THE REVIEW OF **ECONOMIC STUDIES**

Review of Economic Studies (2010) **77**, 339–371 0034-6527/09/00411011$02.00

doi: 10.1111/j.1467-937X.2009.00564.x

Quantile Maximization in Decision Theory*

MARZENA ROSTEK

University of Wisconsin-Madison

First version received August 2007; *final version accepted January* 2009 *(Eds.)*

This paper introduces a model of preferences, in which, given beliefs about uncertain outcomes, an individual evaluates an action by a quantile of the induced distribution. The choice rule of Quantile Maximization unifies maxmin and maxmax as maximizing the lowest and the highest quantiles of beliefs distributions, respectively, and offers a family of less extreme preferences. Taking preferences over acts as a primitive, we axiomatize Quantile Maximization in a Savage setting. Our axiomatization also provides a novel derivation of subjective beliefs, which demonstrates that neither the monotonicity nor the continuity conditions assumed in the literature are essential for probabilistic sophistication. We characterize preferences of quantile maximizers towards downside risk. We discuss how the distinct properties of the model, robustness and ordinality, can be useful in studying choice behaviour for categorical variables and in economic policy design. We also offer applications to poll design and insurance problems.

1. INTRODUCTION

This paper examines the choice behaviour of an individual who, when selecting among uncertain alternatives, chooses the one with the highest quantile of the utility distribution. For example, she might be maximizing median utility, as opposed to mean utility, as she would if she were an Expected Utility maximizer. More generally, she might be comparing alternatives through some other quantile that corresponds to any given number between 0 and 1.

Although largely ignored in decision theory, quantile-based decision criteria have long been influential in economic policy design. Most prominently, quantiles have been used in scenario-based analysis and as order statistics. Quantile-based decision rules have also been common in resource allocation programs and in the design of treatment effects, as they permit distributional targeting and explicit analysis of distributional consequences. Policy decision

*I am grateful to Itzhak Gilboa, Stephen Morris and Ben Polak for their numerous helpful suggestions throughout the project. I also thank Dino Gerardi for his advice; Scott Boorman, Don Brown, Markus Brunnermeier, Chris Chambers, Larry Epstein, Simon Grant, Philippe Jehiel, Barry O'Neill, Jacob Sagi, Bernard Salanie, Barry Schachter, Peter Wakker, Jan Werner, John Weymark, and especially Marek Weretka for helpful discussions. The Editor, Enrique Sentana, and two anonymous referees provided invaluable suggestions. I have also benefitted from comments from seminar participants at Caltech, Columbia, Duke FUQUA, Hebrew University of Jerusalem, Minnesota, Northwestern, Oxford, Rochester, Wisconsin-Madison, Yale; and conference participants at the RUD 2006, the FUR XII 2006, the Midwest Economic Theory Meeting 2005, the XVI Conference on Game Theory in Stony Brook 2005, and the X SMYE 2005. Any errors are mine.

making and forecasting have further benefitted from techniques of robust estimation (Least Absolute Deviations method; Koenker and Bassett, 1978), where quantiles have provided an alternative to classical mean-based estimators, which are sensitive to large errors.

To which properties of quantile-based decision criteria can one attribute their normative appeal over those based on moments, such as the mean? Two key characteristics of quantiles–robustness and ordinality–have proven attractive among practitioners. Quantile-based techniques are robust to fat tails, which are often encountered in practice and offer predictions not driven by outliers. More pragmatically, unlike tools that are based on the mean, quantiles do not require that there exist moments of any order, which is a problem, for example, in non-life insurance, finance and income studies. In terms of ordinality, quantiles have the advantage of not requiring any parametric assumptions about utilities.

To our knowledge, apart from the work of Manski (1988) and a recent contribution by Chambers (2007) on which we elaborate below, quantiles have not been studied in choice theory. There are two famous exceptions: *maxmin* and *maxmax*. Decision makers who select an alternative that offers the highest minimal or maximal payoff can be viewed as maximizing the lowest or the highest quantile, respectively.[1] Maxmax, and especially maxmin, have been applied in game theory, robust control, individual and social choice, bargaining, voting and other areas of economics. These criteria have, however, been commonly criticized for basing choice on what may be extreme and unlikely outcomes. Indeed, maxmin agents would not invest, would not drive, and so on. Surprisingly, there appears to be no model that captures more moderate preferences[2] while preserving the qualitative properties of maxmin and maxmax that the Expected Utility does not exhibit, such as ordinality and robustness. Compared to the extreme maxmin and maxmax, the family of quantile-based criteria incorporates richer information (in terms of outcome and probability) from the uncertain alternatives faced by an agent, thus additionally gaining the attribute of robustness.

One important class of decision problems with which *Quantile Maximization* (but not Expected Utility) can formally deal is that where the alternatives involve categorical, sometimes referred to as qualitative, variables. Many economic and social variables are categorical (e.g., careers, the A–F grading scheme, qualities in online ratings). What makes modelling challenging in this case is that the outcomes are informative only up to their rank–and not distance. O'Neill (2001) surveys the difficulties involved in formalizing choice behaviour for categorical variables.) Stated somewhat informally, applying the Expected Utility would require taking an integral and, hence, assigning real (utility) numbers to categorical values. Since the possible (and, in this setting, arbitrary) utility assignments admit functional forms as diverse as concave and convex, the prediction about choice behaviour would then derive from the assignment rather than ordinal rankings themselves. By offering a tractable technique that

1. Maxmin was formally analysed by Roy (1952, *safety first* rule), Milnor (1954), Rawls (1971, *justice as fairness* theory), Maskin (1979), Barbera and Jackson (1988), Cohen (1992, *security level*), and Segal and Sobel (2002). Formalizations of maxmax include Cohen (1992, *potential level*), Segal and Sobel (2002), and Yildiz (2007, *wishful thinking*). The maxmin in uncertain settings should be distinguished from that in ambiguous environments studied by Gilboa and Schmeidler (1989), where maxmin is taken over expected utilities, each being evaluated by a different prior.

2. Perhaps the closest concept is α-*maxmin*, defined as a convex combination of the minimal and the maximal payoffs with fixed weights α and $1-\alpha$. (The rule was introduced by Hurwicz (1951), and by Arrow and Hurwicz (1972) in a context of "complete ignorance" and subsequently applied also to decision problems under uncertainty.) The α-maxmin rule is, however, not ordinal. Furthermore, like maxmin and maxmax, α-maxmin loses the entire information contained in the prior except for its support. Maxmin and maxmax are useful decision criteria in settings where an analyst has no probabilistic information about events. Nonetheless, such information is often available, especially about part of the domain, which is all that quantiles require.

respects these rankings, Quantile Maximization complements the existing models of choice under uncertainty, virtually all of which imply the use of cardinal properties of utility functions over outcomes. For the same reason, Quantile Maximization does not capture preference for diversification and should not be used in decision problems in which the outcome spread is a concern (see Section 6.1).

Alternatively, even if cardinal information about outcomes is available to an agent, she might want to make a choice that is robust to her own utility. Quantile Maximization thus contributes to the literature on robustifying economic and policy design, which has primarily focused on relaxing the assumption that decision makers know–or act as if they knew–the true probability distribution (e.g., Hansen and Sargent (2007) applying the model of Gilboa and Schmeidler (1989), Klibanoff, Marinacci and Mukerji (2005), and Maccheroni, Marinacci and Rustichini (2006)). The quantile model permits a less explored robustness test that involves relaxing the assumption that decision makers have cardinal as well as ordinal rankings of outcomes, or that cardinal, parametric assumptions about utilities do affect decisions. Likewise, the model can be applied by decision makers who explicitly seek a rule that can apply to a population with heterogeneous preferences and are willing to assume only that people prefer more to less (e.g., distributional targeting in resource allocations, expert recommendations).

1.1. *Results*

This paper formalizes the concept of Quantile Maximization in choice-theoretic language to understand its implications for decision making and provide a foundation both for its practice and for its applications in economic theory. The central theoretical contribution of the paper is to provide the complete behavioural characterization of an agent who, when choosing between uncertain alternatives, evaluates each alternative by the τ-th quantile of the implied distributions and selects the one with the highest quantile payoff. Thus, a decision maker is characterized by a scalar $\tau \in [0, 1]$, subjective beliefs over events π, and a rank order over outcomes. Taking preferences over Savage-style acts (maps from states to outcomes) as a primitive, we jointly axiomatize Quantile Maximization and subjective probabilities with five intuitive conditions. We next describe the main results.

First, the axiomatization is revealing about how quantile maximizers code information about uncertain alternatives. As our central axiom asserts, Quantile Maximization implies that, for any act (uncertain alternative), there exists an event, called a *pivotal event*, such that exchanging outcomes outside of this event in a way that preserves their rank with respect to the outcome on the pivotal event does not affect preferences over acts. The agent thus assesses the realization of the uncertain alternative via a "typical" consequence (scenario); for the purpose of making the choice, it suffices that she categorizes the remaining outcomes as worse or better–with Quantile Maximization, the choice will be robust to the actual realization of these consequences. Crucially, which event the agent considers typical (pivotal) is determined by her preferences. After we derive beliefs, we show that there is a formal sense in which the selection of the scenario considered typical by the agent is governed by her attitude toward downside risk. The model admits an elegant characterization of risk attitudes: τ itself provides a measure and a complete ranking of risk attitudes, ranging from extreme downside risk aversion ($\tau = 0$) to extreme downside risk tolerance ($\tau = 1$).

Furthermore, the axiomatization uncovers an important difference in how beliefs enter the preferences of the extreme and the intermediate quantiles. For all values of τ strictly between 0 and 1, the derived probability measure that represents subjective beliefs is unique (and also convex-ranged and finitely additive). For the extreme values of τ, equal to 0 or 1, we derive a set of non-atomic measures that are monotone but not necessarily finitely additive (capacities). This is intuitive: choices of 0- or 1-maximizers do not depend on beliefs, just on their support; hence, these are consistent with any measure that assigns strictly positive (and less than one) values to the same outcomes. As a by-product, our results axiomatize maxmin and maxmax under uncertainty. While these two rules have been studied extensively, to the best of our knowledge, we are the first to derive and characterize the implied beliefs of maxmin or maxmax agents.

Apart from providing the model with an identification result regarding the uniqueness of τ and beliefs, the axiomatization also yields a set of testable predictions for the model (and, notably, for the additivity of beliefs), which are helpful in understanding the relation between Quantile Maximization and other axiomatic models of preferences.

Perhaps the key implication of the characterization of beliefs is that, despite comparing acts through a single quantile, quantile maximizers are probabilistically sophisticated (in the sense of Machina and Schmeidler, 1992) in that they behave as if they had a probability measure in mind (which is also unique for $\tau \in (0, 1)$). The result and the novel technique we have developed to derive a probability measure from preferences are of independent interest, since an analyst can derive the beliefs of a decision maker without having to deal with a numerical representation of preferences. One direct use of the technique (which is constructive) might be for quantile-based expectations in survey research (see, e.g., the work of Manski and co-authors), as it permits separating the implications of the elicited predictions from those of the decision rule itself. We hope that the derived testable conditions and, in particular, the implied test for additivity of beliefs will direct more economic attention to quantile-based decision criteria. The technique and characterization of beliefs also contribute to the literature on probabilistic sophistication. The goal of that research is to understand when choices of an agent are consistent with her having beliefs that conform to a probability measure without restricting the actual decision rule to an Expected-Utility or other functional form. Yet, the existing alternatives to Savage's derivation of beliefs (Machina and Schmeidler, 1992; Grant, 1995) rely on conditions on preferences that are too restrictive for the quantile model; for instance, a median maximizer would not be probabilistically sophisticated according to these characterizations.[3] In addition, unlike these results, our technique can be used to derive beliefs for lexicographic agents and in decision problems that involve categorical variables. Moreover, our results demonstrate that the commonly used formal definitions of probabilistic sophistication admit preferences that intuitively should not be regarded as being based on probabilistic beliefs.

The main technical contribution of this paper is a new derivation of subjective probabilities that represent agents' beliefs. In axiomatizing expectation-based models, the derivation of beliefs is typically a quick step, as the proof of Savage (1954) can be readily applied. We could not, however, directly use either Savage's or other arguments in the literature: derivation of probabilities involves defining a likelihood relation–a binary relation on events–induced from the preference relation on acts. In the quantile model, the commonly used likelihood relation[4] generates only two equivalence classes on the entire collection of events: all events are judged to be either equally likely to the null set or to the whole state space. Hence, even if there is a probability measure that represents the beliefs of τ-maximizers, the relation would not allow an analyst to recover it from data as rich as recording all choices in all decision problems.

3. Her choices violate all axioms in Machina and Schmeidler, except P1 (Ordering), P4 (Weak Comparative Likelihood), and P5 (Non-degeneracy) and all axioms in Grant except P1 and P5.

4. According to the commonly used definition (employed by Savage and earlier by Ramsey and de Finetti), event E is assessed as *more likely than* event F if, for any pair of outcomes x and y, where x is strictly preferred to y, an individual strictly prefers betting on x when E occurs than when F occurs.

1.2. *Other related literature*

Ordinal representations of preferences have been advocated by Börgers (1993, *pure-strategy dominance*), Chambers (2007, 2009), and earlier by Manski (1988, *quantile utility* and *utility mass* models). Roberts (1980) introduced the idea of *rank-dictatorship* in social decision making, which thus does not involve probabilities. Manski was the first to draw attention to the decision-theoretic attributes of Quantile Maximization and examine risk preferences of quantile maximizers. The result by Chambers (2009) provided a compelling motivation for studying quantile decision criteria. Suppose one is interested in applying an ordinal decision criterion that should not violate weak first-order stochastic dominance. Chambers (2009) showed that, in the class of all such ordinal and weakly monotonic rules, quantiles are an essentially unique decision criterion. Working with the real-valued bounded measurable functions, Chambers (2007) provided several results that illuminated the relation among the functionals satisfying the two conditions. Interestingly, thanks to mild informational requirements, the ordinal and, in particular, fully qualitative (i.e., in terms of outcomes and probabilities) approach to modelling choice has become increasingly popular in the area of artificial intelligence over the past decade (see, e.g., Boutilier (1994), Dubois *et al.* (2000, 2002) and references therein). The primary difference here is that Quantile Maximization is based on probabilistic reasoning.

1.3. *Structure of the paper*

Section 2 presents the model of Quantile Maximization. Section 3 states our axioms, and Section 4 provides the main results, namely, the representation theorem and a characterization of probabilistic sophistication. Section 5 outlines the key steps in the proof. Section 6 examines the properties of risk preferences and discusses the applications of the model. Section 7 offers concluding remarks. All proofs appear in the Appendices.

2. THE QUANTILE MAXIMIZATION MODEL

Let $\mathcal{S}$ denote a set of states of the world $s \in \mathcal{S}$, and let $\mathcal{X}$ be an arbitrary set of *outcomes* $x, y \in \mathcal{X}$. An individual chooses among simple *acts* $f : \mathcal{S} \to \mathcal{X}$,[5] which map from states to outcomes. $\mathcal{F}$ is the set of all such acts. The set of events $\mathcal{E} = 2^{\mathcal{S}}$, with typical elements E and F, is the set of all subsets of $\mathcal{S}$. A collection $\{\mathcal{S}, \mathcal{X}, \mathcal{E}, \mathcal{F}\}$ defines the Savagean model of purely subjective uncertainty. An individual is characterized by a binary relation over acts in $\mathcal{F}$, $\succ$, which will be defined as a strict preference relation and taken to be the primitive of the model. Let $\succ_x$ denote the preference relation over certain outcomes, $\mathcal{X}$, obtained as a restriction of $\succ$ to constant acts. Say that event E is *null* if, for any two acts, f and g, which differ only on E, we have $f \sim g$. Let $\mathcal{P}_0(\mathcal{X})$ denote the set of *simple* probability distributions over the outcomes (lotteries). Finally, δ_x stands for the degenerate lottery $P = (x, 1)$.

Let π denote a probability measure on $\mathcal{E}$, and let u be a utility over outcomes $u : \mathcal{X} \to \mathbb{R}$. For each act, π induces a probability distribution over payoffs, referred to as a *lottery*. For an act f, Π_f denotes the induced cumulative probability distribution of utility $\Pi_f(z) = \pi[s \in \mathcal{S} | u(f(s)) \leq z,\ z \in \mathbb{R}]$. Then, for a fixed act f and $\tau \in (0, 1]$, the τ-th *quantile* of the distribution of the random variable $u(x)$ is a (generalized) inverse of the cumulative distribution at τ. The generalized inverse is defined as the smallest value z, such that the probability that a random variable will be less than z is not smaller than τ:

$$Q^{\tau}(\Pi_f) = \inf\{z \in \mathbb{R} \,|\, \pi[u(f(s)) \leq z] \geq \tau\}, \tag{1}$$

5. An act f is *simple* if its outcome set $f(\mathcal{S}) = \{f(s) | s \in \mathcal{S}\}$ is finite.

while for $\tau = 0$, the quantile is defined as[6]

$$Q^0(\Pi_f) = \sup\{z \in \mathbb{R} \mid \pi[u(f(s)) \leq z] \leq 0\}. \tag{2}$$

Definition 1. A decision maker is said to be a *τ-quantile maximizer* if there exists a unique $\tau \in [0, 1]$, a probability measure π on $\mathcal{E}$, and utility u over outcomes in $\mathcal{X}$, such that for all $f, g \in \mathcal{F}$,

$$f \succ g \Leftrightarrow Q^\tau(\Pi_f) > Q^\tau(\Pi_g). \tag{3}$$

By analogy with the Expected Utility, where the mean is a single statistic via which a distribution is evaluated, when choosing among lotteries a τ-maximizer assesses the value of each lottery by the τ-th quantile realization. We will show that, although generally a correspondence, generically in payoffs the set of optimal choices is a singleton.

The quantile model nests two choice rules famous in the literature of choice under risk: maxmin and maxmax. A decision maker choosing according to *maxmin* picks the act with the highest minimal payoff:

$$f \succ g \Leftrightarrow \min_{\{x \in f(\mathcal{S}) | \pi(x) > 0\}} u(x) > \min_{\{x \in g(\mathcal{S}) | \pi(x) > 0\}} u(x). \tag{4}$$

Maxmax dictates selection of the act with the highest maximal payoff:

$$f \succ g \Leftrightarrow \max_{\{x \in f(\mathcal{S}) | \pi(x) > 0\}} u(x) > \max_{\{x \in g(\mathcal{S}) | \pi(x) > 0\}} u(x). \tag{5}$$

That the maxmin and maxmax decision makers are, respectively, the 0- and 1-quantile maximizers follows from $Q^0(\Pi_f) = \min_{\{x \in f(\mathcal{S}) | \pi(x) > 0\}} u(x)$ and $Q^1(\Pi_f) = \max_{\{x \in f(\mathcal{S}) | \pi(x) > 0\}} u(x)$. Quantile Maximization can thus be viewed as a generalization of those extreme choice rules to any intermediate quantile. While the focus of the paper will be on simple acts, in the example below, it is convenient to illustrate the relation between maxmin, maxmax and Quantile Maximization using infinite-outcome acts.

Example. *Consider an individual who is facing a choice between two acts, f and g. Let π be the probability measure that represents the agent's beliefs. The cdf's induced by acts f and g and the measure π are plotted in Figure 1. The 0-quantile maximizer would choose f, the 1-quantile maximizer would be indifferent, and the median ($\tau = 0.5$) maximizer would prefer g.*

3. AXIOMS

Consider the following five axioms on $\succ$. The numbering is Savage's, the names of his conditions are adapted from Machina and Schmeidler (1992), and the superscript "^{Q}" (for "Quantile") is added to new axioms.

Axiom P1 (Ordering). *Relation $\succ$ is a weak order.*

This standard condition defines $\succ$ as a preference relation. To state the next axiom, for a fixed act $f \in \mathcal{F}$ and event E, such that $f^{-1}(x) = E$ for some $x \in f(\mathcal{S})$, we define the unions

6. Clearly, definitions (1) and (2) are conceptually the same: The quantile operator picks the smallest value z (from the support of a lottery induced by act f and probability π, given utility u), such that the probability of a realization that yields utility less than z is at least τ. The separate formulation for $\tau = 0$ merely ensures that the inverse operator maps to an element from the support of Π_f (given u).

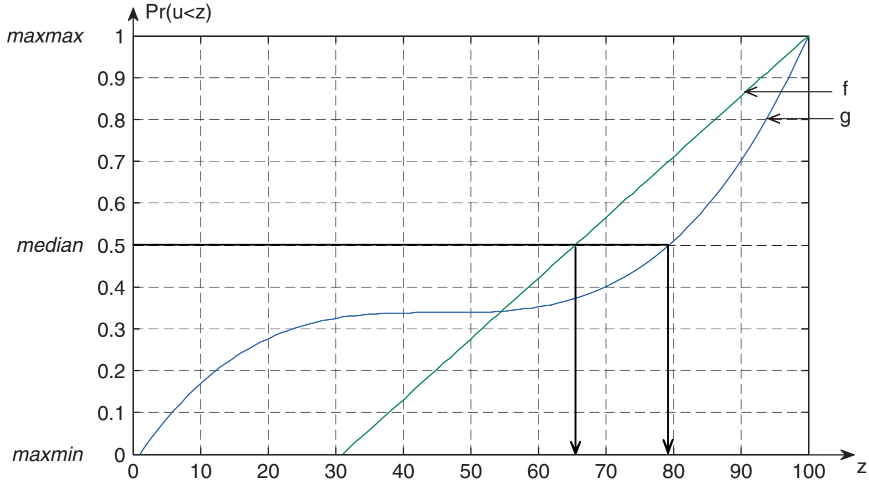

FIGURE 1

Distributions induced by acts in the example

of events which by f are assigned outcomes strictly more and strictly less preferred to x, respectively:[7]

$$E^{+}_{f,x} = \{s \in \mathcal{S} | f(s) \succ x\}, \tag{6}$$

$$E^{-}_{f,x} = \{s \in \mathcal{S} | f(s) \prec x\}. \tag{7}$$

Since the acts are finite-ranged, every act induces a natural partition of the state space $\mathcal{S}$, which is the coarsest partition with respect to which it is measurable. The event E is an element of such a partition. Let the function g^{+}_{x} be any mapping $g^{+}_{x} : E^{+}_{f,x} \to \mathcal{X}$ with $g^{+}_{x}(s) \succsim x$, for all $s \in E^{+}_{f,x}$, and similarly, let g^{-}_{x} be any mapping $g^{-}_{x} : E^{-}_{f,x} \to \mathcal{X}$ with $g^{-}_{x}(s) \precsim x$, for all $s \in E^{-}_{f,x}$.

Axiom P3$^{\mathcal{Q}}$ (Pivotal Monotonicity). *For any act $f \in \mathcal{F}$, there exists a non-null event E, such that $f^{-1}(x) = E$ for some $x \in \mathcal{X}$, and for any outcome y, and subacts g^{+}_{x}, g^{-}_{x}, g^{+}_{y}, and g^{-}_{y}:*

$$\begin{bmatrix} g^{+}_{x} \text{ if } E^{+}_{f,x} \\ x \text{ if } E \\ g^{-}_{x} \text{ if } E^{-}_{f,x} \end{bmatrix} \succsim \begin{bmatrix} g^{+}_{y} \text{ if } E^{+}_{f,x} \\ y \text{ if } E \\ g^{-}_{y} \text{ if } E^{-}_{f,x} \end{bmatrix} \Leftrightarrow x \succsim y. \tag{8}$$

Before we explain the roles that this axiom serves, we first interpret the following key implication: for any act $f \in \mathcal{F}$, there exists a non-null event E, such that $f^{-1}(x) = E$ for some $x \in \mathcal{X}$, and:

$$f \sim \begin{bmatrix} g^{+}_{x} \text{ if } E^{+}_{f,x} \\ x \text{ if } E \\ g^{-}_{x} \text{ if } E^{-}_{f,x} \end{bmatrix}. \tag{9}$$

By P1, (9) holds for all subacts g^{+}_{x}, g^{-}_{x}, g^{+}_{y}, and g^{-}_{y}. Condition (9) states that for a given act, there exists an event, which will be called a *pivotal event*, such that changing outcomes outside

7. For notational clarity, we assume (w.l.o.g.) that the set $\{y \in f(\mathcal{S}) | y \sim x, f(E) = x\}$ is a singleton. Alternatively, the events (6) and (7) could be defined with respect to $f^{-1}(\bigcup_{x \in f(\mathcal{S})} y | y \sim x)$ for some $x \in f(\mathcal{S})$.

of that event in a (weakly) rank-preserving way does not affect preferences over acts–a form of separability. Crucially, what are held fixed during the transformation are the events assigned to outcomes, which in the original act f are either strictly preferred or strictly less preferred to x, the outcome on the pivotal event. After the transformation, these events will still map to outcomes preferred or less preferred, respectively, to x, with a weak preference permitted. The measurability requirement that the act f be constant for the pivotal event ensures that the conditions (8) and (9) are non-trivial; otherwise, the state space could be taken as pivotal for any act.

The behavioural implications of Pivotal Monotonicity are threefold. First, the axiom features a "pick a typical outcome and discard the tails" behaviour, and will be the key to guaranteeing the existence and uniqueness of a number $\tau \in [0, 1]$. Consequently, for a quantile maximizer, the certainty equivalent of any lottery will always be one of the outcomes in the support. This is in stark contrast to the Expected Utility and other cardinal models. Second, as its name suggests, P3^Q also provides preferences over acts with an appropriate, local notion of monotonicity. It states that replacing an outcome y on the pivotal event by a (weakly) preferred outcome x always leads to a (weakly) preferred act. Worth noting is that it suffices that the preference be monotonic on the pivotal event only; the axioms jointly allow for extending the monotonicity to the whole collection of events $\mathcal{E}$. Finally, in the presence of Ordering (P1) and Non-degeneracy (P5), Pivotal Monotonicity is equivalent to the following property: for any pair of acts, replacing the outcomes in their ranges in a weakly rank-preserving way (w.r.t. $\succ_x$) leaves intact the agent's preferences over these acts (Lemma 3, Appendix A.1). Underlying this equivalence is the ordinal nature of the model.

Notice that Pivotal Monotonicity does not require that the pivotal event be unique in a given act; therefore, it does not say how to relate pivotal events across acts. Together with other axioms, P3^Q will render the property of being pivotal state-independent (Lemma 6, Appendix A.4). This will set up a relation between pivotal events across acts. More importantly, the unique number τ in the unit interval [0, 1] can only be pinned down after the measure representation for beliefs is derived, and it is largely the mildness of Pivotal Monotonicity that renders constructing the measure(s) the most challenging part of the characterization.

Axiom P4^Q (Comparative Probability). *For all pairs of disjoint events E and F, outcomes $x^* \succ x$, and subacts g and h,*

$$\begin{bmatrix} x^* \text{ if } s \in E \\ x \text{ if } s \in F \\ g \text{ if } s \notin E \cup F \end{bmatrix} \succ \begin{bmatrix} x^* \text{ if } s \in F \\ x \text{ if } s \in E \\ g \text{ if } s \notin E \cup F \end{bmatrix} \Rightarrow \begin{bmatrix} x^* \text{ if } s \in E \\ x \text{ if } s \in F \\ h \text{ if } s \notin E \cup F \end{bmatrix} \succsim \begin{bmatrix} x^* \text{ if } s \in F \\ x \text{ if } s \in E \\ h \text{ if } s \notin F \cup F \end{bmatrix}. \quad (10)$$

P4^Q asserts that replacing the common subact mapping from $(E \cup F)^c$ does not strictly reverse the likelihood ranking of events E and F. It implies that adding a common complement event to either E or F will not strictly reverse the likelihood ranking between them. In turn, this will provide the representation of the "more likely than" relation over events, to be induced from preferences over acts, with a finitely additive form. Remarkably, the behavioural content of Comparative Probability is precisely the additivity of the beliefs representation of quantile maximizers: a decision maker whose preferences satisfy all axioms but P4^Q can be viewed as a quantile maximizer with respect to capacities. Hence, P4^Q provides a testable condition for whether the expectations of quantile maximizers are probabilistic (additive). The significance of this question has re-emerged in research eliciting expectations from survey respondents (see, e.g., the presidential address by Manski, 2004) as well as via the literature on probabilistic sophistication. Notice that the axiom has no effect in the cases leading to τ equal to 0 or 1; that it does not imply Savage's P4; and that no events are required to be non-null.

Axiom P5 (Non-Degeneracy). *There exist acts f and g, such that $f \succ g$.*

This is the familiar non-triviality condition. By requiring that the individual not be indifferent among all outcomes, P5 assures that both the preference relation and the derived likelihood relation are well-defined (in particular, non-reflexive) weak orders. It also permits establishing the uniqueness of a probability-measure representation of beliefs.

Before we state the final axiom, we introduce conditions that identify two important classes of preferences. Intuitively, these preferences will lead to $\tau = 0$ and $\tau = 1$, respectively.

(**L**, "lowest"): For any act $f \in \mathcal{F}$, the pivotal event maps to an outcome from the least preferred equivalence class w.r.t. $\succ_x$ in the outcome set $\{x \in \mathcal{X} | x \in f(\mathcal{S})\}$.

(**H**, "highest"): For any act $f \in \mathcal{F}$, the pivotal event maps to an outcome from the most preferred equivalence class w.r.t. $\succ_x$ in the outcome set $\{x \in \mathcal{X} | x \in f(\mathcal{S})\}$.

Definition 2. A preference relation over acts $\mathcal{F}$, $\succ$, satisfying P3$^{\mathcal{Q}}$, is called *extreme* if either (**L**) or (**H**) holds. It is called *non-extreme* if neither (**L**) nor (**H**) is satisfied.

Let us define two continuity properties that will be used in the final axiom.

(**P6**$^{\mathcal{Q}_*}$) *For all events $E, F \in \mathcal{E}$, if for any pair of outcomes $x \succ y$,*

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix}, \tag{11}$$

then there exists a finite partition $\{G_1, \ldots, G_N\}$ of $\mathcal{S}$, such that, for all $n = 1, \ldots, N$,

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin F \cup G_n \\ y \text{ if } s \in F \cup G_n \end{bmatrix}. \tag{12}$$

(**P6**$^{\mathcal{Q}^*}$) *For all events $E, F \in \mathcal{E}$, if for any pair of outcomes $x \succ y$,*

$$\begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \notin E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \in F \\ y \text{ if } s \notin F \end{bmatrix}, \tag{13}$$

then there exists a finite partition $\{H_1, \ldots, H_M\}$ of $\mathcal{S}$, such that, for all $m = 1, \ldots, M$,

$$\begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \notin E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \in F \cup H_m \\ y \text{ if } s \notin F \cup H_m \end{bmatrix}. \tag{14}$$

Axiom P6$^{\mathcal{Q}}$ (Event Continuity). *Relation $\succ$ satisfies P6$^{\mathcal{Q}_*}$ for all pairs of events in $\mathcal{E}$ if $\succ$ is non-extreme or (**H**) holds; $\succ$ satisfies P6$^{\mathcal{Q}^*}$ for all pairs of events in $\mathcal{E}$ if (**L**) holds or for a pair of a null event and any event E in $\mathcal{E}$ if $\succ$ is non-extreme.*

For the non-extreme preferences, the main force of this Archimedean axiom comes from the implication that the state space is infinite. Moreover, it ensures that the quantile in the representation is left-continuous. Formulated in terms of two-outcome acts, it has no further implications for risk preferences (i.e., the restriction of the implied lottery preferences to constant lotteries).

3.1. *A useful interpretation*

The conditions (P6$^{\mathcal{Q}_*}$) and (P6$^{\mathcal{Q}^*}$) can be interpreted in terms of likelihood relations–we will use that interpretation in the sequel. Although the definition of likelihood that we employ to construct probabilities differs from the standard one (see Section 5), the standard definition, which is used implicitly in P6^Q, still allows us to retrieve useful information from preferences. Formally, the likelihood relation adopted by Savage (1954), a binary relation $\succ^*$ on $\mathcal{E}$, is defined through Savage's P4, implied by our P1 and P3^Q:

$$E \succ^* F \text{ if for all } x \succ y, \begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \notin E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \in F \\ y \text{ if } s \notin F \end{bmatrix}. \tag{15}$$

We also employ the following definition, which differs from $\succ^*$ in that it maps the events E and F, whose likelihood is being compared, to the *less* preferred outcome:

$$E \succ_* F \text{ if for all } x \succ y, \begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix}. \tag{16}$$

With these definitions, conditions P6^{Q_*} and P6$^{Q^*}$ can be restated appropriately. In all cases leading to $\tau \in (0, 1]$, measure-representations for beliefs will be derived using relation $\succ_*$. The reason for altering the relation from the commonly used $\succ^*$ to $\succ_*$ is that $\succ^*$ would yield right-continuity of the quantile representation functional. We follow the convention in the literature and define (and derive) quantiles as left-continuous. The distinctive formulation of the condition in P6^Q for the subclass (**L**) of the extreme preferences is due to the fact that P6^{Q_*} fails in this case.

4. AXIOMATIC FOUNDATIONS OF QUANTILE MAXIMIZATION

4.1. *Probabilistic sophistication*

This section presents the first of two central theorems of the paper. The result shows that quantile maximizers' preferences $\succ$ over uncertain alternatives are consistent with them having a subjective probability distribution over the states in $\mathcal{S}$, thereby establishing probabilistic sophistication. Since the seminal paper by Machina and Schmeilder (1992), a formal definition of probabilistic sophistication has been evolving. We adopt the following conceptualization, which was first proposed by Grant (1995) and also used in a general result by Chew and Sagi (2006): fix a probability measure π on the set of events $\mathcal{E}$. Each act $f \in \mathcal{F}$ can be mapped to a lottery in $\mathcal{P}_0(\mathcal{X})$ in a natural way, through the mapping $f \to \pi \circ f^{-1}$. We say a decision maker (or, relation $\succ$) is *probabilistically sophisticated* if she is indifferent between two acts that induce identical probability distributions over outcomes. Formally, for all lotteries P, Q in $\mathcal{P}_0(\mathcal{X})$, and all acts f, g in $\mathcal{F}$,

$$(P = Q, \pi \circ f^{-1} = P,\ \pi \circ g^{-1} = Q) \Rightarrow f \sim g. \tag{17}$$

In passing, we define a mapping from acts in $\mathcal{F}$ to lotteries in $\mathcal{P}_0(\mathcal{X})$ using a fixed (possibly non-additive) measure λ. For an act $f \in \mathcal{F}$, rank the outcomes in f's outcome set $f(\mathcal{S})$ w.r.t. $\succ$: $x_1 \precsim x_2 \precsim \cdots \precsim x_N$, $N \in \mathbb{N}_{++}$; next, map the corresponding events $E_1, E_2, \ldots, E_N$ to numbers $p_1, p_2, \ldots, p_N$ in $[0, 1]$ according to: $p_1 = \lambda(E_1)$ and $p_n = \lambda(\bigcup_{m \le n} E_m) - \lambda(\bigcup_{m' \le n-1} E_{m'})$ for $n \in \{2, \ldots, N\}$. As $\sum_{n \le N} p_n = 1$, the mapping $f \to \lambda \circ f^{-1}$ uniquely yields a lottery $P \in \mathcal{P}_0(\mathcal{X})$. With the mapping $f \to \lambda \circ f^{-1}$, a non-additive measure λ (uniquely) implies a cumulative probability distribution for a given act $f \in \mathcal{F}$, denoted by Λ_f. Theorem 1 characterizes subjective beliefs about the likelihood of events for individuals whose preferences over acts satisfy axioms P1–P6^Q.

Theorem 1. *Suppose a preference relation $\succ$ over $\mathcal{F}$ satisfies P1, P3^Q, P4^Q, P5, and P6^Q. Then,*

A. There exists a unique, finitely additive, convex-ranged probability measure π with respect to which relation $\succ$ is probabilistically sophisticated if and only if it is not extreme.

B. If relation $\succ$ is extreme, there exists a set of non-atomic capacities $\Lambda(\mathcal{E})$ on $\mathcal{E}$, such that the condition (17) holds for any capacity $\lambda \in \Lambda(\mathcal{E})$.

The result reveals two interesting behavioural characteristics of beliefs underlying the choice of quantile maximizers. Theorem 1 first unveils that beliefs enter differently into the decision making of agents with extreme versus non-extreme preferences. The result identifies the condition on preferences that satisfy axioms P1–P6^Q under which quantile maximizers, like expected utility maximizers, behave as if they based their choice on a unique probability measure. This is the case as long as their preferences are not extreme. The theorem further asserts that the beliefs of maxmin and maxmax agents, although not additive, are nonetheless monotone with respect to event inclusion, and the agents can hence distinguish among such events whenever the event differences are non-null. Thus, the choices of maxmin and maxmax decision makers reflect more of their beliefs than merely whether events (and their complements in $\mathcal{S}$) are null or not.

Discussion. Theorem 1 has two general implications for modelling probabilistic sophistication. First, the qualitatively different properties of beliefs of extreme and non-extreme quantile maximizers (additivity, uniqueness, and convex-rangedness) invite a question about whether the definition of probabilistic sophistication should not be strengthened. Note carefully that for the extreme preferences, condition (17), which is increasingly used in the literature to define probabilistically sophisticated agents, is satisfied even if each act being compared is evaluated through a different (and non-additive) measure in $\Lambda(\mathcal{E})$. Furthermore, we establish that the preferences of the extreme-quantile maximizers remain intact whenever outcomes are exchanged between arbitrary disjoint non-null events, and not only between equally likely events (Lemma 6). Should probabilistic sophistication permit that? Perhaps, its definition ought to require both the additivity and uniqueness of the measure representing beliefs.[8]

In fact, similar arguments apply to a stronger version of probabilistic sophistication, which is satisfied by the preferences of extreme (as well as not) τ-maximizers: define a relation over lotteries, $\succsim_P$, induced from the underlying preferences over acts $\succsim$,

$$\text{If } \pi\circ f^{-1} = P, \pi\circ g^{-1} = Q \text{ for some } f, g \in \mathcal{F}, \text{ then } (f \succsim g \Rightarrow P \succsim_P Q). \tag{18}$$

Then, $\succsim$ is probabilistically sophisticated if there exists a measure μ on the set of events, inducing a relation $\succsim_P$ over lotteries, such that for all lotteries P, Q in $\mathcal{P}_0(\mathcal{X})$, and all acts f, g in $\mathcal{F}$

$$(P \succsim_P Q,\ \mu\circ f^{-1} = P,\ \mu\circ g^{-1} = Q) \Rightarrow f \succsim g. \tag{19}$$

The stronger definition is equivalent with (17) if $\succ$ is a weak order and π is convex-ranged, which is the case in our model for non-extreme $\succ$. The property entails that preferences $\succsim$ can be recovered from the knowledge of μ and lottery preferences $\succsim_P$ alone. The condition can also be attributed to Grant (1995), who, in addition, demanded uniqueness of μ. In the quantile model, although the choices of extreme-preferences agents are consistent with a set of beliefs, one can recover the agents' entire preference relation over all acts, even with measures that are

8. Even though additivity has no bite for the beliefs of maxmin and maxmax individuals, there exists an additive measure representing their beliefs. If condition (17) was adopted to define probabilistic sophistication with the additional requirement that π be unique, the extreme preferences (and only these preferences in the quantile model) would not be probabilistically sophisticated. As established in the results by Machina and Schmeidler and Chew and Sagi, uniqueness was not required in their definitions of probabilistic sophistication.

not only not additive but also are not convex-ranged. The knowledge of (one measure from) that set and the lottery preferences suffices.

Second, a concern about the developments in probabilistic sophistication, which has not been emphasized thus far, is that (through mixture continuity) these results restrict the set of outcomes from which acts are defined. Formally, the presence of a P6-like axiom imposes a restriction on the set of outcomes $\mathcal{X}$ ($\succ$-denseness of a countable subset), which, along with a weak-order structure, is then equivalent to the existence of a real-valued index on the set $\mathcal{X}$ (Debreu, 1954). Ideally, the existence of subjective beliefs should not hinge on the properties of the set of outcomes; for instance, the restriction excludes categorical variables. Theorem 1 neither assumes nor implies any conditions for that set, or that there is a real-valued utility index providing them with a numerical representation. Instead, we can derive a real-valued probability measure (and a real number $\tau \in [0, 1]$) without having a numerical representation of preferences. One benefit offered by our technique (outlined in Section 5) is that, unlike the results of Machina and Schmeidler or Grant, it can be used to derive beliefs for lexicographic agents and for decision problems that involve categorical variables.

In a recent beautiful result, Chew and Sagi (2006) established probabilistic sophistication under the weakest conditions to date. The result can be applied to derive beliefs in our model (though not for the extreme-τ maximizers), but doing so requires using an exchangeability rather than a likelihood relation. In Appendix 3, we explain the relative merits of both techniques of deriving beliefs. Identifying the minimal sufficient conditions that establish probabilistic sophistication by defining a likelihood relation (as, e.g., in Machina and Schmeidler, Grant, and this paper) rather than inducing it indirectly via exchangeability remains an open question.

4.2. *Representation result*

We now present a complete characterization of choice behaviour in Quantile Maximization. The second main result of the paper states that the preferences of a quantile maximizer satisfy axioms P1–P6^Q, and conversely, an individual whose preferences conform to those axioms can be viewed as a quantile maximizer.

Theorem 2. *Let $\mathcal{X}$ contain a countable $\succ$-order dense subset. Consider a preference relation $\succ$ over $\mathcal{F}$. The following are equivalent:*

(1) $\succ$ satisfies: P1, P3^Q, P4^Q, P5, and P6^Q.
(2) There exist:
   (i) a unique number $\tau \in [0, 1]$;
   (ii) a probability measure π for $\tau \in (0, 1)$ and a set of capacities $\Lambda(\mathcal{E})$ for $\tau \in \{0, 1\}$, as characterized in Theorem 1;
   (iii) an ordinal utility function $u : \mathcal{X} \to \mathbb{R}$ that represents $\succ_x$;

   such that the relation $\succ$ over acts can be represented by the preference functional $\mathcal{V}(f) : \mathcal{F} \to \mathbb{R}$ given by

$$\mathcal{V}(f) = Q^\tau(\Pi_f) \text{ if } \tau \in (0, 1); \tag{20}$$

$$\mathcal{V}(f) = Q^\tau(\Lambda_f) \text{ for any } \lambda \in \Lambda(\mathcal{E}) \text{ if } \tau \in \{0, 1\}. \tag{21}$$

The choice mechanism is thus decomposed into two factors: τ which is assured to be unique, and a probability measure, unique for all $\tau \in (0, 1)$; a set of monotone measures represents beliefs held by quantile maximizers with $\tau = 0$ or $\tau = 1$. Index u is only ordinal, that is, unique modulo strictly increasing transformations.

In the characterization of preferences through Theorems 1 and 2, Comparative Probability has no implications for the derived representation of beliefs of maxmin and maxmax decision makers. When $P4^Q$ is dispensed with, one can still uniquely pin down a number $\tau \in [0, 1]$ and, for any such τ, an agent's beliefs can then be represented by capacities (see Rostek, 2006). The significance of that is twofold. First and remarkably, the sole implication of $P4^Q$ is additivity of the beliefs representation of quantile maximizers. The remaining four conditions axiomatize Quantile Maximization with respect to non-additive measures; $P4^Q$ provides a test of additivity of the beliefs representation. Second, condition (17) would hold even if each act being compared is evaluated by a different (and non-additive) measure. The earlier discussion questioning the aptness of (17) to capture probabilistic sophistication thus extends to all τ-maximizers, $\tau \in [0, 1]$.

As noted in Section 4.1, one novel aspect of our axiomatization is that the axioms do not impose any structure on the set of outcomes $\mathcal{X}$. Given that $\succ$ is a weak order, due to the ordinality property of the quantile-maximization representation, the utility on outcomes, u, depends exclusively on how the decision maker perceives the structure of the set $\mathcal{X}$ (e.g., whether or not the outcomes are categorical). The $\succ$-order denseness condition[9] is added to assure that the representation is numerical. Without it, Theorem 2 could be recast in terms of quantiles of distributions of outcomes x rather than payoffs $u(x)$. The construction of the numerical representation for $\succ$ does not depend on the existence of the best and worst outcomes; again, ordinality provides the reason.

5. SKETCH OF THE PROOF

In the proof sketch, we focus on the heart of the axiomatization, which involves separating beliefs from preferences over $\mathcal{F}$, $\succ$, that is, establishing probabilistic sophistication. We begin by explaining why Savage's (1954) construction cannot be used directly in the Quantile Maximization model. Briefly, in order to derive a probability measure representation for beliefs under the Expected Utility model, Savage first defined a likelihood relation–the binary relation $\succ^*$ over events in $\mathcal{E}$, formulated in (15)–induced from preferences $\succ$ over acts in $\mathcal{F}$; Savage then showed that axioms P1–P6, satisfied by the binary relation over acts, imply conditions on the likelihood relation that are necessary and sufficient for the likelihood relation to admit a unique probability measure that (i) represents it, and (ii) is convex-ranged.[10] These conditions are: **A1** $\varnothing \not\succ^* E$; **A2** $\mathcal{S} \succ^* \varnothing$; **A3** $\succ^*$ is a weak order; **A4** $(E \cap G = F \cap G = \varnothing) \Rightarrow (E \succ^* F \Leftrightarrow E \cup G \succ^* F \cup G)$; **A5** $P6^{Q^*}$. In the quantile model, however, relation $\succ^*$ does not satisfy the above set of axioms. What fails for all $\tau \in (0, 1)$ is axiom A4, which is critical to establishing the additivity of the probability measure. For $\tau \in \{0, 1\}$, A4 is vacuous under relation $\succ^*$. In addition, A5 fails for $\tau \in (0, 1]$, but this problem disappears when our Event Continuity ($P6^Q$) is used instead: **A5′** $P6^Q$.

What underlies the failure of A4 is that the commonly used likelihood relation $\succ^*$ does not discriminate well between events from $\mathcal{E}$, as we now make precise. Consider a median maximizer ($\tau = 0.5$), and suppose that there does exist a probability measure, π, that represents her beliefs. Suppose further that she compares events E and F, such that $\pi(E) = 0.3$ and

9. Natural examples include $\mathcal{X}$ being finite, countably infinite, $\mathcal{X} = \mathbb{R}$.

10. (i) $E \succ^* F$ if and only if $\pi(E) > \pi(F)$, for all $E, F \in \mathcal{E}$; (ii) For any $E \in \mathcal{E}$, and any $\rho \in [0, 1]$, there is $G \subseteq E$, such that $\pi(G) = \rho \cdot \pi(E)$. Equivalent to non-atomicity for countably additive measures, convex-rangedness is stronger for finitely additive measures. (e.g., Bhaskara Rao and Bhaskara Rao, 1983, Ch. 5).

$\pi(F) = 0.2$. Given the measure π, each act in the definition of $\succ^*$, (15), induces a probability distribution. When the median maximizer compares these distributions, she ranks them as indifferent. What this means in terms of relation $\succ^*$ is that the decision maker ranks events E and F as equally likely. In general, under τ-maximization, relation $\succ^*$ ranks as equally likely all pairs of events with probabilities either both smaller than $1-\tau$ or both greater than $1-\tau$; for $\tau = 1$ ($\tau = 0$), the likelihood of no events both being more likely than $\varnothing$ (less likely than $\mathcal{S}$) can be ranked strictly by $\succ^*$. Lemma 1 demonstrates the crudeness of relation $\succ^*$. It generates only two equivalence classes in the collection $\mathcal{E}$; all events are ranked as either being equally likely to the null set or to the state space.

Lemma 1. *$E \succ^* \varnothing \Leftrightarrow E \sim^* \mathcal{S}$; $E \prec^* \mathcal{S} \Leftrightarrow E \sim^* \varnothing$.*

Therefore, even if there is a probability measure that represents the beliefs of a quantile maximizer, relation $\succ^*$ will not allow us to recover that measure from a data-set containing all choices among acts in all possible subsets of $\mathcal{F}$. Nonetheless, we show that the structure embedded in the preference relation over acts $\succ$ is rich enough to reveal the relative likelihoods of events that can be represented by a probability measure. Our approach is to construct a sub-collection of "small" events, $\mathcal{E}_{**} \subset \mathcal{E}$, and define a new binary relation on events, $\succ_{**}$, which although incomplete on $\mathcal{E}$, is complete on the sub-collection. We then derive a unique, convex-ranged and additive measure that represents a decision maker's beliefs about the relative likelihoods of events in $\mathcal{E}_{**} \cup \overline{\mathcal{E}}$, where $\overline{\mathcal{E}} = \{E \in \mathcal{E} | \nexists F$ non-null: $E \backslash F \in \mathcal{E} \backslash \mathcal{E}_{**}\}$. Adding $\overline{\mathcal{E}}$ assures that there are events that serve the role of $\mathcal{S}$ in an appropriate counterpart of A2. Intuitively, collection $\mathcal{E}_{**} \cup \overline{\mathcal{E}}$ contains events whose probabilities will not be greater than $\min\{\tau, 1-\tau\}$. Next, we show that any event from the complement of sub-collection $\mathcal{E}_{**} \cup \overline{\mathcal{E}}$ in $\mathcal{E}$ can be partitioned into events from $\mathcal{E}_{**}$. Appealing to the properties of the measure derived on $\mathcal{E}_{**} \cup \overline{\mathcal{E}}$, we then uniquely extend the measure to all events in $\mathcal{E}$ and use it to derive a unique number τ. As explained below, in the cases leading to $\tau \in \{0, 1\}$, the information contained in $\succ$ does not suffice to permit all these steps.

The sub-collection of "small" events, $\mathcal{E}_{**} \subset \mathcal{E}$, is defined to contain all events that are ranked by relation $\succ_*$ as less likely than their complements:

$$\mathcal{E}_{**} \equiv \{E \in \mathcal{E} | E^c \succ_* E\}. \tag{22}$$

The reason why this construction is helpful, and the make-up of events that comprise the sub-collection, will become clear after we specify a likelihood relation on the collection $\mathcal{E}_{**}$.

Definition 3. Let $E, F \in \mathcal{E}_{**}$.

$$E \succ_{**} F \text{ if } E \cup G \succ_* F \cup G \text{ for some event } G \in \mathcal{E}, \text{ such that } (E \cup F) \cap G = \varnothing. \tag{23}$$

The idea behind the new likelihood relation $\succ_{**}$ is as follows. The events that can be ranked strictly by $\succ_{**}$ are "small" in the sense that there exists an event G in their common complement, such that the unions $E \cup G$ and $F \cup G$ can be ranked strictly by $\succ_*$. In our example in which a 0.5-maximizer compares events E and F, such that $\pi(E) = 0.3$ and $\pi(F) = 0.2$, how will relation $\succ_{**}$ rank these events? Take an event G disjoint with both E and F, and such that $\pi(G) = 0.25$. Roughly, adding G to events E and F will enlarge the magnitudes of probabilities in an additive way (which is to be established) and, thereby, switch the evaluation of the distribution induced by act $[x \text{ if } s \notin E \cup G;\ y \text{ if } s \in E \cup G]$ in the definition of $\succ_*$, (16), from outcome x to y,[11] while maintaining the evaluation of the distribution induced by $[x \text{ if } s \notin F \cup G;\ y \text{ if } s \in F \cup G]$ at x.

For relation $\succ_{**}$ to be well-defined, we need to show that event G in (23) exists, and that there is no other event G' for which the ranking is reversed. That is, for all $E, F \in \mathcal{E}_{**}$, if $E \succ_{**} F$, then $F \not\succ_{**} E$. Lemma 5 (Appendix A.1) establishes the latter. Explaining why the former is true will also clarify the moniker "small": it is key to show that collection $\mathcal{E}_{**}$ consists of all (and only) the events, such that for any $E, F \in \mathcal{E}_{**}$ there exists $G \subseteq (E \cup F)^c$ for which

$$E \cup G \succ_* F \cup G \text{ or } E \cup G \prec_* F \cup G \text{ or } E \cup G \sim_* F \cup G \sim_* \mathcal{S}. \tag{24}$$

Condition (24) establishes the sense in which events in $\mathcal{E}_{**}$ can be compared through their complements by relation $\succ_{**}$. In particular, the sub-collection $\mathcal{E}_{**}$ does not contain events for which $E \cup G \sim_* F \cup G \sim_* \varnothing$ for all events $G \subseteq (E \cup F)^c$. Why this holds can be understood when we further demonstrate that $\mathcal{E}_{**}$ is equal to the equivalence class containing events ranked equally likely to the null set by one of relations $\succ_*$ or $\succ^*$, where "or" is meant exclusively:

$$\mathcal{E}_{**} = \{E \in \mathcal{E} | E \sim_* \varnothing\} \text{ or } \mathcal{E}_{**} = \{E \in \mathcal{E} | E \sim^* \varnothing\}. \tag{25}$$

In hindsight, after the measure and τ are pinned down, we show that the former corresponds to the representation with $\tau < 0.5$ while the latter corresponds to one with $\tau \geq 0.5$. Relations $\succ^*$ and $\succ_*$ should be seen as helpful in retrieving information from preferences in the construction of collection $\mathcal{E}_{**}$; it is relation $\succ_{**}$ that represents the likelihood ranking of events of a quantile maximizer.

In deriving the measure representation, it is essential that disjoint non-null subsets of the state space can be ranked strictly. This cannot be assured when preferences are extreme. For that case, we show that a decision maker's preferences over acts depend only on (and thus can only reveal) whether an event is null, or it is the state space, or nested in another event, all up to differences on null sub-events. Therefore, when preferences are extreme, there cannot exist an event in the common complement of any two disjoint non-null events, so that they can be ranked strictly by $\sim_{**}$. Hence, while all τ-maximizers can compare nested events, these are the only events that can be ranked strictly by 0- and 1-maximizers. It is at this point that the derivation of beliefs for $\tau \in \{0, 1\}$ departs from the general proof.

6. QUANTILE MAXIMIZATION IN APPLICATIONS

In applications, one would like to be able to characterize the attitudes of the quantile maximizers toward risk. Unlike the Expected Utility, characterizing risk attitudes through concavity of utility functions is clearly not available. Do quantile maximizers then exhibit any consistent attitudes towards risk? In Section 6.1, we show that the quantile model admits a notion of comparative risk attitude. Sections 6.2 and 6.3 offer two stylized applications that illustrate the striking properties of quantiles and encourage their more in-depth treatment. Here, we should mention other applications. Mylovanov and Zapechelnyuk (2007) examine information transmission with expert recommendations and use the quantile decision rule to characterize the cost-minimizing contracts. Bhattacharya (2009) studies the optimal peer assignment and applies quantiles as an objective function of a designer.

11. The earlier example referred to the relation $\succ^*$, whereas the new likelihood $\succ_{**}$ builds on the relation $\succ_*$. The change merely ensures left-continuity of the quantile representation to be derived. The logic is intact.

6.1. *Quantifying risk attitudes of quantile maximizers*

In identifying the notions of risk and risk attitudes suitable for Quantile Maximization, it may be worthwhile to comment on the critique of the usefulness of the *Value-at-Risk* (*VaR*) as a measure of riskiness by Artzner *et al.* (1999), who pointed out that VaR is not sub-additive. It might seem that VaR, which is defined as a quantile of the distribution of losses, is an example of Quantile Maximization; however, VaR is typically used by practitioners as a restriction of the domain of choice in the usual *mean*-maximization program. By appealing to violations of sub-additivity, Artzner *et al.* (1999) essentially argued that VaR is not a rich enough measure of risk to capture the risk considerations of *mean*-maximizers; in particular, VaR does not induce preference for diversification, which is embedded in sub-additivity. Clearly, quantile maximizers' assessment of the relative riskiness of gambles does not depend on whether or not these gambles offer diversification opportunities. Therefore, Quantile Maximization should not be used in decision problems in which the outcome spread is a concern.

The (mean-preserving) spread captures just one specific risk consideration, and we now argue that quantile maximizers are instead concerned with downside risk, which is, incidentally, an essential concept in practical risk management. Moreover, defining riskiness in terms of mean-preserving spread or sub-additivity may not be feasible in environments where the quantile model, but not the Expected Utility model, can be applied−a suitable notion of riskiness must be well-defined for non-numerical outcomes and settings where a mean need not exist. Say that distribution $Q \in \mathcal{P}_o(\mathcal{X})$ *crosses* distribution $P \in \mathcal{P}_o(\mathcal{X})$ *from below* if there exists $x \in \mathcal{X}$, such that (i) $Q(y) \leq P(y)$ for all y, such that $y \prec x$ and (ii) $Q(y) \geq P(y)$ for all y, such that $y \succ x$. Consider the class of all pairs of distributions with the single-crossing property, $\mathcal{SC} = \{(P, Q) \in \mathcal{P}_o(\mathcal{X}) \times \mathcal{P}_o(\mathcal{X}) : Q$ crosses P from below$\}$. For any pair (P, Q) in $\mathcal{SC}$, there exists an outcome x such that $P(y \prec z) \geq Q(y \prec z)$ for all $z \precsim x$, and $P(y \succ z) \geq Q(y \succ z)$ for all $z \succsim x$; we will say that P *involves more downside risk than* Q with respect to x. Intuitively, this comparative notion allows ranking the attractiveness of distributions by comparing the likelihood of losses with respect to outcome x. Say that individual A is *more risk-averse than* individual B if, for all pairs of distributions $(P, Q) \in \mathcal{SC}$, whenever B weakly prefers a distribution which involves less downside risk, so does A.

Observation. *In the Quantile Maximization model, $\tau < \tau'$ if and only if a τ-maximizer is weakly more averse toward downside risk than a τ'-maximizer.*[12]

Thus, the lower τ, the weakly more averse with respect to downside risk the decision maker is, with maxmin being the most risk-averse and maxmax the most risk-tolerant. This characterization suggests two ways in which the model studied in the present paper contributes to the description of choice behaviour. On the conceptual side, maxmin agents have been commonly, though informally, referred to as *cautious* (e.g., in game theory); with the general quantile representation, the intuited notion of cautiousness can be linked to an agent's attitude toward an objective property of gambles: downside risk. More on the practical side, τ can serve as a comparative measure or as an index of risk aversion for one agent. The measure is "global" in three ways: (i) it is defined for large as well as small gambles; (ii) it is independent

12. This result also appears in Manski (1988), who additionally defined counterparts of a risk premium and a certainty equivalent for the quantile model. The novelty here is characterization in terms of downside and upside risk, and the connection to maxmin and maxmax.

of wealth; (iii) it permits a complete ranking of agents with respect to their risk attitudes. All of these properties are in sharp contrast to those in the Expected Utility model.

The single-crossing condition admits for all quantiles a more symmetric definition of riskiness in terms of downside risk and upside chance, that is, losses as well as gains. Then, a lower-τ agent protects herself more often against downside risk and takes upside chances less often than a higher-τ agent does. More formally, this holds for any measure on decision problems.

6.2. *Application 1: Poll design*

This section demonstrates how the robustness and ordinality of quantiles might be useful in practice. In the context of a public-good problem with incomplete information, we also argue how quantiles can (and why they should) be employed in designing voting schemes that allow individuals to express the intensity of their preferences, as opposed to restricting them to making "yes/no" choices. Consider a utilitarian policy maker (he) who wishes to build a highway network in an area populated by N citizens. Let the possible density of the network lie in $[0, 1]$, where 0 stands for "no highways" and 1 corresponds to "the maximal highway density permitted by ownership structure". The citizens differ in their preference intensity over different network densities, and each citizen's intensity is her private information. The policy maker designs a poll asking the citizens: how dense a highway network would you prefer? Suppose that, prior to conducting the poll, the planner chooses whether to implement the demand for highway density based on the mean or the median reported preference intensity. After he chooses between the mean and the median, he will then (credibly) announce that the policy will be implemented based on the selected statistic. The central question we will consider is the following: based on which statistic of the distribution of citizens' reports should the policy maker decide about the highway density?

Each citizen n has a utility $u_n(q) = \theta_n q - \frac{1}{2}q^2$, where q is the density of the highway network to be chosen by the planner, common for all n, and θ_n is the citizen's *preference intensity*; the intensities are i.i.d., each being drawn from the uniform distribution F on $[0, 1]$, and the distribution is commonly known. Hence, the citizen's n demand for highway density is equal to $q^* = \theta_n$. Being utilitarian, the policy maker wishes to find the value for q that maximizes $U = \sum_n u_n(q)$. Given the agents' demands, the planner's optimal program gives $\sum_n \theta_n - q^* = 0$. Knowing the average preference intensity $\frac{1}{N}\sum_n \theta_n$ would thus suffice to maximize the planner's objective function. Instead, the policy maker knows only the distribution of preference intensities F.

Implementing a quantile rather than the mean preference intensity leads to higher expected welfare and is thus preferred by the policy maker even if he is utilitarian. This happens because being ordinal and robust to changes at the tails, a quantile-based poll design is immune to manipulation.[13] To see that, suppose the mean report is to be implemented. Then, each citizen n will pick a report q_n^r such that $\frac{1}{N}(q_n^r + \frac{1}{2}(N-1)) = \theta_n$, or 0, or 1. For example, the agent with lower-than-mean intensity optimally manipulates her reported demand q_n^r by adjusting it downwards just enough so that it does not lower the expected sample average below her true demand q_n. The optimal (Bayesian Nash) report of agent n under the mean-based policy

13. I would like to thank Marek Weretka for encouraging me to think about the implications of the robustness of quantiles to manipulation.

(depicted in Figure 2A) is then

$$q_n^r = \begin{cases} 1 \text{ if } \theta_n \geq \frac{1}{2}\left(1 + \frac{1}{N}\right); \\ \frac{1}{2} + N\left(\theta_n - \frac{1}{2}\right) \text{ if } \frac{1}{2}\left(1 + \frac{1}{N}\right) \geq \theta_n \geq \frac{1}{2}\left(1 - \frac{1}{N}\right); \\ 0 \text{ if } \theta_n \leq \frac{1}{2}\left(1 - \frac{1}{N}\right). \end{cases} \tag{26}$$

Manipulation bias, measured by $|\theta_n - q_n^r|$, becomes stronger in larger groups, as a larger individual misreport is required to adjust the expected average. Policies targeting the mean demand are thus susceptible to manipulation.[14] By contrast, if the planner commits to implementing the median report, it is optimal for the agents to reveal their true valuations; the median is determined by the ranks of reports above and below, and it is invariant to under- or over-reporting given the rank. It is clear why a citizen weakly prefers to respond truthfully to the median-based policy. Crucially, in the event that the citizen's report turns out to be the median of the empirical distribution, she strictly prefers not to misreport.

Observation. *Under the mean-based policy, all types other than mean intensities misreport their true demands. Under the median-based policy, truthful revealing is a weakly dominant strategy and the policy maker can thus recover the entire empirical distribution of valuations.*

This observation extends to all quantiles. In principle, which quantile the planner picks depends on how he is willing to trade off the downside and upside risk of the empirical distribution (as explained in Section 6.1). Since the planner is utilitarian, the relevant welfare counterpart for the mean is the median. It is easy to see that neither the mean-based nor the median-based policy is welfare-superior state-by-state; the mean induces misreporting, and the median does not reflect the planner's objective function. Strikingly, the mean can encourage a policy that is strictly lower (or higher) than *all* citizens' preferred demands. This happens if the intensities lie in $(0, \frac{1}{2}(1 - \frac{1}{N}))$ (or in $(\frac{1}{2}(1 + \frac{1}{N}), 1)$). Quantiles ensure that the policy is between the minimal and the maximal desired demand. How do welfare losses for the mean-based and the quantile-based policies compare *ex ante*? Using numerical simulations, we find:

Observation. *The median-based policy leads to strictly higher expected welfare than the mean-based policy. This happens for all population sizes.*

The welfare gap is larger for smaller groups (Figure 2B). Clearly, given the symmetry of distribution F, expected welfare converges to the first-best welfare when N increases; however, observe from the Bayesian best response (26) that, in large populations, even though consistent, the mean report aggregates information crudely, as if recording only "yes/no" votes and losing all information about preference intensities, which happens endogenously. Remarkably, the median-based policy will reflect and implement the median preference intensity. The numerical simulations further suggest that the welfare distribution induced by the mean has fatter tails and a lower perfect score (which is the mode in both cases). In fact, the numerical simulations indicate that the median dominates the mean not only in terms of expected welfare but also in the strong sense of First-Order Stochastic Dominance applied to the resulting welfare distributions.

14. Of course, a clever policy maker would try to infer the true demands from the reports using the agents' optimal strategies, but the agents would then adjust their reports accordingly. Here, the primary purpose is to illustrate the properties of quantiles and means; therefore, we abstract away from the strategic interaction between the agents and the policy maker by assuming he can commit to the announced policy. In any case, if the planner tried to deduce the true valuations from the reports, he would not recover all intensities with the mean-based policy.

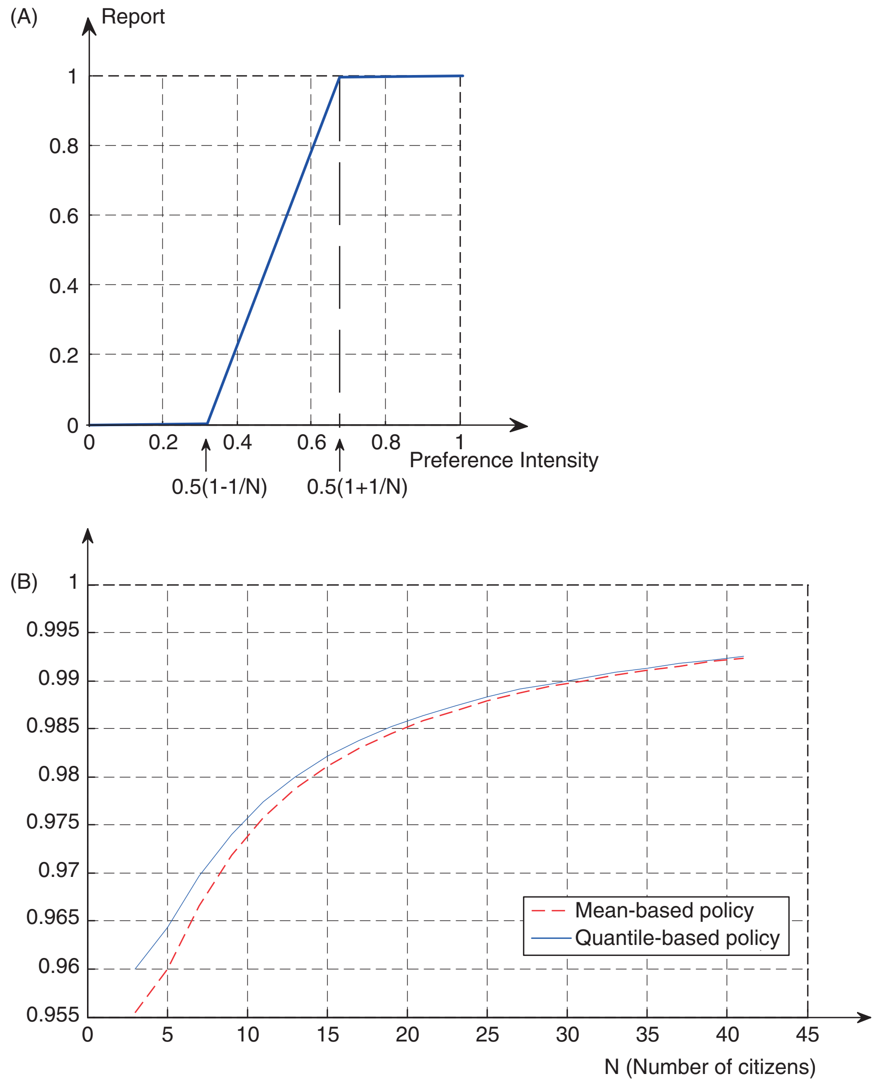

FIGURE 2

A. Report (26); B. Expected welfare relative to first-best

Notes: Panel B is based on numerical simulations for a sample size of 1 million, for each N.

It follows that the median-based policy is more attractive than the mean-based one for a larger class of the policy maker's decision criteria than utilitarian. This holds, for example, for any risk preference of the planner, or when the planner is concerned with the variance of the induced welfare distribution as well as its mean.

One implication of our analysis is that allowing citizens to express their preference intensity rather than merely making binary choices leads to higher expected welfare. The idea that the option selected through voting should reflect the strength of voters' preference extends beyond the public-good setting to electoral competition, jury voting, eliciting expectations, etc., and it is worthy of further exploration.

6.3. *Application 2: Medicare Plan D*

This section applies Quantile Maximization to a prescription drug insurance problem. The application is intended both as a concrete illustration of the model and of the aversion to downside and upside risk, as well as being suggestive of the model's potential in applied work. Medicare Part D is a US federal drug benefit programme administered by private insurance plans, which has been in effect since 1 January 2006. The standard benefit structure requires payment of a \$265 deductible. The beneficiary then pays 25% of the cost of a covered prescription drug, up to an initial coverage limit of \$2400. Once the limit is reached, the beneficiary is subject to another deductible, known as the *coverage gap* (or the "Donut Hole"), in which they must pay the full cost of medicine. Partial insurance is also provided when total out-of-pocket expenses on formulary drugs for the year, including the deductible and co-insurance, exceed a specified limit. Insurance providers offer their own variations of the standard benefit that may eliminate the deductible phase or extend the Initial Coverage limit. The premiums for the enhanced plans are appropriately adjusted.

Consider a stylized model of Medicare Part D in which insurance companies offer a menu of insurance plans to a population of quantile maximizers with heterogeneous τ's. For the sake of illustration, we assume that agents face the same distribution of losses, which is uniform on [0, 1]. The typical contract offered is $(IC, \alpha; P(IC, \alpha))$, where IC is the initial coverage limit, α is the fraction of co-insurance for which the insurer is responsible up to IC, and $P(IC, \alpha)$ is the deductible; catastrophic coverage is normalized to 1. The benefit structure is depicted in Figure 3A. The potential menu of contracts can be viewed as the square $[0, 1] \times [0, 1]$, with each point being a contract. When $\alpha = 1$ (and only then) full insurance is available. Assuming competitive insurers, all agents are offered contracts at the fair price (i.e., the expected loss to the beneficiary is zero).

The familiar prediction of the Expected Utility model asserts that, whenever a contract with $\alpha = 1$ is available, all individuals with a concave utility function over money will choose it to equalize expected wealth across states. Notably, even if all contracts with any (or all) $\alpha < 1$ are available, according to the Expected Utility model, all individuals will insure and they all will choose the maximal IC. Instead, Quantile Maximization predicts that an individual will insure a typical state as opposed to all states. Consequently, according to the quantile model, heterogeneity in contract choice will be observed and types will separate (cf. Figure 3B).

Observation. *Ceteris paribus, (i) lower-τ-maximizers will select contracts with a weakly higher initial coverage IC; and (ii) a weakly higher coinsurance α.*

The key differences in the choice behaviour underlying the Expected Utility and Quantile Maximization can be summarized by four testable predictions. Under the Quantile Maximization hypothesis: (1) Even if full insurance is available, no individuals but $\tau = 0$ will choose to fully insure and some will choose not to insure at all. (2) The separation of types through the contract choice is weakly monotonic in τ: individuals that are more cautious will choose to insure more. (3) Consider a restriction of the menu $\{(IC, \alpha; P(IC))\}_{IC\in[0,1],\alpha\in[0,1]}$ such that all contracts are offered at the same price so that the beneficiaries can trade off the initial coverage IC and co-insurance α. Then, more cautious agents will prefer to insure against downside risk by selecting the plan with a smaller donut hole and accept the upside risk of paying co-insurance for low expenses (i.e., up to IC); the higher-τ agents will, instead, choose to cover smaller amounts of co-insurance for low expenses and tolerate downside risk. (4) The contract selection by higher-τ (lower-τ) agents–and their willingness to pay–is not sensitive to moderate alterations of the co-insurance α (initial coverage IC). These predictions are independent of wealth (as long as the wealth is non-stochastic). Finally, it is worth pointing

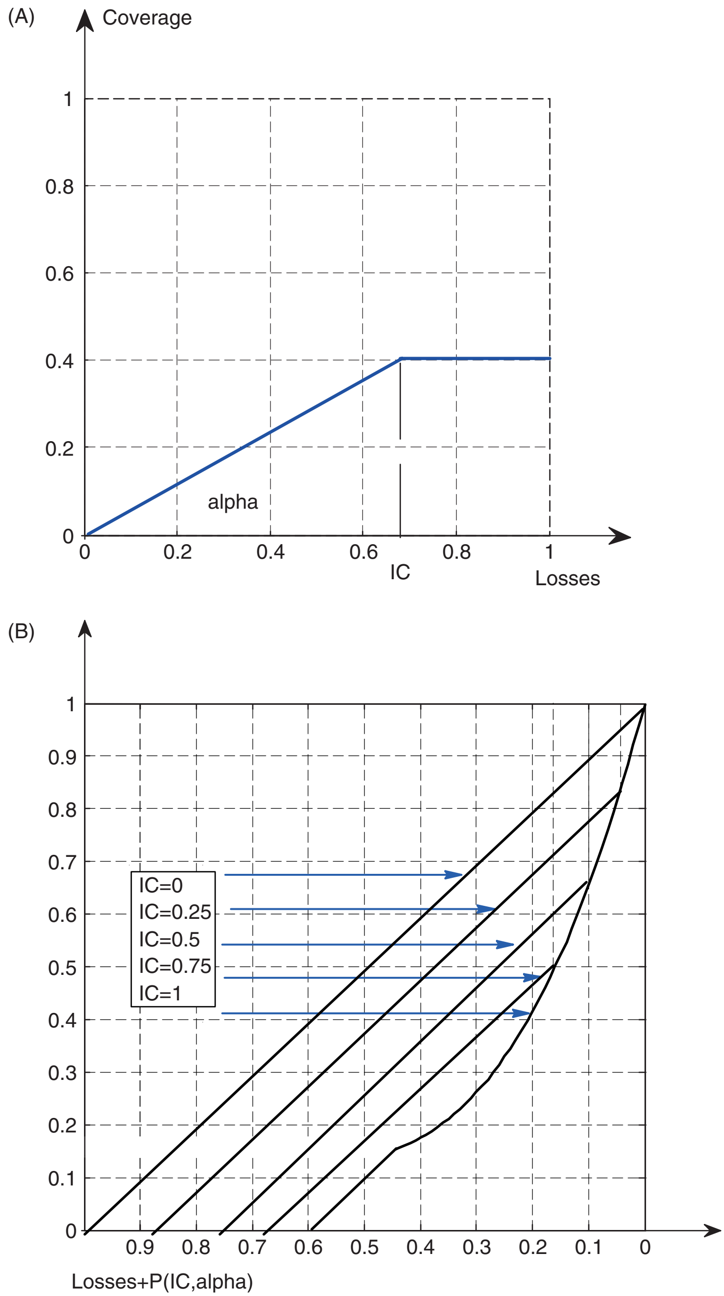

FIGURE 3

A. The benefit structure; B. Distributions of expenses with variable IC and $\alpha = 0.9$

out the difference in the reasoning that is implicit in the mean and the quantile utility models. The Expected Utility requires that, in order to choose optimally, the beneficiary must know all (insurable) levels of losses and their respective probabilities. Instead, Quantile Maximization proposes that the agent compare, say, the median levels of the distributions of wealth ("I am of median health, and I expect to incur a median level of losses"). The agent needs only local information about the loss distribution, relevant to his risk group.

Interestingly, the full-insurance prediction of the Expected Utility model for all individuals with concave utility functions over money coincides with the behaviour implied by the "worst-case scenario", which recommends minimizing the donut hole, irrespective of how skewed the distribution of expenses is towards low values. Quantile Maximization predicts the choice of more moderate insurance plans for all but the most extreme types. Observe also that the equilibrium price of a Medicare Plan D contract $P(IC, \alpha)$ is proportional in co-insurance α and concave in IC. This implies that, in equilibrium, the per-dollar-of-loss cost of a contract will be lower for agents who are more cautious (and insure more).

It is also worth contrasting the data requirements behind the Expected Utility and the Quantile Maximization models. To this end, consider in turn the insurance company taking the Quantile Maximization model to data. Unlike using the Expected Utility, in the quantile model: (1) There is no need to make any parametric assumptions about the client's utility function. (2) To compare agent risk attitudes, one does not have to first recover the concavity of utility function from data; the model allows for the analysis of attitudes toward risk even though the utilities need not be continuous, let alone concave; and risk attitudes can be studied even if outcomes are not measurable on an interval scale (as they are, e.g., for categorical variables). (3) To make policy recommendations based on the quantile model, it suffices to recover a unique parameter τ; this pins down the entire preference ordering $\succ_P$ over lotteries in $\mathcal{P}_0(\mathcal{X})$ (cf. recovering the cardinal Bernoulli utility function−in principle, an infinitely dimensional object). (4) The quantile model is robust to fat tails and works well with distributions that do not possess finite moments, a circumstance often encountered in non-life insurance.

7. CONCLUDING REMARKS

The model suggests several projects for future work. In light of the increasing concerns with model misspecification (e.g., Hansen and Sargent, 2007), an important and natural direction to take would be to permit model uncertainty by studying quantile maximization with multiple priors where the set of priors is endogenously determined. Under appropriate assumptions, for the 0-th quantile, the framework would yield the multiple-prior maxmin by Gilboa and Schmeidler (1989).

For some applications, it may be desirable to extend the model proposed in this paper to more than one quantile.[15] In particular, a choice rule may depend on the "focal" worst-, best-, and typical-case scenarios; or, a range of quantiles that are higher or lower than some threshold may be of interest. For instance, a policy may be targeted at a specific range of income distribution, school attainment, test performance, etc.

From an empirical perspective, free from parametric assumptions and moment restrictions, Quantile Maximization can be an appealing tool in applications. We have already suggested how our results could be useful for measuring expectations in survey research and we have described the comparative advantages of Quantile Maximization (with respect to the Expected Utility) in studying choice over categorical variables, robust economic policy design, and problems of resource allocation and treatment effects. For such applications, Bhattacharya (2009) develops identification and estimation methods for optimal policy design (rather than just policy evaluation). He shows that almost all the insights from mean-maximization carry over to quantile-maximization even if the quantile objective function, unlike the mean problem,

15. I would like to thank Bernard Salanié and Aldo Rustichini for (independently) making this suggestion.

is non-linear in the allocation probabilities. The methods are applicable to randomized and non-randomized settings.

APPENDIX A. PROOF OF THEOREM 1

In Section A.1, we establish several auxiliary results that will be frequently used throughout. Section A.2 presents the proof for the non-extreme preferences. Although the general logic of the proof is essentially the same for the extreme preferences, the derived properties of the relation over acts are distinct and, therefore, the representation results require that alternative arguments be employed. To highlight the differences, we present the proof for the extreme preferences separately in Section A.3. Section A.4 proves Lemmas 6 and 7.

A.1. *Auxiliary results*

The lemmas in this section characterize the binary relations $\succ^*$, $\succ_*$, and $\succ_{**}$ over events in $\mathcal{E}$ (defined in (15), (16) and (23), respectively).

Lemma 2. *$E \succ_* F \Leftrightarrow (\mathcal{S} \sim_* E$ and $F \sim_* \varnothing)$; $E \succ^* F \Leftrightarrow (\mathcal{S} \sim^* E$ and $F \sim^* \varnothing)$.*

Proof. Let $E \succ_* F$, that is by definition (16) and P5,

$$f = \begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix} = g, \text{ for all } x \succ y. \tag{A1}$$

Then, it must be that event E is pivotal for act f and F^c–for g. By P3^Q,

$$f \sim y \sim \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix} \text{ and } g \sim x \sim \begin{bmatrix} x \text{ if } s \in \mathcal{S} \\ y \text{ if } s \notin \mathcal{S} \end{bmatrix}. \tag{A2}$$

That is, by the definition of $\succsim_*$, P1 and P4^Q, $\mathcal{S} \sim_* E$ and $F \sim_* \varnothing$. For the converse, assume $\mathcal{S} \sim_* E$, $F \sim_* \varnothing$. Using the definition of $\succ_*$ in (16) and P5, for all $x \succ y$

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \sim \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix} \sim y \text{ and } \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix} \sim \begin{bmatrix} x \text{ if } s \in \mathcal{S} \\ y \text{ if } s \notin \mathcal{S} \end{bmatrix} \sim x. \tag{A3}$$

Then, by P1, P4^Q and (16), $E \succ_* F$, as desired. The argument is symmetric for $\succsim^*$.‖
Lemma 1 in Section 5 is a direct corollary of Lemma 2:

Lemma 1. *$E \succ_* \varnothing \Leftrightarrow E \sim_* \mathcal{S}$; $E \prec_* \mathcal{S} \Leftrightarrow E \sim_* \varnothing$; $E \succ^* \varnothing \Leftrightarrow E \sim^* \mathcal{S}$; $E \prec^* \mathcal{S} \Leftrightarrow E \sim^* \varnothing$.*

Lemma 3 establishes a useful property of the preferences $\succ$: For any pair of acts, replacing the outcomes in their ranges in a weakly rank-preserving (w.r.t. $\succ_x$) way, does not affect the agent's preferences over these acts. Quantile Maximization could, in fact, be axiomatized with condition (A4), which, in the presence of P1 and P5, is equivalent to Pivotal Monotonicity; the proof is straightforward and omitted here. Consider act $f \in \mathcal{F}$, such that for some disjoint events E and F, $f^{-1}(E) = x^*$ and $f^{-1}(F) = x$. Define g^+_{x*} as a mapping $\mathcal{S} \to \mathcal{X}$, such that $g^+_{x*}(\mathcal{S}) \succsim x^*$, $g^{-}_{x*,}{}^{+}_{x}$ as a mapping, such that $x^* \succsim g^{-}_{x*,}{}^{+}_{x}(\mathcal{S}) \succsim x$, and g^-_x as a mapping, such that $x \succsim g^-_x(\mathcal{S})$.

Lemma 3. *Assume Weak Order (P1), Pivotal Monotonicity (P3^Q) and Non-degeneracy (P5). For all events E and F, all pairs of outcomes $x^* \succ x$ and $y^* \succ y$, and all subacts g^+_{x*}, $g^{-}_{x*,}{}^{+}_{x}$, g^-_x, h^+_{y*}, $h^{-}_{y*,}{}^{+}_{y}$, and h^-_y,*

$$\begin{bmatrix} g^+_{x*} \textit{ if } s \in G_1 \\ x^* \textit{ if } s \in E \\ g^{-}_{x*,}{}^{+}_{x} \textit{ if } s \in G_2 \\ x \textit{ if } s \in F \\ g^-_x \textit{ if } s \in G_3 \end{bmatrix} \succ \begin{bmatrix} g^+_{x*} \textit{ if } s \in G_1 \\ x^* \textit{ if } s \in F \\ g^{-}_{x*,}{}^{+}_{x} \textit{ if } s \in G_2 \\ x \textit{ if } s \in E \\ g^-_x \textit{ if } s \in G_3 \end{bmatrix} \Rightarrow \begin{bmatrix} h^+_{y*} \textit{ if } s \in G_1 \\ y^* \textit{ if } s \in E \\ h^{-}_{y*,}{}^{+}_{y} \textit{ if } s \in G_2 \\ y \textit{ if } s \in F \\ h^-_y \textit{ if } s \in G_3 \end{bmatrix} \succ \begin{bmatrix} h^+_{y*} \textit{ if } s \in G_1 \\ y^* \textit{ if } s \in F \\ h^{-}_{y*,} {<}^{+}_{y} \textit{ if } s \in G_2 \\ y \textit{ if } s \in E \\ h^-_y \textit{ if } s \in G_3 \end{bmatrix}. \tag{A4}$$

The condition (A4) will allow inducing a likelihood relation over events, which it implicitly defines. Because of Lemma 3 the quantifiers in P4^Q can be considerably weakened compared to Savage's P4. Lemma 4 asserts consistency between relations $\succsim_*$ and $\succsim^*$, an important corollary of which is Lemma 5.

Lemma 4. *$E \succ_* F \Rightarrow E \not\prec^* F$; $E \succ^* F \Rightarrow E \not\prec_* F$.*

Proof. There are three relevant cases to the proof of the assertion for relation $\succsim_*$:

If E and F are disjoint: The result follows by P4^Q and Lemma 3.

If $F \subset E$: Define $H = E \backslash F$ and let H be non-null. The assertion will be implied if we show that it cannot be that $E \prec^* F$ or $E \prec_* F$. Suppose $E \prec_* F$. Using the definition of $\succ_*$ and Lemma 3,

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in H \\ y \text{ if } s \in F \end{bmatrix} \sim_{by\ P3Q} \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} = \begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in H \\ y \text{ if } s \in F \end{bmatrix}. \tag{A5}$$

A contradiction to P1. Hence, if $F \subset E$, then $F \precsim_* E$ and $F \precsim^* E$. The argument is symmetric for $E \subset F$.

If $F \backslash E \neq \varnothing$, $E \backslash F \neq \varnothing$ and $F \cap E \neq \varnothing$: Let $I = F \cap E$. The result follows by P4^Q applied to events $F \backslash E$ and $E \backslash F$, and Lemma 3. An analogous argument holds for relation $\succsim^*$.‖

Lemma 5. *If $E \sim_* F$ and there is a non-null event $G \subseteq (E \cup F)^c$, such that $E \cup G \succ_* F \cup G$, then there is no event $G' \subseteq (E \cup F)^c$, such that $E \cup G' \prec_* F \cup G'$.*

A.2. *Proof of Theorem 1 for non-extreme preferences*

Proof. Assume that $\succ$ is not extreme. Step 1 of the proof demonstrates that $\succ_*$ and $\succ^*$ are weak orders. Step 2 characterizes the equivalence classes of $\mathcal{E}$ under $\sim_*$ and $\sim^*$. They are used in Step 3 to derive a subset of $\mathcal{E}$, $\mathcal{E}_{**}$, on which a new and complete likelihood relation is defined, $\succ_{**}$. Step 4 verifies that axioms A1, A3, A4 and A5$'$ (specified in Section 5) hold on $\mathcal{E}_{**}$, which is then employed in Step 5 to derive a unique, convex-ranged and finitely additive probability-measure representation of $\succ_{**}$ on $\mathcal{E}_{**}$, $\tilde{\pi}$. Step 6 constructs a likelihood relation which is complete on the entire set of events, $\mathcal{E}$, and shows that measure $\tilde{\pi}$ on $\mathcal{E}_{**}$ uniquely extends to $\mathcal{E}$; we call the extended measure π. Finally, Step 7 establishes that $\succ$ is probabilistically sophisticated w.r.t. π. We will invoke Lemma 3 without mentioning.

Step 1 ($\succ_*$ and $\succ^*$ are weak orders). The argument is provided for $\succ_*$.

1. Asymmetry of $\succ_*$ is implied by (16), P1 and P5. To show negative transitivity, suppose $E \not\succ_* F$ and $F \not\succ_* G$. With $x \succ y$ (P5), (16) and Lemma 3 give (A6), and P1 then yields (A7)

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \not\prec \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix} \text{ and } \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix} \not\prec \begin{bmatrix} x \text{ if } s \notin G \\ y \text{ if } s \in G \end{bmatrix}, \tag{A6}$$

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \not\prec \begin{bmatrix} x \text{ if } s \notin G \\ y \text{ if } s \in G \end{bmatrix}, \text{ and hence } E \not\succ_* G. \tag{A7}$$

2. $\mathcal{S} \succ_* \varnothing$ and $\mathcal{S} \succ^* \varnothing$. Otherwise, the definitions of $\succ$ and $\succ_*$ lead to a contradiction.

Step 2 (Characterization of equivalence classes of $\mathcal{E}$). Since $\succ_*$ on $\mathcal{E}$ is a weak order, $\sim_*$ is an equivalence relation. By Lemma 2, there are only two equivalence classes on $\mathcal{E}$ under $\sim_*$: $\mathcal{E}|_{\sim_*\varnothing} = \{F \in \mathcal{E} | F \sim_* \varnothing\}$ and $\mathcal{E}|_{\sim_*\mathcal{S}} = \{F \in \mathcal{E} | F \sim_* \mathcal{S}\}$. Similarly, there are only two equivalence classes on $\mathcal{E}$ under $\sim^*$: $\mathcal{E}|_{\sim^*\varnothing} = \{F \in \mathcal{E} | F \sim^* \varnothing\}$ and $\mathcal{E}|_{\sim^*\mathcal{S}} = \{F \in \mathcal{E} | F \sim^* \mathcal{S}\}$. That the sets $\mathcal{E}|_{\sim_*\varnothing}$, $\mathcal{E}|_{\sim_*\mathcal{S}}$ and $\mathcal{E}|_{\sim^*\varnothing}$, $\mathcal{E}|_{\sim^*\mathcal{S}}$ all contain non-null events, follows from the assumption of the non-extreme preferences.

Step 3 (Construction of $\mathcal{E}_{**}$). 1. Define $\mathcal{E}_{**} = \{\overline{E} \in \mathcal{E} | \overline{E} \prec_* \overline{E}^c\}$. Fix $E \in \mathcal{E}_{**}$. Then, by P3^Q, for any $F \in \mathcal{E}_{**}$ disjoint (w.l.o.g.) with E (and hence such that $F \subset E^c$), there exists $G \subseteq (E \cup F)^c$, such that $F \cup G \succ_* \varnothing$. Since the above is true for disjoint E and F, hence for any $E, F \in \mathcal{E}_{**}$ there exists $G \subseteq (E \cup F)^c$, such that $E \cup G \succ_* F \cup G$ or $E \cup G \prec_* F \cup G$ or $E \cup G \sim_* F \cup G \sim_* \mathcal{S}$.

2. We prove that either $\mathcal{E}_{**} = \mathcal{E}|_{\sim_*\varnothing}$ or $\mathcal{E}_{**} = \mathcal{E}|_{\sim^*\varnothing}$. Assume first that there exists an event $E \in \mathcal{E}|_{\sim_*\varnothing}$, such that $E \sim_* E^c \sim_* \varnothing$. Consider $F \in \mathcal{E}|_{\sim^*\varnothing}$. We will show that, for all such events F, $F \sim_* \varnothing$. By the definitions of $\succ_*$ and $\succ^*$ and Lemma 1, $F^c \succ_* \varnothing$, $F \prec^* E$ and $F \prec^* E^c$. Then, by Lemma 4, $F \precsim_* E$ and $F \precsim_* E^c$. Since $\succ_*$ is a weak order, it follows that $F^c \succ_* F \sim_* \varnothing$. The event $F \in \mathcal{E}|_{\sim^*\varnothing}$ was picked arbitrarily, and hence, $\mathcal{E}_{**} = \mathcal{E}|_{\sim^*\varnothing}$. If there is no $E \in \mathcal{E}|_{\sim_*\varnothing}$ for which $E \sim_* E^c \sim_* \varnothing$, then $\mathcal{E}_{**} = \mathcal{E}|_{\sim_*\varnothing}$.

3. On the collection $\mathcal{E}_{**}$, define a binary relation over events, $\succ_{**}$, as in definition 3.
4. By Step 2.2, and axioms P6^{Q*} and P6$^{Q^*}$, $\mathcal{E}_{**}$ and relation $\succ_{**}$ are non-degenerate.

Step 4 (Axioms A1, A3, A4, A5′ hold on $\mathcal{E}_{**}$). Let $E, F, H \in \mathcal{E}_{**}$.
(**A1**) Fix $E \sim_* \varnothing$. By an argument analogous to the one in Lemma 4, there is no event $G \subseteq E^c$, such that

$$f = \begin{bmatrix} x \text{ if } s \notin E \cup G \\ y \text{ if } s \in E \cup G \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin G \\ y \text{ if } s \in G \end{bmatrix} = g. \tag{A8}$$

Hence $E \not\succ_{**} \varnothing$. If E is null, then again $E \not\succ_{**} \varnothing$, as for all $G' \subseteq E^c$

$$\begin{bmatrix} x \text{ if } s \notin E \cup G' \\ y \text{ if } s \in E \cup G' \end{bmatrix} \sim \begin{bmatrix} x \text{ if } s \notin G' \\ y \text{ if } s \in G' \end{bmatrix}. \tag{A9}$$

(**A3**) By Lemma 5, $\succ_{**}$ is asymmetric. Condition (i) in negative transitivity follows from the transitivity of $\sim_*$ $(E, F, H \in \mathcal{E}_{**})$. Suppose that (1) there is no $G \subseteq E^c \cap F^c$ non-null, such that $E \cup G \succ_* F \cup G$; and (2) there is no $G' \subseteq F^c \cap H^c$ non-null, such that $F \cup G' \succ_* H \cup G'$. We need to demonstrate that for no $G" \subseteq E^c \cap H^c$ non-null, $E \cup G" \succ_* H \cup G"$. Observe that (1) extends to all subsets of F^c: for all $G \subseteq F^c$ and all $G' \subseteq H^c$,

$$\begin{bmatrix} x \text{ if } s \notin E \cup G \\ y \text{ if } s \in E \cup G \end{bmatrix} \succsim \begin{bmatrix} x \text{ if } s \notin F \cup G \\ y \text{ if } s \in F \cup G \end{bmatrix} \text{ and } \begin{bmatrix} x \text{ if } s \notin F \cup G' \\ y \text{ if } s \in F \cup G' \end{bmatrix} \succsim \begin{bmatrix} x \text{ if } s \notin H \cup G' \\ y \text{ if } s \in H \cup G' \end{bmatrix}. \tag{A10}$$

$$\text{Thus, for all } G'' \subseteq F^c \cap H^c, \begin{bmatrix} x \text{ if } s \notin E \cup G \\ y \text{ if } s \in E \cup G \end{bmatrix} \succsim \begin{bmatrix} x \text{ if } s \notin H \cup G \\ y \text{ if } s \in H \cup G \end{bmatrix}, \tag{A11}$$

($F^c \cap H^c \neq \varnothing$ by Step 3.1). Applying the above argument again, (A11) holds for all $G'' \subseteq H^c$, and hence for all $G'' \subseteq E^c \cap H^c$. By P1, P5, Lemma 3 and Step 3.1, this proves the assertion.
(**A4**) Assume $E \cap H = F \cap H = \varnothing$, $E \cup H, F \cup H \in \mathcal{E}_{**}$, and $x \succ y$.

($\Leftarrow$) Assume that there is event $G \subseteq E^c \cap F^c \cap H^c$, such that $E \cup H \cup G \succ_* F \cup H \cup G$. Taking $G' = G \cup H$ immediately gives event G' non-null, such that $E \cup G' \succ_* F \cup G'$.

($\Rightarrow$) Assume now that there is event $G'' \subseteq E^c \cap F^c$, such that $E \cup G'' \succ_* F \cup G''$. By Lemma 5, $E \cup H \cup G''' \succsim_* F \cup H \cup G'''$ for all $G''' \subseteq E^c \cap F^c \cap H^c$. It suffices to show that it is not the case that for all $G''' \subseteq E^c \cap F^c \cap H^c$, $E \cup H \cup G''' \sim_* F \cup H \cup G'''$ (note that, by Step 3.1, $E^c \cap F^c \cap H^c$ is non-empty and there exists $\tilde{G} \subseteq E^c \cap F^c \cap H^c$ for which $E \cup H \cup \tilde{G} \sim_* \mathcal{S}$, or $F \cup H \cup \tilde{G} \sim_* \mathcal{S}$, or both). Suppose then that $E \cup H \cup G'' \not\succ_* F \cup H \cup G''$, for otherwise the assertion is delivered. Since, by assumption of $E \cup G'' \succ_* F \cup G''$ and Lemma 2, $E \cup G'' \sim_* \mathcal{S}$, we have that $E \cup H \cup G'' \sim_* F \cup H \cup G'' \sim_* \mathcal{S}$. Observe that, by P6^{Q*}, event H can be partitioned into two non-null events H_1 and H_2 such that

$$\begin{bmatrix} x \text{ if } s \notin F \cup G'' \cup H_1 \\ y \text{ if } s \in F \cup G'' \cup H_1 \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin F \cup G'' \cup H \\ y \text{ if } s \in F \cup G'' \cup H \end{bmatrix} \sim \begin{bmatrix} x \text{ if } s \notin E \cup G'' \cup H \\ y \text{ if } s \in E \cup G'' \cup H \end{bmatrix} \sim \begin{bmatrix} x \text{ if } s \notin E \cup G'' \\ y \text{ if } s \in E \cup G'' \end{bmatrix}. \tag{A12}$$

Using P6^Q and non-nullness of H_1 and H_2, there is a subset of G'', $\tilde{G}$, such that

$$\begin{bmatrix} x \text{ if } s \notin F \cup \tilde{G} \cup H \\ y \text{ if } s \in F \cup \tilde{G} \cup H \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin F \cup G'' \cup H \\ y \text{ if } s \in F \cup G'' \cup H \end{bmatrix} \sim \begin{bmatrix} x \text{ if } s \notin E \cup \tilde{G} \cup H \\ y \text{ if } s \in E \cup \tilde{G} \cup H \end{bmatrix}. \tag{A13}$$

(**A5'**) It follows by P6^Q and the definition of $\succ_*$.

Step 5 (Derivation of π on $\mathcal{E}_{**}$). 1. Axioms A1, A3, A4, and A5′ hold for all subsets of $\mathcal{E}_{**}$. Define a sub-collection of events, $\overline{\mathcal{E}}$, by $\overline{\mathcal{E}} = \{E \in \mathcal{E} | \nexists F \text{ non-null: } E \backslash F \in \mathcal{E} \backslash \mathcal{E}_{**}\}$. By construction, for any event $E \in \overline{\mathcal{E}}$, $E \succ_* \varnothing$ if $\mathcal{E}_{**} = \mathcal{E}|_{\sim_* \varnothing}$ and $E \succ_* \varnothing$ if $\mathcal{E}|_{\nsim * \varnothing}$. Call the latter property A2', a counterpart of Savage's A2 (see Section 5). We will argue that using A1, A2′, A3, A4 and A5′, Fishburn's (1970, Ch.14) proof can be applied to construct a unique, finitely additive and convex-ranged probability measure $\tilde{\pi}$ that represents the likelihood relation $\succ_{**}$ on the collection $\mathcal{E}_{**} \cup \overline{\mathcal{E}}$. For any event $E \in \overline{\mathcal{E}}$, set $\tilde{\pi}(E) = \tau$ if $\mathcal{E}_{**} = \mathcal{E}|_{\sim_* \varnothing}$ and set $\tilde{\pi}(E) = 1 - \tau$ if $\mathcal{E}_{**} = \mathcal{E}|_{\nsim * \varnothing}$. We need to show that Fishburn's argument delivers the desired result when the collection $\mathcal{E}$ is replaced by $\mathcal{E}_{**} \cup \overline{\mathcal{E}}$ (with $\overline{\mathcal{E}}$ serving the role of $\mathcal{S}$).

2. Take an event $E \in \overline{\mathcal{E}}$ and let $\mathcal{E}_E$ be the collection of all of its subsets. Applied to relation $\succ_{**}$ and collection $\mathcal{E}_E$, Fishburn's result yields a unique, finitely additive and convex-ranged probability measure $\tilde{\pi}$ that represents $\succ_{**}$ on $\mathcal{E}_E$. By Step 5.1, the measure $\tilde{\pi}$ on $\mathcal{E}_E$ is normalized to τ if the collection $\mathcal{E}_{**}$, which induced $\overline{\mathcal{E}}$, is equal to $\mathcal{E}|_{\sim_* \varnothing}$ and to $1 - \tau$ if that collection is equal to $\mathcal{E}|_{\nsim * \varnothing}$.

3. Using that $\succ_{**}$ is well defined on all of $\mathcal{E}_{**}$, we extend the measure $\tilde{\pi}$ to the remaining events in $\mathcal{E}_{**} \cup \overline{\mathcal{E}}$ as follows. For an event F in $(\mathcal{E}_{**} \cup \overline{\mathcal{E}}) \backslash \mathcal{E}_E$, let

$$\tilde{\pi}(F) = \begin{cases} \pi(H) \text{ if } F \in \mathcal{E}_{**}, \text{ where } H \in \mathcal{E}_E \text{ and } H \sim_{**} F; \\ \pi(E) \text{ if } F \in \overline{\mathcal{E}}. \end{cases} \tag{A14}$$

For the extension (A14) to be well defined, we need to show that for any event $F \in \mathcal{E}_{**}$ there exists an event H as specified in (A14). This follows by convex-rangedness of $\tilde{\pi}$ on $\mathcal{E}_E$ and Debreu (1954) ($\succ_{**}$ is a weak order and, by Step 4 (A5′), the collection $(\mathcal{E}_{**} \cup \overline{\mathcal{E}}) \backslash \mathcal{E}_E$ contains a countable $\succ_{**}$-dense subset). Thus, one can map all equivalence classes in $(\mathcal{E}_{**} \cup \overline{\mathcal{E}}) \backslash \mathcal{E}_E$ to those in $\mathcal{E}_E$.

4. We demonstrate that the extension $\tilde{\pi}$ to $\mathcal{E}_{**} \cup \overline{\mathcal{E}}$, normalized as in Step 5.1, is unique, and preserves finite additivity and convex-rangedness. Consider an event $F \in (\mathcal{E}_{**} \cup \overline{\mathcal{E}}) \backslash \mathcal{E}_E$ and its two partitions: $\{F_1, \dots, F_M\}$ and $\{H_1, \dots, H_N\}$. Let $\{I_1, \dots, I_L\}$ be the coarsest common refinement of those partitions. Such finite partitions exist by P6^{Q*} applied to $\mathcal{S}$ if $E \in \mathcal{E}_{**}$, and applied directly to E if $E \in \overline{\mathcal{E}}$. Uniqueness of summations $\sum_{n=1,\dots,M} \tilde{\pi}(F_m)$ and $\sum_{m=1,\dots,N} \tilde{\pi}(H_n)$ follows immediately from their each being equal to $\sum_{l=1,\dots,L} \tilde{\pi}(I_l) = \sum_{l=1,\dots,L} \tilde{\pi}(E_l)$ where, for every $l \in \{1, \dots, L\}$, $I_l \sim_{**} E_l \in \mathcal{E}_E$, and since each event is finitely decomposed, $\tilde{\pi}$ is finitely additive on $\mathcal{E}_{**} \cup \overline{\mathcal{E}}$. It is convex-ranged, as for any $\rho \in [0, 1]$ and $F \in \mathcal{E}_{**} \cup \overline{\mathcal{E}}$, $\rho \cdot \tilde{\pi}(F) = \rho \cdot \tilde{\pi}(H) = \tilde{\pi}(H')$, where we used that there is an event H, such that $F \sim_{**} H \in \mathcal{E}_E$ and that, by convex-rangedness of $\tilde{\pi}$ on $\mathcal{E}_E$, there is a sub-event $H' \subseteq H$, such that $\tilde{\pi}(H') = \rho \cdot \tilde{\pi}(H)$. It remains to show that $\tilde{\pi}(H') = \tilde{\pi}(F')$ for some $F' \subseteq F \in (\mathcal{E}_{**} \cup \overline{\mathcal{E}}) \backslash \mathcal{E}_E$. To this end, apply P6^{Q} to construct a partition of $\mathcal{S}$, $\{G_1, \dots, G_K\}$, every event $k \in \{1, \dots, K\}$ of which is such that $\tilde{\pi}(G_k) \leq \tilde{\pi}(E)$. On each of these events, Fisburn's theorem generates a measure with the desired properties. Normalize measures on events G_k according to (A14).

Step 6 (Extending $\tilde{\pi}$ to $\mathcal{E}$). From relation $\succ_{**}$ on $\mathcal{E}_{**}$, we derive a complete relation $\succ_{***}$ on $\mathcal{E}$.

1. We first show that all events in $\mathcal{E} \backslash \mathcal{E}_{**}$ can be partitioned into finitely many events, each of which is in $\mathcal{E}_{**}$. We consider the cases $\mathcal{E}_{**} = \mathcal{E}|_{\sim_* \varnothing}$ and $\mathcal{E}_{**} = \mathcal{E}|_{\sim^* \varnothing}$ separately. First, assume that $\mathcal{E}_{**} = \mathcal{E}|_{\sim_* \varnothing}$ and consider an event $E \in \mathcal{E}|_{\sim_* \mathcal{S}}$. By P6^{Q*}, there exists an N-partition of $\mathcal{S}$, $\{G_1, \dots, G_N\}$, such that $E \succ_* G_n$, for all $n = 1, \dots, N$. By Lemma 2, for all $n = 1, \dots, N$, $G_n \sim_* \varnothing$ and, hence, $G_n^N \in \mathcal{E}_{**}$. Write $E = (\bigcup_{\tilde{n}=1}^m G_{\tilde{n}}^N) \cup (E \backslash \bigcup_{\tilde{n}=1}^m G_{\tilde{n}}^N)$, where $m \in \{1, \dots, N\}$ is such that $\bigcup_{\tilde{n}=1}^m G_{\tilde{n}}^N \subseteq E \subset \bigcup_{\tilde{n}=1}^{m+1} G_{\tilde{n}}^N$. By construction, for all $\tilde{n} = 1, \dots, m$, $G_{\tilde{n}}^N \in \mathcal{E}_{**}$, while $E \backslash \bigcup_{\tilde{n}} G_{\tilde{n}}^N \subset G_{m+1}^N$ and hence also $E \backslash \bigcup_{\tilde{n}} G_{\tilde{n}}^N \in \mathcal{E}_{**}$. Thus, the events in the collection $\mathcal{E} \backslash \mathcal{E}_{**}$ can be decomposed into events from $\mathcal{E}_{**}$. The proof for $\mathcal{E}_{**} = \mathcal{E}|_{\sim^* \varnothing}$ is analogous via P6$^{Q^*}$.

2. Define a relation $\succ_{***}$ on $\mathcal{E}$: $E \succ_{***} F$ if there exists N-partitions of E and F, such that for all $n = 1, \dots, N$, $E_n \succ_{**} F_n$. The condition necessary (and sufficient, as demonstrated below) for the existence of such partitions can be found using the convex-rangedness of $\tilde{\pi}$ on $\mathcal{E}_{**} \cup \overline{\mathcal{E}}$: Pick $E, F \notin \mathcal{E}_{**}$ and let $\{E_1, \dots, E_N\}$ and $\{F_1, \dots, F_N\}$ be partitions of E and F, respectively, into elements in $\mathcal{E}_{**}$. By convex-rangedness of $\tilde{\pi}$, those partitions can be made equi-numbered and such that if $\sum_{n=1,\dots,N} \tilde{\pi}(E_n) > \sum_{n=1,\dots,N} \tilde{\pi}(F_n)$, then for each $n = 1, \dots, N$, $\tilde{\pi}(E_n) > \tilde{\pi}(F_n)$. This also shows that if there is an N-partition of E and F, such that for all $n = 1, \dots, N : E_n \succ_{**} F_n$, then it cannot be for any N'-partition that for all $n' = 1, \dots, N' : E_{n'} \prec_{**} F_{n'}$.

3. Define an extension of $\tilde{\pi}$ to $\mathcal{E} \backslash \mathcal{E}_{**}$, π, representing relation $\succ_{***}$ on the collection $\mathcal{E}$: For each $E \in \mathcal{E} \backslash \mathcal{E}_{**}$ and its finite partition $\{E_1, \dots, E_N\}$, $E_n \in \mathcal{E}_{**}$ for $n \in \{1, \dots, N\}$, let $\pi(E) = \sum_{n=1,\dots,N} \tilde{\pi}(E_n)$. Uniqueness, finite additivity, and convex-rangedness of the extension obtain analogously to Step 5.4.

Step 7 ($\succ$ Is probabilistically sophisticated w.r.t. π). Establishing condition (17) is an application of the argument in Machina and Schmeidler (1992, Theorem 1, Step 5). It suffices to show that the construction employed there can be used. This follows from Lemma 6A. Given that π is convex-ranged, for any $P \in \mathcal{P}_0(\mathcal{X})$, there exists an act $f \in \mathcal{F}$, such that $\pi \circ f^{-1} = P$. Using, in addition, that $\succ$ is a weak order, the stronger version of probabilistic sophistication (19) is also satisfied.‖

A.3. *Proof of Theorem 1 for extreme preferences*

Proof. Each step is assigned the same number as its counterpart in the proof for the non-extreme preferences in Appendix A.2. Some steps are left out as no longer relevant. Let $\succ_*^*$ be a binary likelihood relation on $\mathcal{E}$ defined as follows: $E \succ_*^* F$ if $E \succ^* F$ or $E \succ_* F$.

Step 1: $\succsim_*^*$ is a weak order on $\mathcal{E}$ by an argument as in Step 1, Appendix A.2, and Lemma 4.

Step 2: 1. Given Step 1, Lemma 2 defines three equivalence classes of $\mathcal{E}$ under $\sim_*^*$: $\mathcal{E}|_{\sim_*^* \varnothing} = \{F \in \mathcal{E} | F \sim_*^* \varnothing\}$, $\mathcal{E}|_{\succ_*^* \varnothing}^{\prec_*^* \mathcal{S}} = \{F \in \mathcal{E} | \mathcal{S} \succ_*^* F \succ_*^* \varnothing\}$ and $\mathcal{E}|_{\sim_*^* \mathcal{S}} = \{F \in \mathcal{E} | F \sim_*^* \mathcal{S}\}$. Assume (**H**)–the case (**L**) is analogous. We show that the equivalence classes $\mathcal{E}|_{\sim_*^* \varnothing}$ and $\mathcal{E}|_{\sim_*^* \mathcal{S}}$ contain events that differ from $\varnothing$ and $\mathcal{S}$, respectively, only on a null sub-event, i.e. $\mathcal{E}|_{\sim_*^* \varnothing} = \{E \in \mathcal{E} | E \text{ is null}\}$ and $\mathcal{E}|_{\sim_*^* \mathcal{S}} = \{F \in \mathcal{E} | F = \mathcal{S} \backslash H, H \text{ null}\}$.

Consider an arbitrary non-null event E, such that $\mathcal{S}\backslash E$ is non-null (possible by P6$^{\mathcal{Q}*}$). Given (**H**),

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix} \text{ and } \begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \notin E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix}, \ x \succ y. \tag{A15}$$

By Lemma 3 and the definitions of $\succ_*$ and $\succ^*$, it follows accordingly that $E \prec_* \mathcal{S}$ and $E \succ^* \varnothing$. By Lemma 6B, all events that differ from each $\varnothing$ and $\mathcal{S}$ on a non-null sub-event are contained in a single equivalence class, $\mathcal{E}|^{\prec^*_*\mathcal{S}}_{\succ^*_*\varnothing} = \{G \in \mathcal{E} | G \text{ is non-null and } G^c \text{ is non-null}\}$.

Step 3: Suppose that we wish to define a counterpart of the relation $\succ_{**}$ (definition 3), now based on relation $\succsim^*_*$ rather than $\succ_*$, on the subcollection of $\mathcal{E}$ containing all events E, such that $E \prec^*_* E^c$. When either (**H**) or (**L**) holds, the subset of events in $\mathcal{E}|^{\prec^*_*\mathcal{S}}_{\succ^*_*\varnothing}$ that could be ranked strictly by such a relation contains only nested events that differ on non-null sub-events. For example, assuming (**H**), consider two events $E_1, E_2 \in \mathcal{E}|^{\prec^*_*\mathcal{S}}_{\succ^*_*\varnothing}$, such that $E_1 \subset E_2$ and $E_2\backslash E_1$ is non-null. Then, by Step 2, $E_1 \sim^*_* E_2$ and for $G = \mathcal{S}\backslash E_2$,

$$\begin{bmatrix} x \text{ if } s \notin E_1 \cup G \\ y \text{ if } s \in E_1 \cup G \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix} = \begin{bmatrix} x \text{ if } s \notin E_2 \cup G \\ y \text{ if } s \in E_2 \cup G \end{bmatrix}; \tag{A16}$$

Lemma 3 delivers the conclusion. Since the strict ranking of E_1 and E_2 can only be achieved by adding the complement of the nesting event E_2 (up to null differences), there is no event G' for which the ranking would be reversed. The strict relation cannot, however, be extended to non-nested events F_1 and F_2 for which $F_1\backslash F_2$ and $F_2\backslash F_1$ are non-null. What fails is that, under (**H**) or (**L**), any event can be ranked strictly by relation $\succ^*_*$ only with the events in $\mathcal{E}|_{\sim^*_*\mathcal{S}}$; hence, by Step 2, there is no non-null event in the common complement of any two non-nested events that would enable their strict comparison. Hence, there is no non-null event E, such that $E \prec^*_* E^c$.

Step 4: Establishing that A1, A2, A3 and A5′ hold for relation $\succ^*_*$ on $\mathcal{E}$ is analogous to the proof for the non-extreme relation $\succ$. To verify A4, assume $E \cap H = F \cap H = \varnothing$ and $x \succ y$. By Step 3, relation $\succ^*_*$ can satisfy A4 only for nested events (with non-null set differences), which renders the argument straightforward.

Step 5: 1. Define a set function $\lambda : \mathcal{E} \to [0, 1]$ as follows: let $\lambda(\varnothing) = 0$, $\lambda(\mathcal{S}) = 1$; and whenever $F \subseteq E$, let $\lambda(F) \leq \lambda(E)$ with a strict inequality if E/F is non-null. Notice that any function λ that satisfies the conditions from Step 5.1 represents relation $\succ^*_*$ on $\mathcal{E}$ under (**H**) or (**L**): For all $E \in \mathcal{E}$, $E \succ^*_* F \Leftrightarrow \lambda(E) > \lambda(F)$. Denote by $\Lambda(\mathcal{E})$ the set of all measures λ that represent $\succsim^*_*$ under (**H**) or (**L**).

2. Each measure in $\Lambda(\mathcal{E})$ is non-atomic. We will prove this for (**H**). Fix $\lambda \in \Lambda(\mathcal{E})$ and consider an event $E \in \mathcal{E}|^{\prec^*_*\mathcal{S}}_{\succ^*_*\varnothing}$. By Steps 2 and 5.1, $\lambda(E) > 0$ and $\lambda(E^c) > 0$. Using P5, construct a pair of acts

$$\begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \notin E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix}. \tag{A17}$$

Applying the definition of $\succ_*$, $E^c \prec_* \mathcal{S}$, and by P6$^{\mathcal{Q}*}$ event E can be partitioned into F and $E\backslash F$, such that $E^c \cup F \prec_* \mathcal{S}$. Both F and $E\backslash F$ are necessarily non-null, for otherwise $E^c \cup F \sim_* \mathcal{S}$ or $E^c \cup (E\backslash F) \sim_* \mathcal{S}$–a contradiction to P6$^{\mathcal{Q}*}$. By Step 3 and Step 5.1, $\lambda(E) > \lambda(F)$ and $\lambda(E) > \lambda(E\backslash F)$. The derived measures need not be convex-ranged; take, for example, a monotone measure $\tilde{\lambda}$ that assigns $\tilde{\lambda}(\varnothing) = 0$, $\tilde{\lambda}(\mathcal{S}) = 1$, and a maximum of 0.9 to any event E for which $\mathcal{S}\backslash E$ is non-null; $\tilde{\lambda}$ represents the relation $\succ^*_*$ on $\mathcal{E}$ but it is not convex-ranged.

Step 7: Fix a (possibly non-additive) measure $\lambda \in \Lambda(\mathcal{E})$. Consider lotteries $P, Q \in \mathcal{P}_0(\mathcal{X})$, such that $P = Q$ and $\lambda \circ f^{-1} = P$, $\lambda \circ g^{-1} = Q$ for some $f, g \in \mathcal{F}$. Since the least preferred (w.r.t. $\succ_P$) outcomes assigned some positive probability by lotteries P and Q are identical, and equal to the least preferred (w.r.t. $\succ$) outcomes mapped from non-null events by acts f and g, condition (17) follows for (**L**); and similarly for (**H**). As for non-extreme preferences, the stronger version of probabilistic sophistication (19) is satisfied.‖

A.4. *Proof of Lemmas 6 and 7*

Lemma 6. *A. If the binary relation over acts, $\succ$, is not extreme, then for all events $E, F \in \mathcal{E}_{**}$, such that $E \sim_{**} F$, and all acts $h \in \mathcal{F}$,*

$$\begin{bmatrix} x \textit{ if } s \in E \\ y \textit{ if } s \in F \\ h \textit{ if } s \notin E \cup F \end{bmatrix} \sim \begin{bmatrix} x \textit{ if } s \in F \\ y \textit{ if } s \in E \\ h \textit{ if } s \notin E \cup F \end{bmatrix}. \tag{A18}$$

B. If the binary relation over acts, $\succ$, is extreme, then (A18) holds for all non-null events $E, F \in \mathcal{E}_{**}$, $E \cap F = \varnothing$, and all acts $h \in \mathcal{F}$.

Proof. Assume that $\succ$ is not extreme. Let π be the measure from Theorem 1. Take a pair of equal-probability events $E, F \in \mathcal{E}_{**}$. Define $\widetilde{E} = E \backslash F$ and $\widetilde{F} = F \backslash E$. By additivity of π, $\pi(\widetilde{E}) = \pi(\widetilde{F})$. Consider acts

$$f = \begin{bmatrix} x \text{ if } s \in \widetilde{F} \\ y \text{ if } s \in \widetilde{E} \\ h \text{ if } s \notin \widetilde{E} \cup \widetilde{F} \end{bmatrix} = \begin{bmatrix} h_{x+} \text{ if } s \in G_1 \\ x \text{ if } s \in \widetilde{F} \\ h_{x-,y+} \text{ if } s \in G_2 \\ y \text{ if } s \in \widetilde{E} \\ h_{y-} \text{ if } s \in G_3 \end{bmatrix} \text{ and } g = \begin{bmatrix} x \text{ if } s \in \widetilde{E} \\ y \text{ if } s \in \widetilde{F} \\ h \text{ if } s \notin \widetilde{E} \cup \widetilde{F} \end{bmatrix} = \begin{bmatrix} h_{x+} \text{ if } s \in G_1 \\ x \text{ if } s \in \widetilde{E} \\ h_{x-,y+} \text{ if } s \in G_2 \\ y \text{ if } s \in \widetilde{F} \\ h_{y-} \text{ if } s \in G_3 \end{bmatrix}. \tag{A19}$$

Using the definition of $\sim_{**}$,

$$\widetilde{E} \sim_{**} \widetilde{F} \text{ if for any } x \succ y, \begin{bmatrix} x \text{ if } s \notin \widetilde{E} \cup G \\ y \text{ if } s \in \widetilde{E} \cup G \end{bmatrix} \sim \begin{bmatrix} x \text{ if } s \notin \widetilde{F} \cup G \\ y \text{ if } s \in \widetilde{F} \cup G \end{bmatrix}, \text{ for any } G \subseteq (\widetilde{E}^c \cap \widetilde{F}^c). \tag{A20}$$

Lemma 3, invoked to define the likelihood relation $\sim_{**}$ and to reduce the cardinality of the outcome sets $f(\mathcal{S})$ and $g(\mathcal{S})$ to two, implies $f \sim g$.‖

Lemma 6B.

Proof. Assume that (**H**) holds. (The argument is analogous under (**L**).) Pick two non-null disjoint events E' and F'. Consider the following acts:

$$f = \begin{bmatrix} x \text{ if } s \in F' \\ y \text{ if } s \in E' \\ h \text{ if } s \notin E' \cup F' \end{bmatrix} \text{ and } g = \begin{bmatrix} x \text{ if } s \in E' \\ y \text{ if } s \in F' \\ h \text{ if } s \notin E' \cup F' \end{bmatrix}, \; x \succ y, \; h(\mathcal{S}) \prec x. \tag{A21}$$

Since event F' is pivotal in act f, event E' is pivotal in act g, and $f \sim g$, as desired.‖

For the next two results, we focus on the non-trivial case of the non-extreme preferences.

Lemma 7. *In the coarsest measurable partition of the state space $\mathcal{S}$ induced by act $f \in \mathcal{F}$, there is a unique pivotal event.*

Proof. We first show that the property of being pivotal is state-independent. Next, we establish that, given the coarsest measurable partition induced by an act, the pivotal event is unique.

1. Lemma 6A,B implies the following key assertion: Let π be the probability measure derived in Theorem 1. Consider act $f \in \mathcal{F}$, such that E is the pivotal event of f, and for a disjoint with E event F, $\pi(E) = \pi(F)$ and $f^{-1}(E) = x \nsim f^{-1}(F)$ for some $x \in f(\mathcal{S})$. (Such a pair of events with $\pi(E) = \pi(F)$ and $\pi(E) + \pi(F) \leq 1$ can be constructed by convex-rangedness of π.) Then, swapping outcomes between events E and F yields act g, such that $g^{-1}(F) = x$, $g^{-1}(E) = f^{-1}(F)$ and $g \sim f \sim x$.

2. Take act $f \in \mathcal{F}$ and let E and F be disjoint measurable events that map to non-indifferent outcomes: $f^{-1}(E) = x \nsim y = f^{-1}(F)$. Suppose these events are both pivotal to act f. Applying P3$^{\mathcal{Q}}$ twice to f and invoking P1 yields a contradiction: for $h \in \mathcal{F}$, $h \nsim x$, and $h \nsim y$, we have

$$x \sim \begin{bmatrix} x \text{ if } s \in E \\ x \text{ if } s \in E^c \end{bmatrix} \sim_{\text{Pivotal } E} f = \begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \in F \\ h \text{ if } s \notin E \cup F \end{bmatrix} \sim_{\text{Pivotal } F} \begin{bmatrix} y \text{ if } s \in F^c \\ y \text{ if } s \in F \end{bmatrix} \sim y. \tag{A22}$$

‖

Corollary 1. *Let E be the pivotal event of act $f \in \mathcal{F}$ and $f(E) = x$ for some $x \in f(\mathcal{S})$. Then,*

$$\textit{either } f = \begin{bmatrix} g_{x+} \textit{ if } s \in E^+_{f,x} \\ x \textit{ if } s \in E \\ g_{x-} \textit{ if } s \in E^-_{f,x} \end{bmatrix} \succ \begin{bmatrix} g_{x+} \textit{ if } s \in E^+_{f,x} \\ x \textit{ if } s \in E_1 \\ g_{x-} \textit{ if } s \in E_2 \cup E^-_{f,x} \end{bmatrix} \textit{ or } f \sim \begin{bmatrix} g_{x+} \textit{ if } s \in E^+_{f,x} \\ x \textit{ if } s \in E_1 \\ g_{x-} \textit{ if } s \in E_2 \cup E^-_{f,x} \end{bmatrix}. \tag{A23}$$

APPENDIX B. PROOF OF THEOREM 2

Proof. **Sufficiency**: (1) $\Rightarrow$ (2) Let π be the probability measure derived for the non-extreme preferences, and let $\Lambda(\mathcal{E})$ be the set of capacities derived for the extreme preferences in Theorem 1. $\succ_P$ is the preference relation over lotteries induced by the mapping $f \to v \circ f^{-1}$, where $v = \pi$ for the non-extreme and $v = \lambda \in \Lambda(\mathcal{E})$ for the extreme preferences. Step 1 of the sufficiency part establishes the existence and uniqueness of τ. Step 2 then constructs a preference functional over probability distributions that represents $\succ_P$ as a left-continuous τ-th quantile. Step 3 builds on the derived functional for distributions, $\succ_P$, to derive a representation for the relation over acts, $\succ$.

Step 1 (Existence and uniqueness of τ). For the extreme preferences, set $\tau = 0$ if (**L**) holds and $\tau = 1$ if (**H**) holds. Assume next that relation $\succ$ is not extreme. We will repeatedly use that whenever $F \succ_{***} \varnothing$, then for any $N \in \mathbb{N}_{++}$, there exists a 2^N-partition of F, $\{F_1^{2^N}, \ldots, F_{2^N}^{2^N}\}$, such that $F_1^{2^N} \sim_{***} \cdots \sim_{***} F_n^{2^N} \sim_{***} \cdots \sim_{***} F_{2^N}^{2^N}$. (The relation $\succ_{***}$ is defined in Step 6.2 in the proof of Theorem 1.) Given that axioms A1–A5′ hold on the set $\mathcal{E}$, such a *uniform* 2^N*-partition* of F can be derived by the argument in Fishburn (1970, Ch.14.2). Fix $N \in \mathbb{N}_{++}$, consider a uniform 2^N-partition of the state space $\mathcal{S}$ and associate with it a sequence of acts $\{f(n|N)\}_{n=1,\ldots,2^N}$, where

$$f(n|N) = \begin{bmatrix} x \text{ if } s \in \bigcup\limits_{l=n+1,\ldots,2^N} F_l^{2^N} \\ y \text{ if } s \in \bigcup\limits_{l=1,\ldots,n} F_l^{2^N} \end{bmatrix}, n = 1, \ldots, 2^N, \; x \succ y. \tag{B1}$$

Corollary 1 can be applied recursively to the sequence of acts $\{f(n|N)\}_{n=1,\ldots,2^N}$ to establish that there is a unique $n(N) \in \{1, \ldots, 2^N\}$ for which

$$\begin{bmatrix} x \text{ if } s \in \mathcal{S} \\ y \text{ if } s \notin \mathcal{S} \end{bmatrix} \sim \cdots \sim \begin{bmatrix} x \text{ if } s \notin \bigcup_{l=1,\ldots,n(N)-1} F_l^{2^N} \\ y \text{ if } s \in \bigcup_{l=1,\ldots,n(N)-1} F_l^{2^N} \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin \bigcup_{l=1,\ldots,n(N)} F_l^{2^N} \\ y \text{ if } s \in \bigcup_{l=1,\ldots,n(N)} F_l^{2^N} \end{bmatrix} \sim \cdots \sim \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix}. \tag{B2}$$

Now, construct a sequence of such events, one for every N, $\left\{\bigcup_{l=1,\ldots,n(N)} F_l^{2^N}\right\}_{N \in \mathbb{N}_{++}}$. Applying Lemma 7, these events are nested and weakly decreasing in N. By properly choosing a subsequence, we may define $\tau = \lim_{N\to\infty} \pi\left(\bigcup_{l=1,\ldots,n(N)} F_l^{2^N}\right) = \bigcap_N \pi\left(\bigcup_{l=1,\ldots,n(N)} F_l^{2^N}\right)$, $\tau \in (0, 1)$.

Step 2 (Representation functional for $\succ_P$). 1. By P3^{Q}, there is a one-to-one map between the equivalence classes on $\mathcal{F}$ w.r.t. $\succ$ and the outcome set $\mathcal{X}$ w.r.t. $\succ_x$; that is, for all pairs of acts $f, g \in \mathcal{F}$,

$$f \succ g \Leftrightarrow x \succ y \tag{B3}$$

where x is the outcome mapping from the pivotal event by act f, while y–by act g.

Assume that preferences are not extreme. (B3) is equivalent to $\pi \circ f^{-1} \succ_P \pi \circ g^{-1} \Leftrightarrow \delta_x \succ_P \delta_y$. The set of equivalence classes on $\mathcal{P}_0(\mathcal{X})$ w.r.t. $\succ_P$ can thus be mapped onto the set of equivalence classes on $\mathcal{X}$ (understood as the set of constant distributions) w.r.t. $\succ_P$. Hence, the certainty-equivalence mapping for distributions (each simple distribution is $\succ_P$-indifferent to an outcome in its support) can be used to construct a representation for $\succ_P$ on $\mathcal{P}_0(\mathcal{X})$. The latter can be used to provide a representation for $\succ$ on $\mathcal{F}$.

If preferences are extreme, for any measure $\lambda \in \Lambda(\mathcal{E})$, condition (B3) is equivalent to $\lambda \circ f^{-1} \succ_P \lambda \circ g^{-1} \Leftrightarrow \delta_x \succ_P \delta_y$. For a fixed measure $\lambda \in \Lambda(\mathcal{E})$, define $\mathcal{P}_0(\mathcal{X}, \lambda) = \{P \in \mathcal{P}_0(\mathcal{X}) | P = \lambda \circ f^{-1}$ for some $f \in \mathcal{F}\}$, a subset of lotteries in $\mathcal{P}_0(\mathcal{X})$. The set of equivalence classes on $\mathcal{P}_0(\mathcal{X}, \lambda)$ w.r.t. $\succ_P$ can now be mapped onto the set of equivalence classes on $\mathcal{X}$. Again, the certainty-equivalence mapping for distributions can be used to construct a representation for relation $\succ_P$ on $\mathcal{P}_0(\mathcal{X}, \lambda)$, and back up from it a representation for $\succ$ on $\mathcal{F}$.

2. The remaining steps characterize the certainty equivalence map between lotteries and outcomes as a (generalized) inverse of a distribution and establish that it represents the preference relation $\succ_P$. The unique number $\tau \in [0, 1]$ derived in Step 1 will be used in defining the inverse equal to the τ-th quantile of the distribution. By Step 1, for any act $f \in \mathcal{F}$,

$$f \sim x, \text{ where } x \text{ is such that } \pi(f(s) \precsim x) \geq \tau \tag{B4}$$

and x is (one of) the least preferred outcome(s) in $\left\{y \in f(\mathcal{S}) | \pi(f(s) \precsim y) \geq \tau; \pi(f^{-1}(y)) > 0\right\}$. By the definition of $\succ_P$, P1 and P3^{Q}, it follows that acts implying indifferent τ-th outcomes are indifferent.

3. We verify that the inverse to-be-defined (in Step 2.5) should be left continuous. Let τ be the number from [0, 1], derived in Step 1. Consider the sequence of acts $\{f(n|N)\}_{n=1,\ldots,2^N}$ constructed in Step 1.2. For any $N \in \mathbb{N}_{++}$,

define $\tau_n = \pi(\bigcup_{l=1,\ldots,n(N)-1} F_l^{2^N})$. By properly choosing a subsequence, we obtain $\tau_n \to \tau$ as $N \to \infty$; and for any given $N \in \mathbb{N}_{++}$, $f(\bigcup_{l=1,\ldots,n(N)-1} F_l^{2^N}) = y$ and $f(F_{n(N)}^{2^N}) = y$. Together with the fact that the sequence of events $\left\{\bigcup_{l=1,\ldots,n(N)} F_l^{2^N}\right\}_{N\in\mathbb{N}_{++}}$ used to derive τ is weakly decreasing, that gives that the inverse to-be-defined should be left continuous.

4. Given that $\succ$ on $\mathcal{X}$ is a weak order (P1) and $\mathcal{X}$ contains a countable $\succ$-order dense subset, a standard argument (Debreu, 1954) delivers a real-valued utility function $u(\cdot)$ on $\mathcal{X}$, unique up to a strictly increasing transformation. Let $\mathcal{U}^O$ be the set of all such functions u that represent $\succ_x$.

5. Fix utility $u \in \mathcal{U}^O$, the number $\tau \in [0, 1]$, the measure π for the non-extreme preferences and the set of capacities $\Lambda(\mathcal{E})$ for the extreme preferences. Recall that $\nu = \pi$ if $\succ_P$ is non-extreme and $\nu = \lambda$ if $\succ_P$ is extreme. For any $P = \nu \circ f$, and any $\lambda \in \Lambda(\mathcal{E})$ if $\upsilon = \lambda$, with an outcome set $\{x_1, \ldots, x_N\}$ define $V : \mathcal{P}_0(\mathcal{X}) \to \mathcal{X}$ as

$$V(P) = \begin{cases} \inf\{z \in \mathbb{R} | \upsilon[u(x_n) \le z | x_n \in f(\mathcal{S})] \ge \tau\} \text{ if } \tau \in (0, 1); \\ \sup\{z \in \mathbb{R} | \upsilon[u(x_n) \le z | x_n \in f(\mathcal{S})] \le 0\} \text{ if } \tau = 0; \\ \inf\{z \in \mathbb{R} | \upsilon[u(x_n) \le z | x_n \in f(\mathcal{S})] \ge 1\} \text{ if } \tau = 1. \end{cases} \tag{B5}$$

Step 3 (Representation functional for $\succ$). We can now combine the above steps to define a functional $\mathcal{V} : \mathcal{F} \to \mathcal{X}$ that represents relation $\succ$ on acts: For all $f, g \in \mathcal{F}$, and all $P, Q \in \mathcal{P}_0(\mathcal{X})$, such that $P = \pi \circ f$ and $Q = \pi \circ g$, we have $f \succ g \Leftrightarrow P \succ Q$, which by Steps 1 and 2 is equivalent to $\mathcal{V}(f) = V(P) > V(Q) = \mathcal{V}(g)$. That is, as desired, the preference relation $\succ$ on $\mathcal{F}$ can be represented by evaluating each act $f \in \mathcal{F}$ by the τ-th quantile of the distribution induced by act f and measures π for the non-extreme and $\Lambda(\mathcal{E})$ for the extreme preferences:

$$\mathcal{V}(f) = \begin{cases} \inf\{z \in \mathbb{R} | \upsilon[u(f(s)) \le z] \ge \tau\} \text{ if } \tau \in (0, 1); \\ \sup\{z \in \mathbb{R} | \upsilon[u(f(s)) \le z] \le 0\} \text{ if } \tau = 0; \\ \inf\{z \in \mathbb{R} | \upsilon[u(f(s)) \le z] \ge 1\} \text{ if } \tau = 1. \end{cases} \tag{B6}$$

Fix $\tau \in [0, 1]$. Given a complete pre-order over outcomes $\succsim_x$, say that distribution $Q = (y_1, q_1; \ldots; y_M, q_M)$ τ-*first-order stochastically dominates* (τ-*FOSD*) distribution $R = (y_1, r_1; \ldots; y_M, r_M)$ with respect to $\succsim_x$ if $V(Q) > V(R)$, where $V(\cdot)$ is as defined in (B5). $\succsim_P$ is said to satisfy τ-*first order stochastic dominance* if $P \succ_P Q$ whenever P τ-FOSD Q with respect to $\succsim_x$. τ-FOSD ranks distributions completely.

Necessity: (2) $\Rightarrow$ (1) Assume that the representation $\mathcal{V}(f)$ holds for $\succ$. Fix a number $\tau \in [0, 1]$, a measure π for $\tau \in (0, 1)$ and a set of capacities $\Lambda(\mathcal{E})$ for τ equal to 0 or 1; for the extreme quantiles, the arguments below hold for all capacities in $\Lambda(\mathcal{E})$. Fix utility $u \in \mathcal{U}^O$. Showing that conditions (**L**) and (**H**) hold for the representation with $\tau = 0$ and $\tau = 1$, respectively, is straightforward and is omitted here.

P1 (Ordering). This holds, since there is a real-valued representation of $\succ$.

P3^Q (Pivotal monotonicity). (*if*) Implied by τ-FOSD. (*only if*) Pick an act $f \in \mathcal{F}$. By τ-FOSD, $f \sim x$ for $x \in \mathcal{X}$, such that $\mathcal{V}(f) = x$. That the event to which x is mapped by act f, $f^{-1}(x) = E$, is non-null follows from the representation. By τ-FOSD, E is such that $f \sim [g_x^+ \text{ if } E_{f,x}^+; x \text{ if } E; g_x^- \text{ if } E_{f,x}^-]$. Consider outcome $y \precsim x$. Then, appealing to τ-FOSD again yields $[g_x^+ \text{ if } E_{f,x}^+; x \text{ if } E; g_x^- \text{ if } E_{f,x}^-] \succsim [g_y^+ \text{ if } E_{f,x}^+; y \text{ if } E; g_y^- \text{ if } E_{f,x}^-]$ for any subacts g_x^+, g_x^-, g_y^+, and g_y^-, as desired.

P4^Q (Comparative probability). Assume $\tau \in (0, 1)$. Pick disjoint events E and F, outcomes $x^* \succ x$, and the following acts

$$\begin{bmatrix} x^* \text{ if } s \in E \\ x \text{ if } s \in F \\ g \text{ if } s \notin E \cup F \end{bmatrix} \succ \begin{bmatrix} x^* \text{ if } s \in F \\ x \text{ if } s \in E \\ g \text{ if } s \notin E \cup F \end{bmatrix}. \tag{B7}$$

Define $(E \cup F)_{gx}^{c-} = \{s \in \mathcal{S} | g(s) \prec x\}$. Then, for (B7), τ-FOSD implies that $\pi((E \cup F)_{gx}^{c-}) + \pi(F) < \tau \le \pi((E \cup F)_{gx}^{c-}) + \pi(E)$, and hence, $\pi(E) > \pi(F)$. Reversing the above argument, it must be that

$$\begin{bmatrix} x^* \text{ if } s \in E \\ x \text{ if } s \in F \\ h \text{ if } s \notin E \cup F \end{bmatrix} \succsim \begin{bmatrix} x^* \text{ if } s \in F \\ x \text{ if } s \in E \\ h \text{ if } s \notin E \cup F \end{bmatrix}, \tag{B8}$$

where $(E \cup F)_{hx}^{c-} = \{s \in \mathcal{S} | h(s) \prec x\}$, and we used that $\pi((E \cup F)_{hx}^{c-}) + \pi(F) < \pi((E \cup F)_{hx}^{c-}) + \pi(E)$.

For $\tau \in 0$, (B7) can hold only if event F is null, E is non-null and $g(s) \succ x$ for every $s \in (E \cup F)^c$; similarly, for $\tau \in 1$, (B7) can hold only if event F is null, E is non-null and $g(s) \prec x^*$ for every $s \in (E \cup F)^c$.

P5 (Nondegeneracy). This follows, since the functional $\mathcal{V}: \mathcal{F} \to \mathbb{R}$ is non-constant.

P6^Q (Small-event continuity of $\succsim_l$). Let $\tau \in (0,1)$. Suppose that for all $x \succ y$,

$$\begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \notin E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \in F \\ y \text{ if } s \notin F \end{bmatrix} \tag{B9}$$

By the definition of relation $\succ^*$, $E \succ^* F$, and hence, given the measure π and the definition of relation $\succ_{***}$ which it represents, $\pi(E) > \pi(F)$. Further, the representation of $\succ$ implies that $1 - \pi(F) \geq \tau > 1 - \pi(E)$. By non-atomicity of the measure π, we can partition the state space $\mathcal{S}$ into N events $\{H_1, \ldots, H_N\}$ and choose N, such that $\tau > 1 - \pi(E) + \pi(H_n)$ for all $n = 1, \ldots, N$. The definitions of $\succ^*$ and $\succ_{***}$, then, imply that $E \succ^* F \cup H_n$, for all $n = 1, \ldots, N$. By a similar argument, P6$^{Q^*}$ holds for any event $E \in \mathcal{E}$ and $\varnothing$.

For (**L**), (B9) is satisfied only when F^c is non-null and E^c is null. Then, $\lambda(E) = 1 > \lambda(F)$ for any $\lambda \in \Lambda(\mathcal{E})$. Non-atomicity of measures in $\Lambda(\mathcal{E})$ completes the proof. So does it for (**H**), in which case (B9) is satisfied only when E is non-null and F is null; or, $\lambda(F^c) = 1 > \lambda(E^c)$.‖

APPENDIX C. COMPARISON WITH CHEW AND SAGI (2006)

Here we contrast our technique of deriving beliefs with the one proposed recently by Chew and Sagi (2006, hereafter "CS"). The approach used by CS is based on the notion of exchangeability. Two events are said to be *exchangeable* if the agent is always indifferent to permuting the payoffs assigned to these events.

Definition 4. For any pair of disjoint events E and F, E is exchangeable with F if, for any outcomes x and y, and any act f,

$$\begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \in F \\ f \text{ if } s \notin E \cup F \end{bmatrix} \sim \begin{bmatrix} y \text{ if } s \in E \\ x \text{ if } s \in F \\ f \text{ if } s \notin E \cup F \end{bmatrix}. \tag{C1}$$

The relation of exchangeable events is then used to define the *comparability* relation, $\succsim^C$.

Definition 5. For any events E and F, $E \succsim^C F$ whenever $E \backslash F$ contains a sub-event G that is exchangeable with $F \backslash E$.

Intuitively, exchangeability carries the meaning of "equal likelihood", whereas comparability conveys "greater likelihood". CS find a set of axioms on those relations, *so they can yield a likelihood relation* that can be provided with a probability-measure representation. By contrast, we *define a strict likelihood relation revealed from the preferences that we study* and find conditions on the likelihood relation to derive a measure representation. This raises the question of a comparison between the two approaches, which we will refer to as *exchangeability-based* and *direct-likelihood* (i.e., one that defines the strict likelihood relation from preferences; e.g., Machina and Schmeidler, 1992; Grant, 1995, this paper). Interested in probabilistic sophistication as such, CS accomplished the derivation of beliefs without monotonicity and continuity of preferences by using exchangeability relation. It remains an open question as far as how to derive beliefs without using monotonicity or continuity and without resorting to the exchangeability relation but, instead, directly defining a strict likelihood relation from preferences. Our Theorem 1 is the first in the direct-likelihood literature to dispose of continuity, and it uses the weakest notion of monotonicity (weak FOSD). Note that our monotonicity axiom enters into the derivation of a measure in the quantile model *only* in one step, namely, to show that non-null events are judged more likely than the empty set–this is exactly how Chew and Sagi used their weakening of Savage's monotonicity. Why would a direct-likelihood argument be attractive given the CS result?

One reason is that the link between the likelihood that CS construct and preferences over acts is only through the definition of exchangeable events, a pre-notion of "equally likely". In particular, the transitivity of the comparability relation is proved without any recourse to the strict preference relation over acts, $\succ$ and thus *cannot reveal any information about these preferences*. By contrast, in the direct-likelihood method, the strict "more-likely-than" relation is revealed by preferences over acts, from which it inherits its properties. For example, by maintaining the link between preferences and beliefs, the direct-likelihood approach is quite revealing about what is going on in the quantile model.

The price of deriving beliefs with conditions that do not draw on the strict preference relation is that they do not carry any behavioural content from $\succ$. In fact, as CS point out, applying the exchangeability-based approach to an arbitrary decision-making model might not be warranted–for example, in the famous example of "Machina's Mother"

(Machina, 1989), the exchangeability relation fails to deliver a notion of likelihood (and, hence, also a probability measure). The direct-likelihood approach is immune from that (by construction) and can deliver a probability-measure representation of beliefs even in the example of "Machina's Mother".

In addition, CS *assume* (in an axiom) that exchangeable events exist. In the direct-likelihood method, it is *established* that some events are judged as equally likely or exchangeable (Step 4 in the proof of Theorem 1 in Machina and Schmeidler (1992), Claim 5a in Grant (1995), our Lemma 6A).

While the CS method could, in principle, be used for non-extreme preferences (i.e., those leading to $\tau \in (0, 1)$), it cannot be applied to extreme preferences. In summary, the two techniques represent two conceptually different and complementary approaches to deriving beliefs from preferences.

REFERENCES

ARROW, K. J. and HURWICZ, L. (1972), "An Optimality Criterion for Decision-Making under Ignorance", in Carter, C. F. and Ford, J. L. (eds) *Uncertainty and Expectations in Economics* (Fairfield, NJ: Augustus M. Kelley).

ARTZNER, P., DELBAEN, F., EBER, J.-M. and HEATH, D. (1999), "Coherent Measures of Risk", *Mathematical Finance,* **9** (3), 203–228.

BARBERA, S. and JACKSON, M. (1988), "Maximin, Leximin and the Projective Criterion: Characterizations and Comparisons", *Journal of Economic Theory,* **46**, 34–44.

BHASKARA RAO, K. P. S. and BHASKARA RAO, M. (1983), *Theory of Charges. A Study of Finitely Additive Measures* (London: Academic Press).

BHATTACHARYA, D. (2009), "Inferring Optimal Peer Assignment from Experimental Data", *Journal of the American Statistical Association,* (forthcoming).

BÖRGERS, T. (1993), "Pure Strategy Dominance", *Econometrica,* **61** (2), 423–430.

BOUTILIER, C. (1994), "Towards a Logic for Qualitative Decision Theory", in *Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning*, Bonn, 75–86.

CHAMBERS, C. P. (2007), "Ordinal Aggregation and Quantiles", *Journal of Economic Theory,* **137**, 416–431.

CHAMBERS, C. P. (2009), "An Axiomatization of Quantiles on the Domain of Distribution Functions", *Mathematical Finance,* **19**, 335–342.

CHEW, S. H. and SAGI, J. S. (2006), "Event Exchangeability: Probabilistic Sophistication without Continuity or Monotonicity", *Econometrica,* **74** (3), 771–786.

COHEN, M. (1992), "Security Level, Potential Level, Expected Utility: A Three-criteria Decision Model under Risk", *Theory and Decision,* **33**, 101–134.

DEBREU, G. (1954), "Representation of a Preference Ordering by a Numerical Function", in Thrall, R. M., Coombs, C. H. and Davis, R. L. (eds) *Decision Processes* (New York: Wiley).

DUBOIS, D., FARGIER, H., PRADE, H. and PERNY, P. (2002), "Qualitative Decision Theory: From Savage's Axioms to Nonmonotonic Reasoning", *Journal of the ACM,* **49** (4), 455–495.

DUBOIS, D., GRABISCH, M. M., MODAVE, F. and PRADE, H. (2000), "Relating Decision under Uncertainty and Multicriteria Decision Making Models", *International Journal of Intelligent Systems,* **15** (10), 967–979.

FISHBURN, P. C. (1970), *Utility Theory for Decision Making* (New York: Wiley, reprinted in Huntington, NY: Robert E. Kreiger Publishing Company).

GILBOA, I. and SCHMEIDLER, D. (1989), "Maxmin Expected Utility with Non-unique Prior", *Journal of Mathematical Economics,* **18**, 141–153.

GRANT, S. (1995), "Subjective Probability Without Monotonicity: Or How Machina's Mom May Also Be Probabilistically Sophisticated", *Econometrica,* **63** (1), 159–191.

HANSEN, L. P. and SARGENT, T. J. (2007), *Robustness* (Princeton, NJ: Princeton University Press).

HURWICZ, L. (1951), "Optimality Criteria for Decision Making under Ignorance" (Cowles Commission DP, 370).

KLIBANOFF, P., MARINACCI, M. and MUKERJI, S. (2005), "A Smooth Model of Decision Making Under Ambiguity", *Econometrica,* **73** (6), 1849–1892.

KOENKER, R. and BASSETT, G. (1978), "The Asymptotic Distribution of the Least Absolute Error Estimator", *Journal of the American Statistical Association,* **73**, 618–622.

MACCHERONI, F., MARINACCI, M. and RUSTICHINI, A. (2006), "Ambiguity Aversion, Robustness, and the Varia- tional Representation of Preferences", *Econometrica,* **74** (6), 1447–1498.

MACHINA, M. J. (1989), "Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty", *Journal of Economic Literature,* **27** (4), 1622–1668.

MACHINA, M. and SCHMEIDLER, D. (1992), "A More Robust Definition of Subjective Probability", *Econometrica,* **60** (4), 745–780.

MANSKI, C. F. (1988), "Ordinal Utility Models of Decision Making Under Uncertainty", *Theory and Decision,* **25** (1), 79–104.

MANSKI, C. F. (2004), "Measuring Expectations", *Econometrica,* **72** (5), 1329–1376.

MASKIN, E. (1979), "Decision-Making under Ignorance with Implications for Social Choice", *Theory and Decision,* **11**, 319–337.

MILNOR, J. (1954), "Games against Nature", in Thrall, R. M., Coombs, C. H. and Davis, R. L. (eds) *Decision Processes* (New York: Wiley).

MYLOVANOV, T. and ZAPECHELNYUK, A. (2007), "Contracts for Experts with Opposing Interests" (Working Paper).
O'NEILL, B. (2001), "Risk Aversion in International Relations Theory", *International Studies Quarterly,* **45**, 617–640.
RAWLS, J. (1971), *A Theory of Justice* (Cambridge, MA: Harvard Press).
ROBERTS, K. W. S. (1980), "Possibility Theorems with Interpersonally Comparable Welfare Levels", *Review of Economic Studies,* **47**, 409–420.
ROSTEK, M. (2006), "Quantile Maximization in Decision Theory," PhD Dissertation, Ch. 1, Yale University.
ROY, A. D. (1952), "Safety First and the Holding of Assets,", *Econometrica,* **20** (3), 431–449.
SAVAGE, L. J. (1954), *The Foundations of Statistics* (New York: Wiley, revised and enlarged, New York: Dover, 1972).
SEGAL, U. and SOBEL, J. (2002), "Min, Max, and Sum", *Journal of Economic Theory,* **106** (1), 126–150.
YILDIZ, M. (2007), "Wishful Thinking in Strategic Environments", *Review of Economic Studies,* **74**, 319–344.