int $B_b(\Sigma, u(X))$. This is immediate for $\lambda\vartheta + (1-\lambda)d$; for $\lambda(\vartheta + \delta a) + (1-\lambda)d$, recall that, by construction $\vartheta + \delta a \in \text{int } B_b(\Sigma, u(X))$, so $\inf_s(\vartheta + \delta a), \sup_s(\vartheta + \delta a) \in \text{int } u(X)$ by Lemma 17, and $\inf_s[\lambda(\vartheta + \delta a) + (1-\lambda)d] \geq \lambda \inf_s(\vartheta + \delta a) + (1-\lambda)\inf_s d \in \text{int } u(X)$; similarly, $\sup_s[\lambda(\vartheta + \delta a) + (1-\lambda)d] \in \text{int } u(X)$, so the claim follows again by Lemma 17.

The first inequality follows from Remark C.1, after observing that, if $e \in \text{int } B_b(\Sigma, u(X))$, so is d eventually, and therefore, given any choice of ϵ, if t is small enough we have $d + \frac{t}{1-t}(b + \delta a - \vartheta) \in B_b(\Sigma, u(X))$. The second inequality follows from monotonicity of I and the fact that, as we showed above, $I^0(\vartheta - \delta a) \leq \vartheta$. By Statement 2 in Prop. 19, we conclude that $0 = \sup_{Q \in \partial I(e)} Q(0) = I^{CR}(e;0) \geq I^{CR}(e;-\delta a) = \sup_{Q \in \partial I(e)} Q(-\delta a)$; in other words, for all $Q \in \partial I(e)$, $\delta Q(-a) = Q(-\delta a) \leq 0$, or $Q(a) \geq 0$. Since e was arbitrary, $Q(a) \geq 0$ for all $Q \in D$.

The set in the right-hand side of the displayed equation in the statement, denoted C', is the convex closure of the set of non-zero measures in D, normalized by dividing each Q by $Q(S)$; it is non-empty by the assumptions made. Since f and g were arbitrary, we have shown the following: for all $f, g \in \mathscr{F}$, $P(u \circ f) \geq P(u \circ g)$ for all $P \in C$ iff $P(u \circ f - u \circ g) \geq 0$ for all $P \in C$ iff $Q(u \circ f - u \circ g) \geq 0$ for all $Q \in D$ iff $P(u \circ f) \geq P(u \circ g)$ for all $P \in C'$. Therefore, $C = C'$ by Prop. 3. ■

C.9 Proof of the results in Sec. 6.2

C.9.1 Proposition 9 (smooth ambiguity)

Begin with some preliminary observations. The set $ca_1(\Sigma)$, endowed with the topology induced by $C_b(S)$ (i.e. the "weak-convergence" topology), is a Polish space, and so μ is a regular measure. Henceforth, we let $\mathscr{Q} \equiv ca_1(\Sigma)$, and whenever we refer to $\mathscr{Q}$ as a topological or measurable space, we implicitly assume that it is endowed with its weak-convergence topology and, respectively, with the corresponding Borel sigma-algebra.

The map $Q \mapsto Q(d)$ defined on $\mathscr{Q}$ is Borel measurable for every $d \in B_b(\Sigma, u(X))$ [cf. Theorem 17.24 in Kechris (1995)], and ϕ is continuous, hence Borel measurable; thus, the map $Q \mapsto \phi(Q(d))$, also viewed as a function on $\mathscr{Q}$, is Borel measurable. Furthermore, $|Q(d) - Q(d')| \leq \|d - d'\|$ for all $Q \in \mathscr{Q}$ and $d, d' \in B_b(\Sigma)$. Finally, since ϕ is locally Lipschitz, it satisfies a Lipschitz condition on every compact subset of $u(X)$; this implies that, for every $\alpha, \beta \in \text{int } u(X)$ there is $\ell > 0$ such that $|\phi(\gamma) - \phi(\gamma')| \leq \ell|\gamma - \gamma'|$ whenever $\gamma, \gamma' \in (\alpha, \beta)$.

Since ϕ is a strictly increasing function on a subset of $\mathbb{R}$, $\alpha \in \partial\phi(\gamma)$ for some $\gamma \in \text{int } u(X)$ implies $\alpha \geq 0$ by e.g. Statement 4 in Prop. 19.

This completes the preliminaries. Now fix $e \in \text{int } B_b(\Sigma, u(X))$, so $\min_s e(s), \max_s e(s) \in$

int $u(X)$. Our first objective is to compute $\partial I(e)$. We mimic the proof of Theorem 2.7.2 in Clarke (1983), adapting one key step to our setting.[15]

Choose $\epsilon > 0$ such that $\min_s e(s) - \epsilon, \max_s e(s) + \epsilon \in \text{int } u(X)$, and let $U = B_b\big(\Sigma, (\min_s e(s) - \epsilon, \max_s e(s) + \epsilon)\big)$. Then U is open by Lemma 17, $e \in U$, and for every $d \in U$ and $Q \in \mathscr{Q}$, $Q(d) \in (\min_s e(s) - \epsilon, \max_s e(s) + \epsilon)$. Furthermore, there is $\ell \geq 0$ such that $|\phi(Q(d)) - \phi(Q(d'))| \leq \ell |Q(d) - Q(d')|$ for every $Q \in T$ and $d, d' \in U$, and hence $|\phi(Q(d)) - \phi(Q(d'))| \leq \ell \|d - d'\|$, where the constant ℓ does not depend upon Q. Therefore, for $d, d' \in U$,

$$|I(d) - I(d')| \leq \int_{\mathscr{Q}} |\phi(Q(d)) - \phi(Q(d'))| \, d\mu(Q) \leq \ell \|d - d'\|,$$

i.e. I is Lipschitz on U; thus, the CR differential $\partial I(e)$ is actually a Clarke differential.

Temporarily denote by $\Lambda(e)$ the class of measurable maps $\lambda : \mathscr{Q} \mapsto \mathbb{R}$ such that $\lambda(Q) \in \partial\phi(Q(e))$ for μ-almost all $Q \in \mathscr{Q}$. We first show that

$$\partial I(e) \subset \left\{ \int_{\mathscr{Q}} \lambda(Q) Q \, d\mu(Q) : \lambda \in \Lambda(e) \right\}, \tag{12}$$

where the integrals inside the braces denote the linear functional $a \mapsto \int_{\mathscr{Q}} \lambda(Q) Q(a) \, d\mu$ defined for all $a \in B(\Sigma)$. To do so, fix $L \in \partial I(e)$. By definition, for all $a \in B(\Sigma)$, $L(a) \leq I^c(e; a)$. The assumptions on S, μ and ϕ imply Clarke's Hypotheses 2.7.1, and in particular allow us to invoke Fatou's lemma and conclude that, for any fixed $a \in B(\Sigma)$,

$$\begin{aligned} L(a) &\leq I^c(e; a) = \limsup_{d \to e, t \downarrow 0} \int_{\mathscr{Q}} \frac{\phi(Q(d + ta)) - \phi(Q(d))}{t} \, d\mu(Q) \leq \\ &\leq \int_{\mathscr{Q}} \limsup_{d \to e, t \downarrow 0} \frac{\phi(Q(d + ta)) - \phi(Q(d))}{t} \, d\mu(Q) = \\ &= \int_{\mathscr{Q}} (\phi \circ Q)^c(e; a) \, d\mu(Q) = \int_{\mathscr{Q}} \max_{M \in \partial(\phi \circ Q)(e)} M(a) \, d\mu(Q). \end{aligned}$$

We now follow the strategy of Clarke's "alternate proof sketch" on pp. 77-78, rather than his main proof. However, the key step in Clarke's sketch involves applying a measurable selection theorem to the correspondence $Q \mapsto \partial(\phi \circ Q)(e)$; this requires assumptions that need not hold in our setting. Thus, we exploit the special structure of the functional I instead.

If ϕ is regular, then for every $d \in U$ and $Q \in \mathscr{Q}$, $\partial(\phi \circ Q)(d)) = \{\lambda Q : \lambda \in \partial\phi(Q(d))\}$ by Theorem 2.3.9 part (iii) in Clarke (1983), and $\phi \circ Q$ is also regular. If $-\phi$ is regular, we apply the

[15] Unfortunately, we cannot invoke Clarke's result directly, unless S is finite or $\text{supp}(\mu)$ is countable.

cited result to the map $d \mapsto ((-\phi) \circ Q)(d) = -(\phi \circ Q)(d)$, concluding that $-\phi \circ Q$ is regular and $\partial(-\phi \circ Q)(d) = \{\lambda Q : \lambda \in \partial(-\phi)(Q(d))\}$; but $\partial(-\phi)(\gamma) = -\partial\phi(\gamma)$ and $\partial(-\phi \circ Q) = -\partial(\phi \circ Q)$ by Prop. 2.3.1 in Clarke (1983), so again $\partial(\phi \circ Q)(d) = -\partial(-\phi \circ Q)(d) = \{-\lambda Q : \lambda \in \partial(-\phi)(Q(d))\} = \{-\lambda Q : \lambda \in -\partial\phi(Q(d))\} = \{\lambda' Q : \lambda' \in \partial\phi(Q(d))\}$. Hence, in either case, we can write

$$L(a) \leq \int_{\mathcal{Q}} \max_{\lambda_Q \in \partial\phi(Q(e))} \lambda_Q Q(a)\, d\mu(Q).$$

We now apply a measurable maximum theorem. By Proposition 2.1.5 (d) in Clarke (1983), the map $\gamma \mapsto \partial\phi(\gamma)$ is upper hemicontinuous; hence, the lower inverse (Aliprantis and Border, 2007, p.557) of every closed set is closed (Aliprantis and Border, 2007, Lemma 17.4), so $\partial\phi$ is measurable (hence also weakly measurable) in the sense of Def. 18.1 of Aliprantis and Border (2007) (see also Lemma 18.2 therein). Now consider an open subset V of $\mathbb{R}$: since $\partial\phi$ is weakly measurable, the set $V' = \{\gamma : \partial\phi(\gamma) \cap U \neq \emptyset\}$ is Borel in $\mathbb{R}$; since, as noted above, the evaluation map $Q \mapsto Q(e)$ is Borel measurable on $\mathcal{Q}$, the preimage of V', i.e. $\{Q : Q(e) \in \{\gamma : \partial\phi(\gamma) \cap V \neq \emptyset\}\} = \{Q : \partial\phi(Q(e)) \cap V \neq \emptyset\}$ is Borel in $\mathcal{Q}$. Therefore, the correspondence $Q \mapsto \partial\phi(Q(e))$, defined on the measurable space $\mathcal{Q}$ and taking values in the (separable metric) space $\mathbb{R}$, is weakly measurable. Finally, the map $(Q, \lambda) \mapsto \lambda Q(a)$ is a Caratheodory function: that is, for every $\lambda \in \mathbb{R}$ the map $Q \mapsto \lambda Q(a)$ is Borel-measurable, because it is a constant multiple of an evaluation function on $\mathcal{Q}$, and for every $Q \in \mathcal{Q}$, the map $\lambda \mapsto \lambda Q(a)$ is continuous. Finally, $Q \mapsto \partial\phi(Q(e))$ takes values in a separable metric space, has non-empty compact values, and is weakly measurable. Then, by Theorem 18.19 in Aliprantis and Border (2007), there exists a measurable selection from the argmax correspondence, i.e. a map $\lambda_a \in \Lambda(e)$ such that, for every $Q \in \mathcal{Q}$, $\lambda_a(Q)Q(a) = \max_{\lambda \in \partial\phi(Q(e))} \lambda Q(a)$.

We conclude that, for every $a \in B(\Sigma)$, we can write

$$L(a) \leq \max_{\lambda \in \Lambda(e)} \int_{\mathcal{Q}} \lambda(Q)Q(a)\, d\mu(Q).$$

We now go back to Clarke's proof sketch, and merely adapt his notation and fill in details that Clarke omits. Since in particular equality obtains in case $a = 0$, we can rewrite this as

$$\min_{a \in B(\Sigma)} \max_{\lambda \in \Lambda(e)} \left\{ \int_{\mathcal{Q}} \lambda(Q)Q(a)\, d\mu(Q) - L(a) \right\} = 0.$$

The next step entails applying a "lop-sided minmax theorem" (Aubin, 2000, Theorem 2.7.1); we now verify that the required assumptions hold (cf. Clarke, 1977, proof of Lemma 2). First, the

sets $B(\Sigma)$ and $\Lambda(e)$ are convex; for the latter, this follows from the fact that each $\partial\phi(\gamma)$ is itself convex. Second, $\Lambda(e)$ is a weak-compact subset of $L^1(\mu)$. To see this, note first that, as was shown above, ϕ satisfies a Lipschitz condition with Lipschitz constant ℓ on $[\min_s e(s), \max_s e(s)]$; by Remark 5.2.1 in ACL, this implies that $|\lambda| \leq \ell$ for all $\gamma \in [\min_s e(s), \max_s e(s)]$ and $\lambda \in \partial\phi(\gamma)$; in particular, if $\lambda \in \Lambda(e)$, then $|\lambda(Q)| \leq \ell$ for μ-almost all $Q \in \mathscr{Q}$. Hence, $\Lambda(e) \subset L^1(\mu)$. Furthermore, if $(\lambda^k) \in \Lambda(e)$ converges to some $\lambda \in L^1(\mu)$ in the $L^1(\mu)$ norm, then some subsequence converges pointwise (Aliprantis and Border, 2007, Theorem 13.6), and this implies that $\lambda(Q) \in \partial\phi(Q(e))$ μ-a.e. because $\partial\phi(Q(e))$ is closed. Thus, $\Lambda(e)$ is norm-closed, and hence (because it is convex) also weak-closed (e.g. Dunford and Schwartz, 1957, V.3.13). Finally, since $|\lambda(Q)| \leq \ell < \infty$ for all $\lambda \in \Lambda(e)$ and μ-almost all $Q \in \mathscr{Q}$, the Dunford-Pettis criterion (e.g. Dunford and Schwartz, 1957, Theorem IV.8.9) implies that $\Lambda(e)$ is weakly sequentially compact; finally, since it is also weak closed, it is weak compact by the Eberlein-Smulian theorem, as required (e.g. Dunford and Schwartz, 1957, Theorem V.6.1). The above arguments imply that $\Lambda(e)$ is non-empty. Lastly, the function $(a, \lambda) \mapsto \int_{\mathscr{Q}} \lambda(Q)Q(a)\,d\mu(Q) - L(a)$ defined on $B(\Sigma) \times \Lambda(e)$ is clearly linear in a for every $\lambda \in \Lambda(e)$, and weakly continuous in λ for every $a \in B(\Sigma)$, because the evaluation map $Q \mapsto Q(a)$ is bounded [in the sense that $Q(a) \in [\min_s a(s), \max_s a(s)]$ for all Q] and thus defines a continuous linear functional $\lambda \mapsto \int_{\mathscr{Q}} \lambda(Q)Q(a)\,d\mu$ on $L^1(\mu)$ and hence $\Lambda(e)$. The cited theorem by Aubin then yields $\lambda^* \in \Lambda(e)$ such that

$$\min_{a \in B(\Sigma)} \left\{ \int_{\mathscr{Q}} \lambda^*(Q)Q(a)\,d\mu(Q) - L(a) \right\} = 0.$$

But then it cannot be the case that the quantity in braces is strictly positive for some function a: otherwise, it would be strictly negative for $-a$. Hence, we conclude that

$$\forall a \in B(\Sigma), \quad L(a) = \int_{\mathscr{Q}} \lambda^*(Q)Q(a)\,d\mu(Q),$$

which states that every $L \in \partial I(e)$ has the required representation. Hence, Eq. (12) holds. Observe that, since $\lambda^*(Q) \in \partial\phi(Q(e))$ for μ-almost all Q, $\lambda^*(Q) > 0$ for such Q, which implies that $L(1_S) > 0$ for every $L \in \partial I(e)$.

The "final steps" of Clarke's proof (p. 79) apply almost verbatim. If ϕ is regular, we noted above that, by our assumptions, the map $d \mapsto \phi \circ Q$ is regular for every $Q \in \mathscr{Q}$; Clarke's argument then applies to show that $I'(e;a) = \int_{\mathscr{Q}} (\phi \circ Q)'(e;a)\,d\mu(Q)$ for all $a \in B(\Sigma)$, and that I is

itself regular. Then, if $L = \int_{\mathcal{Q}} \lambda(Q) Q \, d\mu(Q)$ for some $\lambda \in \Lambda(e)$, then by definition $\lambda(Q) \in \partial\phi(Q(e))$ for μ-almost all Q, and so $I^0(e;a) = I'(e;a) = \int_{\mathcal{Q}} (\phi \circ Q)'(e;a) \, d\mu(Q) = \int_{\mathcal{Q}} (\phi \circ Q)^\circ(e;a) \, d\mu(Q) = \int_{\mathcal{Q}} \max_{\lambda \in \partial\phi(Q(e))} \lambda Q(a) \, d\mu(Q) \geq \int_{\mathcal{Q}} \lambda(Q) Q(a) \, d\mu(Q) = L(a)$, where the first and third equalities follow by regularity, so $L \in \partial I(e)$. If instead $-\phi$ is regular, then $d \mapsto -\phi \circ Q$ is regular for every $Q \in \mathcal{Q}$, and since $(-I)(d) = \int_{\mathcal{Q}} (-\phi \circ Q)(d) \, d\mu(Q)$, we conclude that $(-I)'(e;a) = \int_{\mathcal{Q}} (-\phi \circ Q)'(e;a) \, d\mu(Q)$, and $-I$ is regular. Then, if $\lambda(Q) \in \Lambda(e)$ and L are as above, $-\lambda(Q) \in \partial(-\phi(Q(e))$ for μ-almost all Q, and so $(-I)^0(e;a) = (-I)'(e;a) = \int_{\mathcal{Q}} (-\phi \circ Q)'(e;a) \, d\mu(Q) = \int_{\mathcal{Q}} (-\phi \circ Q)^\circ(e;a) \, d\mu(Q) = \int_{\mathcal{Q}} \max_{\lambda \in \partial(-\phi(Q(e)))} \lambda Q(a) \, d\mu(Q) \geq \int_{\mathcal{Q}} -\lambda(Q) Q(a) \, d\mu(Q) = -L(a)$. Thus, $-L \in \partial(-I)(e)$, so $L \in \partial I(e)$. Thus, finally,

$$\partial I(e) = \left\{ \int_{ca_1(\Sigma)} \lambda(Q) \cdot Q \, d\mu(Q) : \lambda \in \Lambda(e) \right\}. \tag{13}$$

Now apply Theorem 5 to obtain the result.

C.9.2 Calculations for the example in Sec. 2

In Eq. (6), assume that ϕ is CARA, i.e. $\phi(x) = -e^{-\frac{x}{\theta}}$, and that $\mathrm{supp}(\mu)$, the support of μ, is finite. Fix an exposed point $Q \in \mathrm{supp}(\mu)$; then there is $a \in \mathrm{int}\, u(X)^{|S|}$ such that $Q(a) < Q'(a)$ for all $Q' \in \mathrm{supp}(\mu) \setminus \{Q\}$. Let $a_{K,\theta} = K\theta a$: then, for every $s \in S$,

$$Q_{a_{K,\theta},\theta}(\{s\}) = \frac{Q(\{s\})\mu(Q) + \sum_{Q' \neq Q} e^{-K[Q'(a)-Q(a)]} Q'(\{s\})\mu(Q')}{\mu(Q) + \sum_{Q' \neq Q} e^{-K[Q'(a)-Q(a)]} \mu(Q')} \to Q(\{s\})$$

as $K \to \infty$. Therefore, $C = \overline{\mathrm{co}}\, \mathrm{supp}(\mu)$.

Now let $S = \{s_1, s_2\}$ and, as in Sec. 2, consider $Q_1, Q_2 \in \Delta(S)$ such that $\frac{1}{2}Q_1 + \frac{1}{2}Q_2$ is uniform. This implies that $\frac{1}{2}Q_1(\{s_1\}) + \frac{1}{2}Q_2(\{s_1\}) = \frac{1}{2}$, so $Q_1(\{s_1\}) = 1 - Q_2(\{s_1\}) = Q_2(\{s_2\}) \equiv q_1$, and similarly $Q_1(\{s_2\}) = Q_2(\{s_1\}) \equiv q_2$. Now let $\mu(\{Q_1\}) = \mu(\{Q_2\}) = \frac{1}{2}$, $X = \mathbb{R}_+$, $u(x) = x$, and $\phi(x) = \frac{x^{1-\gamma}}{1-\gamma}$, with

$\gamma > 0$, again as in Sec. 2. Assume wlog that $q_1 > q_2$. Then, from Eq. (6),

$$\begin{aligned}
C &= \overline{\text{co}}\left(\frac{\frac{1}{2}\left[Q_1(\{s_1\})x+Q_1(\{s_2\})y\right]^{-\gamma}Q_1+\frac{1}{2}\left[Q_2(\{s_1\})x+Q_2(\{s_2\})y\right]^{-\gamma}Q_2}{\frac{1}{2}\left[Q_1(\{s_1\})x+Q_1(\{s_2\})y\right]^{-\gamma}+\frac{1}{2}\left[Q_2(\{s_1\})x+Q_2(\{s_2\})y\right]^{-\gamma}} : x,y\in\mathbb{R}_{++}\right) = \\
&= \overline{\text{co}}\left(\frac{[q_1x+q_2y]^{-\gamma}Q_1+[q_2x+q_1y]^{-\gamma}Q_2}{[q_1x+q_2y]^{-\gamma}+[q_2x+q_1y]^{-\gamma}} : x,y\in\mathbb{R}_{++}\right) = \\
&= \overline{\text{co}}\left(\frac{[q_1+q_2t]^{-\gamma}Q_1+[q_2+q_1t]^{-\gamma}Q_2}{[q_1+q_2t]^{-\gamma}+[q_2+q_1t]^{-\gamma}} : t\in\mathbb{R}_{++}\right) = \\
&= \overline{\text{co}}\left(\frac{1}{1+\left(\frac{q_1+q_2t}{q_2+q_1t}\right)^{\gamma}}Q_1+\frac{\left(\frac{q_1+q_2t}{q_2+q_1t}\right)^{\gamma}}{1+\left(\frac{q_1+q_2t}{q_2+q_1t}\right)^{\gamma}}Q_2 : t\in\mathbb{R}_{++}\right) = \\
&= \left\{\alpha Q_1+(1-\alpha)Q_2 : \alpha\in\left[\frac{1}{1+\left(\frac{q_1}{q_2}\right)^{\gamma}},\frac{\left(\frac{q_1}{q_2}\right)^{\gamma}}{1+\left(\frac{q_1}{q_2}\right)^{\gamma}}\right]\right\}.
\end{aligned}$$

C.9.3 Remark 6.1 (nice preferences)

We first propose a simple sufficient condition that implies that the MBC representation under consideration is nice. Say that an MBC representation (I,u) is **very nice** if, for all $a \in \text{int } B_b(\Sigma,u(X))$, there exist $\bar{t}>0$, $\eta>0$ and $\rho>0$ such that, for all $b\in B_b(\Sigma)$, the conditions $\|a-b\|<\eta$ and $t\in(0,\bar{t})$ imply $b, b-1_S t \in \text{int } B_b(\Sigma,u(X))$ and $I(b-t1_S)\leq I(b)-\rho t$.

The following result contains key facts about nice and very nice MBC representations.

Lemma 22 *If an MBC representation (I,u) is very nice, it is nice. Furthermore, $\partial(-I)(e) = -\partial I(e)$ for every $e\in\text{int } B_b(\Sigma,u(X))$, so if (I,u) is nice, then $Q_0\notin\partial(-I)(e)$; also,*

$$\forall\epsilon>0,\quad \exists e^+\in\text{int } B_b(\Sigma,u(X))\quad s.t.\quad \|e-e^+\|<\epsilon \text{ and } I(e^+)>I(e). \tag{14}$$

Proof: By definition,

$$\begin{aligned}
I^{CR}(e;-1_S) &= \lim_{\epsilon\downarrow 0}\limsup_{d\to e,t\downarrow 0}\inf_{b:\|b-(-1_S)\|<\epsilon}\frac{I(d+tb)-I(d)}{t}\leq \\
&\leq \limsup_{d\to e,t\downarrow 0}\frac{I(d-t1_S)-I(d)}{t}\leq\limsup_{d\to e,t\downarrow 0}\frac{I(d)-t\rho-I(d)}{t}=-\rho<0,
\end{aligned}$$

where the second inequality follows because, as soon as d is sufficiently close to e and t is sufficiently small, a suitable $\rho>0$ can be found. Since $I^{CR}(e;-1_S)=\sup_{Q\in\partial I(e)}Q(-1_S)$, the zero measure cannot belong to $\partial I(e)$.

For the remaining claims, since I is monotonic, by Theorem 1 part (d) in Rockafellar (1979), I is directionally Lipschitz, and thus (cf. p. 333 therein) $\partial(-I)(e) = -\partial I(e)$. In particular, $Q_0 \notin \partial(-I)(e)$, as claimed. Finally, for Eq. (14), Eqs. (1.5) and (1.7) in Rockafellar (1979), applied to $-I$, imply that, since $Q_0 \notin \partial(-I)(e)$, there exists $a \in B(\Sigma)$ and $\rho > 0$ such that, in particular, for every $\eta > 0$ there exists $\lambda_\eta > 0$ such that, for every $t \in (0, \lambda_\eta)$ there is $a'_t \in B(\Sigma)$ with $\|a'_t - a\| < \eta$ and $(-I)(e + t\,a'_t) \leq (-I)(e) - t\,\rho$, or equivalently $I(e + t\,a'_t) \geq I(e) + t\rho > I(e)$. For given $\epsilon > 0$, choose $\eta > 0$ and $\mu > 0$ such that $\|e + t\,a'\| \in \operatorname{int} B_b(\Sigma, u(X))$ and $\|(e + t\,a') - e\| = \|t\,a'\| < \epsilon$ for all $t \in (0, \mu)$ and $a' \in B(\Sigma)$ with $\|a' - a\| < \eta$, then choose $t \in (0, \min(\lambda_\eta, \mu))$ and define $e^+ = e + t\,a'_t$, where a'_t is the function whose existence is implied by Rockafellar's result. ■

Now turn to the proof of Remark 6.1. Variational preferences are characterized by $G(\gamma, Q) = \gamma + c(Q)$, and the infimum is always attained; furthermore, they are translation-invariant. This implies that they are uniformly continuous on $B_0(\Sigma, u(X))$, and hence satisfy Axiom 5, as shown in the proof of Proposition 1. Furthermore, $I(b - t\,1_S) = I(b) - t$ whenever $b, b - t\,1_S \in B_b(\Sigma, u(X))$. Thus, (I, u) is very nice, with $\rho = 1$ and arbitrary $\bar{t}$ and η.

CF preferences are characterized by $G(\gamma, Q) = \frac{\gamma}{\varphi(Q)}$, with $\varphi(Q) \in (0, 1]$, and again the infimum is always attained. Furthermore, the CF functional is Lipschitz continuous (Chateauneuf and Faro, 2009, Lemma 22), and hence satisfies Axiom 5. Finally, if $b, b - t\,1_S \in B_b(\Sigma, u(X))$,

$$I(b - t\,1_S) \leq \frac{Q_b(b - t\,1_S)}{\varphi(Q_b)} = I(b) - \frac{t}{\varphi(Q_b)} \leq I(b) - t\,,$$

where we let $Q_b \in \arg\min_{Q \in ba_1(\Sigma)} \frac{Q(b)}{\varphi(Q)}$. So again (I, u) is very nice.

Smooth ambiguity preferences extend to all of $B_b(\Sigma, u(X))$, as we argued in §6.2.2, and hence are MBC preferences. Furthermore, the assumptions on μ and u ensures that, by Theorem 16 in Cerreia-Vioglio et al. (2008), a UA representation exists; furthermore, since $u(X) = \mathbb{R}$, the infimum in Eq. (9) is attained. Next, recall that, if ϕ is concave, it has a finite and weakly decreasing right derivative ϕ'_- in the interior of its domain $u(X)$. By concavity, $\phi(\gamma - t) \leq \phi(\gamma) - t\,\phi'_-(\gamma)$ whenever $\gamma, \gamma - t \in \operatorname{int} u(X)$. Hence, for $a \in \operatorname{int} B_b(\Sigma, u(X))$ and for any $\eta > 0$, $\bar{t} > 0$ such that $\|b - a\| \leq \eta$ and $t \in (0, \bar{t})$ imply $b, b - 1_S t \in \operatorname{int} B_0(\Sigma, u(X))$, $\int \phi(Q(b - t\,1_S))\,d\mu(Q) \leq \int \phi(Q(b))\,d\mu(Q) - \rho\,t$, where $\rho = \phi'_-(\inf_s a - \eta)$;[16] this choice of ρ ensures that $-\phi'_-(Q(b))t \leq$

[16]Note that, if $\eta > 0$ is such that $\|b - a\| \leq \eta$ (the weak inequality is important) implies $b \in \operatorname{int} B_b(\Sigma, u(X))$, then in

$-\rho t$ whenever $\|a-b\|<\eta$. Thus, (I,u) is a very nice, hence nice MBC representation.

C.9.4 Proposition 12 (UA)

Denote by D the set in parentheses in the statement. To see that $C \subset \overline{\text{co}}\, D$, fix $f,g \in \mathscr{F}$ and suppose $Q(u\circ f) \geq Q(u\circ g)$ for all $Q \in D$. Then, for all $\lambda \in (0,1)$ and $h \in \mathscr{F}$, $Q(u\circ[\lambda f+(1-\lambda)h]) \geq Q(u\circ[\lambda g+(1-\lambda)h])$. Let $Q_f \in \arg\min_{Q\in ba_1(\Sigma)} Q(u\circ[\lambda f+(1-\lambda)h]) \geq Q(u\circ[\lambda g+(1-\lambda)h])$: then $I(u\circ[\lambda f+(1-\lambda)h]) = G(Q_f(u\circ[\lambda f+(1-\lambda)h],Q_f) \geq G(Q_f(u\circ[\lambda g+(1-\lambda)h]),Q_f) \geq \min_{Q\in ba_1(\Sigma)} G(Q(u\circ[\lambda g+(1-\lambda)h]),Q) = I(u\circ[\lambda f+(1-\lambda)h])$, i.e. $\lambda f+(1-\lambda)h \succcurlyeq \lambda g+(1-\lambda)h$. Thus, by definition, $f \succcurlyeq^* g$, and Proposition A.1 in GMM implies that $C \subset \overline{\text{co}}\, D$.

For the opposite inclusion, fix $a \in \text{int}\, B_b(\Sigma,u(X))$ such that $\partial I(a) \neq \emptyset$, and let $S_a = \{b \in B_b(\Sigma,u(X)) : I(b) \geq I(a)\}$. Since I is quasiconcave, S_a is convex. By Lemma 22, $Q_0 \notin \partial(-I)(a) = -\partial I(a)$, and $S_a = \{b : -I(b) \leq -I(a)\}$ is the sublevel set of $-I$ at a. By Theorem 5 in Rockafellar (1979), the (Clarke) normal cone to S_a at a is contained in $\bigcup_{\lambda\geq 0} \lambda\partial(-I)(a)$; since S_a is convex, its normal cone coincides with the set $N(a) = \{Q \in ba(\Sigma) : \forall b \in S_a,\ Q(a) \geq Q(b)\}$: cf. Rockafellar (1980), Theorem 1 and §7. Using the noted fact that $\partial(-I) = -\partial I$, we obtain

$$\{Q \in ba(\Sigma) : \forall b \in S_a,\ Q(b) \geq Q(a)\} \subset \bigcup_{\lambda\geq 0} \lambda\partial I(a). \tag{15}$$

We now show that

$$\arg\min_{Q'\in ba_1(\Sigma)} G(Q'(a),Q') \subset \{Q \in ba(\Sigma) : \forall b \in S_a,\ Q(b) \geq Q(a)\} \tag{16}$$

Fix $Q \in \arg\min_{Q'\in ba_1(\Sigma)} G(Q'(a),Q')$ and consider first $b \in S_a$ such that $I(b) > I(a)$: then

$$G(Q(b),Q) \geq \arg\min_{Q'\in ba_1(\Sigma)} G(Q'(b),Q') = I(b) > I(a) = G(Q(a),Q).$$

Since G is non-decreasing in its first argument, $Q(b) > Q(a)$. Next, consider an arbitrary $b \in S_a$ such that $I(b) \geq I(a)$. Since $a \in \text{int}\, B_b(\Sigma,u(X))$, the property in Eq. (14) yields $a^+ \in \text{int}\, B_b(\Sigma,u(X))$ such that $I(a^+) > I(a)$. Next, for $n \geq 1$, let $b_n = \frac{1}{n}a^+ + \frac{n-1}{n}b$. Since I is quasiconcave, $I(b_n) \geq \min(I(a^+),I(b)) \geq I(a)$; furthermore, by construction $b_n \in \text{int}\, B_b(\Sigma,u(X))$. We then invoke Eq. (14) again to obtain $b_n^+ \in \text{int}\, B_b(\Sigma,u(X))$ such that $\|b_n^+ - b_n\| < \frac{1}{n}$ and $I(b_n^+) > I(b_n)$, so

particular this is true for $b = a - 1_S\eta$, which implies that $\inf_s a - \eta \in \text{int}\, u(X)$ by Lemma 17; thus, ρ is well-defined.

$I(b_n^+) > I(a)$, and therefore $Q(b_n^+) > Q(a)$ by the preceding argument. But $\|b_n^+ - b\| \le \|b_n^+ - b_n\| + \|b_n - b\| \le \frac{1}{n} + \frac{1}{n}\|a^+ - b\| \to 0$, so $Q(b) = \lim_n Q(b_n^+) \ge Q(a)$.

We thus conclude that Eq. (16) holds, so that in light of Eq. (15) we obtain

$$\arg \min_{Q' \in ba_1(\Sigma)} G(Q'(a), Q') \subset \bigcup_{\lambda \ge 0} \lambda \partial I(a).$$

Clearly, $Q \in \arg\min_{Q' \in ba_1(\Sigma)} G(Q'(a), Q)$ satisfies $Q(S) = 1$; since, by Statement 4 in Prop. 19, $Q'(S) \ge 0$ for all $Q' \in \partial I(a)$, and $Q_0 \notin \partial I(a)$, $Q = \lambda Q'$ for $\lambda = 1/Q'(S)$ and $Q' \in \partial I(a)$, i.e. $Q = Q'/Q'(S)$ for some non-zero $Q' \in \partial I(a)$. Now Theorem 5 implies that $Q \in C$. Since a and Q were arbitrary, $D \subset C$.

C.9.5 Proposition 10

Denote by $H : B_b(\Sigma, u(X)) \to \mathbb{R}^n$ the map defined by $H_i(a) = P(\zeta_i a)$ for $0 \le i < n$ and $H_n(a) = P(a)$; also, let $\tilde{W} : \mathcal{E}(P,\zeta) \times u(X) \to \mathbb{R}$ be defined by $\tilde{W}(\varphi, \gamma) = \gamma + A(\varphi)$ for all $\gamma \in u(X)$ and $\varphi \in \mathcal{E}(P,\zeta)$. Then $I = \tilde{W} \circ H$, $\partial H_i = \{M_i\}$ for $0 \le i < n$, $\partial H_n = \{P\}$. If the function A is regular, then $\tilde{W}$ is regular, and furthermore $\partial \tilde{W}/\partial \gamma = \{1\}$; thus, by Clarke (1983, Corollary 1 to Prop. 2.3.3 and Thm. 2.3.9 part (iii)),

$$\partial I(a) = \left\{ P + \sum_{0 \le i < n} \alpha_i M_i : (\alpha_i)_{0 \le i < n} \in \partial A(H_0(a), \ldots, H_{n-1}(a)) \right\},$$

for every $a \in \operatorname{int} B_b(\Sigma, u(X))$, and the result now follows from Theorem 5. If instead $-A$ is regular, then similarly $\partial(-I)(a) = \left\{ P + \sum_{0 \le i < n} \alpha_i M_i + \alpha_n P : (\alpha_i)_{0 \le i < n} \in \partial(-A(H_0(a), \ldots, H_{n-1}(a))) \right\}$; arguing as in the proof of Proposition 9, $\partial(-\tilde{W}(H(a))) = -\partial \tilde{W}(H(a))$ and $\partial(-I)(a) = -\partial I(a)$, so the above equality obtains in this case as well.

C.9.6 Proof of Proposition 11 (mean-dispersion preferences)

Begin with a preliminary result.

Lemma 23 *Let $D \subset ba(\Sigma)$ with $Q \ge 0$ for all $Q \in D$, and $Q(S) > 0$ for at least one $Q \in D$. Let*

$$C_1 \equiv \left\{ \frac{Q}{Q(S)} : Q \in D, Q(S) > 0 \right\}, \quad C_2 \equiv \left\{ \frac{Q}{Q(S)} : Q \in \overline{\text{co}} D, Q(S) > 0 \right\} :$$

then $\overline{\text{co}} C_1 = \overline{\text{co}} C_2$.

Proof: It is obvious that $C_1 \subset C_2$, so it is enough to show that $C_2 \subset \overline{\text{co}} C_1$. Let $P = \sum_{i=1}^n \lambda_i \frac{Q_i}{Q_i(S)}$, with $Q_i \in \overline{\text{co}} D$ and $Q_i(S) > 0$ for each i. It is enough to consider the case $Q_i \in \text{co } D$, because if every Q_i is the limit of a net $(Q_i^{j_i})$ in co D, then eventually $Q_i^{j_i}(S) > 0$, and we can define a net $(P^{j_1,\ldots,j_n})$, with $P^{j_1,\ldots,j_n} = \sum_i \lambda_i [Q_i^{j_i}(S)]^{-1} Q_i^{j_i}$, that converges to P.

Thus, write $Q_i = \sum_{j=1}^{n_i} \rho_j^i Q_j^i$ for each i, where $Q_j^i \in D$. We can also assume that $Q_j^i(S) > 0$ for all i, j, because $Q_j^i(S) = 0$ and $Q_j^i \geq 0$ imply $Q_j^i = 0$ and so we can simply drop it from the summation. We have

$$P = \sum_i \lambda_i \frac{\sum_{j=1}^{n_i} \rho_j^i Q_j^i}{\sum_{j=1}^{n_i} \rho_j^i Q_j^i(S)} = \sum_i \lambda_i \frac{\sum_{j=1}^{n_i} \rho_j^i Q_j^i(S)}{\sum_{k=1}^{n_i} \rho_k^i Q_k^i(S)} \frac{Q_j^i}{Q_j^i(S)} = \sum_{i=1}^n \sum_{j=1}^{n_i} \frac{\lambda_i \rho_j^i Q_j^i(S)}{\sum_{k=1}^{n_i} \rho_k^i Q_k^i(S)} \frac{Q_j^i}{Q_j^i(S)};$$

clearly, for each $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$, $\frac{\lambda_i \rho_j^i Q_j^i(S)}{\sum_{k=1}^{n_i} \rho_k^i Q_k^i(S)} \geq 0$, and these quantities add up to 1. Hence, P is a convex combinations of measures $Q_j^i \in D$ with $Q_j^i(S) > 0$, as required. ■

Proof: Let $\hat{\varphi}(\mu, \rho) = \varphi(\mu, -\rho)$, which by assumption is regular. Note that $\partial \hat{\varphi}(e)$ has non-negative components by Lemma 19 part 4, because it is increasing in both arguments.

Next, the map $a \mapsto a - a \cdot \pi 1_S$ is linear, hence strictly differentiable; the function $\hat{\rho} : \mathbb{R}^n \to \mathbb{R}$ defined by $\hat{\rho}(a) = -\rho(a - a \cdot \pi 1_S)$ is then the composition of a regular function and a scalar-valued, strictly differentiable function; by (iii) in Clarke (1983, Prop. 2.3.9), it is regular, and $\partial \hat{\rho}(e) = \{\sum_s \alpha_s (1_{\{s\}} - \pi) : \alpha \in \partial(-\rho)(e - e \cdot \pi 1_S)\} = \{\sum_s \alpha_s (\pi - 1_{\{s\}}) : \alpha \in \partial \rho(e - e \cdot \pi 1_S)\}$.

Finally, let $\hat{\pi}(e) = \pi \cdot e$, which is linear and hence regular, and observe that $I(e) = \hat{\varphi}(\hat{\pi}(e), \hat{\rho}(e))$. By (i) in the cited Proposition, $\partial I(e) = \overline{\text{co}}\{\beta_1 \pi + \beta_2 \sum_s \alpha_s (\pi - 1_{\{s\}}) : (\beta_1, \beta_2) \in \partial \bar{\varphi}(\hat{\pi}(e), \hat{\rho}(e)), \alpha \in \partial \rho(e - e \cdot \pi 1_S)\}$. Furthermore, $\hat{\varphi}(\mu, \rho) = \varphi(\mu, -\rho)$ implies that $\hat{\varphi}^{CR}(\mu, \rho; \mu', \rho') = \varphi^{CR}(\mu, -\rho; \mu', -\rho')$ and hence that $(\beta_1, \beta_2) \in \partial \hat{\varphi}(\mu, \rho)$ iff $(\beta_1, -\beta_2) \in \partial \varphi(\mu, -\rho)$. Therefore,

$$\partial I(e) = \overline{\text{co}} \left\{ \beta_1 \pi + \beta_2 \sum_s \alpha_s (1_{\{s\}} - \pi) : (\beta_1, \beta_2) \in \partial \varphi(e \cdot \pi, \rho(e - e \cdot \pi 1_S)),\ \alpha \in \partial \rho(e - e \cdot \pi 1_S) \right\}$$

Notice that, for every β_1, β_2, α as above, $\beta_1 \pi \cdot 1_S + \beta_2 \sum_s \alpha_s (1_{\{s\}} - \pi) \cdot 1_S = \beta_1$. The result now follows from Theorem 5 and Lemma 23. ■

References

Charalambos Aliprantis and Kim Border. *Infinite Dimensional Analysis*. Springer Verlag, Berlin, 2007.

Frank J. Anscombe and Robert J. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34:199–205, 1963.

Jean-Pierre Aubin. *Applied Functional Analysis*. Wiley-Interscience, 2000.

Didier Aussel, Jean-Noël Corvellec, and Marc Lassonde. Mean value property and subdifferential criteria for lower semicontinuous functions. *Transactions of the American Mathematical Society*, 347:4147–4161, 1995.

Aurélien Baillon, Olivier L'Haridon, and Laetitia Placido. Ambiguity models and the machina paradoxes. *American Economic Review*, forthcoming.

Antoine Billot, Alain Chateauneuf, Itzak Gilboa, and Jean-Marc Tallon. Sharing beliefs: Between agreeing and disagreeing. *Econometrica*, 68:685–694, 2000.

Nicolas Bourbaki. *General Topology*. Springer Verlag, 1998.

Simone Cerreia-Vioglio, Fabio Maccheroni, Massimo Marinacci, and Luigi Montrucchio. Uncertainty averse preferences. Carlo Alberto Notebook 77, 2008.

Alain Chateauneuf and Josè Heleno Faro. Ambiguity through confidence functions. *Journal of Mathematical Economics*, 45(9-10):535–558, September 2009.

Alain Chateauneuf and Jean-Marc Tallon. Diversification, convex preferences and non-empty core in the Choquet expected utility model. *Economic Theory*, 19:509–523, 2002.

Frank H. Clarke. Multiple integrals of Lipschitz functions in the calculus of variations. *Proceedings of the American Mathematical Society*, 64:260–264, 1977.

Frank H. Clarke. *Optimization and Nonsmooth Analysis*. J. Wiley, New York, 1983.

Arthur P. Dempster. Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

Nelson Dunford and Jacob T. Schwartz. *Linear Operators, Part I*. Wiley Interscience, New York, 1957.

Daniel Ellsberg. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75: 643–669, 1961.

Larry G. Epstein. A definition of uncertainty aversion. *Review of Economic Studies*, 66:579–608, 1999.

Larry G. Epstein and Tan Wang. Intertemporal asset pricing under Knightian uncertainty. *Econometrica*, 62:283–322, 1994.

Paolo Ghirardato and Massimo Marinacci. Ambiguity made precise: A comparative foundation. *Journal of Economic Theory*, 102:251–289, 2002.

Paolo Ghirardato and Marciano Siniscalchi. Monotonic, bernoullian and continuous preferences. Mimeo, June 2010.

Paolo Ghirardato, Fabio Maccheroni, and Massimo Marinacci. Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory*, 118:133–173, 2004.

Paolo Ghirardato, Fabio Maccheroni, and Massimo Marinacci. Revealed ambiguity and its consequences: Unambiguous acts and events. Mimeo, 2005.

Itzhak Gilboa and David Schmeidler. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

Itzhak Gilboa, Fabio Maccheroni, Massimo Marinacci, and David Schmeidler. Objective and subjective rationality in a multiple prior model. *Econometrica*, 78:755–770, 2010.

Simon Grant and Benjamin Polak. Generalized variational preferences. Mimeo, 2007.

Harvey J. Greenberg and William P. Pierskalla. Quasiconjugate function and surrogate duality. *Cahiers du Centre d'Etude de Recherche Operationelle*, 15:437–448, 1973.

Lars P. Hansen and Thomas J. Sargent. Robust control and model uncertainty. *The American Economic Review*, 91:60–66, 2001.

Lars P. Hansen and Thomas J. Sargent. *Robustness*. Princeton University Press Princeton, NJ, 2008.

Lars P. Hansen, Thomas J. Sargent, and Thomas D. Tallarini. Robust permanent income and pricing. *Review of Economic Studies*, 66:873–907, 1999.

Chip Heath and Amos Tversky. Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, 4:5–28, 1991.

Robin M. Hogarth and Hillel J. Einhorn. Venture theory: A model of decision weights. *Management Science*, 36:780–803, 1990.

N. Ju and J. Miao. Ambiguity, learning, and asset returns. *Boston University and Hong Kong University of Science and Technology*, 2009.

Alexander Kechris. *Classical Descriptive Set Theory*. New York, Springer-Verlag, 1995.

Peter Klibanoff, Massimo Marinacci, and Sujoy Mukerji. A smooth model of decision making under ambiguity. *Econometrica*, 73:1849–1892, 2005.

Fabio Maccheroni, Massimo Marinacci, and Aldo Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74:1447–1498, 2006.

Mark J. Machina. Risk, ambiguity, and the rank-dependence axioms. *American Economic Review*, 99:385–392, 2009.

Mark J. Machina and David Schmeidler. A more robust definition of subjective probability. *Econometrica*, 60:745–780, 1992.

Robert E. Megginson. *An Introduction to Banach Space Theory*. Springer-Verlag, New York, 1998.

Sujoy Mukerji. Ambiguity aversion and incompleteness of contractual form. *American Economic Review*, 88:1207–1231, 1998.

Klaus Nehring. Capacities and probabilistic beliefs: A precarious coexistence. *Mathematical Social Sciences*, 38:197–213, 1999.

Klaus Nehring. Imprecise probabilistic beliefs as a context for decision-making under ambiguity. Mimeo, November 2002.

Klaus Nehring. Bernoulli without Bayes: A theory of utility-sophisticated preferences under ambiguity. Mimeo, 2007.

R. Tyrrell Rockafellar. Directionally Lipschitzian functions and subdifferential calculus. *Proceedings of the London Mathematical Society*, 39:331–355, 1979.

R. Tyrrell Rockafellar. Generalized directional derivatives and subgradients of nonconvex functions. *Canadian Journal of Mathematics*, XXXII:257–280, 1980.

David Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57:571–587, 1989.

Marciano Siniscalchi. A behavioral characterization of plausible priors. *Journal of Economic Theory*, 128:91–135, 2006.

Marciano Siniscalchi. Vector expected utility and attitudes toward variation. Mimeo, 2007.

Marciano Siniscalchi. Machina's Reflection example and VEU preferences: a very short note. Mimeo, 2008.

Marciano Siniscalchi. Vector expected utility and attitudes toward variation. *Econometrica*, 77: 801–855, 2009.

Cedric A. B. Smith. Consistency in statistical inference and decision. *Journal of the Royal Statistical Society, Series B*, 23:1–25, 1961.

Jiankang Zhang. Subjective ambiguity, probability and capacity. *Economic Theory*, 20:159–181, 2002.