Department of Economics
and
Institute for Policy Analysis
University of Toronto
150 St. George St.
Toronto, Canada
M5S 1A1

WORKING PAPER
NUMBER UT-ECIPA-EPSTEIN-97-01

# UNCERTAINTY AVERSION

Larry G. Epstein*
University of Toronto and
Hong Kong University of Science and Technology

July 1997


ISSN 0829-4909
Author's email: epstein@chass.utoronto.ca
On-line Version: http://www.epas.utoronto.ca:8080/ecipa/wpa.html

**Abstract**

A definition of uncertainty or ambiguity aversion is proposed. It is argued that the definition is well-suited to modelling within the Savage (as opposed to Anscombe and Aumann) domain of acts. The defined property of uncertainty aversion has intuitive empirical content, behaves well in specific models of preference (multiple-priors and Choquet expected utility) and is tractable. Tractability is established through use of a nonstandard notion of differentiability for utility functions, called eventwise differentiability.



Journal of Economic Literature Classification Numbers: D80,D81


---


[*]The financial support of the Social Sciences and Humanities Research Council of Canada and the hospitality of the Hong Kong University of Science and Technology are gratefully acknowledged. I have also benefitted from discussions with Chew Soo Hong, Massimo Marinacci, Uzi Segal and Jiankang Zhang and from comments by participants at the Chantilly Workshop on Decision Theory, June 1997. The author can be contacted via epstein@usthk.ust.hk.

# 1. INTRODUCTION

## 1.1. Objectives

The concepts of risk and risk aversion are cornerstones of a broad range of models in economics and finance. In contrast, relatively little attention is paid in formal models to the phenomenon of uncertainty that is arguably more prevalent than risk. The distinction between them is roughly that *risk* refers to situations where the perceived likelihoods of events of interest can be represented by probabilities, whereas *uncertainty* refers to situations where the information available to the decision-maker is too imprecise to be summarized by a probability measure. Thus the terms 'vagueness' or 'ambiguity' can serve as close substitutes. Ellsberg, in his famous experiment, has demonstrated that such a distinction is meaningful empirically, but it cannot be accommodated within the subjective expected utility (SEU) model.

Perhaps because this latter model has been so dominant, our formal understanding of uncertainty and *uncertainty aversion* is poor. There exists a definition of uncertainty aversion, due to Schmeidler [21], for the special setting of Anscombe-Aumann (AA) horse-race/roulette wheel acts. Its intuitive appeal within the AA framework is (arguably) open to question. More importantly, though it has been transported and widely adopted in models employing the Savage domain of acts, I feel that it is both less appealing and less useful in such contexts. Because the Savage domain is typically more appropriate and also more widely used in descriptive modelling, this suggests the need for an alternative definition of uncertainty aversion that is more suited to applications in a Savage domain. Providing such a definition is the objective of this paper.

Uncertainty aversion is defined for a large class of preferences. This is done for the obvious reason that a satisfactory understanding of uncertainty aversion can be achieved only if its meaning does not rely on preference axioms that are auxiliary rather than germane to the issue. On the other hand, Choquet expected utility (CEU) theory [21] and its close relative, the multiple-priors model [7], provide important examples for understanding the nature of our definition, as they are the most widely used and studied theories of preference that can accommodate Ellsberg-type behaviour. Recall that risk aversion has been defined and characterized for general preferences, including those that lie outside the expected utility class (see [27] and [3], for example).

There is a separate technical or methodological contribution of the paper. After the formulation and initial examination of the definition of uncertainty aversion, subsequent analysis is facilitated by assuming *eventwise differentiability* of utility. The role of eventwise differentiability may be described roughly as follows: The notion of uncertainty aversion leads to concern with the 'local probabilistic beliefs' implicit in an arbitrary preference order or utility function. These beliefs represent the decision-maker's underlying 'mean' or 'ambiguity-free' likelihood assessments for events. In general, they need not be unique. But they are unique if utility is eventwise differentiable (given suitable additional conditions). Further perspective is provided by recalling the role of differentiability in decision theory under risk, where utility functions are defined on cumulative distribution functions. Much as calculus is a powerful tool, Machina [12] has shown that differential methods are useful in decision theory under risk. He employs Frechet differentiability; others have shown that Gateaux differentiability suffices for many purposes [2]. In the present context of decision making under uncertainty, where utility functions are defined over acts, the preceding two notions of differentiability are not useful for the task of uncovering implicit local beliefs. On the other hand, eventwise differentiability 'works'. Because local probabilistic beliefs are likely to be useful more broadly, so it seems will the notion of eventwise differentiability. It must be acknowledged, however, that eventwise differentiability has close relatives in the literature, namely in [16] and [13].[1] The differences from this paper and the value-added here are clarified later (Section 4.1 and Appendix C). It seems accurate to say that this paper *adds to* the demonstration in [13] that differential techniques are useful also for analysis of decision-making under uncertainty.

The paper proceeds as follows: The current definition of uncertainty aversion is examined first and the choice between a Savage domain and an Anscombe-Aumann domain of acts is discussed. Then, because the parallel with the well understood theory of risk aversion is bound to be helpful, relevant aspects of that theory are reviewed. A new definition of uncertainty aversion is formulated in the remainder of Section 2 and some attractive properties are described in Section 3. In particular, uncertainty aversion is shown to have intuitive empirical content and to admit simple characterizations within the CEU and multiple-priors models. The notion of 'eventwise derivative' and the analysis of uncertainty aversion given eventwise differentiability follow in Section 4. It is shown that eventwise differentiability of utility simplifies the task of checking whether the corresponding preference order is uncertainty averse and thus enhances the tractability of the proposed definition.

[1] After a version of this paper was completed, I learned of a revision of [13], dated 1997, that is even more closely related.

Two important limitations of the analysis should be acknowledged at the start. First, uncertainty aversion is defined relative to an *exogenously* specified collection of events $\mathcal{A}$. Events in $\mathcal{A}$ are thought of as unambiguous or uncertainty-free. They play a role here parallel to that played by constant (or risk-free) acts in the standard analysis of risk aversion. However, whether or not an event is ambiguous is naturally viewed as subjective or derived from preference. Accordingly, it seems desirable to define uncertainty aversion relative to the collection of subjectively unambiguous events. Unfortunately, such a formulation is beyond the scope of this paper. In particular, there exists as yet no satisfactory definition of 'subjectively unambiguous events'.[2] In defense of the exogenous specification of the collection $\mathcal{A}$, observe that Schmeidler [21] relies on a comparable specification through the presence of objective lotteries in the Anscombe-Aumann domain. In addition, it seems likely that in any future success in endogenizing ambiguity, the present analysis of uncertainty aversion relative to a given collection $\mathcal{A}$ will be useful.

The other limitation concerns the limited success in this paper in achieving the ultimate objective of deriving the *behavioural consequences* of uncertainty aversion. The focus here is on the definition of uncertainty aversion. Some behavioural implications are derived but much is left for future work. In particular, applications to standard economic contexts, such as asset pricing or games, are beyond the scope of the paper. However, the importance of the groundwork laid here for future applications merits emphasis - an essential precondition for understanding the behavioural consequences of uncertainty aversion is that the latter term have a precise and intuitively satisfactory meaning. Admittedly, there have been several papers in the literature claiming to have derived consequences of uncertainty aversion for strategic behaviour and also for asset pricing. To varying degrees these studies either adopt the Schmeidler definition of uncertainty aversion or they do not rely on a precise definition. In the latter case, they adopt a model of preference that has been developed in order to accommodate an intuitive notion of uncertainty aversion and interpret the implications of this preference specification as due to uncertainty aversion. (This author is partly responsible for such an exercise [5]; there are other examples in the literature.) There is an obvious logical flaw in such a procedure and the claims made (or the interpretations proposed) are unsupportable without a satisfactory definition of uncertainty aversion.

[2]Zhang [28] is the first paper to propose a definition of ambiguity that is derived from preference, but his definition is problematic. An improved definition is the subject of current research by this author and Zhang.

### 1.2. The Current Definition of Uncertainty Aversion

In order to further motivate the paper, turn to a discussion of Schmeidler's definition of uncertainty aversion.

Fix a state space $(S, \Sigma)$ and outcome set $\mathcal{X}$. Denote by $\mathcal{F}$ and $\mathcal{H}$ the sets of all finite-ranged (simple) and measurable acts $e$ from $(S, \Sigma)$ into $\mathcal{X}$ and $\Delta(\mathcal{X})$, respectively.[3] The domain $\mathcal{F}$ was used by Savage, while $\mathcal{H}$ is the domain of horse-race/roulette-wheel acts used by Anscombe and Aumann. Each such act $h$ involves two stages - in the first, uncertainty is resolved through realization of the horse-race winner $s \in S$ and in the second stage the risk associated with the objective lottery $h(s)$ is resolved. Schmeidler assumes that preference $\succeq$ and the representing utility function $U$ are defined on the larger domain $\mathcal{H}$. He calls $U$ uncertainty averse if it is quasiconcave, that is, if

$$U(e) \geq U(f) \implies U(\alpha e + (1-\alpha) f) \geq U(f), \tag{1.1}$$

for all $\alpha \in [0,1]$, where the mixture $\alpha e + (1-\alpha) f$ is defined in the obvious way. The suggested interpretation (p. 582) is that "substituting objective mixing for subjective mixing makes the decision-maker better off."

This definition is then examined further within Choquet expected utility theory, according to which uncertain prospects are evaluated by a utility function having the following form:

$$U^{ceu}(e) = \int_S u(e)\, d\nu. \tag{1.2}$$

Here, $u$ is a vNM utility index, assumed to be mixture linear on the set of lotteries $\Delta(\mathcal{X})$, $\nu$ is a capacity (or non-additive probability) on $\Sigma$, integration is in the sense of Choquet and other details will be provided later.[4] Schmeidler shows that $U^{ceu}$ is uncertainty averse if and only if the corresponding capacity $\nu$ is *convex*, that is,

$$\nu(A \cup B) + \nu(A \cap B) \geq \nu(A) + \nu(B), \tag{1.3}$$

for all measurable events $A$ and $B$. Additivity is a special case that characterizes uncertainty neutrality (suitably defined).

Subsequently, the identification of convexity of $\nu$ with uncertainty aversion has been widely adopted in the literature developing or applying the CEU model,

[3] $\Delta(\cdot)$ denotes the set of finitely additive probability measures. Unless otherwise specified, $\Sigma$ is assumed to be an algebra.

[4] See Section 3.2 for the definition of Choquet integration.

including many instances where the Savage domain $\mathcal{F}$ is adopted rather than the AA domain.[5] This paper is concerned with situations in which $\mathcal{F}$ is the appropriate domain, that is, in which choice behaviour over $\mathcal{F}$ is the object of study and in which only such behaviour is observable to the analyst. Accordingly, it is assumed that preference $\succeq$ and the representing utility function $U$ are defined on $\mathcal{F}$. Turn to the relevance for such modelling situations of the preceding definition of uncertainty aversion.

Ellsberg's single-urn experiment illustrates the weak connection between convexity of the capacity and behaviour that is intuitively uncertainty averse. The urn is represented by the state space $S = \{R, B, G\}$, where the symbols represent the possible colors, red, blue and green of a ball drawn at random from an urn. The information provided the decision-maker is that the urn contains 30 red balls and 90 balls in total. Thus, while he knows that there are 60 balls that are either blue or green, the relative proportions of each are not given. Let $\succeq$ be the decision-maker's preference over bets on events $E \subset S$. Typical choices in such a situation correspond to the following rankings of events:[6]

$$\{R\} \succ \{B\} \sim \{G\},\ \{B, G\} \succ \{R, B\} \sim \{R, G\}. \tag{1.4}$$

The intuition for these rankings is well known and is based on the fact that $\{R\}$ and $\{B, G\}$ have objective probabilities, while the other events are 'ambiguous', or have 'ambiguous probabilities'. Thus these rankings correspond to an intuitive notion of uncertainty or ambiguity aversion.

Next suppose that the decision-maker has CEU preferences with capacity $\nu$. Then convexity is *neither necessary nor sufficient* for the above rankings. For example, if $\nu(R) = 8/24$, $\nu(B) = \nu(G) = 7/24$ and $\nu(\{B, G\}) = 13/24$, $\nu(\{R, G\}) = \nu(\{R, B\}) = 1/2$, then (1.4) is implied but $\nu$ is not convex. For the fact that convexity is not sufficent, observe that convexity does not even exclude the 'opposite' rankings that intuitively reflect an affinity for ambiguity. (Let $\nu(R) = 1/12$, $\nu(B) = \nu(G) = 1/6$, $\nu(\{B, G\}) = 1/3$, $\nu(\{R, G\}) = \nu(\{R, B\}) = 1/2$.)

This failure of the convexity definition of uncertainty aversion to conform with typical behaviour in the Ellsberg example is troubling because it is precisely such

[5] While the relaxation from additivity to convexity is intended to accommodate behaviour such as that exhibited in the Ellsberg Paradox, convexity is neither necessary nor sufficient for such behaviour (see section 3.4).

[6] In terms of acts, $\{R\} \succ \{B\}$ means $1_R \succ 1_B$ and so on. For CEU, a decision-maker always prefers to bet on the event having the larger capacity.

behaviour that motivates recent generalizations of subjective expected utility. One possible reaction is to suggest that the single-urn experiment is special and that convexity is better suited to Ellsberg's other principal experiment involving two urns, one ambiguous and the other unambiguous.[7] Because behaviour in this experiment is also prototypical of the behaviour that is to be modelled and because it might be unrealistic to expect a single definition of uncertainty aversion to perform well in all settings, good performance of the convexity definition in this setting might restore its appeal. Moreover, such good performance might be expected because the Cartesian product state space that is natural for modelling the two-urn experiment suggests a connection with the horse-race/roulette-wheel acts in the AA domain, where convexity of the capacity has been motivated behaviourally by Schmeidler. According to this view, the state space for the ambiguous urn 'corresponds' to the horse-race stage of the AA acts and the state space for the unambiguous urn 'corresponds' to the roulette-wheel component.

In fact, the performance of the convexity definition is no better in the two-urn experiment than in the single-urn case. Rather than providing specific examples of capacities supporting this assertion, it may be more useful to point out why the grounds for optimism described above are unsound. In spite of the apparent correspondence between the AA setup and the Savage domain with a Cartesian product state space, these are substantially different specifications because, as pointed out by Sarin and Wakker [17], only the AA domain involves *two-stage acts* (the horse-race first and then the roulette-wheel) and in Schmeidler's formulation of CEU, these are evaluated in an iterative fashion. Eichberger and Kelsey [4] show that this difference leads to different conclusions about the connection between convexity of the capacity and attitudes towards randomization. For the same reason the difference in domains leads to different conclusions about the connection between convexity of the capacity and attitudes towards uncertainty. In particular, convexity is not closely connected to typical behaviour in the two-urn experiment.

While the preceding discussion has centered on examples, albeit telling examples, there is a general point that may be worth making explicit in light of the widespread acceptance of convexity as modelling uncertainty aversion within CEU. The general point concerns the practice of transferring to the Savage domain

[7] Each urn contains 100 balls that are either red or blue. For the ambiguous urn this is all the information provided. For the unambiguous urn, the decision-maker is told that there are 50 balls of each colour. The choice problem is whether to bet on drawing a red (or blue) ball from the ambiguous urn versus the unambiguous one.

notions, such as uncertainty aversion, that have been formulated and motivated in the AA framework. The difference between the decision-maker's attitude towards the second-stage roulette-wheel risk as opposed to the uncertainty inherent in the first-stage horse-race is the basis for Schmeidler's definition of uncertainty aversion.. The upshot is that uncertainty aversion is *not* manifested exclusively or primarily through the choice of pure horse-races or acts over $S$. Frequently, however, it is the latter choice behaviour that is of primary interest to the modeller. This is the case, for example, in the Ellsberg experiments discussed above and is the reason for the weak (or non-existent) connection between convexity and intuitive behaviour in those experiments. This is not to deny that convexity may be a useful hypothesis even in a Savage framework nor that its interpretation as uncertainty aversion may be warranted where preferences over AA acts are observable, say in laboratory experiments. Accordingly, this is not a criticism of Schmeidler's definition within his chosen framework. It argues only against the common practice of interpreting convexity as uncertainty aversion outside that framework. (An alternative behavioural interpetation for convexity is provided in [25].)

I conclude with one last remark on the AA domain. The extension of the Savage domain of acts to the AA domain is useful because the inclusion of second-stage lotteries delivers greater analytical power or simplicity. This is the reason for their inclusion by Anscombe and Aumann - to simplify the derivation of subjective probabilities - as well as in the axiomatizations of the CEU and multiple-priors utility functions in [21] and [7] respectively. In all these cases, roulette-wheels are a tool whose purpose is to help in delivering the representation of utility for acts over $S$. Kreps [10, p. 101] writes that this is sensible in a normative application but "is a very dicey and perhaps completely useless procedure in descriptive applications" *if only choices between acts over $S$ are observable.* Emphasizing and elaborating this point has been the objective of this subsection.

# 2. AVERSION TO RISK AND UNCERTAINTY

## 2.1. Risk Aversion

Recall first some aspects of the received theory of risk aversion. This will provide some perspective for the analysis of uncertainty aversion. In addition, it will become apparent that if a distinction between risk and uncertainty is desired, then the theory of risk aversion must be modified.

Because a subjective approach to risk aversion is the relevant one, adapt Yaari's analysis [27], which applies to the primitives $(S, \Sigma)$, $\mathcal{X} \subset \mathcal{R}^N$ and $\succeq$, a preference over the set of acts $\mathcal{F}$.

Turn first to 'comparative risk aversion'. Say that $\succeq^2$ is *more risk averse than* $\succeq^1$ if for every act $e$ and outcome $x$,

$$x \succeq^1 (\succ^1) \, e \implies x \succeq^2 (\succ^2) \, e. \tag{2.1}$$

The two acts that are being compared here differ in that the variable outcomes prescribed by $e$ are replaced by the single outcome $x$. The intuition for this definition is clear given the identification of constant acts with the absence of risk or perfect certainty.

To define absolute (rather than comparative) risk aversion, it is necessary to adopt a 'normalization' for risk neutrality. Note that this normalization is exogenous to the model. The standard normalization is the 'expected value function', that is, risk neutral orders $\succeq^{rn}$ are those satisfying:

$$e \succeq^{rn} e' \iff \int_S e(s) \, dm(s) \succeq^{rn} \int_S e'(s) \, dm(s), \tag{2.2}$$

for some probability measure $m$ on $(S, \Sigma)$, where the $R^N$-valued integrals are interpreted as constant acts and accordingly are ranked by $\succeq^{rn}$. This leads to the following definition of risk aversion: Say that $\succeq$ is *risk averse* if there exists a risk neutral order $\succeq^{rn}$ such that $\succeq$ is more risk averse than $\succeq^{rn}$ . Risk loving and risk neutrality can be defined in the obvious ways.

In the subjective expected utility framework, this notion of risk aversion is the familiar one characterized by concavity of the vNM index, with the required $m$ being the subjective beliefs or prior. By examining the implications of uncertainty aversion for choice between binary acts, Yaari [27] argues that this interpretation for $m$ extends to more general preferences.

Three points from this review merit emphasis. First, the definition of comparative risk aversion requires an *a priori* definition for the absence of risk. Observe that the identification of risklessness with constant acts is *not* tautological. For example, Karni [9] argues that in a state-dependent expected utility model 'risklessness' may very well correspond to acts that are not constant. Thus the choice of how to model risklessness is a substantive normalization that precedes the definition of 'more risk averse'.

Second, the definition of risk aversion requires further an *a priori* definition of risk neutrality.

The final point is perhaps less evident or familiar. Consider rankings of the sort used in (2.1) to define 'more risk averse'. A decision-maker may prefer the constant act because she dislikes variable outcomes even when they are realized on events that are understood well enough to be assigned probabilities (risk aversion). Alternatively, the reason for the indicated preference may be that the variable outcomes occur on events that are ambiguous and because she dislikes ambiguity or uncertainty. Thus it seems more appropriate to describe (2.1) as revealing that $\succeq^2$ is 'more risk *and* uncertainty averse than $\succeq^1$,' with no attempt being made at a distinction. However, the importance of the distinction between these two underlying reasons seems self-evident; it is reflected also in recent concern with formal models of 'Knightian uncertainty' and decision theories that accommodate the Ellsberg (as opposed to Allais) Paradox. The second possibility above can be excluded, and thus a distinction made, by assuming that the decision-maker is indifferent to uncertainty or, put another way, by *assuming that there is no uncertainty* (all events are assigned probabilities). But these are extreme assumptions that are contradicted in Ellsberg-type situations. This paper identifies and focuses upon the uncertainty aversion component implicit in the comparisons (2.1) and, to a limited extent, achieves a separation between risk aversion and uncertainty aversion.

### 2.2. Uncertainty Aversion

Once again, consider orders $\succeq$ on $\mathcal{F}$, where for the rest of the paper the outcome set $\mathcal{X}$ is arbitrary rather than Euclidean. The objective now is to formulate intuitive notions of comparative and absolute uncertainty aversion.

Turn first to comparative uncertainty aversion. It is clear intuitively and also from the discussion of risk aversion that one can proceed only given a prior specification of the 'absence of uncertainty.' This specification takes the form of an exogenous family $\mathcal{A} \subset \Sigma$ of 'unambiguous' events.

Assume throughout the following intuitive requirements for $\mathcal{A}$: It contains $S$ and

$$\begin{aligned} A &\in \mathcal{A} \text{ implies that } A^c \in \mathcal{A}; \\ A_1, A_2 &\in \mathcal{A} \text{ and } A_1 \cap A_2 = \emptyset \text{ imply that } A_1 \cup A_2 \in \mathcal{A}. \end{aligned}$$

Zhang [28] argues that these properties are natural for a collection of unambiguous events and calls such collections $\lambda$-systems. Intuitively, if an event being unambiguous means that it can be assigned a probability by the decision-maker, then the sum of the individual probabilities is naturally assigned to a disjoint union, while the complementary probability is naturally assigned to the complementary event. Note that $\mathcal{A}$ need not be an algebra, because it need not be closed with respect to nondisjoint unions or intersections.[8] Denote by $\mathcal{F}^{ua}$ the set of $\mathcal{A}$-measurable acts, also called *unambiguous acts.*

The following definition parallels the earlier one for comparative risk aversion. Given two orderings, say that $\succeq^2$ is *more uncertainty averse than* $\succeq^1$ if for every unambiguous act $h$ and every act $e$ in $\mathcal{F}$,

$$h \succeq^1 (\succ^1)\, e \quad \Longrightarrow \quad h \succeq^2 (\succ^2)\, e\,. \tag{2.3}$$

There is no loss of generality in supposing that the acts $h$ and $e$ deliver the identical outcomes. The difference between the acts lies in the nature of the events where these outcomes are delivered (some of these events may be empty). For $h$, the typical outcome $x$ is delivered on the unambiguous event $h^{-1}(x)$, while it occurs on an ambiguous event given $e$. Then whenever the greater ambiguity inherent in $e$ leads $\succeq^1$ to prefer $h$, the more ambiguity averse $\succeq^2$ will also prefer $h$. This interpretation relies on the assumption that each event in $\mathcal{A}$ is unambiguous and thus is (weakly) less ambiguous than any $E \in \Sigma$.

Fix an order $\succeq$. To define absolute (rather than comparative) uncertainty aversion for $\succeq$, it is necessary to adopt a 'normalization' for uncertainty neutrality. As in the case of risk, a natural though exogenous normalization exists, namely that preference is based on probabilities in the sense of being *probabilistically sophisticated* as defined in [14]. The functional form of representing utility functions reveals clearly the sense in which preference is based on probabilities. The components of that functional form are a probability measure $m$ on the state space $(S, \Sigma)$ and a functional $W : \Delta(\mathcal{X}) \longrightarrow \mathcal{R}^1$, where $\Delta(\mathcal{X})$ denotes the set of all simple (finite support) probability measures on the outcome set $\mathcal{X}$. Using $m$, any act $e$ induces such a probability distribution $\Psi_{m,e}$. Probabilistic sophistication requires that $e$ be evaluated only through the distribution over outcomes $\Psi_{m,e}$ that it induces. More precisely, utility has the form

$$U^{ps}(e) \;=\; W\left(\Psi_{m,e}\right), \;\; e \in \mathcal{F}. \tag{2.4}$$

[8] To see that closure with respect to intersections is not a natural property for a family of unambiguous events, consider the following example taken from [28]: An urn contains 100 balls in total, with colour composition $R$, $B$, $W$, and $G$, such that $R + B\ =\ 50\ =\ G + B$. Then it is natural to think that $\{R, B\}$ and $\{G, B\}$ are unambiguous, but that $\{B\}$ is ambiguous.

Following Machina and Schmeidler (p. 754), assume also that $W$ is strictly increasing in the sense of first-order stochastic dominance, suitably defined.[9] Denote any such order by $\succeq^{ps}$. A decision-maker with $\succeq^{ps}$ assigns probabilities to all events and in this way transforms any act into a lottery, or pure risk. Such exclusive reliance on probabilities is, in particular, inconsistent with the typical 'uncertainty averse' behaviour exhibited in Ellsberg-type experiments. Thus it is both intuitive and consistent with common practice to identify probabilistic sophistication with uncertainty neutrality. Think of $m$ and $W$ as the 'beliefs' (or probability measure) and 'risk preferences' underlying $\succeq^{ps}$.[10]

This normalization leads to the following definition: Say that $\succeq$ is *uncertainty averse* if there exists a probabilistically sophisticated order $\succeq^{ps}$ such that $\succeq$ is more uncertainty averse than $\succeq^{ps}$. In other words, under the conditions stated in (2.3),

$$h \succeq^{ps} (\succ^{ps})\, e \quad \Longrightarrow \quad h \succeq (\succ)\, e\,. \tag{2.5}$$

The intuition is similar to that for (2.3).

It is immediate that $\succeq$ and $\succeq^{ps}$ agree on unambiguous acts. Further, $\succeq^{ps}$ is indifferent to uncertainty and thus views all acts as being risky only. Therefore, interpret (2.5) as stating that $\succeq^{ps}$ is a 'risk preference component' of $\succeq$. The indefinite article is needed for two reasons - first because all definitions depend on the exogenously specified collection $\mathcal{A}$ and second, because $\succeq^{ps}$ need not be unique even given $\mathcal{A}$. Subject to these same qualifications, the probability measure underlying $\succeq^{ps}$ is naturally interpreted as 'mean' or 'uncertainty-free' beliefs underlying $\succeq$. The formal analysis below does not depend on these interpretations.

It might be useful to adapt familiar terminology and refer to $\succeq^{ps}$ satisfying (2.5) as constituting a *support* for $\succeq$ at $h$. Then uncertainty aversion for $\succeq$ means that there exists a single order $\succeq^{ps}$ supporting $\succeq$ at every unambiguous act. A parallel requirement in consumer theory is that there exist a single price vector supporting the indifference curve at each consumption bundle on the 45° line.

[9] Write $y \geq x$ if receiving outcome $y$ with probability 1 is weakly preferable, according to $U^{ps}$, to receiving $x$ for sure. $\Psi'$ first-order stochastically dominates $\Psi$ if for all outcomes $y$, $\Psi'(\{x \in \mathcal{X} : y \geq x\}) \leq \Psi(\{x \in \mathcal{X} : y \geq x\})$. Thus the partial order depends on the utility function $U^{ps}$, but that causes no difficulties. See [14] for further details.

[10] Subjective expected utility is the special case of (2.4) with $W(\Psi) = \int_{\mathcal{X}} u(x)\, d\Psi(x)$. But more general risk preferences $W$ are admitted, subject only to the noted monotonicity restriction. In particular, probabilistically sophisticated preference can rationalize behaviour such as that exhibited in the Allais Paradox. It follows that uncertainty aversion, as defined shortly, is concerned with Ellsberg-type, and not, Allais-type, behaviour.

Two possible variations of the preceding definition merit mention. First, refer to $\succeq$ as 'locally uncertainty averse' if for every unambiguous $h$ there exists $\succeq^{ps}$, *depending on* $h$, that satisfies (2.5). The innovation here is that the supporting $\succeq^{ps}$ is not required to be the same for all unambiguous acts $h$. Adoption of such a local notion would invalidate most of our results, including the behavioural implications of uncertainty aversion described in Section 3.4, the uniqueness of the support $\succeq^{ps}$ delivered by eventwise differentiability (Section 4.3) and possibly the negative conclusion in the next section regarding the connection between uncertainty aversion of a CEU utility function and convexity of its capacity.[11] However, the weaker local notion seems to have too little empirical content to be of much use or interest.

A second variation would involve comparisons of conditional acts. It is well known that in the context of the theory of risk aversion, comparison with conditional perfect certainty delivers a stronger definition of risk aversion, though the conditional and unconditional definition are equivalent given expected utility theory.[12] A conditional version of (2.5) might take the form: For all $T \in \Sigma$ and acts $g$,

$$(h,T;\, g,T^c) \succeq^{ps} (e,T;\, g,T^c) \quad \Longrightarrow \quad (h,T;\, g,T^c) \succeq (e,T;\, g,T^c)\,,$$

and similarly for strict preference. But this is an intuitively problematic condition. The intuition surrounding (2.5) is that the act preferred by the probabilistically sophisticated order is unambiguous. But here $(h,T;\, g,T^c)$ is ambiguous without further restrictions on $T$ and $g$. Furthermore, there is *no sensible notion of 'conditionally unambiguous'*.[13] The nature of $T^c$ and the outcomes delivered on $T^c$ by $g$ affect the nature of the subacts ($h$ on $T$) and ($e$ on $T$); for example, for some $T$ and $g$, it could be the latter subact that is less ambiguous. This is precisely the message of the Ellsberg Paradox and the violation of the Sure-Thing Principle that it reveals. On other hand, if $T$ and $g$ are restricted so that $(h,T;\, g,T^c)$ is unambiguous, then one obtains a definition that is equivalent to the uncondi-

[11] An intermediate definition that *would* preserve most results would require that for each unambiguous $A$ there exists $\succeq^{ps}$ that serves as a support at both $A$ and $A^c$; otherwise, $\succeq^{ps}$ could vary with $A$.

[12] See [12] and [3] for the relation between these two definitions of 'more risk averse than'. Yaari deals with the weaker unconditional form of risk aversion.

[13] This is in contrast to the case of conditional certainty. The property
$(x,T;\, g,T^c) \succeq^1 (e,T;\, g,T^c) \Longrightarrow (x,T;\, g,T^c) \succeq^2 (e,T;\, g,T^c)$
seems appropriate as a (strong) definition of '$\succeq^2$ is more risk averse than $\succeq^1$'.

tional one. Thus, there does not seem to be an acceptable and distinct conditional formulation of uncertainty aversion.

Turn next to uncertainty loving and uncertainty neutrality. For the definition of the former, reverse the inequalities in (2.5). That is, say that $\succeq$ is *uncertainty loving* if there exists a probabilistically sophisticated order $\succeq^{ps}$ such that, under the conditions stated in (2.3),

$$h \preceq^{ps} (\prec^{ps})\, e \quad \Longrightarrow \quad h \preceq (\prec)\, e \,. \tag{2.6}$$

The conjunction of uncertainty aversion and uncertainty loving is called *uncertainty neutrality*.

As emphasized earlier, the meaning of uncertainty aversion depends on the exogenously specified $\mathcal{A}$. Two extreme specifications for $\mathcal{A}$ illustrate this dependence. If $\mathcal{A} = \{\emptyset, S\}$, then only constant acts are unambiguous. Thus uncertainty aversion is defined by the appropriate form of (2.1). In terms of the discussion at the end of Section 2.1, the comparison (2.1) reflects uncertainty aversion exclusively. At the other extreme where $\mathcal{A} = \Sigma$, all acts in $\mathcal{F}$ are unambiguous. Thus every preference order is indifferent to uncertainty and (2.1) reflects risk aversion exclusively.

Consider further the question of a separation between attitudes towards uncertainty and attitudes towards risk. Suppose that $\succeq$ is uncertainty averse with support $\succeq^{ps}$. Because $\succeq$ and $\succeq^{ps}$ agree on the set $\mathcal{F}^{ua}$ of unambiguous acts, $\succeq$ is probabilistically sophisticated there. Thus, treating the probability measure underlying $\succeq^{ps}$ as objective, one may adopt the standard notion of risk aversion (or loving) for objective lotteries (see [12], for example) in order to give precise meaning to the statement that $\succeq$ is *risk averse* (or *loving*). In the same way, such risk attitudes are well defined if $\succeq$ is uncertainty loving. That a degree of separation between risk and uncertainty attitudes has been achieved is reflected in the fact that all four logically possible combinations of risk and uncertainty attitudes are admissible. On the other hand, the separation is partial: Given two preference orders $\succeq^1$ and $\succeq^2$ that are comparable in the sense of uncertainty aversion, they must agree on $\mathcal{F}^{ua}$ and thus embody the same risk aversion.

# 3. IS THE DEFINITION ATTRACTIVE?

## 3.1. Some Attractive Properties

The definition of uncertainty aversion has been based on the *a priori* identification of uncertainty neutrality (defined informally) with probabilistic sophistication.

Therefore, internal consistency of the approach should deliver this identification as a formal result. On the other hand, because attitudes towards uncertainty have been defined relative to a given $\mathcal{A}$, such a result cannot be expected unless it is assumed that $\mathcal{A}$ is 'large'. Suppose, therefore, that $\mathcal{A}$ is *rich*: There exist $x^* \succ x_*$, such that for every $E \in \Sigma$, there exists $A \in \mathcal{A}$, satisfying

$$(x^*,\, A;\; x_*,\, A^c) \sim (x^*,\, E;\; x_*,\, E^c).$$

A corresponding notion of richness is valid for the roulette-wheel lotteries in the Anscombe-Aumann framework adopted by Schmeidler [21]. The assumption that $\mathcal{A}$ is rich corresponds to the common assumption in risk theory that every act has a certainty equivalent.

The next theorem (proved in Appendix A) establishes the internal consistency of our approach.

**Theorem 3.1.** *If $\succeq$ is probabilistically sophisticated, then it is uncertainty neutral. The converse is true if $\mathcal{A}$ is rich.*

The potential usefulness of the notion of uncertainty aversion depends on being able to check for the existence of a probabilistically sophisticated order supporting a given $\succeq$. This concern with tractability motivates the later analysis of eventwise differentiability. Anticipating that analysis, consider here the narrower question "does there exist $\succeq^{ps}$ that *both* supports $\succeq$ *and* has underlying beliefs represented by the *given* probability measure $m$ on $\Sigma$?" On its own, the question may seem to be of limited interest. But once eventwise differentiability delivers $m$, its answer completes a procedure for checking for uncertainty aversion.

**Lemma 3.2.** *Let $\succeq^{ps}$ support $\succeq$ in the sense of (2.5) and have underlying probability measure $m$ on $\Sigma$. Then:*

*(i) For any two unambiguous acts $h$ and $h'$, if $\Psi_{m,h}$ first-order stochastically dominates $\Psi_{m,h'}$, then $U(h) \geq U(h')$.*

*(ii) For all acts $e$ and unambiguous acts $h$,*

$$\Psi_{m,e} = \Psi_{m,h} \implies U(e) \leq U(h).$$

*The converse is true if $m$ satisfies: For each unambiguous $A$ and $0 < r < mA$, there exists unambiguous $B \subset A$ with $mB = r$.*

The added assumption for $m$ is satisfied if $S = S_1 \times S_2$, unambiguous events are measurable subsets of $S_1$ and the marginal of $m$ on $S_1$ is convex-ranged in the usual sense. The role of the assumption is to ensure that, using the notation surrounding (2.4),

$$\{\Psi_{m,h} : h \in \mathcal{F}^{ua}\} = \Delta(\mathcal{X}).$$

### 3.2. Multiple-Priors and CEU Utilities

The two most widely used generalizations of subjective expected utility theory are CEU and the multiple-priors model. In this subsection, uncertainty aversion is examined in the context of these models.

Say that $\succeq$ is a *multiple-priors preference order* if it is represented by a utility function $U^{mp}$ of the form

$$U^{mp}(e) = \min_{m \in P} \int_S u(e)\, dm\,, \tag{3.1}$$

for some set $P$ of probability measures on $(S, \Sigma)$ and some vNM index $u : \mathcal{X} \longrightarrow \mathcal{R}^1$. Given a class $\mathcal{A}$, it is natural to model the unambiguous nature of events in $\mathcal{A}$ by supposing that all measures in $P$ are identical when restricted to $\mathcal{A}$; that is,

$$mA = m'A \quad \text{for all } m \text{ and } m' \text{ in } P \text{ and } A \text{ in } \mathcal{A}. \tag{3.2}$$

These two restrictions on $\succeq$ imply uncertainty aversion, because $\succeq$ is more uncertainty averse than the expected utility order $\succeq^{ps}$ with vNM index $u$ and any probability measure $m$ in $P$. More precisely, the following result obtains:

**Theorem 3.3.** *Any multiple-priors order satisfying (3.2) is uncertainty averse.*

**Proof.** Let $\succeq^{ps}$ denote an expected utility order with vNM index $u$ and any probability measure $m$ in $P$. Then $h \succeq^{ps} e \iff \int u(h)\, dm \geq \int u(e)\, dm \Longrightarrow U^{mp}(h) = \int u(h)\, dm \geq \int u(e)\, dm \geq U^{mp}(e)$. ■

A commonly studied special case of the multiple-priors model is a Choquet expected utility order with convex capacity $\nu$. Then (3.1) applies with

$$P = core(\nu) = \{m : m(\cdot) \geq \nu(\cdot) \text{ on } \Sigma\}.$$

Thus convexity of the capacity implies uncertainty aversion *given* (3.2).

Focus more closely on the CEU model, with particular emphasis on the connection between uncertainty aversion and convexity of the capacity. The next result translates Lemma 3.2 into the present setting, thus providing necessary and sufficient conditions for uncertainty aversion combined with a prespecified supporting probability measure $m$. For necessity, an added assumption is adopted. Say that a capacity $\nu$ is *convex-ranged* if for all events $E_1 \subset E_2$ and $\nu(E_1) < r < \nu(E_2)$, there exists $E$, $E_1 \subset E \subset E_2$, such that $\nu(E) = r$. This terminology applies in particular if $\nu$ is additive, where it is standard.[14] For axiomatizations of CEU that deliver a convex-ranged capacity, see [6, p. 73] and [17, Proposition A.3]. Savage's axiomatization of expected utility delivers a convex-ranged probability measure.

Use the notation $U^{ceu}$ to refer to utility functions defined by (1.2), restricted to Savage acts, where the vNM index $u : \mathcal{X} \longrightarrow \mathcal{R}^1$ satisfies

$$u(\mathcal{X}) \text{ has nonempty interior in } \mathcal{R}^1.$$

For those unfamiliar with Choquet integration, observe that for simple acts it yields

$$U^{ceu}(e) = \Sigma_{i=1}^{n-1}[\,u(x_i) - u(x_{i+1})]\,\nu\left(\cup_1^i E_j\right) + u(x_n)\,, \tag{3.3}$$

where the outcomes are ranked as $x_1 \succ x_2 \succ ... \succ x_n$ and the act $e$ has $e(x_i) = E_i$, $i = 1, ..., n$.

**Lemma 3.4.** *Let $U^{ceu}$ be a CEU utility function with capacity $\nu$.*

*(a) The following conditions are sufficient for $U^{ceu}$ to be uncertainty averse with supporting $U^{ps}$ having $m$ as underlying probability measure: There exists a bijection $g : [0,1] \longrightarrow [0,1]$ such that*

$$m \in core\left(g^{-1}(\nu)\right); \text{ and} \tag{3.4}$$

$$m(\cdot) = g^{-1}\left(\nu(\cdot)\right) \quad \text{on } \mathcal{A}. \tag{3.5}$$

*(b) Suppose that $\nu$ is convex-ranged and that $\mathcal{A}$ is rich. Then the conditions in (a) are necessary in order that $U^{ceu}$ be uncertainty averse with supporting $U^{ps}$ having $m$ as underlying probability measure.*

*(c) Finally, in each of the preceding parts, the supporting utility $U^{ps}$ can be taken to be an expected utility function* if and only if *in addition $g$ is the identity function.*

[14] See [15]. Given countable additivity, convex-ranged is equivalent to non-atomicity.

See Appendix A for a proof. The supporting utility function $U^{ps}$ that is provided by the proof of (a) has the form (2.4), where the risk preference functional $W$ is

$$W(\Psi) = \int_{\mathcal{X}} u(x)\, d(g \circ \Psi)(x),$$

a member of the rank-dependent expected utility class [2].

Observe first that attitudes towards uncertainty do not depend on properties of the vNM index $u$. More surprising is that given $m$, the conditions on $\nu$ described in (a) are ordinal invariants, that is, if $\nu$ satisfies them, then so does $\varphi(\nu)$ for any monotonic transformation $\varphi$. In other words, $\nu$ and $g$ satisfy these conditions if and only if $\varphi(\nu)$ and $\widehat{g} = \varphi(g)$ do. Consequently, under the regularity conditions in the lemma, the CEU utility function $\int u(e)\, d\nu$ is uncertainty averse if and only if the same is true for $\int u(e)\, d\varphi(\nu)$. The fact that *uncertainty aversion is determined by ordinal properties of the capacity* makes it perfectly clear that uncertainty aversion has little to do with convexity, a cardinal property.

Thus far, only parts (a) and (b) of the lemma have been used. Focus now on (c), characterizing conditions under which $U^{ceu}$ is "more uncertainty averse than some expected utility order with probability measure $m$." Because the CEU utility functions studied by Schmeidler are defined on horse-race / roulette-wheels and conform with expected utility on the objective roulette-wheels, this latter comparison may be more relevant than uncertainty aversion per se for understanding the connection with convexity. The lemma delivers the requirement that $\nu$ be additive on $\mathcal{A}$ and that it admit an extension to the measure $m$ lying in $core(\nu)$.[15] It is well known that convexity of $\nu$ is sufficient for nonemptiness of the core, but that seems to be the extent of the link with uncertainty aversion. An example in the next subsection shows that $U^{ceu}$ may be more uncertainty averse than some expected utility order even though its capacity is not convex.

To summarize, there appears to be no logical connection in the Savage framework between uncertainty aversion and convexity. Convexity does not imply uncertainty aversion, unless added conditions such as (3.2) are imposed (see also the interval beliefs example in Section 4.2). Furthermore, convexity is not necessary even for the stricter notion 'more uncertainty averse than some expected utility order' that seems closer to Schmeidler's notion. As emphasized in the introduction, this is not to say that convexity and the asssociated multiple-priors functional

[15] The following observation may be useful in a future extension that partially endogenizes $\mathcal{A}$: Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two $\lambda$-systems and $m_1$ and $m_2$ probability measures on $\Sigma$, such that $\nu = m_i$ on $\mathcal{A}_i$ and $m_i \in core(\nu)$, $i = 1, 2$. Then $m_1 = m_2$ on $\mathcal{A}_1 \cup \mathcal{A}_2$.

structure that it delivers are not useful hypotheses. Rather, the point is to object to the widely adopted behavioural interpretation of convexity as uncertainty aversion.

### 3.3. Inner Measures

Zhang [28] argues that rather than convex capacities, it is capacities that are inner measures that model uncertainty aversion. These capacities are defined as follows: Let $p$ be a probability measure on $\mathcal{A}$; its existence reflects the unambiguous nature of events in $\mathcal{A}$. Then the corresponding *inner measure* $p_*$ is the capacity given by

$$p_*(E) = \sup\{p(B) : B \subset E, B \in \mathcal{A}\}, \quad E \in \Sigma.$$

The fact that the capacity of any $E$ is computed by means of an inner approximation by unambiguous events seems to capture a form of aversion to ambiguity. Zhang provides axioms for preference that are consistent with this intuition and that deliver the subclass of CEU preferences having an inner measure as the capacity $\nu$.

It is interesting to ask whether CEU preferences with inner measures are uncertainty averse in the formal sense of this paper. The answer is 'sometimes' as described in the next lemma.

**Lemma 3.5.** *Let $U^{ceu}(\cdot) \equiv \int u(\cdot)dp_*$, where $p_*$ is the inner measure generated as above from the probability measure $p$ on $\mathcal{A}$.*

*(a) If $p$ admits an extension to a probability measure on $\Sigma$, then $U^{ceu}$ is more uncertainty averse than the expected utility function $\int u(\cdot)dp$.*

*(b) Adopt the auxiliary assumptions in Lemma 3.4(b). If $U^{ceu}$ is uncertainty averse, then $p$ admits an extension from $\mathcal{A}$ to a measure on all of $\Sigma$.*

**Proof.** (a) $p_*$ and $p$ coincide on $\mathcal{A}$. For every $B \subset E$, $p(B) \leq p(E)$. Therefore, $p_*(E) \leq p(E)$. From the formula (3.3) for the Choquet integral, conclude that for all acts $e$ and unambiguous acts $h$,

$$\int u(h)\, dp_* = \int u(h)\, dp \text{ and } \int u(e)\, dp_* \leq \int u(e)\, dp.$$

(b) By Lemma 3.4 and its proof, $p = p_* = g(m)$ on $\mathcal{A}$ and $m(\mathcal{A}) = [0, 1]$. Therefore, $g$ must be the identity function. Again by the previous lemma, $m$ lies in $core(p_*)$, implying that $m \geq p_* = p$ on $\mathcal{A}$. Because $\mathcal{A}$ is closed with respect to complements, conclude that $m = p$ on $\mathcal{A}$ and hence that $m$ is the asserted extension of $p$. ■

Both directions in the lemma are of interest. Because in general, a probability measure on $\mathcal{A}$ need not admit an extension to $\Sigma$, part (a) shows that the intuition surrounding 'inner approximation' is flawed or incomplete, demonstrating the importance of a formal definition of uncertainty aversion. On the other hand, part (b) provides a class of examples of Choquet expected utility functions that are more uncertainty averse than some expected utility order. These can be used to show that even if this stricter notion of (more) uncertainty averse is adopted, the capacity $p_*$ *need not be convex*. This follows from Zhang's observation that an inner measure is not convex in general.[16] For the convenience of the reader, I repeat here Zhang's counterexample, consisting of an Ellsberg urn with balls of 4 colors: $S = \{R, B, G, W\}$, $\mathcal{A} = \{\emptyset, S, \{B, G\}, \{R, W\}, \{B, R\}, \{G, W\}\}$ and $\Sigma$ is the power set. Let $p$ be the equally likely (counting) probability measure on the power set. Using the restriction of $p$ to $\mathcal{A}$, define the inner measure $p_*$. Then the conditions of part (b) are satisfied, but $p_*$ is not convex because

$$1 = p_*(\{B, G\}) + p_*(\{B, R\}) > p_*(\{B, G, R\}) + p_*(\{B\}) = 1/2.$$

It is worth mentioning also that in this example, $\mathcal{A}$ is rich because for every $E$ there exists some unambiguous $A$ such that $p_*(E) = p_*(A) = p(A)$.

### 3.4. Bets, Beliefs and Uncertainty Aversion

This section examines some implications of uncertainty aversion for the ranking of binary acts. Because the ranking of bets reveals the decision-maker's underlying beliefs or likelihoods, these implications clarify the meaning of uncertainty aversion and help to demonstrate its intuitive empirical content. The generic binary act is denoted $xEy$, indicating that $x$ is obtained if $E$ is realized and $y$ otherwise.

Let $\succeq$ be uncertainty averse with probability sophisticated order $\succeq^{ps}$ satisfying (2.5). Apply the latter to binary acts, to obtain the following relation: For all unambiguous $A$, events $E$ and outcomes $x_1$ and $x_2$,

$$x_1 A x_2 \succeq^{ps} (\succ^{ps})\, x_1 E x_2 \quad \Longrightarrow \quad x_1 A x_2 \succeq (\succ)\, x_1 E x_2.$$

Proceed to transform this relation into a more illuminating form.

[16] $p_*$ is convex if $\mathcal{A}$ is an algebra [23, Theorem 5.1]. But recall that $\lambda$-systems are more appropriate for modelling the class of unambiguous events.

Exclude the uninteresting case $x_1 \sim x_2$ and assume that

$$x_1 \succ x_2.$$

Then $x_1 E x_2$ can be viewed as a *bet* on the event $E$. As noted earlier, $\succeq^{ps}$ necessarily agrees with the given $\succeq$ in the ranking of unambiguous acts and hence also constant acts or outcomes, so $x_1 \succ^{ps} x_2$. Let $m$ be the subjective probability measure on the state space $(S, \Sigma)$ that underlies $\succeq^{ps}$. Then the monotonicity property inherent in probabilistic sophistication implies that

$$x_1 A x_2 \succeq^{ps} (\succ^{ps})\, x_1 E x_2 \iff m(A_1) \geq (>)\, m(E_1).$$

Conclude that uncertainty aversion implies the existence of a probability measure $m$ such that: For all $A$, $E$, $x_1$ and $x_2$ as above,

$$m(A) \geq (>)\, m(E) \implies x_1 A x_2 \succeq (\succ)\, x_1 E x_2.$$

One final rewriting is useful. Define, for the given pair $x_1 \succ x_2$,

$$\nu(E) = U\left(x_1 E x_2\right).$$

Then,

$$mA \geq (>)\, mE \implies \nu A \geq (>)\, \nu E, \tag{3.6}$$

which is the sought-after implication of uncertainy aversion.[17]

In the special case of CEU (1.2), with vNM index satisfying $u(x_1) = 1$ and $u(x_2) = 0$, $\nu$ defined as above coincides with the capacity in the CEU functional form. Refer to $\nu$ more generally as a *capacity*, even when CEU is not assumed.[18] The interpretation is that $\nu$ represents $\succeq$ numerically over bets on various events with the given stakes $x_1$ and $x_2$, or alternatively, that it represents numerically the likelihood relation underlying preference $\succeq$. From this perspective, only the ordinal properties of $\nu$ are significant.[19] An implication of (3.6) is that $\nu$ and $m$ must be ordinally equivalent on $\mathcal{A}$ (though not on $\Sigma$).

[17] This condition is necessary for uncertainty aversion but *not* sufficient, even if there are only two possible outcomes. That is because by taking $h$ in (2.5) to be a constant act, one concludes that an uncertainty averse order $\succeq$ assigns a lower certainty equivalent to any act than does the supporting order $\succeq^{ps}$. In contrast, (3.6) contains information only on the ranking of bets and not on their certainty equivalents. (I am assuming here that certainty equivalents exist.)

[18] Contrary to common usage of the term 'capacity', $\nu$ need not be monotone with respect to set inclusion unless $\succeq$ is suitably monotone.

[19] These ordinal properties are independent of the particular pair of outcomes satisfying $x_1 \succ x_2$ if (and only if) $\succeq$ satisfies Savage's axiom P4.

In other words, uncertainty aversion implies the existence of a probability measure $m$ that *supports* $\{E \in \Sigma : \nu(E) \geq \nu(A)\}$ at each unambiguous $A$, where support is in a sense analogous to the usual meaning, except that the usual linear supporting function defined on a linear space is replaced by an additive function defined on an algebra. Think of the measure $m$ as describing the (not necessarily unique) 'mean ambiguity-free likelihoods' implicit in $\nu$ and $\succeq$. This interpretation and the 'support' analogy are pursued and developed further in Section 4.3 under the assumption that preference is eventwise differentiable.

In a similar fashion, one can show that uncertainty loving implies the existence of a probability measure $q$ on $(S, \Sigma)$ such that

$$q(A) \leq \,(<)\, q(E) \implies \nu(A) \leq \,(<)\, \nu(E), \tag{3.7}$$

for every $E \in \Sigma$ and $A \in \mathcal{A}$. The conjunction of (3.6) and (3.7) imply, under a mild additional assumption, that $\nu$ is ordinally equivalent to a probability measure (see Lemma A.1), which is one step in the proof of Theorem 3.1.

Because choice between bets provides much of the experimental evidence regarding nonindifference to uncertainty, the implication (3.6) is convenient for demonstrating the intuitive empirical content of uncertainty aversion. The Ellsberg urn discussed in the introduction provides the natural vehicle. Consider again the typical choices in (1.4). In order to relate these rankings to the formal definition of uncertainty aversion, adopt the natural specification

$$\mathcal{A} = \{\emptyset, S, \{R\}, \{B, G\}\}.$$

Given this specification, it is easy to see that these rankings *imply* uncertainty aversion - the measure $m$ assigning 1/3 probability to each color is a support in the sense of (3.6).

Equally revealing is that the notion of uncertainty aversion excludes behaviour that is interpreted *intuitively* as reflecting an affinity for ambiguity.[20] To see this, suppose that the decision-maker's rankings are changed by reversing the strict preference '$\succ$' to '$\prec$'. These new rankings contradict uncertainty aversion: Let $m$ be a support as in the implication (3.6) of uncertainty aversion and take $A = \{B, G\}$. Then $\{B, G\} \prec \{R, B\}$ implies that $m(\{B, G\}) < m(\{R, B\})$. Because $m$ is additive, conclude that $m(G) < m(R)$. But then uncertainty aversion applied to the unambiguous event $\{R\}$ implies that $\{R\} \succ \{G\}$, contrary to the hypothesis.

[20] Alternatively, we could show that the rankings in (1.4) are inconsistent with the implication (3.7) of uncertainty loving.

Though a general formal result seems unachievable, there is an informal sense in which these results seem to be valid much more broadly than the specific Ellsberg experiment considered. Typically when choices are viewed as paradoxical relative to probabilistically sophisticated preferences, there is a natural probability measure on the state space that is 'contradicted' by observed choices. This seems close to saying precisely that the measure is a support.

Another revealing implication of uncertainty aversion is readily derived from (3.6). Notation that is useful here and below is, given $A$, write an arbitrary event $E$ in the form

$$E = A + F - G, \text{ where } F = E \backslash A \text{ and } G = A \backslash E. \tag{3.8}$$

Henceforth, $E + F$ denotes both $E \cup F$ and the assumption that the sets are disjoint. Similarly, implicit in the notation $E - G$ is that $G \subset E$. Now let $m$ be the supporting measure delivered by uncertainty aversion. Then for any unambiguous $A'$ and $A$, if $F \subset A' \cap A^c$ and $G \subset A'^c \cap A$,

$$A' - F + G \succeq A' \implies A + F - G \preceq A, \tag{3.9}$$

because the first ranking implies (by the support property at $A'$) that $mF \leq mG$ and this implies the second ranking (by the support property at $A$).[21] In particular, taking $A' = A^c$,

$$A^c - F + G \succeq A^c \implies A + F - G \preceq A, \tag{3.10}$$

for all $F \subset A^c$ and $G \subset A$. The interpretation is that if $F$ seems small relative to $G$ when (as at $A'$) one is contemplating subtracting $F$ and adding $G$, then it also seems small when (as at $A$) one is contemplating adding $F$ and subtracting $G$. This is reminiscent of the familiar inequality between the compensating and equivalent variations for an economic change, or the property of diminishing marginal rate of substitution. A closer connection between uncertainty aversion and such familiar notions from consumer theory is possible if eventwise differentiability of preference is assumed, as in the next section.

[21] A slight strengthening of (3.9) is valid. Suppose that

$$A' - F^i + G^i \succeq A' \text{ all } i \ ,$$

for some partitions $F = \Sigma F^i$ and $G = \Sigma G^i$. Only the trivial partitions were admitted above. Then additivity of the supporting measure implies as above that $mF \leq mG$ and hence that $A + F - G \preceq A$.

# 4. DIFFERENTIABLE UTILITIES

Tractability in applying the notion of uncertainty aversion raises the following question: Is there a procedure for deriving from $\succeq$ all probabilistically sophisticated orders satisfying (2.5), or for deriving from $\nu$ all candidate supporting measures $m$ satisfying (3.6)? We turn now to this question and show that eventwise differentiability of preference provides such a procedure in some cases. More precisely, conditions are provided that *deliver a unique supporting measure from the eventwise derivative of utility.* When combined with Lemmas 3.2 and 3.4, this provides the sought after procedure.

## 4.1. Definition of Eventwise Differentiability

The standard representation of an act, used above, is as a measurable map from states into outcomes. Let $e : S \longrightarrow \mathcal{X}$ be such an act. An alternative representation of this act is by means of the inverse correspondence $e^{-1}$, denoted by $\widehat{e}$. Thus $\widehat{e} : \mathcal{X} \longrightarrow \Sigma$, where $\widehat{e}(x)$ denotes the event $E$ on which the act assumes the outcome $x$. For notational simplicity, it is convenient to write $e$ rather than $\widehat{e}$ and to leave it to the context to make clear whether $e$ denotes a mapping from states into consequences or alternatively from outomes into events.

Henceforth, when examining the decision-maker's ranking of a pair of acts, view those acts as assigning a common set of outcomes to different events. This perspective is 'dual' to the more common one, where distinct acts are viewed as assigning different outcomes to common events. These two perspectives are mathematically equally valid; the choice between them is a matter of convenience. The latter is well suited to the study of risk aversion (attitudes towards variability in outcomes) and, it is argued here, the former is well suited to the study of uncertainty aversion. The intuition is that uncertainty or ambiguity stems from events and that aversion to uncertainty reflects attitudes towards changes in those events.

Because acts are simple,

$$\{x \in \mathcal{X} : e(x) \neq \emptyset\} \quad \text{is finite.} \tag{4.1}$$

In addition,

$$\{e(x) : x \in \mathcal{X},\, e(x) \neq \emptyset\} \text{ partitions } S. \tag{4.2}$$

The set of acts $\mathcal{F}$ may be identified with the set of all maps satisfying these two conditions. In particular, $\mathcal{F} \subset \Sigma^{\mathcal{X}}$, where the latter is defined as the set of all maps from $\mathcal{X}$ into $\Sigma$ satisfying (4.1).

Let $U : \mathcal{F} \longrightarrow \mathcal{R}$ be a utility function for $\succeq$ and define the 'eventwise derivative of $U$'. Because utility is defined on a subset of $\Sigma^{\mathcal{X}}$, it is convenient to define derivatives first for functions $\Phi$ that are defined on all of $\Sigma^{\mathcal{X}}$. Continue to refer to elements $e \in \Sigma^{\mathcal{X}}$ as acts even when they are not elements of $\mathcal{F}$.

The following structure for $\Sigma^{\mathcal{X}}$ is useful. Define the operations '$\cup$', '$\cap$' and 'complementation' ($e \longmapsto e^c$) on $\Sigma^{\mathcal{X}}$ co-ordinatewise; for example,

$$(e \cup f)(x) \equiv e(x) \cup f(x), \text{ for all } x \in \mathcal{X}.$$

Say that $e$ and $f$ are *disjoint* if $e(x) \cap f(x) \equiv \emptyset$ for all $x$, abbreviated $e \cap f = \emptyset$. In that case, denote the above union by $e + f$. The notation $e' \backslash e$ and $e' \Delta e$ indicates set difference and symmetric difference applied outcome by outcome. Similar meaning is given to $g \subset e$.

Say that $\{f^j\}_{j=1}^n$ *partitions* $f$ if $\{f^j(x)\}$ partitions $f(x)$ for each $x$. Define the *refinement* partial ordering of partitions in the obvious way. Given an act $f$, $\left\{\{f^{j,\lambda}\}_{j=1}^{n_\lambda}\right\}_\lambda$ denotes the *net* of all finite partitions of $f$, where $\lambda < \lambda'$ if and only if the partition corresponding to $\lambda'$ refines the partition corresponding to $\lambda$.

A real-valued function $\mu$ on $\Sigma^{\mathcal{X}}$ is called *additive* if it is additive across disjoint acts. Refer to such a function as a (signed) measure even though that terminology is usually reserved for functions defined on algebras, while $\Sigma^{\mathcal{X}}$ is not an algebra.[22] Expected utility functions, $U(e) = \Sigma_x u(x)\, p(e(x))$, are additive and hence measures in this terminology. For any additive $\mu$, $\mu(\emptyset) = 0$ and

$$\mu(e) = \Sigma_x \mu_x (e(x)), \tag{4.3}$$

where $\mu_x$ is the *marginal* measure on $\Sigma$ defined by $\mu_x(E) =$ the $\mu$-measure of the act that assigns $E$ to the outcome $x$ and the empty set to every other outcome.

Apply to each marginal the standard notions and results for finitely additive measures on an algebra (see [15]). In this way, one obtains a decomposition of $\mu$,

$$\mu = \mu^+ - \mu^-,$$

where $\mu^+$ and $\mu^-$ are non-negative measures. Define

$$|\mu| = \mu^+ + \mu^-.$$

Say that the measure $\mu$ is *bounded* if

$$\sup_f |\mu|(f) = \sup \left\{\Sigma_{j=1}^{n_\lambda} |\mu(f^{j,\lambda})| : f \in \Sigma^{\mathcal{X}}, \lambda\right\} < \infty. \tag{4.4}$$

[22] In particular, $\Sigma^{\mathcal{X}}$ is *not* the product algebra on $S^{\mathcal{X}}$ induced by $\Sigma$. However, the operations we have defined make $\Sigma^{\mathcal{X}}$ a ring, that is, it is closed with respect to unions and differences.

Call the measure $\mu$ on $\Sigma^{\mathcal{X}}$ *convex-ranged* if for every $e$ and $r \in (0, \mid \mu \mid (e))$, there exists $b$, $b \subset e$ such that $\mid \mu \mid (b) = r$, where $e$ and $b$ are elements of $\Sigma^{\mathcal{X}}$. Lemma B.1 summarizes some properties of convex-ranged measures on $\Sigma^{\mathcal{X}}$.

Define differentiability for a function $\Phi : \Sigma^{\mathcal{X}} \longrightarrow \mathcal{R}^1$.

**Definition 4.1.** *$\Phi$ is* ***(eventwise) differentiable*** *at $e \in \Sigma^{\mathcal{X}}$ if there exists a bounded and convex-ranged measure $\delta\Phi(\cdot; e)$ on $\Sigma^{\mathcal{X}}$, such that: For all $f \subset e^c$ and $g \subset e$,*

$$\Sigma_{j=1}^{n_\lambda} \mid \Phi\left(e + f^{j,\lambda} - g^{j,\lambda}\right) - \Phi\left(e\right) - \delta\Phi(f^{j,\lambda}; e) + \delta\Phi(g^{j,\lambda}; e) \mid \longrightarrow_\lambda 0. \quad (4.5)$$

The requirement of convex range for $\delta\Phi(\cdot; e)$ is not needed everywhere below, but is built into the definition for ease of exposition.

Any utility function $U$ is defined on the proper subset $\mathcal{F}$ of $\Sigma^{\mathcal{X}}$. Define $\delta U(\cdot; e)$ as above, with the exception that the perturbations $f^{j,\lambda}$ and $g^{j,\lambda}$ are restricted so that $e + f^{j,\lambda} - g^{j,\lambda}$ lies in $\mathcal{F}$. Say that $U$ is *eventwise differentiable* if the derivative exists at each $e$ in $\mathcal{F}$.

To clarify the notation, suppose that $e$ is an act in $\mathcal{F}$ that assumes the outcomes $x_1$ and $x_2$ on $E$ and $E^c$ respectively. Let $f$ assume (only) these outcomes on events $F \subset E^c$ and $G \subset E$, while $g$ assumes (only) $x_1$ and $x_2$ on $G$ and $F$ respectively. Then $f$ and $g$ lie in $\Sigma^{\mathcal{X}}$, $f \subset e^c$, $g \subset e$ and $e + f - g$ is the act in $\mathcal{F}$ that yields $x_1$ on $E + F - G$ and $x_2$ on its complement. Further if $\{F^{j,\lambda}\}$ and $\{G^{j,\lambda}\}$ are partitions of $F$ and $G$ and if $f^{j,\lambda}$ and $g^{j,\lambda}$ are defined in fashion paralleling the definitions given for $f$ and $g$, then $\{f^{j,\lambda}\}$ and $\{g^{j,\lambda}\}$ are partitions of $f$ and $g$ that eneter into the definition of $\delta U(\cdot; e)$.

The rest of this subsection is devoted to clarifying and interpreting eventwise differentiability. The essence of the definition can be more easily understood by adapting it to the case where the domain of $\Phi$ is $\Sigma$. Then each act $e$ is simply an event $E$. One can think of $\Phi$ as a capacity and of $\delta\Phi(\cdot; E)$ as its derivative at $E$. The following discussion deals with the general case, but some readers may wish to keep the special case in mind.

For each $\lambda$, $\{f^{j,\lambda}\}_{j=1}^{n_\lambda}$ and $\{g^{j,\lambda}\}_{j=1}^{n_\lambda}$ are partitions of $f$ and $g$ respectively and these partitions are fine when $\lambda$ is large. Therefore, differentiability at $e$ states that the difference $\Phi(e+f-g) - \Phi(e)$ can be approximated by $\delta\Phi(f; e) - \delta\Phi(g; e)$ for suitably 'small' $f$ and $g$, where the small size of the perturbation "$f - g$" is in the sense of the fineness of the partitions as $\lambda$ grows. Naturally, it is important that the approximating functional $\delta\Phi(\cdot; e)$ is additive (a signed measure). There is an apparent parallel with Gateaux (directional) differentiability of functions defined on a linear space - "$f - g$" represents the 'direction' of perturbation and the additive approximation replaces the usual linear one. Note that the perturbation from $e$ to $e + f - g$ is perfectly general; any $e'$ can be expressed (uniquely) in the form $e' = e + f - g$, with $f \subset e^c$ and $g \subset e$ (see (3.8)).

Though I use the term derivative, $\delta\Phi(\cdot; e)$ is actually the counterpart of a *differential.* The need for a signed measure arises from the absence of any monotonicity assumptions. If $\Phi(\cdot)$ is monotone with respect to inclusion $\subset$, then each $\delta\Phi(\cdot; e)$ is a non-negative measure.

The notion of eventwise differentiablity is nonstandard, (for example, its definition does not refer to anything resembling a difference quotient), and some readers may wonder about tractability, (that is, can we compute derivatives?). In the hope of addressing these concerns, a stronger form of differentiability (similar to that in [13]) is described in Appendix C. The examples to follow should also help in this respect. Finally, observe that for a function $\varphi : \mathcal{R}^1 \longrightarrow \mathcal{R}^1$ that is differentiable at some $x$ in the usual sense, elementary algebraic manipulation of the definition of the derivative $\varphi'(x)$ yields the following expression paralleling (4.5):

$$\Sigma_{i=1}^{N} \left[\varphi(x + N^{-1}) \,-\, \varphi(x) \,-\, N^{-1}\varphi'(x)\right] \longrightarrow_{N \longrightarrow \infty} 0.$$

These clarifications and 'justifications' for our definition are largely mathematical. At least as important is that the eventwise derivative of utility $U$ and expressions involving the derivative that play a role below admit behavioural interpretations. The suggested interpretation is that $\delta U(\cdot; e)$ represents the 'mean' or 'uncertainty free' assessment of acts implicit in utility, as viewed from the perspective of the act $e$. It may help to recall that in the theory of expected utility over objective lotteries or risk, if the vNM index is differentiable, then utility is linear to the first order and hence preference is risk neutral for small gambles. The suggested parallel here is that a differentiable utility is additive (rather than linear) and uncertainty neutral (rather than risk neutral) to the 'first-order'.

To support this suggestion, consider inequalities of the form

$$\delta U(f; e) \,<\, (>)\, \delta U(g; e), \tag{4.6}$$

where $e \in \Sigma^{\mathcal{X}}$, $f \subset e^c$ and $g \subset e$. Such inequalities play a prominent role below. That they have behavioural significance is apparent from the fact that they are invariant to (suitable) monotonic transformations of $U$; that is, consider the ordinally equivalent utility function $\varphi \circ U$, where $\varphi$ is continuously differentiable. Then the Chain Rule for eventwise differentiability (Theorem B.2) implies that

$$\delta(\varphi \circ U)(\cdot; e) \,=\, \varphi'(U(e))\, \delta U(\cdot; e).$$

Thus (4.6) is satisfied also by the derivative of $\varphi \circ U$.

The behavioural meaning of these inequalities may be described as follows: Given (4.6), then

$$\forall\lambda_0 \;\; \exists\lambda > \lambda_0 \text{ such that } U(e + f^{j,\lambda} - g^{j,\lambda}) \; < \; (>) \; U(e), \quad j = 1, ..., n_\lambda. \tag{4.7}$$

(Let $\delta U(f;e) > \delta U(g;e)$. Because $\delta U(\cdot;e)$ is convex-ranged, Lemma B.1 implies that there exist partitions, as fine as desired, satisfying $\delta U(f^{j,\lambda};e) > \delta U(g^{j,\lambda};e)$, for all $j$. Now apply the definition of eventwise differentiability.) Conversely, (4.7) implies $\delta U(f;e) \leq (\geq) \, \delta U(g;e)$ by the definition of eventwise derivative. In other words, overlooking the distinction between weak and strict inequalities, (4.7) is the *behavioural characterization* of the inequalities (4.6).[23]

In the case of preference over lotteries or risk, it is well known that a decision-maker with suitably Gateaux differentiable utility function will always (strictly) choose one side of an actuarially non-neutral bet or random variable if she can also choose the scale of the bet. This applies in particular to an expected utility maximizer with differentiable vNM index. As a result, if the gamble is actuarially favourable and is represented by the random variable $X$, then $X$ can be decomposed into a sum $\Sigma X^j$ of small gambles ($X^j = tX$ for some small scalar $t$), such that *each single* $X^j$ would be accepted by the decision-maker. The parallel with (4.7) is apparent, though the meanings of "actuarially non-neutral' and 'small gamble' and the nature of the decomposition are different. The reason for the noted property of preference in the case of risk is that, if the vNM index is differentiable, then utility is linear to the first order and hence preference is risk neutral for small gambles.[24] Thus, expected value alone dictates the direction of preference for small gambles. Similarly here, the direction of preference for sufficiently small gambles is determined by the eventwise derivative of utility. If one views such small gambles as uncertainty free, then $\delta U(\cdot;e)$ must reflect uncertainty-free assessments for small gambles. But by additivity, $\delta U(\cdot;e)$ is completely determined by the values it assigns to small gambles. This completes the justification for the behavioural interpretation of (4.6).[25]

[23] Some readers may feel that the definition of eventwise differentiability is somewhat arbitrary and that there exist alternative mathematical definitions, seemingly as plausible, that might be adopted. The behavioural characterization provided for (4.6) may serve to distinguish between alternative definitions on economic grounds. We are not aware of alternative definitions of differentiability consistent with the characterization (4.7).

[24] For a recent discussion see [22]. They study also preferences that are risk averse even for small gambles, a property that they term 'first-order risk aversion.'

[25] This may not be totally convincing because of the identification of additivity (here of

A natural question is "how restrictive is the assumption of eventwise differentiability?" In this connection, the reader may have noted that the definition is formulated for an arbitrary state space $S$ and algebra $\Sigma$. However, eventwise differentiability is potentially interesting only in cases where these are both infinite. That is because if $\Sigma$ is finite, then $\Phi$ is differentiable if and only if it is additive. Examples and further discussion below should clarify the meaning and scope of the assumption of eventwise differentiability.

Another question concerns the uniqueness of the derivative. The limiting condition (4.5) has at most one solution, that is, the derivative is unique if it exists: If $p$ and $q$ are two measures on $\Sigma^{\mathcal{X}}$ satisfying the limiting property, then for each $g \subset e^c$, $\mid p(g) - q(g) \mid \leq \Sigma_{j=1}^{n_\lambda} \mid p(g^{j,\lambda}) - q(g^{j,\lambda}) \mid \longrightarrow_\lambda 0$. Therefore, $p(g) = q(g)$ for all $g \subset e$. Similarly, prove equality for all $f \subset e^c$ and then apply additivity.

The marginals of $\delta\Phi(\cdot; e)$ are defined as in (4.3). The marginal $\delta_x\Phi(\cdot; e)$ is a measure on $\Sigma$, that is non-negative under suitable monotonicity for $\Phi$. Its interpretation is roughly that $\delta_x\Phi(F; e)$ describes the 'first-order' change induced in the value of $\Phi(e)$ as a result of a small perturbation of $e$ in the $x$-component only and in the 'direction' $F \subset (e(x))^c$; similarly $-\delta_x\Phi(G; e)$ describes the effect of a perturbation in the 'direction' $G \subset e(x)$. Roughly, $\delta_x\Phi(\cdot; e)$ is a partial derivative (or differential) with respect to the $x$ component.

**Remark 1.** *Eventwise differentiability is inspired by Rosenmuller's notion, but there are differences between them. Rosenmuller deals with convex capacities defined on $\Sigma$, rather than with utility functions defined on acts. Even within that framework, his formulation differs from (4.5) and relies on the assumed convexity. Moreover, he restricts attention to 'one-sided' derivatives, that is, where the inner perturbation $g$ is identically empty (producing an outer derivative), or where the outer perturbation $f$ is identically empty (producing an inner derivative). Finally, Rosenmuller's application is to co-operative game theory rather than to decision theory.*

### 4.2. Examples

Turn to some examples that illustrate both differentiability and uncertainty aversion. All are special cases of the CEU model (3.3), though other examples are

$\delta U(\cdot; e))$ with uncertainty neutrality. While expected utility functions are additive over acts, there are many probabilistically sophisticated (and hence uncertainty neutral) utility functions that are not. The answer seems to be that their non-additivity vanishes in the small.

readily constructed. Because the discussion of differentiability dealt with functions defined on $\Sigma^{\mathcal{X}}$ rather than just $\mathcal{F}$, rewrite the CEU functional form here using this larger domain. If the outcomes satisfy $x_1 \succ x_2 \succ ... \succ x_n$ and the act $e$ has $e(x_i) = E_i$, $i = 1, ..., n$, then

$$U^{ceu}(e) = \Sigma_{i=1}^{n-1}[\,u(x_i) - u(x_{i+1})]\,\nu\left(\cup_1^i E_j\right) + u(x_n)\,\nu\left(\cup_1^n E_j\right) .$$

Suppose that the capacity $\nu$ is eventwise differentiable with derivative $\delta\nu(\cdot;E)$ at $E$; naturally, differentiability is in the sense of the last section (with $\mid \mathcal{X} \mid = 1$). Then $U^{ceu}(\cdot)$ is eventwise differentiable with derivative

$$\delta U(e';e) = \Sigma_{i=1}^{n-1}[\,u(x_i) - u(x_{i+1})]\,\delta\nu\left(\cup_1^i E_j';\cup_1^i E_j\right) + u(x_n)\,\delta\nu\left(\cup_1^n E_j';\cup_1^n E_j\right), \tag{4.8}$$

where $e'(x_i) = E_i'$. (This follows as in calculus from the additivity property of differentiation.) Because differentiability of utility is determined totally by that of the capacity, it is enough to consider examples of differentiable (and nondifferentiable) capacities. In each case where the capacity is differentiable, (4.8) describes the corresponding derivative of utility.

The CEU case demonstrates clearly that eventwise differentiability is distinct from more familiar notions, such as Gateaux differentiability. It is well-known that a CEU utility function is not (two-sided) Gateaux differentiable, even if the vNM index is smooth, unless it is an expected utility function. In contrast, many CEU utility functions are eventwise differentiable, regardless of the nature of $u(\cdot)$.

Verification of the formulae provided for derivatives is possible using the definition (4.5). Alternatively, verification of the stronger $\mu$-differentiability (see Appendix C) is more straightforward. (Define $\mu$ by (C.2) and $\mu_0 = p$ in the first two examples, $= q$ in the third example and $= \ell^*/\ell^*(S)$ in the final example, where only 'one-sided' derivatives exist.)

**Example (Probability measure):** Let $p$ be a convex-ranged probability measure. Then $\delta p(\cdot;E) = p(\cdot)$, the same measure for all $E$. Application of (4.8) yields

$$\delta U(e';e) = \Sigma_{i=1}^n\, u(x_i)\, pE_i'.$$

Thus the $j^{th}$ partial derivative is $\delta_j U(\cdot;e) = u(x_j)\,p(\cdot)$.

**Example (Probabilistic sophistication** within CEU**):** Let

$$\nu = g(p), \tag{4.9}$$

where $p$ is a convex-ranged probability measure and $g : [0,1] \longrightarrow [0,1]$ is increasing, onto and continuously differentiable. The corresponding utility function lies in the rank-dependent-expected-utility class of functions studied in the case of risk where $p$ is taken to be objective. (See [2] and the references therein.) Then

$$\delta\nu(\cdot; E) = g'(pE)\, p(\cdot) \quad \text{and}$$

$$\delta U(e'; e) = \ \Sigma_{i=1}^{n} [\, u(x_i) - u(x_{i+1})]\, g'(p(\cup_1^i E_j))\, p(\cup_1^i E_j'),$$

where $u(x_{n+1}) \equiv 0$.

**Example (Quadratic capacity):** Let

$$\nu(E) = p(E)\, q(E),$$

where $p$ and $q$ are convex-ranged probability measures with $p << q$. Then

$$\delta\nu(\cdot; E) = p(E)\, q(\cdot) \,+\, p(\cdot)\, q(E),$$

a formula that is reminiscent of standard calculus.

Direct verification shows that $\nu$ is convex. As for uncertainty aversion, if $p$ and $q$ agree on $\mathcal{A}$, then the probability measure on $\Sigma$ defined by

$$m(\cdot) = \delta\nu(\cdot; A) \,/\, \delta\nu(S; A) \ = \ [q(\cdot) \,+\, p(\cdot)] \,/\, 2,$$

serves as a support in the sense of (3.6). That the implied CEU utility function is uncertainty averse in the full sense of (2.5) may be established by application of Lemma 3.4. Observe that $\nu = p^2 = m^2$ on $\mathcal{A}$; thus $g(t) = t^2$. Then $m$ lies in the core of $(pq)^{1/2}$, because $[p(\cdot) + q(\cdot)]^2 \geq 4\, p(\cdot)\, q(\cdot)$. The probabilistically sophisticated supporting utility function $U^{ps}$ is

$$U^{ps}(e) \ = \ \int_S u(e)\, dm^2.$$

**Example (Interval beliefs):** Let $\ell_*$ and $\ell^*$ be two non-negative, convex-ranged measures on $(S, \Sigma)$, such that

$$\ell_*(\cdot) \leq \ \ell^*(\cdot) \,\text{and}\ \ 0 < \ \ell_*(S) \ < \ 1 < \ \ell^*(S).$$

Define $\xi = \ \ell^*(S) - 1$ and

$$\nu(E) = \max\{\ell_*(E),\, \ell^*(E) - \xi\}. \tag{4.10}$$

Then $\nu$ is a convex capacity on $(S, \Sigma)$ and has the core

$$core(\nu) = \{p \in M(S, \Sigma) : \ell_*(\cdot) \leq p(\cdot) \leq \ell^*(\cdot) \text{ on } \Sigma\}.$$

This representation for the core provides intuition for $\nu$ and the reason for its name. See [26] for details regarding this capacity and its applications in robust statistics.

Because the capacity is "piecewise additive", one can easily see that though it has 'one-sided derivatives', $\nu$ is generally not eventwise differentiable at any $E$ such that $\ell_*(E) = \ell^*(E) - \xi$.

It follows from Theorem 3.3 and the nature of $core(\nu)$ that a CEU utility $U^{ceu}$ with capacity $\nu$ is uncertainty averse for any class $\mathcal{A}$ such that $\ell_*(\cdot) = \ell^*(\cdot)$ on $\mathcal{A} \backslash \{S\}$. Because any such class $\mathcal{A}$ excludes events that are 'close to' $S$, such an $\mathcal{A}$ cannot be rich.

In fact, one can show using Lemma 3.4, that it is impossible for $U^{ceu}$ to be uncertainty averse relative to any rich class of unambiguous events, unless $U^{ceu}$ is probabilistically sophisticated, providing another illustration of the lack of a connection between uncertainty aversion and convexity. The proof may be outlined as follows: Let $m$ be a supporting measure. Assume that $\mathcal{A}$ is rich. Lemma 3.4(b) (and its proof showing $m(\mathcal{A}) = [0,1]$) imply that $m = \ell^* / \ell^*(S)$ on $\mathcal{A}$ and that $\nu = g(m)$ there where

$$g(t) = \max\{\ell_*(S)\, t,\, \ell^*(S)\, t - \xi\}.$$

Again by Lemma 3.4, for all $E$ in $\Sigma$,

$$\max\{\ell_*(S)\, m(E),\, \ell^*(S)\, m(E) - \xi\} \geq \max\{\ell_*(E),\, \ell^*(E) - \xi\}.$$

For sufficiently 'small' events $E$, both maxima are achieved at the first term in the respective brackets because the other terms are negative. Thus for such events

$$\ell_*(S)\, m(E) \geq \ell_*(E), \text{ or } m(\cdot) \geq \ell_*(\cdot)/\ell_*(S).$$

By additivity, this obtains for all events $E$. But the latter inequality involves two probability measures and thus implies equality throughout. In a similar fashion one can show that

$$\ell^*(\cdot) \,/\, \ell^*(S) = m(\cdot) = \ell_*(\cdot) \,/\, \ell_*(S) \text{ on } \Sigma,$$

implying that $\nu$ is ordinally equivalent on $\Sigma$ to $m$, that is, $U^{ceu}$ is probabilistically sophisticated.

### 4.3. Uncertainty Aversion under Differentiability

To begin this section, the discussion will be restricted to binary acts; that is, uncertainty aversion will refer to (2.6), or equivalently, to (3.6). Implications are then drawn for uncertainty aversion in the full sense of general acts and (2.5).

The relevant derivative is $\delta\nu(\cdot;E)$, where $\nu E \equiv U(x_1Ex_2)$ and $U$ need *not* be a CEU function. Assume that $\nu E$ is increasing with $E$. Thus $\delta\nu(\cdot;E)$ is a non-negative measure, though not necessarily a probability measure. The suggested interpretation from Section 4.1, specialized to this case, is that $\delta\nu(\cdot;E)$ represents the 'mean' or 'uncertainty free' likelihoods implicit in $\nu$, as viewed from the perspective of the event $E$. This interpretation is natural given that $\delta\nu(\cdot;E)$ is additive over events and hence ordinally equivalent to a probability measure on $\Sigma$.

Turn to the relation between differentiability and uncertainty aversion. When $\nu$ is differentiable, analogy with calculus might suggest that the support at any event $A$, in the sense of (3.6), should be unique and given by $\delta\nu(\cdot;A)$, perhaps up to a scalar multiple. Though the analogy with calculus is imperfect, it is nevertheless the case that, under additional assumptions, differentiability provides information about the set of supports.

The principal additional assumption may be stated as follows: $\mathcal{A}^0 \equiv \{A \in \mathcal{A} : \nu(S) > \max\{\nu A, \nu A^c\}\}$, the set of unambiguous events $A$ such that $A$ and its complement are each strictly less likely than $S$. Say that $\nu$ is *coherent* if there exists a positive real-valued function $\kappa$ defined on $\mathcal{A}^0$, such that

$$\delta\nu(\cdot;A) = \kappa(A)\,\delta\nu(\cdot;A^c) \quad \text{on } \Sigma, \tag{4.11}$$

for each $A$ in $\mathcal{A}^0$. Coherence is satisfied by all the differentiable examples in Section 4.2. By the Chain Rule for eventwise differentiability (Theorem B.2), coherence is invariant to suitable monotonic transformations of $\nu$ and thus is an assumption about the preference ranking of binary acts. It is arguably an expression of the unambiguous nature of events in $\mathcal{A}$. To see this, it may help to consider first the following addition to (3.10):

$$A + F - G \preceq A \implies A^c - F + G \succeq A^c.$$

This is a questionable assumption because the events $A^c - F + G$ and $A + F - G$ are both ambiguous. Therefore, there is no reason to expect the perspective on the change 'add $F$ and subtract $G$' to be similar at $A^c$ as at $A$. However, if $F$ and $G$ are both 'small', then only mean likelihoods matter and it is reasonable that the relative mean likelihoods of $F$ and $G$ be the same from the two perspectives. In fact, such agreement seems to be an expression of the existence of 'coherent' ambiguity-free beliefs underlying preference. This condition translates into the following restriction on derivatives:

$$\delta\nu(F;A) \leq \delta\nu(G;A) \implies \delta\nu(F;A^c) \leq \delta\nu(G;A^c).$$

By arguments similar to those in the proof of the theorem, this implication delivers (4.11) under the assumptions in part (b). (Observe that the reverse implication follows from (3.10)).

We prove the following in Appendix A:[26]

**Theorem 4.2.** *Let $\nu$ be eventwise differentiable.*

*(a) If $\nu$ is uncertainty averse, then for all $A \in \mathcal{A}$, $F \subset A^c$ and $G \subset A$,*

$$\delta\nu(F;A^c) \leq \delta\nu(G;A^c) \implies \nu(A+F-G) \leq \nu(A). \tag{4.12}$$

*(b) Suppose further that $\Sigma$ is a $\sigma$-algebra and that $m$ and each $\delta\nu(\cdot, A)$, $A \in \mathcal{A}^0$, are countably additive, where $m$ is a support in the sense of (3.6). Then for each $A$ in $\mathcal{A}^0$,*

$$\delta\nu(F;A)\,m(G) \;\leq\; \delta\nu(G;A)\,m(F) \quad \textit{and} \tag{4.13}$$

$$\delta\nu(G;A^c)\,m(F) \;\leq\; \delta\nu(F;A^c)\,m(G). \tag{4.14}$$

*(c) Suppose further that $\mathcal{A}^0$ is nonempty and that $\nu$ is coherent. Then the unique countably additive supporting probability measure $m$ is given by $m(\cdot) = \delta\nu(\cdot;A)\,/\,\delta\nu(S;A)$, for any $A$ in $\mathcal{A}^0$.*

When division is permitted, the inequalities in (b) imply that

$$\frac{\delta\nu(F;A)}{\delta\nu(G;A)} \leq \frac{m(F)}{m(G)} \leq \frac{\delta\nu(F;A^c)}{\delta\nu(G;A^c)}, \tag{4.15}$$

which suggests an interpretation as an interval bound for the 'marginal rate of substitution at any $A$ between $F$ and $G$'.

The relation (4.12) states roughly that for each $A$, $\delta\nu(\cdot;A^c)$ serves as a support at $A$. Given our earlier interpretation for the derivative, it states that if the decision-maker would rather bet on $A+F-G$ than on $A$ when ambiguity is

[26]It would be desirable to express assumptions for $\delta\nu$ and $m$ in terms of the primitive $\nu$, but we have not succeeded in doing so.

ignored and when mean-likelihoods are computed from the perspective of $A^c$, then she would make the same choice also when ambiguity is considered. That is because the former event is more ambiguous and the decision-maker dislikes ambiguity or uncertainty.

Finally, part (c) of the theorem describes conditions under which the parallel with calculus is valid - the (countably additive) supporting measure is unique and given essentially by the derivative of $\nu$. Note that the support property in question here is global in that the same measure 'works' at each unambiguous $A$, and not just at a single given $A$.[27] This explains the need for the coherence assumption, which helps to ensure that $\delta\nu(\cdot;A)\,/\,\delta\nu(S;A)$ is independent of $A$.

Even given coherence, uncertainty averse preferences may exhibit the following property:

$$\delta\nu(F;A) > \delta\nu(G;A) \text{ and } \nu(A+F-G) < \nu(A). \tag{4.16}$$

In words, a decision-maker may decline to bet on the event $A+F-G$, in spite of its attractiveness in terms of 'mean likelihoods', *because of its ambiguity*. Of course, another possible explanation for the ranking $\nu(A+F-G) < \nu(A)$ is that $F$ has lower 'mean likelihood' than $G$, that is, $\delta\nu(F;A) < \delta\nu(G;A)$. But one can distinguish between these two explanations - only in the former case is it true that the decision-maker would accept any small portion of 'add $F$ and subtract $G$' (see (4.7)). Recall that sufficiently small portions would be attractive because they involve very little ambiguity. It is hoped that this observation can be developed into a partial explanation of the international portfolio diversification puzzle; that is, that the seemingly suboptimal diversification into foreign assets may be due to the greater ambiguity associated with their returns.

Turn to uncertainty aversion for general nonbinary acts, that is, in the sense of (2.5). Lemma 3.2 characterizes uncertainty aversion for preferences or utility functions, assuming a given supporting measure. Theorem 4.2 delivers the uniqueness of the supporting measure under the stated conditions. Combining these two results produces our most complete characterization of uncertainty aversion.

**Theorem 4.3.** *Let $U$ be a utility function, $x_1 \succ x_2$, $\nu(E) \equiv U(x_1Ex_2)$ and suppose that $\nu$ is eventwise differentiable. Suppose further that each $\delta\nu(\cdot,A)$,*

[27] Even given (4.11), the supporting measure at a given *single* $A$ is not unique, contrary to the intuition suggested by calculus. If the support property "$mF \leq mG \Longrightarrow \nu(A+F-G) \leq \nu A$", is satisfied by $m$, then it is also satisfied by any $m'$ satisfying $m(\cdot) \leq m'(\cdot)$ on $\Sigma \cap A^c$ and $m(\cdot) \geq m'(\cdot)$ on $\Sigma \cap A$. For example, let $m'$ be the conditional of $m$ given $A^c$.

*$A \in \mathcal{A}^0$, is countably additive, $\mathcal{A}^0$ is nonempty and $\nu$ is coherent. Then (1) implies (2), where:*
*(1) $U$ is uncertainty averse with countably additive supporting probability measure.*
*(2) $U$ satisfies conditions (i) and (ii) of Lemma 3.2 with measure $m$ given by*

$$m(\cdot) = \delta\nu(\cdot; A) \, / \, \delta\nu(S; A), \quad \textit{for any } A \textit{ in } \mathcal{A}^0. \tag{4.17}$$

*Conversely, if $\delta\nu(\cdot; A)$ is convex-ranged on $\mathcal{A}$ for any $A$ in $\mathcal{A}^0$, then (2) implies (1).*

The combination of Theorem 4.2 with Lemma 3.4 delivers a comparable result for CEU utility functions. Its statement is omitted in the interest of brevity.

## References


[1] F. Anscombe and R.J. Aumann, "A Definition of Subjective Probability," *Ann. Math. Stat.* 34 (1963), 199-205.

[2] S.H. Chew, E. Karni and Z. Safra, "Risk Aversion in the Theory of Expected Utility with Rank Dependent Probabilities," *J. Econ. Theory* 42 (1987), 370-381.

[3] S.H. Chew and M.H. Mao, "A Schur Concave Characterization of Risk Aversion for Non-Expected Utility Preferences," *J. Econ. Theory* 67 (1995), 402-435.

[4] J. Eichberger and D. Kelsey, "Uncertainty Aversion and Preference for Randomization," *J. Econ. Theory* 71 (1996), 31-43.

[5] L.G. Epstein and T. Wang, "Uncertainty, Risk-Neutral Measures and Security Price Booms and Crashes," *J. Econ. Theory* 67 (1995), 40-82.

[6] I. Gilboa, "Expected Utility with Purely Subjective Non-Additive Probabilities," *J. Math. Econ.* 16 (1987), 65-88.

[7] I. Gilboa and D. Schmeidler, "Maxmin Expected Utility With Nonunique Prior," *J. Math. Econ.* 18, (1989), 141-153.

[8] P.R. Halmos, *Measure Theory*, Springer-Verlag, 1974.

[9] E. Karni, "Risk Aversion for State-Dependent Utility Functions: Measurement and Applications," *Int. Ec. Rev.* 24 (1983), 637-647.

[10] D. M. Kreps, *Notes on the Theory of Choice*, Westview, 1988.

[11] D. Landers, "Connectedness Properties of the Range of Vector and Semimeasures," *Manuscripta Math.* 9 (1973), 105-112.

[12] M. Machina, "Expected Utility Analysis Without the Independence Axiom," *Econometrica* 50 (1982), 277-323.

[13] M. Machina, "Local Probabilistic Sophistication," mimeo, 1992.

[14] M. Machina and D. Schmeidler, "A More Robust Definition of Subjective Probability," *Econometrica* 60 (1992), 745-780.

[15] K.P. Rao and M.B. Rao, *Theory of Charges*, Academic Press, 1983.

[16] J. Rosenmuller, "*Some Properties of Convex Set Functions, Part II*," *Methods of Oper. Research* 17 (1972), 277-307.

[17] R. Sarin and P. Wakker, "A Simple Axiomatization of Nonadditive Expected Utility," *Econometrica* 60 (1992), 1255-1272.

[18] L. Savage, *The Foundations of Statistics*, John Wiley, 1954.

[19] D. Schmeidler, "Cores of Exact Games," *J. Math. Anal. and Appl.*, 40 (1972), 214-225.

[20] D. Schmeidler, "Integral Representation without Additivity," *Proc. Amer. Math. Soc.* 97 (1986), 255-261.

[21] D. Schmeidler, "Subjective Probability and Expected Utility Without Additivity," *Econometrica* 57 (1989), 571-587.

[22] U. Segal and A. Spivak, "'First-Order' versus 'Second-Order' Risk Aversion," *J. Econ. Theory* 51 (1990), 111-125.

[23] G. Shafer, "Allocations of Probability," *Ann. Prob.* 7 (1979), 827-839.

[24] H. Volkmer and H. Weber, "Der Wertebereich Atomloser Inhalte," *Archive der Mathematik* 40 (1983), 464-474.

[25] P. Wakker, "Preference Conditions for Convex and Concave Capacities in Choquet Expected Utility," mimeo, 1996.

[26] L. Wasserman, "'Bayes' Theorem for Choquet Capacities," *Ann. Stat.* 18 (1990), 1328-1339.

[27] M. Yaari, "Some Remarks on Measures of Risk Aversion and on Their Uses," *J. Econ. Theory* 1 (1969), 315-329.

[28] J. Zhang, "Subjective Ambiguity, Probability and Capacity," U. Toronto, mimeo, 1997.

# A. APPENDIX: Proofs

**Proof of Lemma 3.2:** $U^{ps}$ and $U$ agree on $\mathcal{F}^{ua}$. Therefore, (i) follows from (2.4) and the monotonicity assumed for $W$. That $U^{ps}$ supports $U$ implies by (2.5) that for all $e \in \mathcal{F}$ and $h \in \mathcal{F}^{ua}$,

$$W(\Psi_{m,e}) \leq W(\Psi_{m,h}) \implies U(e) \leq U(h).$$

This implies (ii).
For the converse, define $\succeq^{ps}$ as the order represented numerically by $U^{ps}$,

$$U^{ps}(e) = W(\Psi_{m,e}), \quad e \in \mathcal{F},$$

where $W : \Delta(\mathcal{X}) \longrightarrow \mathcal{R}^1$ is defined by

$$W(\Psi) = U(h) \text{ for any } h \in \mathcal{F}^{ua} \text{ satisfying } \Psi_{m,h} = \Psi.$$

Part (i) ensures that $W(\Psi)$ does not depend on the choice of $h$, making $W$ well-defined. The assumption added for $m$ ensures that this defines $W$ on all of $\Delta(\mathcal{X})$. Then $U^{ps}$ supports $U$. ■

**Proof of Lemma 3.4:** (b) $U^{ceu}$ and $U^{ps}$ must agree on $\mathcal{F}^{ua}$, implying that $\nu$ and $m$ are ordinally equivalent on $\mathcal{A}$. Because $\nu$ is convex-ranged and $\mathcal{A}$ is rich, $\nu(\mathcal{A}) = \nu(\Sigma) = [0,1]$. Conclude that $m(\mathcal{A}) = [0,1]$ also. Thus (3.5) is proven.

Lemma 3.2(ii) implies that for all acts $e$ and unambiguous acts $h$,

$$U^{ceu}(e) = \Sigma_{i=1}^{n-1}[u(x_i) - u(x_{i+1})]\,\nu\left(\cup_1^i e(x_j)\right) + u(x_n) \leq$$

$$\Sigma_{i=1}^{n-1}[\,u(x_i) - u(x_{i+1})]\,\nu\left(\cup_1^i h(x_j)\right) + u(x_n) = U^{ceu}(h)$$
$$= \Sigma_{i=1}^{n-1}[\,u(x_i) - u(x_{i+1})]\,g \circ m\left(\cup_1^i h(x_j)\right) + u(x_n)$$

if $m\,(e(x_j)) = m\,(h(x_j))$ for all $j$. Because this inequality obtains for all $u(x_1) > ... > u(x_n)$ and these utility levels can be varied over an open set containing some point $(u(x), ..., u(x))$, it follows that

$$g\left(m\left(\cup_1^i e(x_j)\right)\right) = g\left(m\left(\cup_1^i h(x_j)\right)\right) \geq \nu\left(\cup_1^i e(x_j)\right),$$

for all $e$ and $h$ as above. Given $E \in \Sigma$, let $e(x_1) = E$ and $e(x_2) = E^c$, $x_1 \succ x_2$. There exists unambiguous $A$ such that $mE = mA$. Let $h(x_1) = A$ and $h(x_2) = A^c$. Then $g\,(m(E)) \geq \nu(E)$ follows, proving (3.4).
The sufficiency portion (a) can be proven by suitably reversing the preceding argument. ■

**Proof of Theorem 3.1 :** The following lemma is of independent interest because of the special significance of bets as a subclass of all acts. Notation from Section 3.4 is used below.

**Lemma A.1.** *Suppose that $\mathcal{A}$ is rich, with outcomes $x^*$ and $x_*$ as in the definition of richness. Let $\nu(E) \equiv U(x^* E x_*)$. Then the conjunction of (3.6) and (3.7) implies that $\nu$ is ordinally equivalent to a probability measure on $\Sigma$ (or equivalently, $\nu$ satisfies (4.9)). A fortiori, the conclusion is valid if $\succeq$ is both uncertainty averse and uncertainty loving.*

**Proof.** Let $m$ and $q$ be the hypothesized supports. Their defining properties imply that

$$mF \leq mG \Longrightarrow qF \leq qG,$$

for all $A \in \mathcal{A}$, $F \subset A^c$ and $G \subset A$. But if this relation is applied to $A^c$ in place of $A$, noting that $A^c \in \mathcal{A}$, then the roles of $F$ and $G$ are reversed and one obtains

$$mF \geq mG \Longrightarrow qF \geq qG.$$

In other words,

$$mF \leq mG \iff qF \leq qG,$$

for all $A \in \mathcal{A}$, $F \subset A^c$ and $G \subset A$. Conclude from (3.6) and (3.7) that

$$mF \leq mG \iff \nu(A + F - G) \leq \nu A$$

for all $A \in \mathcal{A}$, $F \subset A^c$ and $G \subset A$; or equivalently, that for all $A \in \mathcal{A}$,

$$mE \leq mA \iff \nu E \leq \nu A.$$

In other words, every indifference curve for $\nu$ containing some unambiguous event is also an indifference curve for $m$. The stated hypothesis regarding $\mathcal{A}$ ensures that every indifference curve contains some unambiguous $A$ and therefore that $\nu$ and $m$ are ordinally equivalent on all of $\Sigma$. ■

Complete the proof of Theorem 3.1. Denote by $\succeq^{ps}$ and $\succeq^{ps}_*$ the probabilistically sophisticated preference orders supporting $\succeq$ in the sense of (2.5) and (2.6), respectively, and having underlying probability measures $m$ and $q$ defined on $\Sigma$. From the proof of the Lemma,

$$m \text{ and } q \text{ are ordinally equivalent on } \Sigma.$$

Claim: For each act $e$, there exists $h \in \mathcal{F}^{ua}$ such that

$$e \sim^{ps} h \text{ and } e \sim^{ps}_* h.$$

To see this, let $e = ((x_i, E_i)_{i=1}^n)$. By the richness of $\mathcal{A}$, there exist unambiguous events $H_i$, such that, $x^* H_i x_* \sim x^* E_i x_*$, $i = 1, ..., n$; or, in the notation of the Lemma, $\nu(H_i) = \nu(E_i)$ for all $i$. But since $\nu$ and $m$ are ordinally equivalent,

$$m(H_i) = m(E_i), \text{ all } i.$$

By the ordinal equivalence of $m$ and $q$,

$$q(H_i) = q(E_i), \text{ all } i.$$

Let $h = ((x_i, H_i)_{i=1}^n)$. The claim now follows immediately from the nature of probabilistic sophistication.

¿From (2.5), $\succeq$ and $\succeq^{ps}$ agree on $\mathcal{F}^{ua}$. Similarly, $\succeq$ and $\succeq^{ps}_*$ agree on $\mathcal{F}^{ua}$. Therefore, $\succeq^{ps}$ and $\succeq^{ps}_*$ agree there. From the claim, it follows that they agree on the complete set of acts $\mathcal{F}$. The support properties (2.5) and (2.6) thus imply that

$$h \succeq^{ps} e \iff h \succeq e, \quad \text{for all } h \in \mathcal{F}^{ua} \text{ and } e \in \mathcal{F}.$$

In particular, every indifference curve for $\succeq^{ps}$ containing some unambiguous act is also an indifference curve for $\succeq$. But the qualification can be dropped because of the claim. It follows that $\succeq$ and $\succeq^{ps}$ coincide on $\mathcal{F}$. ■

**Proof of Theorem 4.2:** (a) Let $m$ satisfy (3.6) at $A$. Show first that

$$mF \leq mG \Longrightarrow \delta\nu(F;A) \leq \delta\nu(G;A), \tag{A.1}$$

for all $F \subset A^c$ and $G \subset A$: $\delta\nu(F;A) > \delta\nu(G;A) \Longrightarrow$ (by (4.7)) $\exists\lambda$ such that $A + F^{j,\lambda} - G^{j,\lambda} \succ A$ for all $j \Longrightarrow mF^{j,\lambda} > mG^{j,\lambda}$ all $j \Longrightarrow mF > mG$.

Replace $A$ by $A^c$, in which case $F$ and $G$ reverse roles and deduce that

$$mF \geq mG \Longrightarrow \delta\nu(F;A^c) \geq \delta\nu(G;A^c)$$

or equivalently,

$$\delta\nu(F;A^c) \leq \delta\nu(G;A^c) \implies mF \leq mG. \tag{A.2}$$

Because $m$ is a support, this yields (4.12).

(b) Let $A \in \mathcal{A}$ satisfy

$$S \succ A \text{ and } S \succ A^c. \tag{A.3}$$

Claim 1: $\delta\nu(A^c;A) > 0$. If it equals zero, then $\delta\nu(A^c;A) = \delta\nu(\emptyset;A)$ implies, by (4.12), that $A + A^c \preceq A$, or $S \sim A$, contrary to (A.3).
Claim 2: $mA^c > 0$. If not, then $mS \leq mA = 1$ and (3.6) implies that $S \sim A$, contrary to (A.3).
Claim 3: $\delta\nu(A;A^c) > 0$ and $mA > 0$. Replace $A$ by $A^c$ above.
Claim 4: $\delta\nu(A^c;A^c) > 0$. If it equals zero, then $\delta\nu(A;A^c)\, mA^c = 0$ by (4.13), contradicting Claim 3.
Claim 5: For any $G \subset A$, $\delta\nu(G;A) = 0 \Longrightarrow mG = 0$: Let $F = A^c$. By Claim 1, $\delta\nu(F;A) > 0$. Therefore, (4.7) implies that $\forall\lambda_0\ \exists\lambda > \lambda_0$, $\delta\nu(F^{j,\lambda};A) > 0 = \delta\nu(G;A)$ for all $j$. By (A.1), $\forall\lambda_0\ \exists\lambda > \lambda_0$, $m(F^{j,\lambda}) > m(G)$ for all $j$, and thus also $mF > \Sigma_{j=1}^{n_\lambda}(mG)$. This implies $mG = 0$.
Claim 6: For any $F \subset A^c$, $mF = 0 \Longrightarrow \delta\nu(F;A) = 0$: $mF = 0 \Longrightarrow$ (by (A.1)) $\delta\nu(F;A) \leq \delta\nu(G;A)$ for all $G \subset A$. Claim 4 implies $\delta\nu(G;A) > 0$ if $G = A$. Therefore, $\delta\nu(\cdot;A)$ convex-ranged implies (Lemma B.1) that $\delta\nu(F;A) = 0$.
Claim 7: $m$ is convex-ranged: By Claim 5, $m$ is absolutely continuous with respect to $\delta\nu(\cdot;A)$ on $A$. The latter measure is convex-ranged. Therefore, $m$ has no atoms in $A$. Replace $A$ by $A^c$ and use the convex range of $\delta\nu(\cdot;A^c)$ to deduce in a similar fashion that $m$ has no atoms in $A^c$. Thus $m$ is non-atomic. Because it is also countably additive by hypothesis, conclude that it is convex-ranged [15, Theorem 5.1.6].

Turn to (4.13); (4.14) may be proven similarly. Define the measures $\mu$ and $p$ on $A^c \times A$ as follows:

$$\mu \,=\, m \qquad \delta\nu(\cdot\,; A), \;\; p \,=\, \delta\nu(\cdot\,; A) \qquad m.$$

Claims 5 and 6 prove that $p << \mu$. Denote by $h \equiv dp/d\mu$ the Radon-Nikodym density. (Countable additivity is used here.)
Claim 8: $\mu\,\{(s,t) \in A^c \times A :\, h(s,t) \,>\, 1\} \,=\, 0$: If not, then there exist $F_0 \subset A^c$ and $G_0 \subset A$, with $\mu(F_0 \times G_0) \,>\, 0$, such that

$$h \,>1 \quad \text{on } F_0 \times G_0.$$

Case 1: $mF_0 = mG_0$. Integration delivers $\int_{F_0} \int_{G_0} [h(s,t) \,-\, 1]\, d\mu \,>\, 0$, implying that

$$\delta\nu(F_0; A)\, mG_0 \,-\, mF_0\, \delta\nu(G_0; A) \,>\, 0.$$

Consequently, $mF_0 = mG_0$ and $\delta\nu(F_0; A) \,>\, \delta\nu(G_0; A)$, contradicting (A.1).
Case 2: $mF_0 \,<\, mG_0$. Because $m$ is convex-ranged (Claim 7), there exists $G_1 \subset G_0$ such that $mG_1 \,=\, mF_0$ and $\mu(F_0 \times G_1) \,>\, 0$. Thus the argument in Case 1 can be applied.
Case 3: $mF_0 \,>\, mG_0$. Similar to Case 2.

This proves Claim 8. Finally, for any $F \subset A^c$ and $G \subset A$, $\delta\nu(F; A)\,(mG) \,-\, (mF)\,\delta\nu(G; A) \,=\, \int_F \int_G\, (h-1)\, d\mu \,\leq\, 0$, proving (4.13).

(c) Though at first glance the proof may seem obvious given (4.15), some needed details are provided here. Let $A \,\in\, \mathcal{A}^0$. Multiply through (4.13) by $\delta\nu(G; A^c)$ to obtain that

$$\delta\nu(F; A)\,\delta\nu(G; A^c)\, mG \,\leq\, \delta\nu(G; A)\,\delta\nu(G; A^c)\, mF,$$

for all $F \subset A^c$ and $G \subset A$. Similarly, multiplying through (4.14) by $\delta\nu(G; A)$ yields

$$\delta\nu(G; A)\,\delta\nu(G; A^c)\, mF \,\leq\, \delta\nu(G; A)\,\delta\nu(F; A^c)\, mG,$$

for all such $F$ and $G$. Conclude from coherence that

$$\delta\nu(G; A)\,\delta\nu(G; A^c)\, mF \,=\, \delta\nu(G; A)\,\delta\nu(F; A^c)\, mG, \tag{A.4}$$

for all $F \subset A^c$ and $G \subset A$.

Take $G = A$ in (A.4) to deduce

$$\delta\nu(F; A^c) \;=\; \delta\nu(A; A^c)\, m(F) \,/\, m(A), \text{ for all } F \subset A^c. \tag{A.5}$$

Next take $F = A^c$ in (A.4). If $\delta\nu(G; A) > 0$, then

$$\delta\nu(G; A^c) \;=\; \delta\nu(A^c; A^c)\, m(G) \,/\, m(A^c), \text{ for all } G \subset A. \tag{A.6}$$

This equation is true also if $\delta\nu(G; A) = 0$, because then (4.12), with $F = A^c$, implies $\delta\nu(A^c; A)\, m(G) = 0$, which implies $mG = 0$ by Claim 1.

Substitute the expressions for $\delta\nu(F; A^c)$ and $\delta\nu(G; A^c)$ into (A.4) and set $F = A^c$ and $G = A$ to derive

$$\delta\nu(A^c; A^c) \,/\, m(A^c) \;=\; \delta\nu(A; A^c) \,/\, m(A) \;\equiv\; \alpha(A) \;> 0.$$

Thus

$$\delta\nu(\cdot\,; A^c) \;=\; \begin{cases} \alpha(A)\, m(\cdot) & \text{on } \Sigma \cap A^c \\ \alpha(A)\, m(\cdot) & \text{on } \Sigma \cap A. \end{cases}$$

By additivity, it follows that $\delta\nu(\cdot\,; A^c) \;=\; \alpha(A)\, m(\cdot)$ on all of $\Sigma$. Thus $\delta\nu(\cdot\,; A) = \kappa(A)\, \alpha(A)\, m(\cdot)$, completing the proof. ■

## B. APPENDIX: Miscellaneous

The following implications of convex range for a measure on $\Sigma^{\mathcal{X}}$ are used often and are collected here for the convenience of the reader. See [15, pp.142-3] for comparable results for measures on an algebra. In [15], property (b) is referred to as strong continuity.

**Lemma B.1.** *Let $\mu$ be a measure on $\Sigma^{\mathcal{X}}$. Then the following statements are equivalent:*

*(a) $\mu$ is convex-ranged.*

*(b) For any act $f$, with corresponding net of all finite partitions $\{f^{j,\lambda}\}_{j=1}^{n_\lambda}$, and for any $\epsilon > 0$, there exists $\lambda_0$ such that*

$$\lambda > \lambda_0 \implies \mid \mu \mid (f^{j,\lambda}) \;<\; \epsilon, \quad \text{for } j = 1, ..., n_\lambda \;.$$

*(c) For any acts $f$, $g$ and $h \equiv f + g$, if $\mu(f) \;>\; \mu(g)$, then there exists a partition $\{h^{j,\lambda}\}_{j=1}^{n_\lambda}$ of $h$, such that $\mu(h^{j,\lambda}) \;<\; \epsilon$ and $\mu\left(h^{j,\lambda} \cap f\right) \;>\; \mu\left(h^{j,\lambda} \cap g\right)$, $j = 1, ..., n_\lambda$.*

Next I describe a Chain Rule for eventwise differentiability.

**Theorem B.2.** *Let $\Phi : \Sigma^{\mathcal{X}} \longrightarrow \mathcal{R}^1$ be eventwise differentiable at $e$ and $\varphi : \Phi(\Sigma^{\mathcal{X}}) \longrightarrow \mathcal{R}^1$ be strictly increasing and continuously differentiable. Then $\varphi \circ \Phi$ is eventwise differentiable at $e$ and*

$$\delta(\varphi \circ \Phi)(\cdot; e) = \varphi'(\Phi(e))\, \delta\Phi(\cdot; e)$$

**Proof.** Consider the sum whose convergence defines the eventwise derivative of $\varphi \circ \Phi$. By the Mean Value Theorem,

$$\varphi \circ \Phi\left(e + f^{j,\lambda} - g^{j,\lambda}\right) - \varphi \circ \Phi(e) = \varphi'(z^{j,\lambda})\left[\Phi\left(e + f^{j,\lambda} - g^{j,\lambda}\right) - \Phi(e)\right]$$

for suitable real numbers $z^{j,\lambda}$. Therefore, it suffices to prove that

$$\Sigma_{j=1}^{n_\lambda} \mid \Phi\left(e + f^{j,\lambda} - g^{j,\lambda}\right) - \Phi(e) \mid\mid \varphi'(z^{j,\lambda}) - \varphi'(\Phi(e))) \mid \longrightarrow_\lambda 0.$$

By the continuity of $\varphi'$, the second term converges to zero uniformly in $j$. Eventwise differentiability of $\Phi$ implies that given $\epsilon$, there exists $\lambda_0$ such that $\lambda > \lambda_0$ $\Longrightarrow$

$$\Sigma_{j=1}^{n_\lambda} \mid \Phi\left(e + f^{j,\lambda} - g^{j,\lambda}\right) - \Phi(e) \mid \leq \epsilon + \Sigma_{j=1}^{n_\lambda} \mid \delta\Phi\left(f^{j,\lambda}; e\right) - \delta\Phi\left(g^{j,\lambda}; e\right) \mid$$

$$\leq \epsilon + \Sigma_{j=1}^{n_\lambda}[\mid \delta\Phi\left(f^{j,\lambda}; e\right) \mid + \mid \delta\Phi\left(g^{j,\lambda}; e\right) \mid] \leq K,$$

for some $K < \infty$ that is independent of $\lambda$, $f$ and $g$, as provided by the boundedness of the measure $\delta\Phi(\cdot; e)$. ■

# C. APPENDIX: $\mu$-Differentiability

For the reasons given in the text, a strengthening of eventwise differentiability is described here. Machina [13] introduces a very similar notion. But because it is new and still unfamiliar and because our formulation is somewhat different and arguably more transparent, a detailed description seems in order.[28]

[28] As mentioned earlier, after a version of this paper was completed, I learned of a revision of [13], dated 1997, in which Machina provides a formulation very similar to that provided in this subsection. The connection with the more general 'partitions-based' notion of eventwise differentiability, inspired by [16], is not observed by Machina.

To proceed, adopt as another primitive a non-negative, bounded and convex-ranged measure $\mu$ on $\Sigma^{\mathcal{X}}$. Assume that it is bounded and convex-ranged. This measure serves the 'technical role' of determining the distance between acts. To be precise, if we identify $e$ and $e'$ whenever $\mu(e \Delta e') = 0$, then

$$d(e, e') \;=\; \mu(e \,\Delta\, e') \tag{C.1}$$

defines a metric on $\Sigma^{\mathcal{X}}$; the assumption of convex range renders the metric space path-connected (by [24], see also [11, Lemma 4]).

One way in which such a measure can arise is from a convex-ranged probability measure $\mu_0$ on $\Sigma$. Given $\mu_0$, define $\mu$ by

$$\mu(e) = \Sigma_x \, \mu_0 \left( e(x) \right). \tag{C.2}$$

Once again let $\Phi : \Sigma^{\mathcal{X}} \longrightarrow \mathcal{R}^1$. Because acts $e$ and $e'$ are identified when $\mu(e \Delta e') = 0$, $\Phi$ is assumed to satisfy the condition

$$\mu(e \,\Delta\, e') = 0 \implies \Phi(e \cup f) \;=\; \Phi(e' \cup f), \text{ for all } f. \tag{C.3}$$

In particular, acts of $\mu$-measure 0 are assumed to be 'null' with respect to every function $\Phi$.

**Definition C.1.** *$\Phi$ is $\mu$-**differentiable** at $e \in \Sigma^{\mathcal{X}}$ if there exists a bounded and convex-ranged measure $\delta\Phi(\cdot; e)$ on $\Sigma^{\mathcal{X}}$, such that for all $f \subset e^c$ and $g \subset e$,*

$$\mid \Phi \left( e + f - g \right) - \Phi \left( e \right) - \; \delta\Phi(f; e) + \delta\Phi(g; e) \mid \, / \, \mu(f + g) \quad \longrightarrow \; 0 \tag{C.4}$$

*as $\mu(f + g) \;\longrightarrow\; 0$.*

The presence of a 'difference quotient' makes the appearance of this definition more standard and permits a standard interpretation. Think in particular of the case ($\mid \mathcal{X} \mid = 1$) where the domain of $\Phi$ is $\Sigma$ (see the comments following (4.5)).

It is easy to see that $\delta\Phi(\cdot; e)$ is absolutely continuous with respect to $\mu$ for each $e$. (Use additivity of the derivative and (C.3).)

Fix an outcome $x$ and suppose that $\mu_0$ and $\delta_x\Phi(\cdot; e)$ (the partial derivative or marginal in the sense of (4.3)) are both countably additive and that $\Sigma$ is a $\sigma$-algebra. Then there exists a Radon-Nikodym derivative $h_x(\cdot; e)$ in $L^1(\mu)$, such that

$$\delta_x\Phi(B; e) \;=\; \int_B h_x(\cdot; e) \, d\mu(\cdot), \;\text{ for each } B \subset S. \tag{C.5}$$

There is a natural sense in which $h_x(\cdot; e)$ can be interpreted as the *gradient* of $\Phi$ at $e$ in the $x$-component.

We have not distinguished notationally between eventwise and $\mu$-derivatives because they coincide whenever they both exist.

**Lemma C.2.** *If $\Phi$ is $\mu$-differentiable at some $e$ in $\Sigma^{\mathcal{X}}$, then $\Phi$ is also eventwise differentiable at $e$ and the two derivatives coincide.*

**Proof.** Let $\delta\Phi(\cdot; e)$ be the $\mu$-derivative at $e$, $f \subset e^c$ and $g \subset e$. Given $\epsilon > 0$, there exists (by $\mu$-differentiability) $\epsilon' > 0$ such that

$$\mid \Phi(e + f' - g') \,-\, \Phi(e) \,-\, \delta\Phi(f'; e) \,+\, \delta\Phi(g'; e) \mid < \epsilon\; \mu(f' + g'), \tag{C.6}$$

if $\mu(f' + g') \,<\, \epsilon'$. By Lemma B.1 applied to the convex-ranged $\mu$, there exists $\lambda_0$ such that

$$\mu(f^{j.\lambda} + g^{j,\lambda}) \,<\, \epsilon', \quad \text{for all } \lambda > \lambda_0.$$

Therefore, one can apply (C.6) to the acts $(f', g') \,=\, (f^{j,\lambda}, g^{j,\lambda})$. Deduce that

$$\Sigma_{j=1}^{n_\lambda} \mid \Phi\left(e + f^{j,\lambda} - g^{j,\lambda}\right) - \Phi\left(e\right) - \,\delta\Phi(f^{j,\lambda}; e) + \delta\Phi(g^{j,\lambda}; e) \mid <$$

$$\epsilon\; \Sigma_{j=1}^{n_\lambda}\, \mu(f^{j,\lambda} + g^{j,\lambda}) \,=\, \epsilon\, \mu(f + g) \,<\epsilon \text{ sup } \mu(\cdot).$$

■

A consequence is that the $\mu$-derivative of $\Phi$ is independent of $\mu$; that is, if $\mu_1$ and $\mu_2$ are two measures satisfying the conditions in the lemma, then they imply the identical derivatives for $\Phi$. This follows from the uniqueness of the eventwise derivative noted earlier. Such invariance is important in light of the exogenous and ad hoc nature of $\mu$. This result is evident *because of* the deeper perspective afforded by the notion of eventwise differentiability and reflects its superiority over the notion of $\mu$-differentiability.

Finally, under a slight strengthening of $\mu$-differentiability, one can 'integrate' back to $\Phi$ from its derivatives. That is, a form of the Fundamental Theorem of Calculus is valid. (It remains to determine if there exists a counterpart result for eventwise differentiability.)

**Lemma C.3.** *Let $\Phi$ be $\mu$-differentiable and suppose that the convergence in (C.4) is uniform in $e$. For every $\epsilon > 0$, $f \subset e^c$ and $g \subset e$, there exist finite partitions $f = \Sigma f^j$ and $g = \Sigma g^j$ such that $\epsilon >$*

$$\mid \Phi(e{+}f{-}g){-}\Phi(e){-}\Sigma_i\delta\Phi(f^i; e{+}\mathcal{F}^{i-1}{-}\mathcal{G}^{i-1}){+}\Sigma_i\delta\Phi(g^i; e{+}\mathcal{F}^{i-1}{-}\mathcal{G}^{i-1}) \mid \tag{C.7}$$

*where $\mathcal{F}^i = \Sigma_{j=1}^i f^j$ and $\mathcal{G}^i = \Sigma_{j=1}^i g^j$.*

**Proof.** $\mu$-differentiability and the indicated uniform convergence imply that

$$\begin{aligned}&| \ \Phi(e+\mathcal{F}^{i-1}-\mathcal{G}^{i-1}+f^i-g^i)-\Phi(e+\mathcal{F}^{i-1}-\mathcal{G}^{i-1}) \\ &\qquad -\delta\Phi(f^i;e+\mathcal{F}^{i-1}-\mathcal{G}^{i-1}) \ + \ \delta\Phi(g^i;e+\mathcal{F}^{i-1}-\mathcal{G}^{i-1}) \ | \quad < \ \epsilon \ \mu(f^i+g^i),\end{aligned}$$

for any partitions $\{f^j\}$ and $\{g^j\}$ such that $\mu(f^j+g^j)$ is sufficiently small for all $j$. But the latter can be ensured by taking the partitions $\{f^{j,\lambda}\}$ and $\{g^{j,\lambda}\}$ for $\lambda$ sufficiently large. The convex range assumption for $\mu$ enters here; use Lemma B.1. Therefore, the triangle inequality delivers $| \ \Phi(e+f-g)-\Phi(e)-\Sigma\delta\Phi(f^i;e+\mathcal{F}^{i-1}-\mathcal{G}^{i-1})+\Sigma\delta\Phi(g^i;e+\mathcal{F}^{i-1}-\mathcal{G}^{i-1}) \ | \ \leq \ \epsilon\Sigma_i\mu(f^i+g^i) = \ \epsilon\mu(f+g)\,.$ ■