Reasoning in Strategic and Non-Strategic Interactions

A Dissertation

Presented to the Faculty of the Graduate School of

Yale University

in Candidacy for the Degree of

Doctor of Philosophy

by

Marzena Joanna Rostek

Dissertation Director: Professor Benjamin Polak

December 2006

Abstract

Reasoning in Strategic and Non-Strategic Interactions

Marzena Joanna Rostek

2006

The first chapter introduces a model of preferences in which an individual compares uncertain alternatives through a quantile of the induced utility distributions. The choice rule of Quantile Maximization unifies maxmin and maxmax and generalizes them to any intermediate quantile. Taking preferences over acts as a primitive, we axiomatize Quantile Maximization in a Savage setting. We derive probability measure(s) representing subjective beliefs and a unique quantile that is maximized by the induced preferences over probability distributions. Importantly, the probability measure is unique for all levels of quantile strictly between 0 and 1. Our axiomatization provides a novel characterization of probabilistic sophistication. We also discuss applications in modeling individual, social and strategic choice, studying risk measures and robust economic policy design.

The second chapter develops identification conditions for Quantile Maximization for finite data sets. We first ask which actions will be observed if individuals are quantile maximizers and show that model predictions differ from several leading alternatives, including expected utility. We then investigate how much information can be inferred about the unobservables of the model from payoff structure and choices. When agents face known probabilities, e.g. in a lab, we show that one can construct decision problems that identify the quantile exactly by observing a single choice. When the beliefs about the likelihood of events cannot be observed *a priori*, we derive bounds that can be placed on the unobservable quantile and on the beliefs from the data. We illustrate how these conditions can be applied to strategic (multiple-agent) settings.

One of the most robust findings on experimental play in normal form games is that the observed outcomes are consistent with small finite levels of reasoning (typically 1-2). At the same time, the equilibrium notions in game theory imply that the set of outcomes must lie weakly within the rationalizable set, defined by the *common* knowledge of rationality and the game structure. To predict outcomes in such interactions, a model should thus explain why players *optimally* choose not to reason further. The received non-equilibrium models (e.g., k-level models), however, rely on bounded-rationality arguments. In the third chapter, we formulate a model that allows the separation of optimal reasoning levels from cognitive limitations. The implied set of outcomes is shown to be parameterized by the players' endogenously derived actual levels of reasoning.

Contents

List of Figures

List of Matrices

Acknowledgements

I am indebted to Itzhak Gilboa, Stephen Morris and Ben Polak for constantly challenging and encouraging me during my years at Yale. Thank you for how much I have learned from your mentorship!

I also thank Dino Gerardi for his advice and many helpful suggestions. Throughout the research project, I greatly benefited from stimulating conversations with Scott A. Boorman, Don Brown, Markus Brunnermeier, Larry Epstein, Simon Grant, Philippe Jehiel, Barry O'Neill, Jacob Sagi, Bernard Salanie, Barry Schachter, Jan Werner, John Weymark. I also wish to acknowledge the comments of my colleagues at Yale, in particular, Rüdiger Bachmann, Pei-Yu Lo and Colin Stewart.

Above all, I would like to thank my husband Marek Weretka for countless hours of discussion and his unshakable enthusiasm for economic theory.

I would also like to acknowledge numerous useful suggestions that I received from seminar participants at Yale, Caltech, Northwestern, Minnesota, Wisconsin-Madison, Columbia, Duke FUQUA, Rochester and Oxford; and the conference participants at the Risk Utility and Decision (Paris, June 2006), the Foundations of Risk and Uncertainty XII (Rome, June 2006), the Midwest Economic Theory Meeting (Lawrence, Kansas, October 2005), the XVI International Conference on Game Theory (Stony Brook, July 2005), and the X Spring Meeting of Young Economists (Geneva, April 2005). I wish to thank them all.

To my mom Krystyna for her loving support

Chapter 1

Quantile Maximization in Decision Theory

1.1 Introduction

Consider an individual buying a product of uncertain quality on the Internet. Today, many companies such as Amazon.com, MyTravelGuide.com, PCWorld.PriceGrabber.com, and BizRate.com, employ on-line customer ratings that provide potential buyers with users' different quality assessments. How should a buyer choose when only distributions of ranks but not of absolute evaluations of the quality of the products are available?

Consider further a policy maker who sets policies while facing inherent uncertainty about the economy fundamentals. Analysis according to a worst-case (precautionary) scenario is commonplace in economic policy design. Compared to rules based on expectations, this decision criterion has the important advantage in that it does not require any parametric assumptions about utilities. Some critics, however, have raised a concern that basing policy choices on the worst-case outcome gives too much importance to what may be very unlikely outcomes (e.g., Svensson (2000]). Is it possible to design robust policies and accommodate the two tenets that a policy should be precautionary but not entirely dependent on or sensitive to unlikely and extreme outcomes?

The Subjective Expected Utility and its variants may not be suitable in those contexts. Their evaluations, and hence the resulting choices, are not only sensitive to outliers, but also to marginal changes in any probability or payoff; they require not just ranks but cardinal comparisons; and their policy implications crucially depend on the assumed specification of agents' utility functions. In short, they are not robust, and they rely on cardinality assumptions about utilities, which are in many settings inadequate or unnecessarily strong, though they may still drive predictions. In economic theory, there are two famous decision criteria, maxmin and maxmax (choosing an alternative with the highest minimal and maximal outcome, respectively). These have ordinal properties and are robust to changes in distributions, although not in their supports. Evaluations according to the worst- and best-case scenarios are examples of their use in practice. Maxmin and maxmax have been popular in games, bargaining, social choice, voting and other areas in economics. The preferences they model are, however, very extreme. Surprisingly, there is no framework that captures less-extreme choice behavior, preserving qualitative properties of maxmin and maxmax that Expected Utility does not exhibit.

This paper notes that maxmin and -max can be viewed as maximizing, respectively, the lowest and the highest quantile of beliefs distributions. Building on this idea, we model a decision maker who, given her beliefs about events, maximizes a quantile of the induced distributions. Thus the model unifies maxmin and maxmax, and generalizes them to any intermediate quantile. Unlike in Expected Utility, in Quantile Maximization, the choices are invariant to arbitrary transformations of payoffs. The rule can model agents who are only able to rank outcomes in lotteries but not to assess by "how much" they prefer one over another. Further, the optimal decisions remain robust to changes of distributions outside of the quantile.

To our knowledge, the only reference to formal analysis of quantile maximization in decision making is the work of Manski (1988). Advocating an ordinal-utility approach to modelling choice under uncertainty, Manski examined risk preferences of an agent maximizing a quantile of the distribution of utility (*quantile utility* model), and an agent maximizing the probability that the realization will exceed some level (*utility mass* model). Although largely ignored in choice theory literature, quantiles are present in many applied areas of economics: Value-at-Risk, one of the most popular measures of risk in finance, is defined as a quantile of the distribution of losses; in econometrics, quantiles are used in techniques of robust estimation and quantile regression; they are also applied in measurement (population-based poverty lines), as order statistics, etc. This paper formalizes the concept of Quantile Maximization in choice-theoretic language and studies its implications for decision making.

Results. We model an individual choosing between uncertain alternatives who evaluates each alternative by the τ^{th}-quantile of the induced distributions and selects the one with the highest quantile payoff. Thus, under Quantile Maximization, a decision maker is characterized by a given level of $\tau \in [0, 1]$ that we call an *anticipation level*; subjective beliefs over events π; and a rank order over outcomes. The central theoretical contributions of the paper are axiomatization and characterization of probabilistic sophistication. We describe the main results.

We provide an exact characterization of the model by jointly axiomatizing Quantile Maximization and subjective probabilities in a Savage setting. That is, taking preferences over acts (maps from states to outcomes) as a primitive, we find conditions that are necessary and sufficient for those preferences to admit a quantile representation. We derive probability measure(s) representing subjective beliefs, and a unique quantile that is maximized by the induced preferences over probability distributions. As expected, utilities on outcomes are ordinal, unique up to all strictly increasing transformations. An important insight delivered by the axiomatization is that the probability measure is unique (and also convex-ranged and finitely additive) for all levels of quantile strictly between 0 and 1. For the extreme τ's equal to 0 or 1, we derive a set of nonatomic measures which are not necessarily finitely additive. This is intuitive: choices of 0- or 1-maximizers are consistent with any measure that assigns strictly positive (and less than one) values to the same outcomes.

We characterize Quantile Maximization through five axioms. Compared to Savage's (1954), our set of conditions retains P1 (Ordering) and a slightly weaker version of P5 (Nondegeneracy). We drop all his remaining axioms, including P2 (the Sure-Thing Principle) and P3 (Eventwise Monotonicity). Savage's P4 (Weak Comparative Probability) is in fact implied by our axioms, but dispensing with P2 requires adding an additional condition that ensures additivity of derived probability measures. This is achieved by the condition that provides the likelihood judgement, induced from preferences over acts, with a weak-order structure, which we call *Comparative Probability* ($P4^Q$). Our central axiom that leads to existence and uniqueness of τ is a new monotonicity condition. The key implication of this axiom is that for any act, there exists an event, called a *pivotal event*, such that exchanging outcomes outside of this event in a way that preserves its rank does not affect preferences over acts. Intuitively, the induced lottery preferences remain unaffected by exchanging parts of cumulative distributions below and above some quantile. We dub this axiom *Pivotal Monotonicity*, $P3^Q$. Finally, due to the ordinality property of the model, the Archimedean axiom typically employed in a Savage setting, P6 (Small Event Continuity), is too strong for our model, as it implies mixture continuity. We weaken it just enough to retain its implications for nonatomicity of probability measures and to ensure that the quantile is left-continuous (*Event Continuity*, $P6^Q$).

While Pivotal Monotonicity seems suggestive about how the quantile may obtain through comparing probability distributions (when probabilities are derived), it does not say that the pivotal event is unique in an act or across acts. Nonetheless, the main challenge in axiomatizing Quantile Maximization was to derive a probability-measure representation for subjective beliefs. We could not directly use either Savage's (1954) or other derivations in the literature, since they rely on some form of mixture continuity and monotonicity which are not present in the model. A starting point in our construction is a definition of a likelihood relation over events induced by the preference relation over acts. According to the commonly used definition by Ramsey (1931), event E is judged 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. In our model, this likelihood relation generates only two equivalence classes: all events are judged either equally likely to the null set or to the whole state space. (For example, think of a median maximizer comparing events with probabilities 0.7 and 0.9.) Our approach is to define a new likelihood relation complete only on a subset of "small" events. The relation embeds enough structure to allow us to derive a probability measure on that subset. We then extend it to the set of all events. In constructing the measure, it is essential that disjoint non-null subsets of the state space can be strictly ranked. For the class of preferences leading to $\tau \in (0, 1)$, we show that this holds. This is not possible for two extreme cases, yielding $\tau = 0$ and $\tau = 1$. As a result, while for $\tau \in (0, 1)$ preferences over acts are not affected by exchanging outcomes on equally likely events, for $\tau = 0$ or $\tau = 1$ they are invariant to swapping outcomes on any disjoint non-null events. As hinted above, choices of 0- and 1-maximizers depend on, and hence reveal, less structure in the primitive acts than do those of individuals with $\tau \in (0, 1)$. For $\tau = 0$ and $\tau = 1$, we derive a set of probability measures.

The axiomatization contributes in two ways to the growing body of literature on probabilistic sophistication initiated by Machina and Schmeidler (1992). To put our results in perspective, the goal of this line of research is to understand when choices of a decision-maker using some decision rule are consistent with her having beliefs that conform to a unique probability measure. Motivated by an observation that Savage's (1954) derivation of subjective probabilities depends on axioms that lead to an expected-utility functional, Machina and Schmeidler (1992) characterized an individual whose choice is based on probabilistic beliefs, but does not necessarily comply with the expected utility hypothesis. Grant

(1995) observed that Machina and Schmeidler's definition and proof still restrict a class of preferences by requiring that the induced lottery preferences satisfy continuity and monotonicity properties: mixture continuity and monotonicity with respect to stochastic dominance. Grant (1995) postulated that the notion of probabilistic sophistication as such should be dissociated from extraneous properties of the induced lottery preferences, and hence of the utility representation of preferences. Nevertheless, Grant's (1995) derivation does use a weaker continuity property (two-outcome mixture continuity). He concludes: "Ideally then, it would be nice to characterize probabilistically sophisticated preferences without requiring the induced risk preferences to exhibit any specific properties save perhaps some form of continuity ..." (p.177). Our characterization of probabilistic sophistication achieves that. In addition, only a weak notion of monotonicity of risk preferences is used: weak stochastic dominance.

As an illustrative example, consider a median maximizer. Her choices violate all axioms in Machina and Schmeidler, except P1 (Ordering), P4 (Weak Comparative Likelihood) and P5 (Nondegeneracy) and all axioms in Grant (1995) except for P1 and P5'. Thus the median maximizer would not be probabilistically sophisticated according to their characterizations.

Another related concern about the developments in probabilistic sophistication, which has not been emphasized so far, is that they impose restrictions on the set of outcomes from which acts are defined. Admittedly, the existence of subjective beliefs about events should not depend on the properties of the set of outcomes. Our result neither assumes nor implies any conditions on the outcome set. One advantage is that it can be used to characterize beliefs of agents without well-defined utility functions. Our results are useful in clarifying some customary interpretations in the literature on probabilistic beliefs.

Applications. Compared to the commonly used choice rules, for instance Expected Utility, the Quantile Maximization model exhibits different theoretical properties (robustness, ordinality, one-dimensional information about preferences) and therefore, it can complement them in applications. We discuss settings in which those properties are desirable (e.g., policy design). From the empirical perspective, the considerably weaker requirements for the knowledge of utility functions and robustness can make the quantile decision rule a useful tool in empirical applications. For example, Quantile Maximization allows a researcher to study risk attitudes without needing first to characterize the concavity of utilities from data.

Related Literature. Maxmin models of choice have been developed by Roy (1952, *safety first* rule), Milnor (1954), Rawls (1971, *justice as fairness* theory[1]), Maskin (1979), Barbera and Jackson (1988), Cohen (1992, *security level*), Segal and Sobel (2002), and others. Studies that formalize maxmax include Cohen (1992, *potential level*), Segal and Sobel (2002) and Yildiz (2004, *wishful thinking*).

The idea of modeling preferences lying between maxmin and maxmax is not new. Hurwicz (1951) and Arrow and Hurwicz (1972) introduced the α-*maxmin* rule, defined as a weighted average of the minimal and maximal outcomes. Unlike in our model, the α-maxmin preferences depend just on the extremes, not on any intermediate realizations; are not based on probability values; and do not preserve the ordinality property of maxmin and maxmax. Other related models involve a combination of Expected Utility and maxmin (Gilboa [1988], Jaffray [1988], Cohen [1992]) or, like *neo-additive capacities* by Chateauneuf, Eichberger and Grant (2002), put a fixed weight on the extreme outcomes and apply Expected Utility for those in between. These concepts were intended to explain deviations from Expected Utility, and they still rely on the expected-utility operator. As a result they behave very differently than the original maxmin and maxmax.

For multiple priors, maxmin was axiomatized by Gilboa and Schmeidler (1989) in what has come to be known as *maxmin expected utility* (*MEU*), and for multiple utilities by Maccheroni (2002). MEU can be thought of as formalizing Wald's (1950) *minimax crite-*

[1]It proposes that behind the veil of ignorance, individuals should choose to maximize wealth of the least well-off.

rion. α-maxmin rule was axiomatized by Ghirardato, Maccheroni, and Marinacci (2004) for multiple priors (α-*MEU*), and in a dual model by Ghirardato (2001).

Manski (1988, *quantile utility* model, *utility mass* model) and Börgers (1993, *pure-strategy dominance*) are rare examples of ordinal concepts in economics. We should mention that in the field of Artificial Intelligence, the ordinal approach to modeling choice has been put forward as a research agenda over the past decade. The goal is to develop decision rules that require less information about utilities and beliefs (*qualitative decision theory*), and can be implemented by information systems, such as recommender systems. Some of the decision criteria that have already been proposed can be interpreted as modeling preferences lying between maxmin and maxmax. (See for example, Boutilier [1994], Dubois *et al.* [2000], Dubois *et al.* [2002] and references therein.)

Recently, Chambers (2005) has studied properties of bounded measurable functions that characterize quantile functions.

Although we have restricted attention to theoretical literature, similar concepts have often appeared in policy or applied settings. We discuss them in the paper.

STRUCTURE OF THE PAPER. The paper proceeds as follows: Section 1.2 defines the decision criterion of Quantile Maximization and characterizes its properties. Section 1.3 states our axioms, provides the main results: the representation theorem and a characterization of probabilistic sophistication, and examines properties of risk preferences. Section 1.4 outlines the proofs for the two results. Section 1.5 relates the axiomatization and our characterization of probabilistic sophistication to the literature. Section 2.2 presents the results for identification with finite data. Section 7 discusses applications of the model. Finally, Section 8 offers concluding remarks. All proofs, unless otherwise noted in the text, appear in Appendices.

1.2 The Quantile Maximization Model

In this section, we formally define Quantile Maximization, intuitively explain how quantile maximizers make choices and describe the key properties of the decision rule.

1.2.1 Model

Let $\mathcal{S} = \{..., s, ...\}$ denote a set of states of the world, and let $\mathcal{X} = \{..., x, y, ...\}$ be an arbitrary set of *outcomes*. An individual chooses among finite-outcome *acts*,[2] maps from states to outcomes. $\mathcal{F} = \{..., f, g, ...\}$ is the set of all such acts. The set of events $\mathcal{E} = 2^{\mathcal{S}} = \{..., E, F, ...\}$ 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 $\mathcal{F}$, which will be defined to be a preference relation and taken to be the primitive of the model. As it will become clear in the sequel, it is convenient to work with the strict binary relation $\succ$. Indifference and weak preference will be defined as usual (here and for all strict binary relations throughout): $f \sim g \Leftrightarrow f \not\succ g$ and $f \not\prec g$, $f \succsim g \Leftrightarrow f \succ g$ or $f \sim g$. Let $\succ_x$ denote the preference relation over certain outcomes, $\mathcal{X}$, obtained as a restriction of $\succ$ to constant acts. We will say that an event E is *null* if for any two acts, $f,\ g$ which differ only on E, we have $f \sim g$.

Define the set of *simple* (finite-outcome) probability distributions over outcomes (lotteries):

$$\mathcal{P}_0(\mathcal{X}) = \left\{ P = (x_1, p_1, ..., x_N, p_N) \middle| \sum_{n=1,...,N} p_n = 1,\ x_n \in \mathcal{X},\ p_n \geq 0, n \in \mathbb{N}_{++} \right\}. \quad (1)$$

Finally, δ_x denotes the degenerate lottery $P = (x, 1)$.

Let π stand for 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

[2] An act f is said to be *finite-outcome* if its outcome set $f(\mathcal{S}) = \{f(s) | s \in \mathcal{S}\}$ is finite.

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 random variable $u(x)$ 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:

$$Q^{\tau}(\Pi_f) = \inf\{z \in \mathbb{R}|\pi[u(f(s)) \leq z]{\geq}\tau\}, \tag{2}$$

while for $\tau = 0$, it is defined as

$$Q^{0}(\Pi_f) = \sup\{z \in \mathbb{R}|\pi[u(f(s)) \leq z]{\leq}0\}. \tag{3}$$

(Cf. Denneberg [1994].)

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

$$f \succ g \Leftrightarrow Q^{\tau}(\Pi_f) > Q^{\tau}(\Pi_g). \tag{4}$$

By analogy with Expected Utility, where the mean is a single estimate of the induced distribution, when choosing among lotteries, a τ-maximizer assesses the value of each lottery by the τ^{th} quantile realization. Put differently, she anticipates that the τ^{th} quantile will be realized. We will call τ an *anticipation level.* Notice that although in general a correspondence, generically in payoffs the set of optimal choices is a singleton.

The model nests two choice rules famous in the literature of choice under risk, namely maxmin and maxmax. Choosing according to *maxmin*, a decision maker will find the minimal outcome in every act she is facing and select the one with the highest minimal outcome:

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

According to *maxmax*, she will pick the act with the highest maximal outcome:

$$f \succ g \Leftrightarrow \max_{\{x \in f(\mathcal{S}) | \pi(x) > 0\}} u(x) > \max_{\{x \in g(\mathcal{S}) | \pi(x) > 0\}} u(x). \quad (6)$$

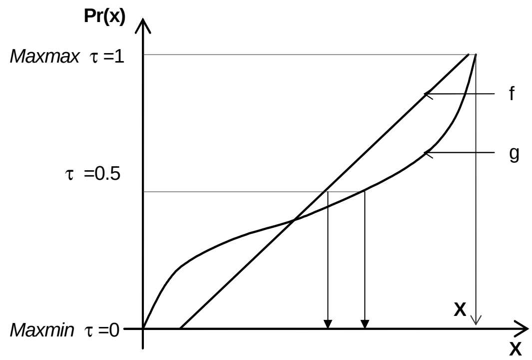
Since the minimal outcome in the support of Π_f is the lowest-quantile outcome and, similarly, the maximand under maxmax is the highest-quantile outcome:

$$Q^0(\Pi_f) = \min_{\{x \in f(\mathcal{S}) | \pi(x) > 0\}} u(x), \ Q^1(\Pi_f) = \min_{\{x \in f(\mathcal{S}) | \pi(x) > 0\}} u(x), \quad (7)$$

the maxmin and maxmax decision makers are, respectively, the 0- and 1-quantile maximizers. Quantile Maximization can therefore be viewed as a generalization of those extreme choice rules to any intermediate level of quantile. While the main focus of the paper will be on finite-outcome acts, in Example 1, we illustrate the relation between maxmin, maxmax and Quantile Maximization using infinite-outcome acts.

Example 1 *Consider an individual facing the two lotteries, induced by some acts f and g, whose cdf's are plotted in Figure 1.1. The* 0*-quantile maximizer would choose f, the* 1*-quantile maximizer would be indifferent, and the median- ($\tau = 0.5$) maximizer would prefer g.*

Figure 1.1 Distributions induced by acts in Example 1



Quantile Maximization is itself an extreme criterion in the sense that the choices result from comparing a single quantile of the distribution. However, the optimal decisions depend on the values of all probabilities and the ranking of all outcomes, while under maxmin and -max they are based only on the support of distribution and on the minimal or maximal outcomes. Consequently, the optimal decisions are not sensitive to outliers, or exclusively based on the extreme outcomes, as they are in maxmin and maxmax.

1.2.2 Properties of the Quantile Maximization rule

In this section, we characterize the main properties of the Quantile Maximization choice rule and contrast it with the Expected Utility approach. We begin with an example, which also explains how τ-maximizers choose among finite-outcome lotteries, which will be the main focus of the paper.

Example 2 *An individual is choosing among acts* f_1, f_2 *and* f_3, *represented by rows in Matrix M1. Let her subjective beliefs about events* E_1 *and* E_2, *represented by columns, be given by* π *and* $1-\pi$, *and let* u *denote her utility function on outcomes. Each entry in the matrix contains a payoff* $u(x)$.

Matrix M1

	E_1	E_2
f_1	11	1
f_2	4	8
f_3	10	3

Figure 1.2 Conditional cdf's for acts in Matrix M1

	1-π	π
f_1	1	11

	π	1-π
f_2	4	8

	1-π	π
f_3	3	10

Suppose that $\pi = \frac{1}{3}$ *and* $1-\pi = \frac{2}{3}$ *and the agent is a median* ($\tau = \frac{1}{2}$) *maximizer. In order to find her optimal choice, we need to transform each conditional lottery induced by each* *act into a cdf (Figure 1.2). In this case, the decision maker would anticipate payoffs of 1, 8, and 3 from* f_1, f_2 *and* f_3*, respectively. Therefore, she would choose* f_2*.*

Several important features of Quantile Maximization differentiate it from other choice rules, for instance Subjective Expected Utility.

(1) *Ordinality:* The optimal decision is invariant to an arbitrary positive monotone transformation of payoffs, because the outcomes affect the choice only through their ranking. In Matrix M1, for any increasing function v, replacing all payoffs $u(x)$ in all lotteries with $v(u(x))$ leaves the choice intact.

(2) *Robustness:* The optimal choice is not affected if, for the fixed τ, the distribution of any induced lottery is perturbed outside of the quantile in a way that preserves the ranking of outcomes. For example, if a payoff of 4 is replaced by -1 or by k ($k \geq 1$) payoffs lower than 8 and jointly assigned probability $\frac{1}{3}$, f_2 will still be optimal.

(3) *One-dimensional information about preferences:* Given the ranking of outcomes, τ specifies the entire preference ordering over lotteries for any beliefs.

(4) *Outcomes need not be objectively measurable:* Risk preferences under Quantile Maximization are well defined even if the outcomes corresponding to payoffs {1, 3, 4, 8, 10, and 11} are replaced with not readily measurable restaurants, movies, political parties, or {war, cold war, treaty, peace, union and cooperation}, and any outcome would be strictly preferred to the one immediately preceding it. There need not be any other relation among them.

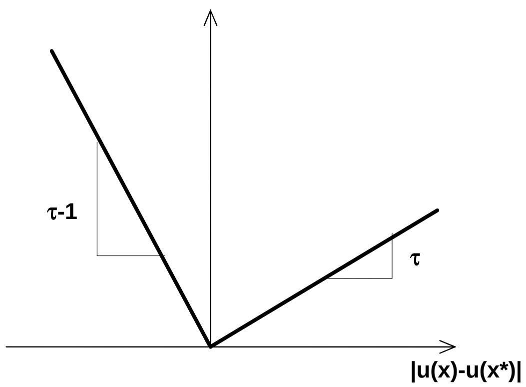
In contrast, under Expected Utility, (1) the choices remain unaffected only by affine transformations of Bernoulli utility over outcomes; (2) any marginal change in payoffs or probabilities affects the value of a lottery; hence, it changes preferences in the sense of changing an equivalence class; (3) specifying (concavity of) Bernoulli utility function is needed to determine preferences over lotteries; and (4) outcomes must be measurable on an interval scale to analyze risk aversion.[3]

To elucidate where these qualitative differences come from, we consider a τ-quantile maximizer evaluating a given lottery or act. It is well known (see e.g. Koenker [2005, Ch. 1]) that this evaluation can be seen as minimizing the loss function (8), defined as a sum of absolute deviations; the solution to this minimization problem is the τ^{th} quantile of the distribution Π, $u(x^*)$. Evaluation (of a given lottery or act) through Expected Utility can be similarly viewed as minimizing a squared symmetric loss function.

$$(1-\tau)\int_{-\infty}^{u(x^*)}[u(x^*)-u(x)]d\Pi(x)+\tau\int_{u(x^*)}^{\infty}[u(x)-u(x^*)]d\Pi(x). \tag{8}$$

An asymmetric weighting function is depicted in Figure 1.3.

Figure 1.3 Weighting function for the loss function (8)



It is useful to think of the solution to minimizing (8) as the agent's prediction or estimation of the realization from the lottery. The lower τ is, the more the agent predicting a realization

[3] An interval scale assumes that the distance between adjacent points on the scale is equal. Section 3.5 provides more examples.

of a lottery is concerned about underpredictions relative to overpredictions - hence, the more she cares about the lower-tail outcomes relative to the higher-tail outcomes. The case when the piecewise linear value function is symmetric corresponds to the median.[4]

That the quantile solves the minimization of absolute rather than quadratic loss function makes it less affected by outliers. This property has been explored in econometrics in robust estimation. Mean-based estimators, such as OLS, are very sensitive to large errors and to asymmetric distributions, often met in practice. A popular alternative estimator, which is more robust, is based on the median or quantile (Least Absolute Deviations, or LAD, method; Koenker and Bassett [1978b]). Just as the mean estimator minimizes the sum of squared errors, the quantile-based estimator minimizes a weighted sum of absolute deviations.

To further examine which properties of the quantile have been appealing in practice, we mention another famous example of Quantile Maximization being used by practitioners (possibly supplemented by other information): *Value-at-Risk* (*VaR*). Given a confidence level $a \in (0,1)$, VaR is the loss in market value that is exceeded with probability $1-\alpha$, which is the $(1-\alpha)^{th}$ quantile of the loss distribution. VaR was developed to provide a single number that could aggregate the several components of risk to convey information about the risk in a portfolio, could be easily calculated and conveniently interpreted, and would focus attention on the so-called "normal market condition." (See Duffie and Pan [1997], Artzner *et al.* [1999], and VaR's Web site: http://www.GloriaMundi.org.)

Practitioners find it a more attractive measure of risk than variance:

> "Modern Portfolio Theory ("MPT"), as taught in business schools, tells us that the risk in a portfolio can be proxied by the portfolio standard deviation, a

[4] In addition, such formulation provides another way of showing that the optimal decision of a quantile maximizer is not exclusively based on the extreme outcomes, but rather depends on the ranking of all outcomes and the values of all probabilities, unlike in maxmin and maxmax.

measure of spread in a distribution. That is, standard deviation is all you need to know in order to (1) encapsulate all the information about risk that is relevant, and (2) construct risk-based rules for optimal risk "management" decisions. [...] standard deviation loses its appeal found in MPT. First, managers think of risk in terms of dollars of loss, whereas standard deviation defines risk in terms of deviations, either above or below expected return and is therefore not intuitive. Second, in trading portfolios deviations of a given amount below expected return do not occur with the same likelihood as deviations above, as a result of positions in options and option-like instruments, whereas the use of standard deviation for risk management assumes symmetry." (Schachter [1997, p. 19])

In addition, standard deviation requires that the second moment of distribution be finite, which is a problem, for example, in non-life insurance.[5]

1.3 Axiomatic foundations of Quantile Maximization

In this section, we derive restrictions on preferences implied by Quantile Maximization. Specifically, the challenge is to formulate axioms on preferences over acts that are necessary and sufficient for there to exist a unique number τ in $[0, 1]$, probability measure(s) π rep-

[5] Although used long before, VaR became popular among trading institutions during the 1990s with the influential report on derivatives practices of the Group of 30 in 1993, the RiskMetrics service launched by JP Morgan in 1994 to promote the use of VaR, and the market risk capital requirements set for banks by the Basel Committee on Banking Supervision 1995. Today, it is being advocated by the Federal Reserve Bank and the Securities and Exchange Commission, the Bank for International Settlements, and it is a widespread risk management tool in finance, banking and insurance.

Value-at-Risk is used in setting position limits for traders, in capital allocation, to incentivize traders through risk-capital charges based in VaR not to take on excessive risk, and in adjusting the performance of risk. It is used to compare risky activities in diverse markets, but the total risk of the firm can also be broken down into "incremental" Value-at-Risk to uncover positions contributing most to total risk.

Clearly, VaR does not capture all relevant information about market risk. It does not say what the largest losses could be. Nor does it measure "event" (e.g., market crash) risk, and for that reason it is supplemented with stress tests. VaR is based on a particular forecast horizon, whereas a disadvantageous economic environment may extend beyond that horizon. In addition, VaR models typically assume that the portfolio under consideration is constant over the forecast horizon (there is ongoing research on dynamic VaR, e.g. Rogachev [2002]). Another strong assumption is that the past data used to construct the VaR estimate contains information useful in forecasting the distribution of losses.

resenting beliefs, and utilities on outcomes u, all derived from preferences over acts, that guide the behavior of a quantile maximizer.

1.3.1 Axioms

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

AXIOM P1 (**ORDERING**): *The relation $\succ$ is a weak order.*

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

$$E_{fx+} = \{s \in \mathcal{S} | f(s) \succ x\} \tag{9}$$

$$E_{fx-} = \{s \in \mathcal{S} | f(s) \prec x\} \tag{10}$$

Note that since the acts are finite-ranged, every act induces a natural partition of the state space 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 $g_{x+} : E_{fx+} \to \mathcal{X}$ with $g_{x+}(s) \succsim x$, for all $s \in E_{fx+}$ and similarly, let g_{x-} be any map $g_{x-} : E_{fx-} \to \mathcal{X}$ with $g_{x-}(s) \precsim x$, for all $s \in E_{fx-}$.

AXIOM P3^Q (**PIVOTAL MONOTONICITY**): *For any act $f \in \mathcal{F}$, there exists a non-null event E_f such that $f^{-1}(x) = E_f$ for some $x \in f(\mathcal{S})$, and for any outcome y, and* *subacts* $g_{x+},\ g_{x-},\ g_{y+},$ *and* g_{y-} :

$$\begin{bmatrix} g_{x+} \text{ if } E_{f+} \\ x \text{ if } E_f \\ g_{x-} \text{ if } E_{f-} \end{bmatrix} \succsim \begin{bmatrix} g_{y+} \text{ if } E_{f+} \\ y \text{ if } E_f \\ g_{y-} \text{ if } E_{f-} \end{bmatrix} \Leftrightarrow x \succsim y. \tag{11}$$

Before we explain the many roles this axiom serves, we first interpret the following important implication: for an act $f \in \mathcal{F}$, event E such that $f^{-1}(x) = E$ for some $x \in f(\mathcal{S})$, and all subacts $g_{x+},\ g_{x-},\ g_{y+},$ and g_{y-} define

$$f_E = \begin{bmatrix} g_{x+} \text{ if } E_{fx+} \\ x \text{ if } E \\ g_{x-} \text{ if } E_{fx-} \end{bmatrix}; \tag{12}$$

It follows from $\mathbf{P3}^Q$ that for any act $f \in \mathcal{F}$, there exists a non-null event E_f such that $f^{-1}(x) = E_f$ for some $x \in f(\mathcal{S})$ and

$$f \sim f_{E_f}. \tag{13}$$

The last condition states that for a given act, there exists an event, which will be called a *pivotal event*, such that changing outcomes outside of that event in a (weakly) rank-preserving way does not affect preferences over acts, a form of separability. Crucially, what is fixed during this transformation are the events assigned to outcomes which in the original act f are either strictly preferred or less preferred to x, the outcome on the pivotal event. The requirement that the act f be constant for the pivotal event ensures that the conditions (11) and (13) are non-trivial. (Otherwise, the state space could be taken as pivotal for any act.) This axiom will be the key to guaranteeing the existence and uniqueness of a quantile $\tau \in [0, 1]$. Intuitively, it implies that the induced preferences over lotteries will not be changed by replacing parts of the cumulative probability distributions below and above some quantile.

Moreover, as the name suggests, **P3**^{Q} provides preferences over acts with an appropriate, local notion of monotonicity. It states that replacing any outcome y on the pivotal event by a (weakly) preferred outcome x always leads to a (weakly) preferred act. It is noteworthy that it suffices that the preference is monotonic on the pivotal event only, since the axioms jointly allow for extending the monotonicity to the whole state space.

In addition, together with other axioms, Pivotal Monotonicity will imply that the utility over outcomes, $u(x)$, is event independent. Informally, this ensures that how an outcome in an act is assessed by a decision maker depends only on the likelihood of the event to which the outcome is assigned, and not on the event itself. In a similar manner, **P3**^{Q} will render the property of being pivotal state-independent.

Finally, together with **P1**, **P3**^{Q} gives rise to the "more likely than" judgment of events, derived from preferences over acts.

AXIOM P4^Q (**COMPARATIVE PROBABILITY**): *For all pairs of disjoint events E and F, outcomes $x^* \succ x$, and subacts g and h,*

$$\begin{bmatrix} x^* \text{ if } s \in E \\ x \text{ if } s \in F \\ g \text{ if } s \notin E \cup F \end{bmatrix} \succ \begin{bmatrix} x^* \text{ if } s \in F \\ x \text{ if } s \in E \\ g \text{ if } s \notin E \cup F \end{bmatrix} \Rightarrow \begin{bmatrix} x^* \text{ if } s \in E \\ x \text{ if } s \in F \\ h \text{ if } s \notin E \cup F \end{bmatrix} \succsim \begin{bmatrix} x^* \text{ if } s \in F \\ x \text{ if } s \in E \\ h \text{ if } s \notin F \cup F \end{bmatrix}. \tag{14}$$

P4^Q ensures that the likelihood relation over events, induced from preferences over acts, is a weak order[6] and it provides its representation with a finitely additive form. Notice that

[6] We note that the standard likelihood relation, defined below in (21), inherits the weak-order property directly from **P1** and **P3**^{Q}. **P4**^{Q} gives the weak-order structure to the likelihood relation used in this paper, defined in Section 1.4. We further note that when **P4**^{Q} is dispensed with, the derived represenations of beliefs will be capacities.

no events are required to be non-null.

AXIOM P5 (**NONDEGENERACY**): *There exist acts f and g such that $f \succ g$.*

This is a standard non-triviality condition.[7] 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 leads to the uniqueness of a probability measure.

Before we state the final axiom, we identify an interesting class of preferences. It is convenient to define the following two cases:

(**L**, "lowest"): For any act $f \in \mathcal{F}$, the pivotal event assigns the least-preferred outcome in $\{x \in \mathcal{X} | x \in f(\mathcal{S})\}$. (15)

(**H**, "highest"): For any act $f \in \mathcal{F}$, the pivotal event assigns the most-preferred outcome in $\{x \in \mathcal{X} | x \in f(\mathcal{S})\}$. (16)

Intuitively, these preferences will lead to $\tau = 0$ and $\tau = 1$, respectively.

Definition 2 *A preference relation over acts $\mathcal{F}$, $\succ$, satisfying **P3**^{Q}, is called* extreme *if either **L** or **H** holds.*

We define two continuity properties that will be used in the axiom.

[7]Strictly speaking, it is weaker than the commonly used **P5** of Savage (see Appendix 1), but the two conditions are equivalent in the presence of other axioms.

($\mathbf{P6}^{Q_*}$) *Fix a pair of events* $E, F \in \mathcal{E}$. *If for any pair of outcomes such that* $x \succ y$,

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix}, \tag{17}$$

then there exists a finite partition $\{G_1, ..., G_N\}$ *of* $\mathcal{S}$ *such that*

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin F \cup G_n \\ y \text{ if } s \in F \cup G_n \end{bmatrix} \tag{18}$$

for all $n = 1, ..., N$.

($\mathbf{P6}^{Q^*}$) *Fix a pair of events* $E, F \in \mathcal{E}$. *If for any pair of outcomes such that* $x \succ y$,

$$\begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \notin E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \in F \\ y \text{ if } s \notin F \end{bmatrix}, \tag{19}$$

then there exists a finite partition $\{H_1, ..., H_M\}$ *of* $\mathcal{S}$ *such that*

$$\begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \notin E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \in F \cup H_m \\ y \text{ if } s \notin F \cup H_m \end{bmatrix} \tag{20}$$

for all $m = 1, ..., M$.

AXIOM P6^Q (**EVENT CONTINUITY**): *For non-extreme preferences, the relation* $\succ$ *satisfies* $\mathbf{P6}^{Q_*}$ *for all events in* $\mathcal{E}$ and $\mathbf{P6}^{Q^*}$ *for any event* E *in* $\mathcal{E}$ *and* $\varnothing$. *If* **H** *holds,* $\succ$ *satisfies* $\mathbf{P6}^{Q_*}$, *while if* **L** *holds,* $\succ$ *satisfies* $\mathbf{P6}^{Q^*}$.

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

We close this section by remarking that the conditions ($\mathbf{P6}^{Q_*}$) and ($\mathbf{P6}^{Q^*}$) can be interpreted in terms of likelihood relations. Although the definition of likelihood we adopt differs from the commonly used one (see Section 1.4), the standard definition still appears to reveal useful structure, which is employed in $\mathbf{P6}^{Q}$. Formally, as defined by Ramsey (1931) and adopted by Savage (1954), the likelihood relation $\succ^*$, a binary relation on $\mathcal{E}$, is implicitly defined by Savage's **P4** (Appendix 1), implied by our **P1** and $\mathbf{P4}^{Q}$:

$$E \succ^* F \text{ if for all } x \succ y, \quad \begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \notin E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \in F \\ y \text{ if } s \notin F \end{bmatrix}. \tag{21}$$

We also use the following definition, which maps the events whose likelihood is assessed to the *less* preferred outcome:

$$E \succ_* F \text{ if for all } x \succ y, \quad \begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix}. \tag{22}$$

Given these definitions, ($\mathbf{P6}^{Q_*}$) and ($\mathbf{P6}^{Q^*}$) can be restated as follows:

($\mathbf{P6}^{Q_*}$, restated): *Fix a pair of events $E, F \in \mathcal{E}$. If $E \succ_* F$, then there exists a finite partition $\{G_1, ..., G_N\}$ of $\mathcal{S}$ such that $E \succ_* F \cup G_n$, for all $n = 1, ..., N$.*

($\mathbf{P6}^{Q^*}$, restated): *Fix a pair of events $E, F \in \mathcal{E}$. If $E \succ^* F$, then there exists a finite partition $\{H_1, ..., H_M\}$ of $\mathcal{S}$ such that $E \succ^* F \cup H_m$, for all $m = 1, ..., M$.*

In all cases leading to $\tau \in (0,1)$, definition (22) is used. The reason for altering the definition is that the commonly used definition (21) would yield right-continuity of the quantile representation functional; we follow the convention in the literature and define quantiles as left-continuous. The distinctive formulation of the condition in $\mathbf{P6}^Q$ for the subclass of extreme preferences is due to the fact that, in this case, $\mathbf{P6}^{Q^*}$ fails. (Section 1.4 clarifies this point.)

1.3.2 Probabilistic sophistication

This section presents the first of two central theorems of the paper. The result identifies a condition on preferences that satisfy axioms **P1**, $\mathbf{P3}^Q$, $\mathbf{P4}^Q$, **P5**, $\mathbf{P6}^Q$ under which those preferences are probabilistically sophisticated.

THE MEANING OF PROBABILISTIC SOPHISTICATION. We begin with clarifying the concept of probabilistic sophistication. The discussion serves three purposes. First, it explicates the conditions implicit in the properties in terms of which probabilistic sophistication is defined. Second, the qualifications we make are important in results that state probabilistic sophistication as an assumption.[8] Finally, it motivates our definition of probabilistic sophistication.

The original question asked by Machina and Schmeilder (1992) that stimulated research on probabilistic sophistication was:

> "What does it take for choice behavior that does not necessarily conform to the expected utility hypothesis to nonetheless be based on probabilistic beliefs [i.e., beliefs that conform to a unique probability measure, M.R.]?" (p.747)

Although there is a consensus[9] that derivation and hence definition of probabilistic sophistication should be independent from the conditions implying some specific properties of

[8] E.g. Grant (1995), Propositions 2.1 and 4.1; Propositions 3 and 4 in this paper.
[9] See e.g. Grant (1995), Epstein and Zhang (2001), Grant and Polak (2005), Chew and Sagi (2005).

the utility representation functional, such as continuity or monotonicity, a formal definition has been evolving. In the literature, the following conceptualizations or interpretations of probabilistic sophistication are customary:

(1)[10] Fix a probability measure π on the set of events. 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 $f \to \pi \circ f^{-1}$. A decision maker is probabilistically sophisticated if she is indifferent between two acts that induce identical probability distributions over outcomes. Formally,

$$\left(P = Q,\ \pi \circ f^{-1} = P,\ \pi \circ g^{-1} = Q\right) \Rightarrow f \sim g. \tag{23}$$

(2)[11] Define a preference relation over lotteries, $\succsim_P$, induced from the underlying preferences over acts $\succsim$:

$$\text{If } \pi \circ f^{-1} = P,\ \pi \circ g^{-1} = Q \text{ for some } f, g \in \mathcal{F}, \text{ then } (f \succsim g \Rightarrow P \succsim_P Q). \tag{24}$$

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

$$\left(P \succsim_P Q,\ \pi \circ f^{-1} = P,\ \pi \circ g^{-1} = Q\right) \Rightarrow f \succsim g. \tag{25}$$

The property (2) entails that the preferences over lotteries $\succsim_P$ contain all and only the information about the preferences over acts $\succsim$. Thus preferences over acts $\succsim$ can be recovered from the knowledge of π and lottery preferences $\succsim_P$ alone.

It is important to note that both properties (1) and (2) entail that preferences and

[10] This interpretation (possibly different from the formal definition emloyed) is present in Epstein and Le Breton (1993, p.8), Grant (1995, p. 163), Chew and Sagi (2005, p.2).
[11] Grant (1995, p. 162), Abdellaoui and Wakker (2005, pp. 18-19).

utilities be state-independent. Moreover, although under some assumptions, (1) and (2) are equivalent, in general (2) implies (1).[12] On the other hand, if the model satisfies the extra conditions implicit in definition (2), the interpretation of probabilistic sophistication as informational equivalence between the content of $\succsim$ and $\succsim_P$ given π provides a convenient tool. (See Grant [1995], Grant and Polak [2005].)

The property (1) establishes a mapping between subsets of equivalence classes of $\mathcal{F}$ under $\sim$ and lotteries in $\mathcal{P}_0(\mathcal{X})$. The property (2) sets up a bijection between equivalence classes of $\mathcal{F}$ under $\sim$ and of $\mathcal{P}_0(\mathcal{X})$ under $\sim_P$. Under definition of $\succsim_P$ in (24), condition (25) ensures that the map is one-to-one, while convex-rangedness of the measure π makes it onto. Respectively, this means that the agent is indifferent between any pair of acts that are mapped to indifferent lotteries, and that given the measure π, the preferences over lotteries can usefully characterize preferences over acts.[13]

Motivated by those considerations, we adopt (1), with its implicit assumptions that are satisfied in our model, to formally define probabilistic sophistication. The definition is free from extraneous implications about preferences or measures as well as from requirements for the utility functional, and helps answer the original question posed by Machina and Schmeidler (1992).[14]

[12] To see that, consider the class of all preference relations over acts that imply existence of a u n i q u e probability-measure representation of agent's beliefs, π. (Below, we comment on this restriction.) What (2) does and (1) does not require to be well defined is that the probability measure be convex-ranged not just nonatomic so that the mapping from $\mathcal{F}$ to $\mathcal{P}_0(\mathcal{X})$ is onto. If the relation $\succ$ is a weak order and π is convex-ranged, it is straightforward to show that (1) is equivalent to (2). This, however, imposes restrictions on preferences over lotteries and measure which are not required in (1) a priori.

[13] We stress that separation of beliefs from risk preferences, one of the original motivations of Ramsey, Anscombe and Aumann, and Savage, concerns yet another bijection. Namely, it concerns the bijection between the set of equivalence classes of $\mathcal{P}_0(\mathcal{X})$ under $\sim_P$ and the equivalence set of $\mathcal{X}$ under $\sim_x$, with a slight abuse of notation. Again, this would require that risk preferences are beliefs-independent, which is implied by state-independence. This implicit assumption is apparent in the following interpretation of probabilistic sophistication: "beliefs affect choices only through their subjective likelihood."

[14] Unlike in many papers, we chose not to incorporate the conditions that the measure be finitely additive and convex-ranged or nonatomic in the definition of probabilistic sophistication. Whether they hold or not depends on the structural properties of the primitives, which are not essential to uniqueness and hence to probabilistic sophistication.

Characterization. Theorem 1 derives subjective beliefs of individuals whose preferences over acts satisfy **P1**, $\mathbf{P3}^Q$, $\mathbf{P4}^Q$, **P5**, $\mathbf{P6}^Q$.

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

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

(2) If the relation $\succ$ is extreme, there exists a set of capacities $\Pi(\mathcal{E})$ on $\mathcal{E}$ such that the conditions (23), and ((24),(25)) hold for any $\tilde{\pi} \in \Pi(\mathcal{E})$.

Our result gives more insight into the relation between the notion of probabilistic sophistication and properties (1) and (2). Theorem 1 demonstrates that conditions (23), and ((24), (25)) hold under both non-extreme and extreme preferences. Condition (1) is satisfied even if each act is evaluated through a different probability measure in $\Pi(\mathcal{E})$. For (2), although the choices of agents with extreme preferences are consistent with a set of beliefs, $\Pi(\mathcal{E})$, the knowledge of that set and their lottery preferences does enable us to recover the agents' entire preference relation over all acts, even with measures that are only nonatomic and not convex-ranged. To see why, as we already hinted above, the extreme preferences will lead to cases with $\tau = 0$ (maxmin) and $\tau = 1$ (maxmax). Roughly speaking, the choices of those quantile-maximizers will not change as long as the probability measure assigns strictly positive values to the same events. Any such measure will represent the same preferences.[15]

Consequently, individuals with extreme preferences would be probabilistically sophisticated if the property (1) or (2) was adopted as a definition without the uniqueness requirement.

[15] Our proof of Theorem 1 characterizes some additional properties of the likelihood relation (and thus the set of measures) for the extreme-preference agents: (1) their preferences can tell not only whether an event is null (measure-zero) but also whether it differs from $\mathcal{S}$ on a non-null event, (2) they can rank nested events. In Section 1.4, we explain at level of preferences what causes the differences in the properties of the probability measures representing beliefs of individuals with non-extreme compared to extreme preferences.

Discussion. In all of the literature, the concept of probabilistic sophistication involves a unique probability measure. Uniqueness is also explicit in the question raised by Machina and Schmeilder (1992), cited above. Clearly, by itself, uniqueness of the measure-theoretic representation is not meaningful as a definition of probabilistic sophistication. (Take $V(F) = W(f) + \int \pi = W(f) + 1$.) Nor is it the goal to merely represent (exogenous) relative likelihood of events by a probability measure. Rather, being a property of the preference relation over acts, probabilistic sophistication is a choice-theoretic concept in that beliefs only exist insofar as they are revealed in choice behavior. At the least, it is uniqueness jointly with some other property that gives probabilistic sophistication empirical content (e.g., conditions (23), or (24) and (25)). Theorem 1 suggests that an alternative definition of probabilistic sophistication might be worth considering: under some conditions on the relation $\succsim$, there is a set of probability measures such that for any member of this set, the property (1) or (2) holds. Any such measure would be consistent with the same preference ordering on $\mathcal{F}$.

Moreover, if the requirement of uniqueness is dropped, definition (2) above may accommodate state-dependent utilities (see Karni [1996], Karni and Schmeidler [1993]). The case in which utilities are state-dependent provides another argument in favor of dispensing with the uniqueness condition. As pointed out by Karni (1996), even if individuals act upon their beliefs, and these beliefs can be represented by probabilities, they need not coincide with the choice-theoretic beliefs derived within the Subjective Expected Utility model. This is because, according to Savage's theory, the definition of probabilities assumes that utility functions are state-independent, and the derived probabilities and utilities are unique only when derived jointly.

It seems that independence of tastes from states and beliefs should not be essential for probabilistic sophistication, with or without the uniqueness restriction. This is consistent with the idea that under probabilistic sophistication the events are distinguished only by their subjective probabilities. **P3** (Eventwise Monotonicity: "tastes are independent from states and beliefs") only enters into Savage's derivation of a probability measure to show that non-null events are judged more likely than is the empty set. In fact, this is exactly how Chew and Sagi (2006) use their weakening of **P3**.

1.3.3 Representation theorem

This section presents the second main result of the paper: providing a complete characterization of a quantile maximizer. So far, no restriction has been imposed on the set $\mathcal{X}$. What is worth emphasizing is that, unlike in any other available representation of preferences over acts, the necessity part in Theorem 1 (or, our axioms) does not imply any structure for $\mathcal{X}$. In the next theorem, we need to add the condition that $\mathcal{X}$ contains a countable $\succ$-order dense subset.[16] This assumption is needed only for the existence of a utility function on $\mathcal{X}$, which is, in turn, needed for a numerical representation of $\succ$.[17]

Theorem 2 *Consider $\succ$ a preference relation over $\mathcal{F}$. The following are equivalent:*

*(1) $\succsim$ satisfies the axioms **P1**, **P3**^{Q}, **P4**^{Q}, **P5**, **P6**^{Q},*

(2) There exist:

(i) a number $\tau \in [0, 1]$,

(ii) probability measure(s) π from Theorem 1;

(iii) a utility function on $\mathcal{X}$, u, which represents $\succ_x$, where u is unique up to strictly increasing transformations;

such that the relation $\succ$ over acts can be represented by the preference functional

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

[17] In fact, without the additional assumption, the conditions in Theorem 1 give a non-numerical version of both necessity and sufficiency parts of Theorem 2.

Given a probability measure on $\mathcal{E}$, π, for a fixed act f, let $\Pi_f^\tau \in \mathcal{X}$ be the τ^{th}-quantile of the cumulative probability distribution Π_f, where the outcomes in the domain of Π_f are ordered by $\succ_x$. Let u be a utility function that represents $\succ_x$. Then, $u \circ \Pi_f^\tau \in \mathbb{R}$ is a numerical representation of the quantile Π_f^τ, $Q^\tau(\Pi_f)$.

Theorem 2 can now be restated with (2) replaced by: There exist (i) a number $\tau \in [0, 1]$ and (ii) probability measure(s) π from Theorem 1 such that for any $f, g \in \mathcal{F}$

$$\Pi_f^\tau \succ \Pi_g^\tau \;\Leftrightarrow\; f \succ g.$$

$\mathcal{V}(f) : \mathcal{F} \to \mathbb{R}$ *given by*

$$\mathcal{V}(f) = Q^{\tau}(\Pi_f). \tag{26}$$

The result states that the preferences of a quantile maximizer satisfy the axioms **P1***,* **P3**^{Q}*,* **P4**^{Q}*,* **P5***,* **P6**^{Q}*,* and conversely, an individual whose preferences conform to those axioms can be viewed as a quantile maximizer according to definition 1. Thereby, given the utility u, the choice mechanism is decomposed in two factors: an anticipation level, τ, which is assured to be unique, and probability measure(s) π, unique for all $\tau \in (0,1)$. Quantile maximizers with $\tau = 0$ or $\tau = 1$ are not probabilistically sophisticated (as defined in (23)).

1.3.4 Properties of lottery preferences

Using Theorem 1, each act in $\mathcal{F}$ can be mapped to a lottery in $\mathcal{P}_0(\mathcal{X})$. Since the derived measure is convex-ranged for non-extreme preferences, this mapping is onto. In this section, we characterize the substitution, continuity and monotonicity properties of the induced lottery preferences. We look for the properties of the relation $\succ_P$ on $\mathcal{P}_0(\mathcal{X})$ which are tight in the sense that they are equivalent to some properties (axioms) on the relation $\succ$ on $\mathcal{F}$. Throughout, we delineate differences between Quantile Maximization and the natural benchmark of Subjective Expected Utility. Savage's (1954) axioms are listed in Appendix 2.

Relation to Savage's (1954) axioms. The condition that Quantile Maximization literally shares with the Subjective Expected Utility is the purely structural axiom **P1** (Ordering). Our set of axioms weakens the monotonicity (**P3**, Eventwise Monotonicity) and continuity (**P6**, Small-Event Continuity) properties of preferences. We drop the Sure-Thing Principle (**P2**), which results in a strengthening of the axiom leading to the derived likelihood relation **P4** (Weak Comparative Probability). We also use a slightly weaker version of Nondegeneracy, **P5**. These axioms are relaxed just enough to incorporate conditions leading to the existence of a unique quantile.

Not surprisingly, **P2** (the Sure-Thing Principle) is too strong for quantile maximization.[18] Similarly to **P3** (Eventwise Monotonicity), **P2** fails when a change in the common subact affects how other outcomes rank in an act. Precisely, what fails is the quantification "for all subacts." Even though the implications of the original axioms **P2** and **P3** are very different, after modifying the class of subacts allowed in **P3**, the amended **P2** and **P3** would be implied and equivalent. Our new axiom, **P3**^{Q} (Pivotal Monotonicity), weakens yet another quantifier of **P3**: "for all events." In the presence of other axioms, **P2** still holds, even after the second relaxation of quantifiers, but it has no independent implications. This weakening of **P2** does not, however, preserve the structure in preferences that was used in the Subjective Expected Utility model to obtain additivity of the probability measure. To recover additivity, we strengthen **P4** (Weak Comparative Probability) to Comparative Probability (**P4**^{Q}).

The Archimedean axiom from Subjective Expected Utility theory, **P6** (Small-Event Continuity), does not hold under Quantile Maximization. To see why, fix $\tau = 1$ and $f \succ g$. Then, taking $x > \max_{n=1,\ldots,N}\{x_n \in f(\mathcal{S})\}$, gives

$$f \prec \begin{bmatrix} x \text{ if } s \in E_n \\ g \text{ if } s \notin E_n \end{bmatrix} \tag{27}$$

irrespective of how small E_n (and $\pi(E_n)$) is.[19] The original **P6** ensures both that no consequence is infinitely desirable or undesirable, as well as that the derived probability

[18] For example, consider three equally likely events E_1, E_2 and E_3. A median maximizer ($\tau = 0.5$) prefers act $\begin{bmatrix} 3 \text{ if } E_1 \\ 2 \text{ if } E_2 \\ 0 \text{ if } E_3 \end{bmatrix}$ to act $\begin{bmatrix} 4 \text{ if } E_1 \\ 1 \text{ if } E_2 \\ 0 \text{ if } E_3 \end{bmatrix}$, but she prefers $\begin{bmatrix} 5 \text{ if } E_1 \\ 4 \text{ if } E_2 \\ 1 \text{ if } E_3 \end{bmatrix}$ to $\begin{bmatrix} 5 \text{ if } E_1 \\ 3 \text{ if } E_2 \\ 2 \text{ if } E_3 \end{bmatrix}$.

[19] Analogous counter-examples can be constructed for any level of τ. We provide a sketch of the construction here, though the exact argument relies on results proved in the sequel. For an intermediate value of τ, $\tau \in (0,1)$, take an act with the pivotal event E_f, $f(E_f) = x$ and such that $\pi(E_f \cup E_{f-}) = \tau$. On a nonnull subevent of E_f, $\hat{E}_f$, replace x with $z \succ x$ to obtain the preference reversal $f \prec \begin{bmatrix} z \text{ if } s \in \hat{E}_f \\ g(\cdot) \text{ if } s \notin \hat{E}_f \end{bmatrix}$.

measure is nonatomic. Crucially, what τ-maximization violates is the former but not the latter. Therefore, we can weaken the axiom **P6** to **P6**^{Q} (Event Continuity), which retains only the continuity implications for probabilities.

(Lack of) continuity. Interestingly, the implied risk preferences will not be mixture continuous,[20] not even for 2-outcome lotteries.[21] Mixture continuity is typically implied by **P6** (Small-Event Continuity), which is too strong for Quantile Maximization. Our **P6**^{Q} removes any continuous structure on mixture lotteries from risk preferences.

Substitution. Given that in the restricted class of subacts and events in **P3**^{Q}, the implications of **P2** and **P3** are equivalent, it is especially interesting to ask what is the property of the induced lottery preferences they jointly engender. Failure of **P2** implies that the induced lottery preferences need not obey the *Independence Axiom*. Since **P3** also fails, they need not exhibit the substitution axiom of Grant, Kajii and Polak (1992), the *Axiom of Degenerate Independence*, **ADI**. We relegate the precise relation to Appendix 5A.

Monotonicity. Strong monotonicity with respect to first-order stochastic dominance[22] need not hold under Quantile Maximization. (One way to show it is to use the result that $\succsim_P$ respects **ADI** if and only if it satisfies first-order stochastic dominance. See for example,

[20] $V : P_0(X) \to \mathbb{R}$ is said to be *mixture continuous* if for any lotteries P, Q and R in $\mathcal{P}_0(\mathcal{X})$, the sets $\{\lambda \in [0,1] |\ V(\lambda P + (1-\lambda)Q) > V(R)\}$ and $\{\lambda \in [0,1] |\ V(R) > V(\lambda P + (1-\lambda)P)\}$ are open.

$\succ_P$ is *mixture continuous if* for all distributions P, Q and R in $\mathcal{P}_0(\mathcal{X})$, the sets $\{\lambda \in [0,1] |\ \lambda P+(1-\lambda)Q \succ_P R\}$ and $\{\lambda \in [0,1] |\ R \succ_P \lambda P + (1-\lambda)Q\}$ are open.

[21] See Section 1.5.

[22] For an arbitrary outcome set $\mathcal{X}$, given a complete preorder over outcomes $\succsim_x$, $P = (x_1, p_1; ...; x_N, p_N)$ *weakly first order stochastically dominates (FOSD)* $Q = (y_1, q_1; ...; y_M, q_M)$ with respect to $\succsim_x$ if

$$\sum_{\{n|x \succsim_x x_n\}} p_n \leq \sum_{\{m|x \succsim_x x_m\}} q_m \text{ for all } x \in \mathcal{X} \tag{28}$$

and if in addition (28) holds with strict inequality for some $y \in \mathcal{X}$, then P *strictly FOSD* Q with respect to $\succsim_x$.

It is said that $\succsim_P$ is *monotonic with respect to first order stochastic dominance* if $P \succsim_P (\succ_P) Q$ whenever P weakly (strictly) stochastically dominates Q.

Grant [1995].) To characterize the appropriate monotonicity property for the model, it is useful to find the analogs of axioms **P1-P6**^{Q} for the case of a known and unique probability (risk). We only need to state the counterpart of Pivotal Monotonicity (**P3**^{Q}), which we call *Rankwise Monotonicity.*[23] For a fixed lottery $P \in \mathcal{P}_0(\mathcal{X})$ and outcome x such that x belongs to the support of lottery P, $x \in supp\{P\}$, define

$$P_{x+} = \sum_{\{n|x_n \succ_P x\}} p_n, \; P_{x-} = \sum_{\{m|x_m \prec_P x\}} p_m \tag{29}$$

and let Q_{x+} and Q_{x-} be any sublotteries on P_{x+} and P_{x-} with supports such that $supp\{Q_{x+}\} \succsim_P x$ and $supp\{Q_{x+}\} \precsim_P x$, respectively.

AXIOM P3$^{Q^P}$ (**RANKWISE MONOTONICITY**) For all simple lotteries $P \in \mathcal{P}_0(\mathcal{X})$, there is an outcome $x \in supp\{P\}$ such that for any outcome y, and sublotteries Q_{x-}, Q_{x+}, Q_{y-}, Q_{y+}:

$$x \succsim_x y \Leftrightarrow (Q_{x-}, P_{x-},; \; x, p_x; \; Q_{x+}, P_{x+}) \succsim_P (Q_{y-}, P_{x-}; \; y, p_x; \; Q_{y+}, P_{x+}). \tag{30}$$

The intuition behind this axiom is similar to that of **P3**^{Q}. By **P1**^{P} and **P5**^{P} denote axioms of weak order and nondegeneracy (as defined by **P1** and **P5**) for the binary relation $\succ_P$. For a given lottery $P \in \mathcal{P}_0(\mathcal{X})$ and utility on outcomes u, let $Q^\tau(P)$ be the τ^{th} quantile of the cumulative probability distribution corresponding to P.[24] The following Corollary of Theorem 2 axiomatizes Quantile Maximization under risk.

Corollary 1 *A binary relation on the set of lotteries $\mathcal{L}$ satisfies **P1**^{P}, **P3**$^{Q^P}$ and **P5**^{P} if and only if there exists a number $\tau \in [0,1]$ and a function $u : \mathcal{X} \to \mathbb{R}$ such that the relation $\succ_P$ over simple probability distributions can be represented by the preference functional*

[23] Cf. Ordinal Independence axiom in Jullien and Green (1988) and Irrelevance Axiom in Segal (1989).

[24] In this section we do not constrain the quantile to be left- or right- continuous. Also, we assume that $\mathcal{X}$ contains a countable $\succ_x$-dense subset.

$W : \mathcal{P}_0(\mathcal{X}) \to \mathbb{R}$ given by

$$W(P) = Q^{\tau}(P). \tag{31}$$

where u is unique up to strictly increasing transformations.

It is straightforward to show that Rankwise Monotonicity implies that $\succsim_P$ will respect weak first-order stochastic dominance. Clearly, this is not a tight notion of monotonicity in that it is not equivalent to $\mathbf{P3}^{Q^P}$. (The weak first-order stochastic dominance gives only a partial ordering of distributions, since it involves a simultaneous comparison of all quantiles.) Rather, a local version of it which only compares quantiles is tight. (In Appendix 5A, we state and relate it precisely to axioms on acts.) Thus one could argue that the notion of monotonicity we used in deriving probabilities is, in a sense, the weakest monotonicity property. Even if two distributions P and Q coincide only at one point (at one quantile) and P first-order stochastically dominates Q otherwise, it may still be that $Q \sim_P P$.

We conclude with an interesting result that gives the model proposed in this paper a broader perspective. Although by itself monotonicity with respect to first-order stochastic dominance does not imply Quantile Maximization (again, without constraining the quantile to be left- or right- continuous), it is equivalent to $\mathbf{P3}^{Q^P}$ if one also requires that the following property of *ordinal invariance* holds: write each distribution $R \in \mathcal{P}_0(\mathcal{X})$ as a tuple of ordered vectors of outcomes and probabilities $(\mathbf{x}_R, \mathbf{p}_R)$; for any pair $P, Q \in \mathcal{P}_0(\mathcal{X})$,

$$(\mathbf{x}_P, \mathbf{p}_P) \succsim_P (\mathbf{x}_Q, \mathbf{p}_Q) \Leftrightarrow (\phi \circ \mathbf{x}_P, \mathbf{p}_P) \succsim_P (\phi \circ \mathbf{x}_Q, \mathbf{p}_Q) \tag{32}$$

for any mapping $\phi : \mathcal{X} \to \mathcal{X}$ such that if $x \succ_x y$ then $\phi \circ x \succ_x \phi \circ y$, where $\phi \circ \mathbf{x}_R$ is defined element by element.

Proposition 1 *The following sets of axioms are equivalent for a binary relation on the set of lotteries $\mathcal{L}$, $\succ_P$:*

(1) $\boldsymbol{P1}^P$*,* $\boldsymbol{P3}^{Q^P}$ *and* $\boldsymbol{P5}^P$*,*

(2) $\boldsymbol{P1}^P$*, monotonicity with respect to FOSD, ordinal invariance,* $\boldsymbol{P5}^P$*.*

The implication of Proposition 1 is twofold. First, it provides an equivalent characterization of our model in risk settings. Second, it justifies Quantile Maximization as a unique ordinal choice rule (that is, a rule satisfying ordinal invariance).[25]

1.3.5 Risk attitudes

In applications, one is often interested in agents' attitudes towards risk. Since under Quantile Maximization the choices are consistent with any utility function over certain outcomes provided only that it preserves their ordering, risk attitudes will not be characterized by concavity of these functions. Still, one can use a model-free definition and ask how a quantile maximizer chooses between a lottery P and the expected return from this lottery, $\bar{x}_P$:

$$P \underset{\prec}{\overset{\succ}{\sim}} \bar{x}_P = \int x_P \pi(x) dx. \tag{33}$$

To take a specific example, consider a half-half bet between outcomes 1 and 3 versus a certain outcome of 2. All τ-maximizers with $\tau \leq \frac{1}{2}$ will strictly prefer the gamble's average face value of 2, while those with $\tau > \frac{1}{2}$ will choose to gamble. However, when probabilities are modified to $\frac{1}{3}$ on 1 and $\frac{2}{3}$ on 3, all quantile maximizers with $\tau \in (\frac{1}{3}; \frac{1}{2}]$ will switch. This illustrates two more general features of the model:

Remark 1 *Under definition (33):*

(1) Quantile maximizers do not exhibit any global *(that is, for all lotteries* P*) risk attitude, except in the extreme cases,* $\tau = 0$ *(risk aversion) and* $\tau = 1$ *(risk loving).*

[25] To aid intuition, we note the relation to results in social choice literature, which our axiomatization yields as a special case. Corollary 1 generalizes the model of *rank-dictatorship* (Gevers [1979]), also known as *positional dictatorship* (Roberts [1980]), which predicts that the person with the k^{th} level of wealth in the society will be a dictator. The characterization for the original model, which obtains for the uniform probability, also characterizes choice behavior based on ranking of outcome vectors (e.g., order statistics). The risk version of our result can be usefully interpreted for c l a s s e s of citizens, ranked according to wealth.

(2) For a given *lottery, quantile maximizers with* $\tau \leq \Pr(x \leq \bar{x}_P)$ *are weakly more risk averse.*

Thus, definition (33) of risk aversion is not informative in an ordinal framework. Nonetheless, as suggested by the second remark, τ-maximizers do exhibit some attitude toward risk.

Under Quantile Maximization, it is not the randomness/riskiness property of lottery distributions but rather the relative (with respect to probabilities) ranking of outcomes that affects the choice of an agent.[26] To compare agent's risk attitudes, we suggest two notions for any fixed utility u, which parallel the concepts under Subjective Expected Utility and have very simple and natural expressions in the Quantile Maximization model. They share their properties with the quantile.[27]

The *τ-certainty equivalent of* P, $CE^{\tau}(P)$, is the amount of money for which a decision-maker is indifferent between lottery P and $CE^{\tau}(P)$ with certainty.

Clearly, the τ-certainty equivalent is the quantile outcome,

$$CE^{\tau}(P) = Q^{\tau}(P). \tag{34}$$

For any fixed amount of money, $x \in \mathbb{R}$, Manski [1988] characterizes an agent's risk preference through the amount that would make her indifferent between x and any distribution

[26] It is useful to invoke the interpretation, suggested in (8), that a quantile maximizer evaluates each lottery by a loss function defined as a weighted sum of absolute deviations. The lower the τ the more the decision maker cares about the lower-tail outcomes relative to the higher-tail outcomes; in other words, the more cautious she is. In the literature, the maxmin ($\tau = 0$) and maxmax ($\tau = 1$) behavior are informally referred to as "cautiousness/pessimism" and "wishful thinking/optimism," respectively. We formalize these intuitions for the general quantile representation.

[27] We only define comparative measures of risk attitudes. It seems tempting to define an absolute attitude towards risk by how an agent ranks a lottery and the median rather than the mean return; that is, to call a decision maker *cautious* (*incautious*) if, for any lottery $P \in \mathcal{P}_0(\mathcal{X})$, the degenerate lottery that yields $Q^{0.5}(P)$ with certainty is strictly preferred (weakly less preferred) to P. Then, while it would not be true that the higher the quantile, the more risk-loving an individual is in the sense of (33), it would be true that she is less cautious. Under this definition, quantile τ itself would provide a natural measure of cautiousness. It seems desirable to provide a formal justification for this approach.

$\in \mathcal{P}_0(\mathcal{X})$. He defines a *risk premium* as the value $\mu_P(x, \tau) \in \mathbb{R}$, $P \in \mathcal{P}_0(\mathcal{X})$, that solves

$$u(x - \mu_P(x, \tau)) = Q^\tau(P). \quad (35)$$

Letting $u(x) = x$, we have that the unique and finite risk premium is $\mu_P(x, \tau) = x - Q^\tau(P)$.

The following characterization is straightforward and it is stated without proof.[28]

Claim 1

The following are equivalent:

(i) An individual 1 *is more cautious than an individual* 2.

(ii) $CE^{\tau_1}(P) \leq CE^{\tau_2}(P)$*, for any* $P \in \mathcal{P}_0(\mathcal{X})$.

(iii) $\mu_P(x, \tau_1) > \mu_P(x, \tau_2)$ *for any* x *and* $P \in \mathcal{P}_0(\mathcal{X})$.

(iv) Whenever $P \succ_{P1} x$*, then* $P \succ_{P2} x$*, for any* $P \in \mathcal{P}_0(\mathcal{X})$ *and* x.

1.4 Sketch of the proof of Theorems 1 and 2

This section lays out in detail our axiomatization of Quantile Maximization (Theorems 1 and 2). Our task is to separate probability from preferences, possibly establishing probabilistic sophistication, and a quantile from probability.

[28] It is worth restating the properties of the quantile, now in the context of measuring cautiousness, to contrast them with those of more traditional measures associated with definition (33). First, unlike risk attitudes based on (33), cautiousness is preserved under strictly increasing transformations of utilities. As a result, the quantile can rank both risk and risk attitude. Second, it is less sensitive to small-probability outcomes than is risk aversion. While a risk averse expected-utility maximizer will choose the expected return from a gamble irrespective of how small the probability of the loss is, a quantile maximizer may strictly prefer the gamble. Third, in contrast to any comparative measure of risk aversion based on (33), which only partially ranks utility functions, a complete and transitive ranking of attitudes can be given in the quantile model. When distributions being compared are objective, global risk attitudes can be assessed by means of a single observation.

An important attribute of the risk measures for quantile maximizers is that, unlike definition (33), they do not require that outcomes be measured on an interval scale, which many economic and social variables lack (such as choosing between two jobs or careers, grades A and C versus two B's, receiving two job offers on the same day or on two different days). Traditional measures of risk attitudes are simply not defined when outcomes are not money or other objectively measurable quantities, or when accurate measurement is difficult, or when outcomes involve more than monetary payoffs. Providing a measure of risk attitudes in these contexts has been a challenge. For an extensive survey and discussion of the problems with responses to scale problem, see O'Neill (2001).

The proof proceeds as follows: The first step is to separate probabilities from preferences over $\mathcal{F}$, $\succ$. Formally, it involves deriving the likelihood ranking revealed by the preference relation $\succ$, and showing that it can be represented by a subjective probability measure π over $\mathcal{E}$. This part is the heart and central contribution of the proof. Further, we demonstrate that there exists a unique number $\tau \in [0,1]$ of the cumulative probability distribution implied by π such that acts are indifferent if and only if they imply the same τ^{th}- quantile outcome. Next, we derive (a family of) utility functions over certain outcomes u and use them to construct a functional that represents $\succ$. Finally, we establish that our axioms on the relation $\succ$, sufficient for obtaining the representation, are also necessary.

Probability measures. The very idea of separation of subjective beliefs from preferences is as old as theory of choice under uncertainty. (It has appeared in Ramsey [1931], Savage [1954], Anscombe and Aumann [1963] and has been used in many different settings.) And yet, we cannot directly use any of the available approaches, as we will now discuss. The difficulty is threefold.

First, prior research has focused on axiomatizing expected- and non-expected-utility functional forms, where the properties of risk preferences such as mixture continuity and monotonicity were heavily used in deriving probabilities. (See also Sections 1.3.2 and 1.5.) Under Quantile Maximization, risk preferences are not mixture continuous and only weakly satisfy stochastic dominance.

Second, in this model the commonly used likelihood relation $\succ^*$, defined in (21), does not discriminate between a significant subset of events from $\mathcal{E}$. For instance, with $\pi(E) = 0.4$ and $\pi(F) = 0.6$, a 0.3-maximizer would be indifferent between the acts f and g in (21). In general, under τ-maximization, definition (21) would rank as equally likely any events with probabilities either both smaller than τ or both greater than τ.[29] For $\tau = 1$ ($\tau = 0$),

[29] Allowing for events $E, F \in \mathcal{E}$ to be compared either directly or through the likelihood ranking of their complements (which already imposes some additivity) would still render "equally likely" all events with probabilities $\pi(E)$, $\pi(F)$ either both smaller than $\min\{\tau, 1-\tau\}$, or both greater than $\max\{\tau, 1-\tau\}$, or both between $\min\{\tau, 1-\tau\}$ and $\max\{\tau, 1-\tau\}$.

likelihoods of no events both more likely than $\varnothing$ (less likely than $\mathcal{S}$) will be ranked strictly by (21). The following lemma demonstrates how coarse the relation $\succ^*$ is: there are only two equivalence classes of $\mathcal{E}$ under $\sim^*$.

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

Third, and related, for quantile maximizers definition (21) does not satisfy the set of axioms on the binary relation over events[30] that are necessary and sufficient for the likelihood relation to admit a unique probability-measure that (i) represents it and (ii) is convex-ranged.[31] These axioms 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 $\mathbf{P6}^{Q^*}$.

What fails for all $\tau \in (0,1)$ is axiom **A4**. For $\tau = 0, 1$, it is vacuous under definition (21). In addition, **A5** fails for $\tau \in (0,1]$, but this problem disappears when our $\mathbf{P6}^{Q}$ is used instead.

A5' $\mathbf{P6}^{Q}$

It is important to recognize, however, that the major difficulty is not the failure of the

[30] In Savage (1954), they are implied by the conditions on the binary relation over acts, **P1**-**P6**.

[31] Respectively, (i) $E \succ^* F$ if $\pi(E) > \pi(F)$, for any $E, F \in \mathcal{E}$, and (ii) for any $E \in \mathcal{E}$, and any $\rho \in [0,1]$, there is $G \subseteq E : \pi(G) = \rho \cdot \pi(E)$. While equivalent to nonatomicity for countably additive measures, the property (ii) is stronger for finitely additive measures, which are derived here (see Bhaskara Rao and Bhaskara Rao [1983], Ch. 5).

For the necessity and sufficiency argument, see, for example Kreps (1988), Theorem 8.10.

axioms itself. Nor is it lack of fineness or tightness of the likelihood relation.[32] Rather, the structure from the relation over acts $\succ$ embedded in the relation over events defined in (21) is not rich enough to capture differences in the likelihood of events.

Nonetheless, there is more structure in preferences over acts that reveals the likelihood of events. We will proceed as follows. We will define a new binary relation on events, $\succ_{**}$, which although incomplete on $\mathcal{E}$, will lead to a complete relation on the subset of "small" events, $\mathcal{E}_{**}$. We will show that the axioms **A1-A5'** hold on $\mathcal{E}_{**}$. This will allow us to derive a probability measure that represents a decision maker's beliefs about the relative likelihoods of events in $\mathcal{E}_{**}$. Then, we will build up from $\mathcal{E}_{**}$ to construct a likelihood relation which will be complete and satisfy **A1-A5'** on the whole set of events, $\mathcal{E}$. This will enable us to extend the measure on $\mathcal{E}_{**}$ to $\mathcal{E}$ as well as to derive a unique quantile τ. As explained below, in the cases leading to $\tau = 0, 1$ there is not enough information in $\succ$ to permit all of these steps.

The idea behind the new likelihood relation $\succ_{**}$ is as follows. The events that can be ranked as strictly more or less likely according to $\succ_{**}$ 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 strictly ranked by $\succ_*$.

Definition 3 *$E \succ_{**} F$ if*

$$\begin{array}{ll} (i) & E \sim_* F, \\ (ii) & \textit{There is an event } G \textit{ such that } (E \cup F) \cap G = \varnothing \textit{ and } E \cup G \succ_* F \cup G. \end{array} \tag{36}$$

Lemma 10 (Appendix 3A) shows that there cannot be any other event G' for which the ranking is reversed. Condition (i) plays a role in identifying a subset of events for

[32] Given definition of $\succ^*$, a likelihood relation is *fine* if for all $E \succ^* \varnothing$, there is a finite partition of $\mathcal{S}$, $\{G_1, ..., G_N\}$, such that $E \succ^* G_n$, $n = 1, ..., N$; it is *tight* if whenever $E \succ^* F$, there is an event H such that $E \succ^* F \cup H \succ^* F$. Strictly speaking, given the failure of **A4**, finness and tightness cannot be assured (cf. Proposition 8.9 in Kreps [1988]). Yet, our point is that the reason for it is not structural, but rather it lies in weakness of the definition.

which $\succ_{**}$ (with $\sim_{**}$) will be complete, $\mathcal{E}_{**} \subset \mathcal{E}$.[33] The subset $\mathcal{E}_{**}$ is defined to contain all events for which there exists another event in that subset such that (i) and (ii) hold simultaneously. Roughly speaking, it contains events whose probabilities will not be greater than $\min\{\tau, 1-\tau\}$.

Now comes an important property of the likelihood relation. In deriving the measure representation, it is essential that disjoint non-null subsets of the state space can be strictly ranked. This property does not hold if preferences are extreme. We show that in those cases, a decision maker's preferences over acts only depends on (and thus can only tell) whether an event is null, or it is the state space, or nested in another event, up to differences on null subevents. Formally, we establish the following invariance properties of extreme and non-extreme preferences:

Lemma 2

A. *If preferences over acts are not extreme, an individual is indifferent to exchanging outcomes on events equally likely according to* $\sim_{**}$.

B. *If preferences over acts are extreme, a decision maker is indifferent to exchanging outcomes on disjoint non-null 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 strictly ranked by $\sim_{**}$. Thus, intuitively, while all τ-maximizers, $\tau \in [0,1]$, can compare nested events, these are the only events that can be strictly ranked by 0- and 1-maximizers. It is at this point that the derivation for $\tau = 0$ and $\tau = 1$ departs from the general proof.

For extreme preferences, it is essential to combine strict likelihood judgments from $\succ^*$ and $\succ_*$. The reason is that in that case, it is the extended definition that makes it possible to distinguish the likelihoods of $\varnothing$, $\mathcal{S}$ and of events E which differ from $\varnothing$ and $\mathcal{S}$ on a non-

[33] After excluding a subset of $\mathcal{E} \times \mathcal{E}$ including events that can be strictly ranked by $\succ_*$, there remain two subsets, events within each of which are ranked as $\sim_*$. Lemma 1 characterizes those sets as containing events judged $\sim_* \varnothing$ and $\sim_* \mathcal{S}$, respectively.

null set. The combined judgment is captured through the following relation, demonstrated to be consistent in Lemma 8 (Appendix 3A):

Definition 4

$$E \succ_*^* F \text{ if for any } x \succ y, \quad \begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix} \text{ or } \begin{bmatrix} x \text{ if } s \in F \\ y \text{ if } s \notin F \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \notin E \end{bmatrix}. \tag{37}$$

If preferences are not extreme, we prove that the axioms **A1-A5'** hold on $\mathcal{E}_{**}$, which is then used to derive a unique, convex-ranged and finitely additive probability measure π. Then a complete relation on $\mathcal{E}$ can be constructed from $\succ_{**}$, which is done through *N-partitions*, partitions of the state space $\mathcal{S}$ into N elements. This relation is used to extend the measure π to all events in $\mathcal{E}$. Thus all quantile maximizers with non-extreme preferences are probabilistically sophisticated, as defined in (23).

Since the extreme preferences only distinguish between $\varnothing$, $\mathcal{S}$ and nested events, up to differences on null subevents, any (normalized) measure that respects monotonicity will still represent the same preferences. Therefore, in this case, the relation $\succ$ on $\mathcal{F}$ is consistent with a set of measures which are nonatomic but not necessarily finitely additive.

Quantile. Having derived probability measures, we prove the existence and uniqueness of a quantile $\tau \in [0,1]$. The following result, which relies on Lemma 2, provides the key assertion:

Lemma 3

(i) Given the partition of $\mathcal{S}$ induced by act $f \in \mathcal{F}$, there is a unique pivotal event.

(ii) Let acts $f, g \in \mathcal{F}$ be such that for any $s, s' \in \mathcal{S}$, $f(s') \succ f(s) \Rightarrow g(s') \succsim g(s)$. Then, $E_f \Delta E_g$ cannot be pivotal ($E_f \cap E_g \neq \varnothing$).

Given the derived measures, we can use Lemma 2 to map the set of acts onto the set of simple lotteries, $\mathcal{P}_0(\mathcal{X})$, through $f \rightarrow \pi \circ f^{-1}$. Then, to establish the existence and uniqueness of τ for the non-extreme preferences, we construct a sequence of equi-partitions of $\mathcal{S}$ (finite partitions whose elements are equally likely) with 2^N elements. In the $(N+1)^{th}$ sequence, the pivotal event from the N^{th} sequence (unique by Lemma 3) is partitioned into 2 elements. Then, we obtain τ from $(0,1)$ by approximating it with probabilities of the union of the pivotal event and events assigned to less-preferred outcomes. For the extreme preferences, the set of derived probability measures is applied to N-partitions (not equi-partitions) to yield $\tau = 0$ or $\tau = 1$.

Utilities. Given that $\succ$ is a weak order (**P1**), due to the ordinality property of the quantile-maximization representation, the utility on outcomes, u, depends exclusively on the properties of the set $\mathcal{X}$. The assumption that $\mathcal{X}$ contains a countable $\succ$-dense subset (used only in the final step) can be used together with **P1** ($\succ$ is a weak order) to apply Debreu's (1954) theorem and derive a real-valued utility on $\mathcal{X}$. We note that construction of V does not depend on the existence of the best and worst outcomes - again, the reason being ordinality.

1.5 Related literature

In this section, we relate our results to the literature on probabilistic sophistication. We first compare our axioms, derivation and properties of risk preferences with those in two milestone developments: Machina and Schmeidler (1992), and Grant (1995). We then relate our results to the recent and illuminating paper by Chew and Sagi (2006). All axioms are listed in Appendix 1.

Machina and Schmeidler (1992, hereafter MS). Setting as their goal liberation of the derivation of subjective probability in the Savage world from the Expected Utility hypothesis, MS drop **P2** (the Sure-Thing Principle). However, dispensing with it removes more than the Marschak-Samuelson independence property implying an expected-utility form functional. As mentioned above, Savage used **P2** also to obtain additivity of the probability measure. MS strengthened **P4** (Weak Comparative Probability, to Strong Comparative Probability, **P4***), otherwise using Savage's axioms.

Our class of preferences does not, however, satisfy **P4***,[34] though in the presence of **P1** and $\mathbf{P3}^Q$ it does satisfy the original **P4** of Savage. Strictly speaking, our new axiom $\mathbf{P4}^Q$ is weaker than **P4***. $\mathbf{P4}^Q$ suffices to ensure additivity in our model.

Overall, our set of conditions shares with theirs only **P1**, and it weakens **P3**, **P4***, **P5** and **P6**. The functional form in the representation theorem of MS is mixture continuous and monotonic with respect to first-order stochastic dominance. While encompassing many functional forms, those properties are crucially used in the derivation of probabilities. Our proof does not rely on any form of mixture continuity, and monotonicity with respect to stochastic dominance holds only weakly (i.e., strict first-order stochastic dominance implies only weak preference over distributions).

MS essentially use Savage's (1954) derivation of probability. Although, as mentioned, in general **P2** fails in their case, it does hold for two-outcome acts. Since all axioms of Savage hold for such gambles, it follows from his theorem that there exists a unique, finitely additive and convex-ranged probability measure. MS then use **P4*** to extend the measure to the set of all acts.

Grant (1995). With a novel interpretation of probabilistic sophistication, Grant (1995) obtains a derivation of probabilities without **P2** and **P3** (Eventwise Monotonicity). After relaxing the latter condition, two-outcome gambles cannot be used to infer the relative like-

[34] The following example illustrates that for **P4***. Let $\tau = \frac{1}{3}$, $\pi(E) = \frac{1}{3} + \varepsilon$, $\pi(F) = \frac{1}{3} - \varepsilon$, $\pi(G) = \frac{1}{3}$. Then $\begin{bmatrix} 2 \text{ if } E \\ 1 \text{ if } F \\ 5 \text{ if } G \end{bmatrix} \succ \begin{bmatrix} 1 \text{ if } E \\ 2 \text{ if } F \\ 5\text{if } G \end{bmatrix}$ but $\begin{bmatrix} 2 \text{ if } E \\ 1 \text{ if } F \\ 0 \text{ if } G \end{bmatrix} \sim \begin{bmatrix} 1 \text{ if } E \\ 2 \text{ if } F \\ 0 \text{ if } G \end{bmatrix}$.

lihood of events. (The ranking of constant acts needs not agree with the conditional ranking of two outcomes.) Still, with a modification of **P3** to Conditional Upper or Lower Eventwise Monotonicity (**P3**^{CU}, **P3**^{CL}), conditional preference between two outcomes can be used to draw inference about the likelihood of events from the preference over conditional gambles that involve these two outcomes. This identifies a set of acts on which the hypotheses of MS (1992, Theorem 1) hold, which yields a probability-measure representation. This measure is then extended through continuity of preferences (**P6**†) to the whole state space.

Our lottery preferences need not satisfy either of Grant's (1995) two-outcome mixture continuity[35] or conditional monotonicity.

Chew and Sagi (2006, hereafter CS). The new approach taken recently by Chew and Sagi (2006) is based on the notion of exchangeability. Two events are said to be *exchangeable* if the individual is always indifferent to permuting her payoffs.

Definition 5 *For any pair of disjoint events $E, F \in \mathcal{E}$, E is exchangeable with F if for any outcomes $x, y \in \mathcal{X}$, and any act $f \in \mathcal{F}$,*

$$\begin{bmatrix} x \text{ if } E \\ y \text{ if } F \\ f \text{ if } (E \cup F)^c \end{bmatrix} \sim \begin{bmatrix} y \text{ if } E \\ x \text{ if } F \\ f \text{ if } (E \cup F)^c \end{bmatrix}. \tag{38}$$

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

Definition 6 *For any events $E,\ F \in \mathcal{E}$, $E \succsim^C F$ whenever $E \backslash F$ contains a subevent G that is exchangeable with $F \backslash E$.*

[35] $V : P_0(X) \to \mathbb{R}$ is said to be *mixture continuous for two-outcome sublotteries* if for any pair of outcomes $x,\ y$ in $\mathcal{X}$, any $\gamma \in (0,1]$, and any pair of lotteries $P,\ Q$, the sets $\{\lambda \in [0,1] |\ V(\gamma(\lambda\delta_x + (1-\lambda)\delta_y) + (1-\gamma)P) > V(Q)\}$ and $\{\lambda \in [0,1] |\ V(Q) > V(\gamma(\lambda\delta_x + (1-\lambda)\delta_y) + (1-\gamma)P)\}$ are open.

$\succ_P$ is *mixture continuous for two-outcome sublotteries* if for all pairs of outcomes $x,\ y$ in $\mathcal{X}$, all $\gamma \in (0,1]$, and distributions $P,\ Q$ in $\mathcal{P}_0(\mathcal{X})$, the sets $\{\lambda \in [0,1] |\ \gamma(\lambda\delta_x + (1-\lambda)\delta_y) + (1-\gamma)P \succ_P Q\}$ and $\{\lambda \in [0,1] |\ Q \succ_P \gamma(\lambda\delta_x + (1-\lambda)\delta_y) + (1-\gamma)P\}$ are open.

Intuitively, exchangeability carries the meaning of "equal likelihood," while comparability conveys "greater likelihood." CS find a set of axioms on those relations so that they lead to a likelihood relation having a probability-measure representation.

CS's Theorem 1 can also be used in our model. For non-extreme preferences, their axioms **A**, **C** and **N** follow from **P1**, $\mathbf{P3}^Q$, $\mathbf{P4}^Q$, **P5**, and $\mathbf{P6}^Q$. This raises a more general question of a comparison between the two approaches: *exchangeability-based* and *direct likelihood* (defining the strict likelihood relation from preferences, e.g., Machina and Schmeidler [1992], Grant [1995], this paper). First, 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.[36] Second, 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, completeness of the comparability relation (pre-notion of "more likely than") is assumed. Transitivity is proved without any recourse to the strict relation over acts, $\succ$. 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.

As CS point out, assuming, independently from preferences over acts, that the exchangeability-based likelihood is well defined might not be warranted. (They illustrate this point using the example of "Machina's Mother" (Machina [1989].)[37]

1.6 Applications

"No theory exists to show that VaR is the appropriate measure upon which to build optimal decision rules." Schachter (1997, p. 19) As one application, our axiomatization provides foundations to decision making based on Value-at-Risk. In addition, Proposition 2 shows that regulatory policy decreasing the confidence level α (increasing τ) restricts the set of

[36] Step 4 in the proof of Theorem 1 in Machina and Schmeidler (1992), Claim 5a in Grant (1995), our Lemma 2A.

[37] We conjecture that for a large class of models the exchangeability approach can be applied without loss of generality. Rostek (2006) will address this problem in more detail.

actions that can be taken by risk managers. This paper also suggests that VaR can serve not only as a measure of risk but also (comparative) risk attitude.

In this section, we revisit the properties of the quantile choice rule and discuss how it can provide an attractive alternative in applied work. We consider an ordinal approach to modeling choice (Section 1.6.1) and an application to robust policy design (Section 1.6.2). In a companion paper, we apply Quantile Maximization to strategic settings.

1.6.1 Qualitative decision theory

Virtually all existing models of choice under uncertainty (from Savage [1954] through Gilboa and Schmeidler [1989] to recent developments in Ghirardato, Maccheroni and Marinacci [2004], and Klibanoff, Marinacci and Mukerji [2005]) characterize preferences that imply cardinal properties of utility functions over outcomes. In particular, their policy implications crucially depend on concavity of utilities. In many economic settings, however, the assumptions behind cardinality may not be appropriate. We discuss three reasons.

(1) When preferences have a representation with cardinal properties, that agents act as if they were able not only to rank certain alternatives (ordinally) but also to compare differences in "pleasure" (cardinally) in a consistent way.[38] This assumption is far from innocuous in settings with little learning or experience opportunity, in which people do not have sufficient information about all possible realizations or do not know their preferences sufficiently well to be able to quantify them. The need to account for lack of cardinality has become even more pronounced with recent changes in economic environments such as markets with increasing availability of new goods and fast-changing technology. Another

[38] Put another way, cardinal properties of a utility function over outcomes (i.e., properties preserved under positive affine transformations of utility), which is part of the representation of preferences over acts or lotteries, correspond to some conditions on those preferences which require that individuals can assess differences in outcomes on which those acts or lotteries are defined. Those differences are reflected in differences in valuations. Let u be a utility function over outcomes, $u : \mathcal{X} \to \mathbb{R}$, and let v be any increasing linear function on $\mathbb{R}$, $v : \mathbb{R} \to \mathbb{R}$. Then for $v = \alpha u + \beta,\ \alpha > 0$, $\frac{u(x)-u(y)}{u(z)-u(w)} = \frac{v(x)-v(y)}{v(z)-v(w)}$. We refer to "cardinality" in this sense, as opposed to the meaning used in studying welfare implications. (See for example Weymark [2005].)

point is that economic applications often require a criterion which is objective in the sense that it does not depend on individual preferences over outcomes, just on their ranking.[39]

(2) At the level of primitives, cardinality presumes that individuals behave as if they could compare differences in outcomes on which the uncertain alternatives are defined. Again, this is a strong condition for many goods and services. For example, it has become more and more common for companies to provide potential buyers with on-line quality assessments of their customers; thus a buyer only knows distributions of ranks but not of absolute evaluations of the product quality. Further, sometimes information is naturally or optimally given in the comparative rather than absolute form; e.g., when information must be conveyed but restricting expert's incentive to exaggerate in absolute statements is desired.[40]

(3) In settings in which decisions are typically made once (e.g., tourism, choosing a medical treatment), only a single realization of distribution will occur. Using an expectation-based evaluation seems much less appealing than with repeated decisions, when there is some compensation through averaging coming from repetition.[41]

To conclude, the cardinal properties of utilities (or, more precisely, conditions on prefer-

[39] For example, quality standards, expert recommendations, curve-based grading schemes, agency problems (when a principal delegates a task with uncertain outcome to an agent, she aims to set a standard of performance independent of the agent's preferences over outcomes), etc.

[40] Rubinstein (1996) notes that comparative statements are relatively more common in natural language. He justifies their optimality by formalizing their three properties: ability to indicate elements of the set, e.g., by means of order, ability to accurately convey information, and easiness to communicate the content to others.

Chakraborty and Harbaugh (2005) demonstrate that comparative cheap-talk statements can be credible when absolute statements are not (e.g., a professor ranking students for a prospective employer; an analyst's claims about the likely returns to a stock might not be credible, but the statement that one stock is better than another might be; a seller auctioning goods). In the context of multi-object auction, Chakraborty, Gupta and Harbaugh (2002) show that a seller's incentive to lie may be diminished or eliminated when only comparative statements are allowed.

[41] A similar point was made by Roy (1952): "... an ordinary man has to consider the possible outcomes of a given course of action on one occasion only and the average (or expected) outcome, if this conduct were repeated a large number of times under similar conditions, is irrelevant (p. 431)" and "Is it reasonable that real people have, or consider themselves to have, a precise knowledge of all possible outcomes of a given line of action, together with their respective probabilities or *potential surprise*? Both introspection and observation suggest that expectations are generally framed in a much more vague manner" (p. 432, emphasis original). Recently, this argument was also raised in the artificial intelligence literature (e.g., Dubois *et al.* [2000], Dubois *et al.* [2002]).

ences captured by those properties) appear too demanding or unnecessarily strong in many settings of interest for economists.

While ordinality under certainty is well understood, little attention has been devoted to it under risk or uncertainty.[42] In the Quantile Maximization model, the properties of robustness, one-dimensional information about preferences and no need for measurability requirement on outcomes are all a consequence of ordinality. It is thus interesting to ask how rich the class of ordinal decision rules is. In Rostek (2006), we characterize all ordinal choice rules in terms of conditions on preferences over acts. Specifically, we ask: What are the necessary and sufficient conditions for a class of preferences to admit an ordinal (appropriately defined) functional? Here, Proposition 1 gives one reason why Quantile Maximization may be appealing as an ordinal model: it is a unique ordinal choice rule if one also requires that a weak monotonicity property of first-order stochastic dominance holds (assuming probabilistic sophistication).

1.6.2 Robust decision making

The need for robustness has been recognized in the burgeoning literature on robustifying economic and policy design. Many studies have focussed on relaxing the assumption that decision makers know or act as if they know the true probability distribution (e.g., Hansen and Sargent [2004] applying the model by Gilboa and Schmeidler [1989], Klibanoff, Marinacci and Mukerji [2005]). Another and less explored robustness test involves relaxing the assumption that decision makers have cardinal (as well as ordinal) rankings of outcomes; or, that cardinal, parametric assumptions about utilities affect decisions.

One prominent ordinal decision rule that is commonplace in policy design is choosing

[42] The ordinal rules studied in the latter context include maxmin, maxmax, pure-strategy dominance by Börgers (1993) and, as we mentioned earlier, Manski (1988) argued for an ordinal approach to modeling choice, suggesting quantile maxmization (quantile-utility model) and maximizing probability that the outcome will be higher than some level (utility mass model).

according to the "worst case scenario". It is often justified by arguments supporting cautious policy (e.g., the "worst-case scenario" rule by Environmental Protection Agency and the Department of Justice; *Precautionary Principle* in the European Commission food and agricultural biotechnology policy; Walsh [2004] in the context of monetary policy). Critics have argued that such an extreme criterion may inhibit economic development (Gollier [2001]), may delay innovations that are safe and effective, and, more generally, that it gives too much importance to unlikely outcomes (e.g., Svensson [2000]).[43] The quantile representation studied in this paper can help address those concerns and complement best- (and worst-)case scenario analysis.[44]

1.7 Concluding remarks

For some applications, it might be desirable to extend the model proposed in this paper to more than one quantile. For example, a choice rule may depend on the "focal" worst-, best- and typical- case scenarios; or, only a range of quantiles higher or lower than some threshold may be of interest (e.g., an employer may be targeting candidates from a specific range of quality; when purchasing a good, a buyer may care about a high- or low- range of realizations of quality).

An important and natural direction to go would be to model quantile maximization with multiple priors. One compelling motivation is given by the need to address robustness concerns in modeling, complementary to those in Section 1.6.2. Under some assumptions on the state space, for the 0^{th} quantile, the framework would yield the multiple-prior maxmin rule by Gilboa and Schmeidler (1989).[45]

[43] Warning against making worst-case estimates public, Viscusi (1997) reports that people give excessive attention to the alarmist scenarios when facing a range of conflicting estimates.

[44] In the policy context, distributional consequences of policies are of interest much beyond average statistics. Quantiles have been used to assess social policies and treatment effects, to compare unemployment duration and distributions of wages, etc. Their use in formal empirical studies has been spurred by Quantile Regression (Koenker and Bassett [1978]), in which the classical least squares estimation of conditional mean is replaced by estimation of conditional quantile functions (Chamberlain [1994], Buchinsky [1995], Koenker and Hallock [2001] provide comprehensive surveys).

[45] In the context of multiple priors, Svensson (2000) argues that "the worst possible model is on the

In addition, the results in this paper suggest two interesting projects for future work. The first concerns exploring the property of ordinality, as explained in Section 1.6.1. Second, by thinking through the meaning of probabilistic sophistication in the context of the particular class of ordinal preferences, this paper obtains new insights into probabilistic sophistication. An interesting project is then to abstract from our particular representation and ask more generally what the minimal requirements for probabilistic sophistication are.

boundary of feasible set of models, and hence depends crucially on the assumed feasible set of models. If the worst possible model somehow ended up in the interior of the feasible set of models, one could perhaps argue that the outcome is less sensitive to the assumptions about the feasible set of models" (p. 6-7).

Chapter 2

Identification of Quantile Maximization with Finite Data Sets

2.1 Introduction

In Rostek (2006), we introduced a model of preferences in which a decision maker compares uncertain alternatives through a quantile of utility distributions. For example, she might be maximizing the median utility, as opposed to the mean utility, as she would if she were an expected-utility maximizer. More generally, she might be comparing lotteries through some other quantile that corresponds to any given number between 0 and 1.

Of primary interest in Rostek (2006) was uncovering complete testable implications of Quantile Maximization. We provided a Savage-style representation theorem for quantile maximizers in a decision-theoretic framework. Taking preferences over acts as a primitive, we found conditions on these preferences under which there is an ordinally unique utility index over outcomes, and a unique probability measure over the underlying state space, such that the utility of an act is some quantile of the utility index. Thus, the axiomatization established the following two results: first, it showed that the Quantile Maximization generates non-trivial restrictions that have to be satisfied by observed behavior; second, it provided the model with an identification result, "identification" understood as determining whether the model fundamentals (utilities u, probability measure π, and a number in $[0, 1]$, τ) compatible with the observed behavior are unique. Axiomatization uses the richest data set by assuming that one can observe all choices in all decision problems. Hence, while being an ultimate identification test for the model, axiomatization does not yield identification conditions that can be readily used on small finite data sets in empirical work. Providing such conditions is the goal of this paper.

Specifically, we investigate how much can be inferred about the unobservables of the model from finitely many choices. To investigate the empirical content of the model, we also compare, by means of examples, predictions of Quantile Maximization with those of other models of choice.[46] We model a decision problem as a matrix, in which a column represents an event and a row is an uncertain alternative (act).[47] If one can only observe the event structure (represented by a matrix) and the selected row, but not individual beliefs and τ, is it possible to test whether the players choose according to Quantile Maximization? Assuming Quantile Maximization, in turn, is it possible to separate agents with different risk attitudes as measured by τ?[48]

To investigate the first question, we further ask which actions will be observed if individuals are quantile maximizers and show that model predictions differ from several leading alternatives, including expected utility. Suppose in a given decision problem we observe the choice of a quantile maximizer with unknown anticipation level τ and unknown beliefs π. We show that the sets of choices that are optimal for this agent for some beliefs (i.e., the undominated choices) are weakly nested with respect to τ. That is, although all actions

[46] The answer to the question whether the Quantile Maximization hypothesis generates any testable predictions (i.e., restrictions that have to be satisfied by observables) is trivially positive. It is said that the Quantile Maximization model is *testable* if there is a decision problem F such that (i) $\mathcal{R}_F$, defined in (44), is nonempty (that is, the model is consistent), and (ii) $F \backslash \mathcal{R}_F$ is nonempty (that is, there are actions that can refute the model). By the property of monotonicity, an agent will not choose an act with outcomes that for all events are strictly less preferred to outcomes in some other act.

[47] A matrix-representation of decision problems is without loss of generality as long as alternatives involve finitely many outcomes. Axiomatization in Chapter 1 considered such finite-outcome acts.

[48] In the companion paper, we show that τ provides a comparative measure of risk attitude: the lower τ, the weakly more risk averse the agent is. Thus, $\tau \in (0, 1)$ represents intermediate risk attitudes between maxmin ($\tau = 0$) and maxmax ($\tau = 1$), which are the least and the most risk averse, respectively.

that may be chosen by high-τ individuals may also be chosen for some beliefs by those with lower τ, low-τ decision makers may choose actions that will never (for any beliefs) be selected by high-τ agents. This property, which we call *nestedness*, turns out to be key to most of the identification questions that we ask.

We next examine how much information about the unobservables of the model can be identified from the data (payoff structure and choices). In many applications, beliefs are part of the data set. For instance, they have been induced in a controlled environment like a lab or given as probabilities in a decision problem. Therefore, we also study settings in which beliefs are observed. When agents face known probabilities, we show that one can identify the quantile exactly from observing a single choice. The task of inferring τ becomes more involved if a researcher does not know the agent's beliefs about the likelihood of events. For that problem, we derive bounds that can be placed on the unobservable quantile and on the beliefs from data. The results are constructive in that they suggest how a revealing decision problem can be designed. We conclude by applying Quantile Maximization to normal-form games to illustrate how the conditions derived for single-agent problems can be used in strategic settings.

2.2 Set-up

Let $\mathcal{S} = \{..., s, ...\}$ denote a set of states of the world, and let $\mathcal{X} = \{..., x, y, ...\}$ be an arbitrary set of *outcomes*. An individual chooses among finite-outcome *acts*,[49] defined as maps from states to outcomes. $\mathcal{F} = \{..., f, g, ...\}$ is the set of all such acts. The set of events $\mathcal{E} = 2^{\mathcal{S}} = \{..., E, F, ...\}$ is the set of all subsets of $\mathcal{S}$. An individual is characterized by a preference relation over acts in $\mathcal{F}$, with $f \sim g \Leftrightarrow f \nsucc g$ and $f \nprec g$, $f \succsim g \Leftrightarrow f \succ g$ or $f \sim g$.

Let π stand for 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

[49] An act f is said to be *finite-outcome* if its outcome set $f(\mathcal{S}) = \{f(s) | s \in \mathcal{S}\}$ is finite.

$\Pi_f(z) = \pi[s \in \mathcal{S} | u(f(s)) \leq z],\ z \in \mathbb{R}$. Then, for a fixed act f and $\tau \in (0, 1]$, the τ^{th} *quantile* of the random variable $u(x)$ 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:

$$Q^\tau(\Pi_f) = \inf\{z \in \mathbb{R} | \pi[u(f(s)) \leq z] \geq \tau\}, \tag{39}$$

while for $\tau = 0$, it is defined as

$$Q^0(\Pi_f) = \sup\{z \in \mathbb{R} | \pi[u(f(s)) \leq z] \leq 0\}. \tag{40}$$

The unobservables of the model are the anticipation level τ and beliefs π. We assume that (with or without observable beliefs) a researcher can observe decision problems $F \subset \mathcal{F}$ that consist of a finite number of acts, the ranking of outcomes to which those acts map, and the optimal choices for every F, denoted by f^*. Thus, a pair $\{F, f^*\}$ constitutes one observation in a data set. Due to the ordinality property of the model, without loss of generality, outcomes can be assumed to be monetary. Throughout, we assume that tastes τ and beliefs π are constant over the set of N observations. We also assume probabilistic sophistication, which we take to mean that the agent's beliefs can be represented by a unique finitely additive probability measure.

We use the following notation: Fix a decision problem F and utility function on outcomes u. Define the sets of quantiles that may be associated with the optimal choice for some beliefs $\pi \in \Delta(\mathcal{S})$, given τ,

$$Q^*_F(\tau) = \{z \in \mathbb{R} | z = \max_{f \in F} Q^\tau(\Pi_f) \text{ for some } \pi \in \Delta(\mathcal{S})\} \tag{41}$$

and for some π and some τ,

$$Q^*_F = \bigcup_{\tau \in [0,1]} Q^*_F(\tau). \tag{42}$$

Analogously to payoffs, we define the sets of optimal actions of a τ-maximizer given τ,

$$\mathcal{R}_F(\tau) = \{\tilde{f} \in F | \tilde{f} = \arg\max_{f \in F} Q^{\tau}(\Pi_f) \text{ for some } \pi \in \Delta(\mathcal{S})\}, \tag{43}$$

and for some τ and some π,

$$\mathcal{R}_F = \bigcup_{\tau \in [0,1]} \mathcal{R}_F(\tau), \tag{44}$$

the last set being *the set of actions undominated under Quantile Maximization* in a decision problem F.

2.3 Identification with finite data sets

We first present a result that significantly simplifies the testability and identification exercise by characterizing the set of all actions that may be selected by any τ and any beliefs π, $\mathcal{R}_F$. The proposition also reveals an important feature of the model.

Proposition 2 *For any decision problem F, and any pair of anticipation levels $\tau, \tau' \in [0,1]$,*

$$\tau' > \tau \Rightarrow \mathcal{R}_F(\tau) \supseteq \mathcal{R}_F(\tau'). \tag{45}$$

For any given decision problem, the sets of undominated actions are related by weak set inclusion with respect to τ: although all actions that can be chosen by high-τ agents may be chosen by low-τ agents for some beliefs, low-τ decision makers may choose actions that will not be chosen by those with high τ for any beliefs.

By Proposition 2, in order to find the set of (weakly) undominated actions in the quantile model, $\mathcal{R}_F$, one only needs to look for the set of undominated actions for the 0-quantile maximizers, $\mathcal{R}_F = \bigcup_{\tau \in [0,1]} \mathcal{R}_F(\tau) = \mathcal{R}_F(0)$. As a special case, we state the following useful claim.

Claim 2 *Let $\tilde{F}$ be a decision problem such that all acts are maps from two events E_1 and E_2. Then $Q^*_{\tilde{F}}$ contains at most three payoffs*

$$Q^*_{\tilde{F}} = \{z \in \mathbb{R} | z = \max_{f\in\tilde{F}} u(x_{f1}) \text{ or } z = \max_{f\in\tilde{F}} u(x_{f2}) \text{ or } z = \max_{f\in\tilde{F}} \min_{k=1,2} u(x_{fk})\}. \tag{46}$$

Thus, to find the set of actions that may be optimal for any τ and π in a matrix with two events, one only has to consider actions associated with the maximal payoff in each column or the maxmin payoff of the decision problem.[50]

Next, we set Quantile Maximization against an alternative restriction on preferences: Expected Utility. Let $\mathcal{R}^{EU}_F$ be the set of acts undominated under the Expected Utility hypothesis in a decision problem F, $\mathcal{R}^{EU}_F = \{f \in F | EU_\pi(f) \geq EU_\pi(f')$ for all $f' \neq f$ and some beliefs $\pi \in \Delta(\mathcal{S})\}$.

Example 3 *Consider two decision problems that differ only in one act (f_3). Elements of sets $Q^*_M(0)$, M=M1, M2, are in bold; no other payoff may be anticipated in the optimal choice for any τ and π. Using Claim 2, we observe that while acts f_1 and f_2 could be optimal for some beliefs under both Quantile Maximization and Expected Utility, for no beliefs could f_3 be justified in M1 by the former or in M2 by the latter model. On the other hand, f_3 is in $\mathcal{R}_{M1}$ and in $\mathcal{R}^{EU}_{M2}$.*

Matrix M1

	E_1	E_2
f_1	**11**	1
f_2	**4**	**8**
f_3	10	3

$f_3 \in (\mathcal{R}_{M1})^c \cap \mathcal{R}^{EU}_{M1}$

Matrix M2

	E_1	E_2
f_1	**11**	1
f_2	4	**8**
f_3	6	**5**

$f_3 \in \mathcal{R}_{M2} \cap \left(\mathcal{R}^{EU}_{M2}\right)^c$

[50] For the general case of $K_i \times K_{-i}$ matrices, at the two extremes, the maximal possible set $Q_F(0)$ is equal to the outcome set less the outcomes strictly less preferred to the maxmin outcome, while the minimal set is achieved in matrices with comonotonic payoff structure.

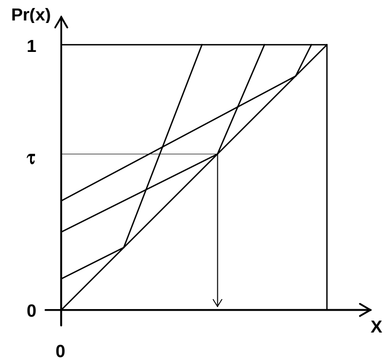
Hence the sets of actions that are undominated under Expected Utility and Quantile Maximization are not related by direct set inclusion. They are not disjoint either; it is easy to show that the set of actions that may be chosen by 1-quantile maximizers, $\mathcal{R}_F(1)$, consists of choices of expected-utility maximizers which are optimal to beliefs putting mass one on one of the events.

We now examine whether it is possible to uniquely pin down (or, ideally, recover)[51] τ with finite data sets. We consider a problem in which beliefs are and are not known. After providing identification results, we answer the following testability question: For a finite data set $\{\{F_n, f_n^*\}_{n=1,\dots,N}\}$, can we recover the set of anticipation levels τ and beliefs π that rationalize it under Quantile Maximization?

Identification under risk. Consider data with an objective probability distribution over events. We are interested in two questions: (1) How much information can be inferred about the quantile τ for a given data set? and (2) How rich a data set is needed to identify the quantile τ? The answer to the former question will follow from the general conditions developed for an uncertain environment. For the latter question, we demonstrate that there is a decision problem with a continuum of lotteries that identifies τ exactly with a single observation ($N = 1$). Consider an individual is choosing a lottery from the set depicted by piecewise-linear distributions in Figure 2.1. The distributions are parameterized by kinks that correspond to points in the unit interval $[0, 1]$.

[51] Identification requires that the unobserved model fundamentals consistent with observed behavior be unique. Recoverability obtains if the argument for identification is also constructive.

Figure 2.1 Recoverability of τ when beliefs are known



The key feature of the test is that for a given τ, there is a unique distribution with the cumulative probability at the kink outcome equal to τ such that all distributions with lower or higher kink outcomes are strictly less preferred. Thus, when asked to select a distribution, a decision maker will truthfully reveal her τ through her choice. Though simple the observation is not without interest because it implies that the model is not only identified under risk, but also that a constructive argument for identification can be provided; that is, that preferences can be recovered.

Identification under uncertainty. We now turn to the more challenging problem that involves inferring τ when beliefs are subjective and unobserved. We begin with the smallest data set, one that contains a single payoff matrix and choice, $\{F, f^*\}$. Observe that Proposition 2 implies that after observing some actions, it may be possible to bound τ from above. We established such bounds as a function of the parameters of the decision problem F (number of actions, number of events, and conditions on payoff structure) and the observed choice, a^*. In deriving them, we take a decision problem as given, but the argument we provide indicates what conditions should hold for a revealing matrix.[52] The following example is suggestive about the source of the bite for identification in the model.

Example 4 *For Matrix M2 in Example 3, we look for the optimal choices of a quantile maximizer for any values of the unobservable τ and any π. Figure 2.3 presents the resulting decision map.*

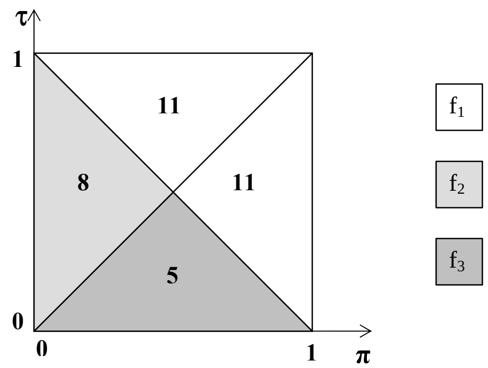
Figure 2.2 Conditional cdf's for acts in Matrix M2

	$1-\pi$	π
f_1	*1*	*11*

	π	$1-\pi$
f_2	*4*	*8*

	$1-\pi$	π
f_3	*5*	*6*

Figure 2.3 Choice correspondence for Matrix M2



For $\tau \leq \min\{\pi, 1-\pi\}$, the individual anticipates payoffs 1, 4 and 5 from acts f_1, f_2 and f_3, respectively. She will choose f_3.

For $\tau > 1-\pi$, she anticipates payoff 11 from act f_1. As this is the highest payoff in the matrix, she will choose f_1.

Finally, for $\pi < \tau \leq 1-\pi$, she anticipates payoffs 1, 8 and 5 from the respective acts, and hence she will choose f_2.

[52] That the conditions are derived for decision problems in the form of a matrix (in particular, for acts with identical partitions of the state space) is without loss of generality. Any decision problem F can be represented by a matrix by taking the coarsest common refinement of the (coarsest) partitions implied by the acts in $\mathcal{F}$.

The following four properties of the correspondence in Figure 2.3 will be important for identification:

1. For the same and fixed beliefs, quantile maximizers with different τ may choose different actions.

2. Even if they choose the same action, they may select it anticipating different payoffs.

3. For some fixed τ, there are payoffs that will never (for any beliefs) be anticipated in an optimal choice; e.g., for $\tau > 0.5$, when f_2 is chosen, 8 is anticipated, while when f_3 is chosen, 11 is anticipated; for no beliefs, will 5 be anticipated.

4. There are outcomes that will not be anticipated in the optimal choice for any τ and any beliefs π (1, 4, 6).

We now establish a property of the model that, along with Proposition 2, will provide the basis for inference. For a given decision problem F, different payoffs enter the set of payoffs associated with optimal actions, $Q^*_F(0)$, for beliefs with different support. For example, in Matrix 1 in Example 3, for beliefs putting mass one on one of the events, only payoffs 11 or 8 can be anticipated in the optimal actions by any $\tau \in [0, 1]$. For beliefs with 2-event support, 4 enters the set $Q^*_{M1}(0)$, but it takes beliefs with cardinality 2 for the payoff 4 to be anticipated under optimal choice.[53] It turns out that this pattern is a general characteristic of the model, proved inductively in Lemma 4: higher-cardinality belief supports admit less-preferred payoffs in $Q^*_F(0)$.

For payoffs in $Q^*_F(0)$, we define $\underline{c}(x)$ as the minimal cardinality of beliefs support for which $u(x) \in Q^*_F(0)$.

Lemma 4 *Fix outcomes $x, y \in f(\mathcal{S})$ and an act $f \in F$.*

$$\underline{c}(x) > \underline{c}(y) \Rightarrow u(x) < u(y). \tag{47}$$

[53] Note that a version of this argument can be used to prove Claim 2.

This sets up a correspondence between each payoff x in $f(\mathcal{S})$ such that $u(x) \in Q^*_F(0)$ and the minimal cardinality of beliefs support under which $u(x)$ enters $Q^*_F(0)$.

With Lemma 4, the inference procedure for a given data set $\{F, f^*\}$ can be summarized in three steps.

Step 1: Verify whether the matrix enables any inference at all. A trivial example of an uninformative matrix is one with comonotonic payoffs across choices; with payoff ranks identical across events, the sets of choices that might be optimal for some beliefs, $\mathcal{R}_F(\tau)$, coincide for all quantiles, $\mathcal{R}_F(\tau) = \mathcal{R}_F(1)$ for all $\tau \in [0, 1]$. Identification requires a stronger condition than excluding comonotonic payoff structures. A necessary condition for a matrix to reveal any information about the quantile τ is that

$$\underline{c}(x) > 1 \text{ for all } x \text{ such that } u(x) \in Q^*_F(0), \tag{48}$$

which boils down to the requirement that the payoffs with the same rank r_f in the outcome sets $f(\mathcal{S})$, $f \in \mathcal{F}$, cannot be associated with the same event.

Step 2: Find the set of outcomes that may be anticipated under optimal choice for any τ and any π. By Proposition 2, this set is equal to $Q^*_F(0)$. Its complement in the outcomes set of F can be ignored, because it contains outcomes that will not affect choices of any τ-maximizer for any beliefs π she may hold. As a practical matter, Lemma 4 suggests that $Q^*_F(0)$ can be found as a union of subsets of payoffs that enter $Q^*_F(0)$ for beliefs with a different cardinality of support.

Step 3: For the observed choice f^*, apply Lemma 5 below to outcomes in $\{x \in f^*(\mathcal{S}) | u(x) \in Q^*_F(0)\}$.

Lemma 5 shows how to place a bound on τ from data containing only a single matrix and the choice of a decision-maker.

Lemma 5 *Consider a data set* $\{F, f^*\}$. *For all* $\tau > \bar{\tau}_F(f^*)$, *where*

$$\bar{\tau}_F(f^*) = \max_{x_{f^*} \in \{x \in f^*(\mathcal{S}) | u(x) \in Q_F^*(0)\}} \left\{ \frac{1}{\underline{c}(x_{f^*})} \right\}, \tag{49}$$

there is no belief $\pi \in \Delta(\mathcal{S})$ *for which* f^* *is optimal. For any* $\tau \leq \bar{\tau}_F(f^*)$, *there exists a belief for which* f^* *is chosen.*

The (proof of the) lemma also implies that after observing a choice of f^* in F, it is possible to make inferences about beliefs. Specifically, the choice may reveal not only whether beliefs have full-support, or the support is of lower cardinality, but also reveal the beliefs assigned to payoffs with ranks lower than the payoffs with the maximal rank in $\{x_{f^*} \in f^*(\mathcal{S}) | u(x) \in Q_F^*(0)\}$. Corollary 2 formalizes these conditions.

Corollary 2 *In a data set* $\{F, f^*\}$,

(i) $\quad |supp\{\pi\}| \geq \min_{x_{f^*} \in \{x \in f^*(\mathcal{S}) | u(x) \in Q_F^*(0)\}} \{\underline{c}(x_{f^*})\}$,

(ii) $$\sum_{\left\{ l \in \mathbb{N}_{++} \middle| l \leq \max_{\left\{x_{f^*,l} \in f^*(\mathcal{S}) | u(x) \in Q_F^*(0)\right\}} r_{f^*}(x_{f^*,l}) \right\}} \pi_l \geq \bar{\tau}_F(f^*).$$

We illustrate Lemma 5 and Corollary 2 in an example.

Example 5 *Suppose a decision maker chose* f_3 *in Matrix M3, where elements of the set*

$Q^*_{M3}(0)$ *are in bold.*

Matrix M3

	E_1	E_2	E_3
f_1	*1*	***6***	***9***
f_2	***8***	*2*	***6***
f_3	*7*	*4*	***3***
f_4	*6*	***5***	*2*

Using Lemma 5, we can conclude that under the Quantile Maximization hypothesis, τ *of the decision maker is not greater than* $\frac{1}{3}$*. Corollary 2 further implies that the agent holds full-support beliefs.*

We conclude with a few implications of the derived bounds.

Remark 2 *1. Among all matrices with* K *events, the maximal* $\hat{\tau}$ *such that all* τ*-maximizers with* $\tau \leq \hat{\tau}$ *may (for some beliefs) choose an act not justified by the Expected Utility hypothesis is equal to* $1 - \frac{1}{K}$*.*

2. Let λ *be a uniform measure on beliefs that an agent may hold,* $\Delta(\mathcal{S})$*. The decision makers with* τ *close to a multiple of* $\frac{1}{K}$ *are more likely (with respect to* λ*) to compare payoffs across rather than within ranks*[54]*; e.g., in Figure 2.3, Example 4, quantile maximizers with* τ *close to 0.5 are more likely to compare payoffs across ranks (when* $\pi < \tau \leq 1 - \pi$ *or* $1 - \pi < \tau \leq \pi$*) than are individuals with* τ *close to 0 or 1.*

The conditions we provided until now can be derived using the smallest data set (one matrix and choice). With a sequence of observed matrices and choices, one could apply the derived conditions on beliefs to place a l o w e r bound on τ and to tighten the upper bound. Furthermore, the results in this section offer testable conditions for finite amount of choice behavior to be consistent with Quantile Maximization.

[54] One might be interested in implementing a desired action. From the conditions derived in this Section, one could learn how to affect the optimal choice by redesigning a decision problem.

2.4 Strategic implications of Quantile Maximization

In this section, we illustrate how the conditions derived for a single-agent setting can be used in games. The application is of independent interest because Expected Utility dominates decision-theoretic foundations of games. Thus, we suggest how the differing properties of Quantile Maximization may contribute to modeling strategic interactions.

First, it is common in economic life that interactions are not anonymous; the players often know their opponents' attitudes toward risk. To capture this in the Expected Utility framework, the knowledge of concavity of the entire Bernoulli utility function must be assumed. In Quantile Maximization, information about risk attitude (alternatively, level of optimism-pessimism) is embedded in the single parameter τ. Second, in many games such as those involving one-shot interactions, it seems more plausible to assume that the players know each other's risk attitude than that they know each other's beliefs. (See also part (3) in the previous section.)

We now exploit the implications of players having (some) beliefs and knowing each other's anticipation levels, τ, and payoff ranking. We show that taking this view, identifying restrictions on choice behavior under Quantile Maximization that come from (possibly different orders of) knowledge of τ without imposing any conditions on beliefs, can increase the predictive power of the decision-making model in games.

Suppose we let I quantile maximizers interact strategically in a game defined as

$\Gamma = \{I,\ \{F_i\}_{i=1,...,I},\ \{\tau_i\}_{i=1,...,I}\}$. Define the rationalizable set for Quantile Maximization.

Definition 7 *Fix a vector* $\boldsymbol{\tau} = (\tau_1, ..., \tau_I)$*. Set* $\mathcal{R}^0_{\Gamma,i}(\tau_i) \equiv F_i$*,* $i = 1, ..., I$*, and for* $n \in \mathbb{N}_{++}$ *recursively define*

$$\mathcal{R}^n_{\Gamma,i}(\tau_i) = \{f_i \in \mathcal{R}^{n-1}_{\Gamma,i}(\tau_i) | \exists \pi_i \in \Delta(\times_{j \neq i}\ \mathcal{R}^{n-1}_{\Gamma,j}(\tau_j)) : f_i \in \arg\max_{\tilde{f} \in F_i} Q^{\tau_i}(\Pi_{\tilde{f}})\}. \qquad (50)$$

The τ_i-rationalizable set for player i *is the limit set*

$$\mathcal{R}^{\infty}_{\Gamma,i}(\tau_i) = \cap_{n\in\mathbb{N}}\mathcal{R}^{n}_{\Gamma,i}(\tau_i) \tag{51}$$

and τ-rationalizable set for the game Γ *is equal to* $\mathcal{R}^{\infty}_{\Gamma}(\boldsymbol{\tau}) = \cap_{i\in\mathcal{I}}\ \mathcal{R}^{\infty}_{\Gamma,i}(\tau_i)$.[55]

Remark 3

1. The common knowledge assumption implicit in the definition of τ*-rationalizability is considerably weaker than in an expectation-based model. Specifically, for* $\mathcal{R}^{\infty}_{\Gamma}$*, the knowledge of utility functions is not required, only that of the pure-strategy preference ranking is. When* $\boldsymbol{\tau}$ *is fixed, the sets* $\mathcal{R}^{\infty}_{\Gamma}(\boldsymbol{\tau})$ *are defined by the common knowledge of (i) quantile-maximization rationality, (ii) outcome ranking and (iii)* $\boldsymbol{\tau}$*.*

2. Fix a game Γ*. Proposition 2 and monotonicity of the relation defined in (50) imply that for quantile maximizers with different* τ*'s, the rationalizable sets are weakly nested with respect to* τ*:*

$$\mathcal{R}^{\infty}_{\Gamma}(1) \subseteq ... \subseteq \mathcal{R}^{\infty}_{\Gamma}(\tau) \subseteq ... \subseteq \mathcal{R}^{\infty}_{\Gamma}(0). \tag{52}$$

It follows that $\mathcal{R}_{\Gamma}(0)$ *is the rationalizable set under Quantile Maximization.*

Consequently, the rationalizable set can be partitioned into subsets parameterized in τ. Example 6 illustrates this point.

Example 6 *Consider a two-player symmetric game in Matrix M4. Applying Lemma 5, one can conclude that if quantile maximizers with* $\tau > 0.5$ *play this game, then the outcome of such an interaction will occur in the subset* $\{f_1^1, f_2^1\} \times \{f_1^2, f_2^2\}$*. Conversely, knowing the outcomes, can one characterize players' risk attitudes (*τ*'s) from the sole observation of the outcomes? If outcome* $\{f_3^1, f_3^2\}$ *is observed, for example, then under the Quantile Maximization hypothesis, one can conclude that* τ *of the players must be weakly less than*

[55] Notice that the formulation of the iterative procedure ensures that the anticipated payoffs are in the limit set.

0.5.

Matrix M4

	f^2_1	f^2_2	f^2_3
f^1_1	*5,5*	*0,3*	*6,2*
f^1_2	*3,0*	*7,7*	*0,2*
f^1_3	*2,6*	*2,0*	*2,2*

We further note that although our axiomatization in Rostek [2006] gives foundations for a fixed τ, the framework of Quantile Maximization can naturally incorporate dependence of choice rules on a particular subset of actions (through τ). Examples of applications include modeling effects of experience, framing (e.g., a player might be more cautious when playing less experienced actions, or when considering deviations). In an analogous manner, event-dependence can be allowed in studying the impact of context.

Chapter 3

Uncertainty about Rationality

3.1 Motivation

What is the impact of how people perceive the rationality of others on equilibrium outcomes? The very notion of equilibrium used in economics implicitly assumes that the impact is none: Equilibrium outcomes are always weakly within the rationalizable set, where rationality is common knowledge. The weakest equilibrium notion that captures the implications of players rationality and the game structure was identified by Brandenburger and Dekel (1987); they showed that a slight strengthening of subjective correlated equilibrium (Aumann [1974]), *a posteriori* equilibrium, attains the rationalizable set of outcomes.[56]

At the same time, a large body of experimental research on play in normal-form games finds that outcomes systematically occur outside of the rationalizable set in a given game (Stahl and Wilson [1994, 1995]; Nagel [1995]; Ho, Camerer, and Weigelt [1998]; Goree and Holt [2001]; Costa-Gomes, Crawford, and Broseta [2001]; Bosch-Domènech et al. [2002]; Camerer, Ho, and Chong [2004]; Costa-Gomes and Crawford [2006]; and Crawford and

[56] Recently, Dekel, Fudenberg and Morris (2006) proved that an analogous equivalence holds in incomplete-information settings.

Iriberri [2006a]). Moreover, invoking the standard recursive characterization of the rationalizable set (Pearce [1984]), players' actions are typically consistent[57] with only 1-2 rounds of deletion of dominated strategies. In that sense, the observed play is heterogenous (as opposed to requiring consistency with infinitely many rounds for all players), and the resulting outcomes occur far from that maximal equilibrium prediction. Whether or not the iterative description of reasoning is adequate, these findings are important in that they question the ability of equilibrium tools to explain outcomes in at least a range of strategic settings. The starting point of our analysis then is an observation that in order to predict outcomes in such interactions, a model should explain why players optimally choose non-rationalizable actions. For example, a common explanation of the (non-rationalizable) outcomes in guessing ("beauty contest") games is that the players perform a small number of iterations because they believe others will do so too and not because the players are not able to reason further or that they believe others are not able to reason as well. Unlike the standard game-theoretical analysis, there the players are actually better off not completing the infinite regress of deletion, given that their opponents do not complete it. And yet, such an explanation does not have a clear foundation in the models to date, because optimality and cognitive abilities are confounded. Providing both decision-theoretic and learning foundations is the objective of this paper.

A remark qualifying the settings for which our analysis is intended is in order here. The classic examples of games that involve explicit iterative reasoning are dominance-solvable games, such as the guessing games mentioned above or traveler's dilemma. More generally, this paper is concerned with interactions that are strategic and yet involve little or no opportunity to learn either through experimenting or from the environment; thus, not even the so-called "steady-state" concepts based on (implicit or explicit) learning would be appropriate.[58] The need for providing foundations for such games, examples of which

[57] The evidence is in fact stronger than mere consistency of outcomes with levels of iterations, as many of these studies back up the choice observations with data on the subjects' information searches.

[58] *Self-confirming equilibirum* by Fudenberg and Levine (1993), *subjective equilibirum* by Kalai and Lehrer

include one-shot interactions (e.g. court cases, business negotiations, auctions, military actions), has often been acknowledged (Kalai and Lehrer [1995, p. 127], Rubinstein and Wolinsky [1994, p. 299], Goree and Holt [2001, p. 1419], among others).

Before we describe our model and relate it to previous developments, we will make precise the sense through which we will refer to the notion of equilibrium as "standard" as well as the relation between equilibrium and rationalizability.

Equilibrium versus rationalizability. To make our point, it suffices to abstract away the specific requirements of different solution concepts and focus instead on the mildest restriction on beliefs implicit in any "equilibrium" concept, as commonly understood. In geometric terms, the condition stipulates that players' joint best-response set coincides with the support of joint beliefs about the strategies of others. We will call the condition a *weak closure* property, to distinguish it from a stronger and more common condition: The joint distribution of players' beliefs exactly coinciding with the joint distribution of players' strategies (*strong closure*; holds, e.g., in Nash Equilibrium). Since that fixed-point property of the best-response mapping implies that there can be no action in the support not justified by some conjecture, the best response set cannot lie outside of the rationalizable subset of a game.

Alternatively, in epistemic terms, one could apply the results of Brandenburger and Dekel (1987) and Zambrano (2005) to show that, when equipped with epistemic language, first, the above condition entails that players (mutually) know the support of others' conjectures; and, second, that (together with a mutual knowledge of rationality and payoffs) the mutual knowledge of supports is precisely the epistemic condition characterizing rationalizable subsets of a game. Thus, even though the sufficient epistemic conditions for an equilibrium need not require common knowledge of rationality, behavior that is non-

(1993, 1995), *conjectural equilibrium* by Battigalli (1987), in Gilli (1987), Battigalli-Guaitoli (1988) and Rubinstein-Wolinsky (1994).

rationalizable (consistent with less than infinite rounds of elimination of strictly dominated strategies) is not an admissible equilibrium behavior. It follows then that in equilibrium, players must be certain (believe with probability one) that their opponents are rational.

Therefore, in order to account for the finite levels of reasoning in the data, one has to relax the condition that beliefs supports are equal to the joint best-response set. For games with more than two players, this relaxation will further remove the requirement that beliefs supports are identical across players (on the same domains). It is worth noticing that moving away from "equilibrium" has already proved successful in accounting for the winner's curse in common-value auctions, overbidding in independent-private-value auctions, trade in textbook no-trade settings, and other commonly observed phenomena (Eyster and Rabin [2005, *cursed equilibrium*], Jehiel and Koessler [2006, *analogy-based expectations*], Battigali and Siniscalci [2003, k-rationalizability], Crawford and Iriberri [2006b, k-level model]).[59] We next argue, however, that in previous studies, relaxing the weak closure property was achieved by essentially assuming bounded rationality rather than by providing an optimality argument.[60]

Received literature. Numerous models have been advanced to explain the observed outcomes. We will introduce ours by comparing it to the most closely related, structural approach,[61] which has also been the most popular in empirical studies. It assumes that individuals perform only k (as opposed to ∞) steps of reasoning, where each step corre-

[59] Likewise, evolutionary dynamics does support eliminating strictly dominated strategies. Hofbauer and Sandholm (2006) show that any evolutionary dynamic that satisfies three mild requirements (continuity, positive correlation, and innovation) does not eliminate strictly dominated strategies in all games. The class of dynamics they consider includes not only well-known dynamics from the evolutionary literature, but also slight modifications of the dynamics under which elimination is known to occur.

[60] Relaxing strong closure has been advocated by the learning ("steady-state") literature cited above. The common idea, shared by the present paper, is that in equilibrium the players' beliefs should not be contradicted, even if those beliefs could possibly be incorrect. Unlike the conditions we propose, however, the "steady-state" concepts require that the distributions of beliefs match those of actions at terminal nodes (the distributions need not match otherwise). Hence, these concepts have no bite in one-shot games, and they cannot account for evidence on normal-form games, which is the focus of our paper.

[61] (Unstructured) implications of relaxing common knowledge of rationality have been analyzed in Börgers (1993), Mukerji (1997), Blume, Brandenburger and Dekel (1991).

sponds to one round of deletion of strictly dominated strategies (k*-level models* proposed by Stahl and Wilson [1994, 1995] and Nagel [1995]).[62] This approach presupposes that each k-step thinker conjectures all others to reason to exactly $k-1$ steps with certainty. In their *Cognitive Hierarchy* model, Camerer, Ho and Chong (2004) allowed player uncertainty about her opponents' *depth of reasoning* k: A k-thinker assumes others' levels of reasoning are distributed from 0 to $k-1$. Their model is the closest to ours and the connection is important since the Cognitive Hierarchy performs at least as well as several alternatives do in explaining data.

In either approach, consistency between the observed outcomes and the conjectures can be viewed as modeled through the following two conditions:

(i) each player maximizes her expected utility given her beliefs (optimization),

(ii) the observed outcome is in the support of the players' joint conjectures (no *ex post* surprise).

Both approaches, however, place additional exogenous restrictions on beliefs:

(1) To justify finite depth levels, $k < \infty$, they assume that a k-player knows the actual (conditional) distribution of $\{0, 1, ..., k-1\}$ (in the Cognitive Hierarchy model) or believes that all other players perform $k-1$ steps (in the k-level model). This assumption is arguably strong, given the intended departure from equilibrium. In addition:

• The consistency requirement on beliefs does not admit the case where a k-player might be aware that others perform k or more steps. Symmetric levels (k, k) can be justified only if both players are boundedly rational at k, which in the Cognitive Hierarchy implies that the players think of others as at-most-$(k-1)$ thinkers, or as exactly $(k-1)$-thinkers in the k-level model.

• More importantly, if repeated play were allowed, it is not clear how updating beliefs would look, given the extra condition on beliefs. In either approach, beliefs are anchored in

[62] The 0-level players are typically assumed to choose at random. Our specification will allow for all possible distributions.

the fixed conjectured depths of others. In the k-level model, these conjectures are simply set at the opponents' action subset, corresponding to the $(k-1)^{th}$ level; in the Cognitive Hierarchy, they are equal to rescaled actual frequencies of reasoning types, truncated up to $k-1$ for a k-player. Repeated interactions would thus not naturally yield learning.

(2) Relatedly, as in the first generation of iterative cognitive models, in the Cognitive Hierarchy players' actual levels of reasoning are exogenously fixed with respect to the predetermined beliefs.

Consequently, predictions are observationally equivalent whether these levels are determined by players' cognitive limitations or players find it optimal not to reason further.[63] In that sense, the proposed explanation of why players do not reason further essentially involves limited cognitive abilities instead of optimality to stop at small finite levels. The latter is a common and intuitive explanation informally given for the outcomes in guessing games. Nevertheless, such an explanation does not appear consistent with the existing models. Providing both decision-theoretic and learning foundations is the goal of this paper. We would like to view our model as providing an optimality argument for Camerer, Ho and Chong's (2004) framework and the earlier iterative models, which can be viewed as specialized versions of our model. We will then suggestively call our model an *Optimal Cognitive Hierachy*.

[63] In addition, if objectivity of distribution were taken seriously, then bounded rationality of players would be necessary to justify finite depth levels, $k < \infty$. To see that , consider a 2-player one-shot game (e.g., the "beauty contest"). In the Cognitive Hierarchy model:

- Asymmetric levels $(k, k+l)$ can be justified only if the k-player is boundedly rational at k. In particular, the model does not admit that she might find it optimal not to reason further even though she would be able to do so. This feature of the model is due to the requirement that the distribution of cognitive types is objectively known.
- Not only is the k-player's choice constrained by bounded rationality at k, but also in the $(k, k+l)$ case both players must think of each other as boundedly rational. Again, the model does not allow the players to believe that their opponent does not find it optimal to stop. This inconsistency will not arise if it is assumed that the players only mutually know the types distribution, but do not know that others know.

Model. In this paper, the apparent bounded reasoning will be optimal, in a sense made precise next, and endogenous with respect to players' uncertainty about the rationality of others, as captured by their conjectures. Specifically, we propose that each player holds a conjecture about the level of rationality of her opponents and responds optimally given her conjecture. To distinguish between her subjective assessment of others' optimal k and her possibly limited cognitive abilities, we also explicitly account for her maximal depth of reasoning. Even if no player is boundedly rational, that player still might (optimally) choose to reason up to small finite levels. This is because the player might conjecture that others will not reason further (cf. guessing games). A formal optimality argument for the observed behavior and the separation of optimality from limited cognitive ability are the key novel features of our model. The discussion above suggests how separating optimality and cognitive limitations qualitatively affects predictions.

Consistency between the observed outcomes and the conjectures is modelled through the conditions that we identified as common to both k-level models and the Cognitive Hierarchy: (i) each player optimizes given her beliefs and (ii) the observed outcome is in the support of the joint conjectures of players. But now, players' beliefs must be consistent with their own cognitive abilities and conjectured levels of others, which in turn must be not be contradicted by the observed outcomes. That second consistency condition crucially differs from fixing the levels of conjectures in k-level models or fixing distributions in the Cognitive Hierarchy.

Conceptually, our conditions are different from the commonly used "equilibrium" conditions in that they are imposed prior-to-play and post-play, respectively. That is, instead of requiring optimality of s t r a t e g i e s g i v e n s t r a t e g i e s o f o t h e r s (as in "equilibrium"), leaving cognitive constraints aside, we require optimality of s t r a t e g i e s and o u t c o m e s g i v e n c o n j e c t u r e s about k. The change is meant as an alternative to the standard inter-play[64] view of modeling non-cooperative behavior, which forces consistency

[64]The names, "prior-to-play," "intra-play" and "post-play," are intended as evocative of the three stages

of the distributions of beliefs and actions.

Results. We derive and characterize the maximal set of outcomes determined by (1) and (2). Only to facilitate comparisons with the k-level and the Cognitive Hierarchy models, we treat players as effectively certain about their opponents' reasoning depths. This allows us to focus on the smallest among the maximal (given $\mathbf{k}$) sets of outcomes that can be implemented with players who reason at most up to $\mathbf{k}$.[65] Our central result shows equivalence of the set to a counterpart of the rationalizable set whose size can be parameterized by the actual levels of reasoning that the players use. In other words, uncertainty about rationality shapes the set of outcomes.

We show that if a player conjectures that the maximal depth of reasoning reached by others is k, she will reason $\min\{k+1, k^a\}$ levels. In the language of iterative reasoning, each k-step thinker in our model conjectures that others reason up to at most k levels. While the *Cognitive Hierarchy* model by Camerer, Ho and Chong (2004) makes a similar prediction, in their model "each player assumes that his strategy is the most sophisticated" (p. 861; emph. MR). We allow for a different mechanism: the finite depths $\{k_i\}_{i\in\mathcal{I}}$ arise because the players are not willing to reason further subject to their cognitive abilities. Due to the distinction between optimality and reasoning ability in our model, the derived levels of k actually used, and hence also the predicted outcome sets, are generically different than those in the Cognitive Hierarchy.

Freeing players' conjectures about the depth of reasoning of others and tying these conjectures to the observed outcomes instead allows us to ask further what happens to the predicted set of outcomes when players interact repeatedly and (Bayes-)update their

in a strategic interaction that are separated by the act of choosing a strategy and the outcome realization. These stages should not be taken for *ex ante*, *interim* and *ex post* stages, which refer to the revelation of private information. For an argument on how powerful the "prior-to-play" view is epistemically, see Barelli (2006).

[65] The set-prediction allows for tightening of the customary interpretation of players as having models of their oponent's behavior: A model involves a single-parameter (depth of reasoning) rather than a measure over actions.

beliefs. We find sufficient conditions under which the outcome set converges and show that it converges to the set achieved by the Cognitive Hierarchy (if $k_i^a < k_{ij}^c$ for each $i, j \in \mathcal{I},\ i \neq j$).

3.2 Characterization result

This section derives a precise relation between the optimal depth of reasoning used by the players and the resulting outcome set. Consider a normal-form game in which $\mathcal{I}$ is the set of players $\mathcal{I} = \{1, ..., I\}$, A_i is a finite set of actions of player i and $U_i : \prod_{i \in \mathcal{I}} A_i \to \mathbb{R}$ is her (von Neumann Morgenstern) utility function. We now introduce a language to talk about depths of reasoning: We will distinguish between a player's reasoning abilities, her conjectured depths of reasoning actually used by other players, her own actual beliefs consistent with her cognitive ability as well as the conjectured levels of others.

Let $\mathbf{k}_{i.}^c = (k_{i,1}^c, ..., k_{i,i-1}^c, k_{i,i+1}^c, ..., k_{i,I}^c)$ be a vector of player i's *conjectures about the actual levels of rationality* that the other players (are able or willing to) use. Define $k_{.i}^c = \max_{j \neq i}\{k_{j,i}^c\}$. If k_i^a is the level of i's *cognitive ability*, then the *actual level of reasoning* she uses given her conjectures $\mathbf{k}_{i.}^c$ and ability k_i^a will be denoted by k_i. This is where her own reasoning ability might be binding. Formally, the game is defined by $\Gamma = \{\mathcal{I}, \{A_i\}_{i \in \mathcal{I}}, \{U_i\}_{i \in \mathcal{I}}, \{k_i^a, \mathbf{k}_{i.}^c\}_{i \in \mathcal{I}}\}$.

For a set A_{-i}, $\Delta(A_{-i})$ is the set of probability measures on A_{-i}. We can now iteratively define the set of actions of player i remaining after k_i rounds of deletion of strict never-best responses. For each $i \in \mathcal{I}$, let $\mathcal{R}_i^0 = A_i$ and for $k_i \geq 1$, $k_i \in \mathbb{N}$,

$$\mathcal{R}_i^{k_i}(\Gamma) = \left\{ a_i \in A_i | a_i \text{ is a b.r. to some } \lambda_i \in \Delta\left(\mathcal{R}_{-i}^{k_i - 1}(\Gamma)\right)\right\}. \tag{53}$$

Call $\mathcal{R}_i^{k_i}(\Gamma)$ a k_i^{th}*-order rationalizable set* of player i's actions. An action profile $a = (a_1, ..., a_{\mathcal{I}})$ is said to be k_i^{th}-order rationalizable if each a_i is k_i^{th}-order rationalizable and $\mathbf{k} = (k_1, ..., k_I)$. Let $\mathcal{R}^{\mathbf{k}}(\Gamma) = \times_{i \in I} \mathcal{R}_i^{k_i}(\Gamma)$. Analogous definitions hold for payoffs (see Section 3.3).

The support of players' beliefs $\lambda_i : A_{-i} \to \Delta(A_{-i})$ depends on i's conjectured reasoning levels of others $\mathbf{k}^c_{i\cdot}$ and her reasoning ability k^a_i, for which k_i is a sufficient statistic. Formally, we will say that player i's beliefs λ_i are *consistent with her actual level of reasoning* k_i if $supp\{\lambda_i\} \subseteq \mathcal{R}^{k_i}_{-i}(\Gamma)$. Hereafter, we assume that beliefs λ_i are admissible in that sense and write $\lambda_i(k_i) \in \Delta(A_{-i})$.

We need to include information structure into the description of the game. Thus, let Ω be a finite state space, $\lambda_i : \Omega \to \Delta(\Omega)$ - player i's probability measure on Ω[66], and $\mathcal{H}_i$ - her partition of Ω. The states in Ω include payoff but not action information (cf. Aumann [1987]). Unless stated otherwise, best responses will be meant to be *interim*, that is, after the players learned their private information.[67] A strategy of player i is then an $\mathcal{H}_i$ -measurable mapping $s_i : \Omega \to A_i$.

We close the model by specifying the consistency conditions that define optimal cognitive behavior.

Definition 8 *A strategy-belief profile* $(\mathbf{f}^*, \boldsymbol{\lambda}^*(\mathbf{k}))$ *is* optimal cognitive *in* Γ *if, for each* $i \in \mathcal{I}$ *and for each* $\omega \in \Omega$,

$$\begin{aligned} &\text{(i)} \quad \textstyle\sum_{a_{-i}\in A_{-i}} \lambda_i(k_i)[\{\omega' : \mathbf{f}_{-i}(\omega') = \mathbf{a}_{-i}\}|\mathcal{H}_i(\omega)]U_i(f(\omega)_i, \mathbf{a}_{-i}) \geq \\ &\qquad \textstyle\sum_{a_{-i}\in A_{-i}} \lambda_i(k_i)[\{\omega' : \mathbf{f}_{-i}(\omega') = \mathbf{a}_{-i}\}|\mathcal{H}_i(\omega)]U_i(a_i, \mathbf{a}_{-i}), \ \forall a_i \in A_i; \qquad (54) \\ &\text{(ii)} \quad \lambda^*_i(k_i)[\mathbf{f}^*_{-i}(\omega') = \mathbf{a}^*_{-i}|\mathcal{H}_i(\omega)] > 0. \end{aligned}$$

The conditions can be interpreted, respectively, as "no unilateral incentive to deviate," or "maximization," and "no *ex post* surprise." In particular, the equilibrium outcomes might be affected by players' beliefs over non-equilibrium actions. Condition (ii) guarantees that

[66] For the discussion when (if at all) it is legitimate to call this probability measure a "prior", see Aumann (1998), Gul (1998), and Dekel and Gul (1997).

[67] This is for convenience. In Section 3.3 we show that the results hold if analysis is carried at the *ex ante* stage.

these beliefs are never falsified. We are interested in characterizing the smallest of the maximal (given $\mathbf{k}$) set of optimal cognitive outcomes of the game Γ, which we call *optimal cognitive hierarchy* of Γ, $\mathcal{OCH}(\Gamma, \mathbf{k})$.

Remark 4 *In their characterization of subjective correlated equilibrium, Brandenburger and Dekel (1987) rely on the equivalence of the following two conditions on a strategy profile* $\mathbf{f}(\omega) = (f_1(\omega), ..., f_{\mathcal{I}}(\omega))$ *to define a posteriori (interim subjective correlated) equilibrium: for every state* $\omega \in \Omega$,

$$\begin{aligned} \sum_{\omega' \in \Omega} \lambda_i[\{\omega\prime\}|\mathcal{H}^i(\omega)]U_i(\mathbf{f}_i(\omega), \mathbf{f}_{-i}(\omega')) &\geq \\ \sum_{\omega' \in \Omega} \lambda_i[\{\omega\prime\}|\mathcal{H}^i(\omega)]U_i(a_i, \mathbf{f}_{-i}(\omega')), \forall a_i &\in A_i; \end{aligned} \tag{55}$$

and, appealing to a change of the variable, for every state $\omega \in \Omega$,

$$\begin{aligned} \sum_{a_{-i} \in A_{-i}} \lambda_i[\{\omega' &: \mathbf{f}_{-i}(\omega') = a_{-i}\}|\mathcal{H}^i(\omega)]U_i(a_i, a_{-i}) \geq \\ \sum_{a_{-i} \in A_{-i}} \lambda_i[\{\omega' &: \mathbf{f}_{-i}(\omega') = a_{-i}\}|\mathcal{H}^i(\omega)]U_i(\tilde{a}_i, a_{-i}), \ \forall \tilde{a}_i \in A_i. \end{aligned} \tag{56}$$

Yet, the equivalence (in particular, the former definition) presupposes that, for every possible state of the world that may occur, each player knows which actions will be taken by her opponents for some of their beliefs. Knowing the range of the other players' strategy functions (knowledge of $\mathbf{f}_{-i}$ *is sufficient but not necessary) renders the equilibrium condition necessarily consistent with common knowledge of rationality. This is an intuitive restatement of the sufficient epistemic conditions for rationalizability (Zambrano [2005]). Thus, uncertainty about whether other players use the full extent of infinite-step reasoning or certainty about less-than-rationalizable level rationality of others is not allowed in equilibrium.*

Let $\bar{\mathbf{k}}_{i.} = \max_{j \neq i}\{\hat{n}_{i,j}\}$, $\bar{\mathbf{k}}_{.\mathcal{I}} = (\bar{k}_{.1}, ..., \bar{k}_{.I})$, $\bar{\mathbf{k}}_{\mathcal{I}.} = (\bar{k}_{1.}, ..., \bar{k}_{I.})$, $\mathbf{k}^a = (k_1^a, ..., k_I^a)$ and $\mathbf{k}_{\mathcal{I}} = (k_1, ..., k_I)$. Finally, define $\mathbf{k}^*$ as $\max\{\mathbf{k}_{\mathcal{I}}, \bar{\mathbf{k}}_{.\mathcal{I}}\}$ (a pairwise comparison). This is a vector of reasoning levels consistent with each player's ability and perceptions of the others as well as others' perceptions of player i. The following result shows that the reasoning levels $\mathbf{k}$ yield the set of outcomes of an optimal cognitive hierarchy consistent with and equivalent to the $\mathbf{k}^{*th}$-rationalizable set: $\mathcal{R}^{\mathbf{k}^*}(\Gamma) = \mathcal{OCH}(\Gamma, \mathbf{k})$.

Proposition 3 *Fix a game* Γ. *The set of* $\mathbf{k}^{*th}$*-rationalizable actions is equal to the set of cognitively optimal actions in* Γ, *where for each* $i \in \mathcal{I}$, $k_i = \min\{k_i^a, k_{i.}^c\}$.[68]

The "$\mathcal{R}^{\mathbf{k}^*}(\Gamma) \supseteq \mathcal{OCH}(\Gamma, \mathbf{k})$" direction shows that the optimal cognitive outcomes are contained within the set of a size that is a function of the levels at which people reason. The more interesting converse "$\mathcal{R}^{\mathbf{k}^*}(\Gamma) \subseteq \mathcal{OCH}(\Gamma, \mathbf{k})$" demonstrates that the subset of Γ whose size is determined by the players' actual levels of reasoning is cognitively optimal.

The proof uses the familiar mediator construct in an amended canonical game. As in the standard implementation literature, this version of the canonical game still allows the mediator to recommend only a specific action to each player. Without assuming the knowledge of the entire strategy function, the recommendation can nonetheless capture the implications for outcomes being consistent with players' perception of the level of rationality of others. This result requires the mediator to know the vector $\mathbf{k}$. We also consider a setting where the mediator has no means to find out what $\mathbf{k}$ is. A new canonical game is proposed, in which the mediator can recommend to each player a subset of his action space. Again, more than $\mathcal{R}^{\infty}(\Gamma)$ can be implemented.

Proof:

$\mathcal{R}^{\mathbf{k}^*}(\Gamma) \subseteq \mathcal{OCH}(\Gamma, \mathbf{k})$: The result is an application of the Revelation Principle.

[68] This will follow from optimality. Writing out, $k_i = \begin{cases} k_i^a & \text{if } k_i^a = \bar{k}_i^c \\ \bar{k}_{i.}^c + 1 & \text{if } k_i^a > \bar{k}_i^c \end{cases}$. Unlike in the Cognitive Hierarchy model, it is possible that $k_i^a = \bar{k}_i^c$. Notice that the "min" operator is consistent with the possibility that a player realizes that others might reason further than his ability ($k_i = k_i^a$ if $k_i^a < \bar{k}_i^c$); there is no such distinction in the Cognitive Hierarchy, which assumes that $k_i^a > \bar{k}_i^c$.

Let $U^{\mathcal{R}^{\mathbf{k}^*}}$ denote the set of interim $\mathbf{k}^{*th}$-rationalizable payoffs.

If the mediator knows the vector $\mathbf{k}^*$*:* Fix a vector of $\mathbf{k}^{*th}$-rationalizable payoffs $u \in U^{\mathcal{R}^{\mathbf{k}^*}}$.[69] Assume that for each $i \in \mathcal{I}$ and each $a_i \in \mathcal{R}_i^{\mathbf{k}^*}(\Gamma)$, there exists beliefs $\lambda_i(k_i)$ such that a_i is a best reply to $\lambda_i(k_i)$ (proved to be true in Step 2). Consider the following canonical game: The mediator randomly selects an action profile $a \in \times_{i \in \mathcal{I}} \mathcal{R}_i^{\mathbf{k}^*}(\Gamma)$ and recommends each player i to play a_i in a (without revealing recommendations for other players). By construction, the action recommended to player i is consistent with her cognitive ability, k_i^a, as well as her perceptions about others' actual levels, $k_{i.}^c + 1$. The definition also ensures that the realized outcome will be consistent with interim beliefs: $\lambda_i^*(k_i)[\mathbf{f}_{-i}^*(\omega') = \mathbf{a}_{-i}^* | \mathcal{H}_i(\omega)] > 0$. That there exist such beliefs $\lambda_i^*(k_i)' \in \Delta\left(\mathcal{R}_{-i}^{k_i - 1}\right)$ and a corresponding best-response action $a_i' \in \mathcal{R}_i^{\mathbf{k}^*}(\Gamma)$ for payoff $u \in U^{\mathcal{R}^{\mathbf{k}^*}}$ follows form the fact that u_i is a $\mathbf{k}^{*th}$-rationalizable payoff to i. Thus, If player i is recommended to play a_i, then her conditional belief about the mediator's choice in $\mathcal{R}_{-i}^{k_i - 1}(\Gamma)$ is $\lambda_i(k_i)$. For $\hat{a}_i \neq a_i$ choose $\hat{\lambda}_i \in \Delta\left(\mathcal{R}_{-i}^{k_i - 1}\right)$ such that $\hat{a}_i$ is a best reply. Hence, the condition (i) holds. By the definition of $\mathbf{k}^*$, so does condition (ii) hold, which implements a_i as an optimal cognitive action.

If the mediator does not know $\mathbf{k}^*$ *(e.g. he cannot receive information from the players, but he can only send it):* The mediation procedure in which the mediator recommends a particular action to each player can no longer be used. The problem would arise if the mediator tried to recommend an action in $\mathcal{R}_i^{\check{k}_i}(\Gamma)$ to a player for whom $k_i > \check{k}_i$. Consider an alternative communication mechanism in the canonical game:

Stage 1: The mediator randomly selects an outcome a in the game.

Stage 2: He recommends to each player what to play according to the following procedure: Let k_i^R, $i \in \mathcal{I}$, be the minimum level of reasoning that is consistent with rationalizable play. If the randomly drawn profile in Stage 1 assigns a rationalizable action to player i,

[69] We will repeatedly use Pearce's (1984) Proposition 1, which implies that the latter set is nonempty and contains at least one pure strategy for each player, $\mathbf{k} \in \mathbb{N}_+^I$.

$a_i \in \mathcal{R}^{k_i^R}$, that action is recommended. If it is $\check{k}_i^{th}$-rationalizable but not rationalizable, the mediator recommends i to play an action in the set $\{a_i^R, \check{a}_i\}$, where $a_i^R \in \mathcal{R}^{n_i^R}$ (again, randomly selected) and $\check{a}_i \in \mathcal{R}^{\check{k}_i}$. In addition, he informs the player that the expected play of her opponents can be described by a probability measure $\check{\lambda}_i \in \Delta\left(\mathcal{R}_{-i}^{\check{k}_i-1}\right)$, so that should the player reason at a level not higher than $\check{k}_i$ and choose an action from $\{a_i^R, \check{a}_i\}$, then her best response would be $\check{a}_i$.

With the set-recommendation, if $\check{k}_i > k_i$, player i will choose to play $\check{a}_i$. The proof can be completed in a manner analogous to the case when the mediator knows n.

$\mathcal{R}^{\mathbf{k}^*}(\Gamma) \supseteq \mathcal{OCH}(\Gamma)$: Let players' cognitive abilities and their conjectured levels of reasoning of other players be $\mathbf{k}^a$ and $\mathbf{k}^c$, respectively. Given, $\mathbf{k}^a$ and $\mathbf{k}^c$, fix a vector of interim payoffs that corresponds to an optimal cognitive action profile $\mathbf{f}^* \in \mathcal{OCH}(\Gamma)$. Next, fix an action profile $\mathbf{a} \in \{\mathbf{a} \in A | a_i = f_i(\omega'),\ i \in \mathcal{I},\ \omega' \in \Omega\}$. Choose ω such that $\mathbf{a} = \mathbf{f}^*(\omega)$. Since for each i, f_i^* is an optimal cognitive choice for i, there are beliefs $\lambda_i \in \Delta(\mathcal{R}_{-i}^{k_i})$ for which a_i is a weak expected-payoff maximizer and these beliefs are consistent with i's given level of reasoning, k_i. Hence, $a_i \in \mathcal{R}_i^{k_i}(\Gamma)$. Using that $\mathbf{f}_{-i}^*$ is optimal cognitive for players $j \neq i$, the action a_i is consistent with their conjectured reasoning level of i. That is, for each $j \neq i$,

$$\lambda_j^*(k_j)[f_i^*(\omega') = a_i^* | \mathcal{H}_j(\omega)] > 0.$$

It follows that $a_i^* \in \mathcal{R}_i^{\bar{k}_{.i}}(\Gamma)$. Hence, $a_i \in \mathcal{R}_i^{k_i}(\Gamma) \cap \mathcal{R}_i^{\bar{k}_{.i}}(\Gamma) = \mathcal{R}_i^{\max\{k_i, \bar{k}_{.i}\}}(\Gamma) = \mathcal{R}_i^{k_i^*}(\Gamma)$ and i's optimal cognitive payoff given information $\mathcal{H}^i(\omega)$ is k_i^{*th}-order rationalizable and equal to $\sum_{a_{-i} \in A_{-i}} \lambda_i(k_i)[\{\omega' : \mathbf{f}_{-i}(\omega') = \mathbf{a}_{-i}\} | \mathcal{H}^i(\omega)] U_i(a_i, \mathbf{a}_{-i})$ ■.

Thus, a finite and less-than-rationalizable level of knowledge (certainty) of rationality provides justification for equilibrium analysis. The result in Brandenburger and Dekel (1987) obtains when there is common knowledge of rationality; then $\mathcal{R}^\infty(\Gamma) = \mathcal{OCH}(\Gamma)$.

3.3 Discussion

(1) Correlated vs. independent beliefs. One might argue that allowing for correlating devices through correlating beliefs may not be suitable for lab settings, where the advantage is precisely in the explicit control of any correlating devices. Correlation in beliefs does not, however, drive the results. Assuming correlation away leads to the notion of independent rationalizability.

Proposition 4 *Fix a game* Γ. *The set of* $\mathbf{k}^{*th}$*-rationalizable actions is equal to the set of cognitively optimal actions taken under conditionally independent beliefs in* Γ*, where for each* $i \in \mathcal{I}$*,* $k_i = \min\{k_i^a, k_{i.}^c\}$*.*[70]

(2) Robustness. Here, we ask whether the main result is robust to altering three assumptions implicit in its formulation.

• Would the equivalence in Proposition 3, expressed for actions, hold when analyzed in terms of payoffs?

• Would the equivalence result, stated for *interim* actions (and payoffs), hold for *ex ante* actions (and payoffs)?

• Would the result, formulated for Bayesian rationality (subjective expected utility), hold if other restrictions on rationality were adopted?

The answer to all three questions is positive. Specifically, the first assumption can be altered if and only if the second can, which is a consequence of the convexity of sets of payoffs in the Optimal Cognitive Hierarchy, $\mathcal{OHC}(\Gamma, \mathbf{k})$, and k-rationalizability, $\mathcal{R}^{\mathbf{k}}$. This property will be important for characterizing the outcome set to which repeated play converges.

Proposition 5 *Propositions 3 and 4 hold for actions as well as the sets of interim and ex ante payoff vectors.*

[70] Conditionally independent beliefs are understood as in Brandenburger and Dekel (1987). The proof is analogous to the proof of the previous proposition and is omitted.

The third question can be answered using the method proposed by Epstein (1997).

(3) Product-set predictions.

The derived optimally cognitive outcome (and strategy) sets and the considered beliefs supports are product subsets of the strategy space. This feature would not change if beliefs supports were instead defined on non-product subsets of the strategy space. Indeed, one could show that for any fixed non-product beliefs support in a given game Γ, the maximal $\mathcal{OCH}(\Gamma)$ set is equal to the $\mathcal{OCH}(\Gamma)$ set generated by beliefs whose product support is the smallest common coarsening.

(4) Non-tight closeness under rational behavior. A notion of optimality close to ours has been studied by Basu and Weibull (1991). They proposed to call a set *closed under rational behavior* (*curb*) if the set contains all its best responses. Curb sets have been applied by Rostek (2003) and Tercieux (2004). Since any game is trivially curb, the applications only used *tight* curb sets, that is, sets that coincide with their best replies. It is immediate from our discussion in Section 3.1, that tightness implies rationalizability (Rostek [2003] gives a formal proof).

Remark 5 *If a subset of the action space $\tilde{A} \subseteq A$ is tight, then $\tilde{A}$ is rationalizable.*

So far, no tool has been proposed to model closure under rational behavior in a non-trivial way without imposing the tightness property (and thereby leading to predictions necessarily consistent with rationalizability). Given Proposition 3, the conditions studied in this paper can be thought as defining such a tool.[71]

[71] In practice, a more general notion of curb sets has been used: a set is $(p_1, ..., p_I)$-*curb* if when every player believes with probability at least p_i, $i = 1, ..., I$, that the opponents will choose an action from this set, all her best responses are there. Observe that introducing p-belief does not alter the result that tight p-curb sets are rationalizable. Our sets can be interpreted as $p_{\mathbf{K}}(\mathbf{k})$-curb, where $\mathbf{k} \in \mathbb{N}^I$ is the vector of reasoning levels (optimally) used by the players, and $\mathbf{K} \in \mathbb{N}^I$ is the cardinality of the the action space in belief supports. In particular, $p_{\mathbf{K}}(\mathbf{K})$-curb sets are p-curb.

(5) LEARNING: Suppose the players play the game repeatedly, Bayes-updating their beliefs. We show that if the play is long enough, the outcome set converges to the one defined by the players' cognitive abilities, $\mathbf{k}^a$. The latter is the endogenous counterpart of the Cognitive Hierarchy.[72] If players are not boundedly rational, the play converges to subjective correlated equilibria (or rationalizability). The next proposition offers a formal statement. We first define a notion of closeness of distributions as follows: for a fixed $\varepsilon > 0$ and two probability distributions μ and λ on A, say that λ is *ε-close* to μ if for any subset $\tilde{A}$ in A, $\left|\lambda\left(\tilde{A}\right) - \mu\left(\tilde{A}\right)\right| \leq \varepsilon$. If a strategy is optimally congitive for ε-close beliefs, we say that it is *ε-cognitively optimal.* For every history of length t, beliefs $\{\boldsymbol{\lambda}_\mathbf{s}\}_{s=1,\ldots,t}$ and distributions of outcomes $\{\boldsymbol{\mu}_\mathbf{s}\}_{s=1,\ldots,t}$ induce a strategy profile $\mathbf{f}_h$.

Proposition 6 *Fix Γ. Assume that for each player $i \in \mathcal{I}$,*

(i) a strategy f_i is optimal given beliefs λ_i, initial conjectured reasoning levels $\mathbf{k}^c_{i.}$ and cognitive ability k^a_i;

(ii) $\lambda^*_{it}(k_i)[\mathbf{f}^*_{-i}(\omega') = \mathbf{a}^*_{-i}] > 0$, for every t.

Then, for any $\varepsilon > 0$, there is a T such that for all $h > T$, the players' strategies $\mathbf{f}_h$ are ε-cognitive optimal relative to beliefs $\boldsymbol{\lambda}_\mathbf{t}$, with $k^c_{ij} = \min\{k^a_i, k^a_j\}$, $j \neq i$.

Proof. Beliefs $\boldsymbol{\lambda} = \{\boldsymbol{\lambda}_\mathbf{t}\}_{t=1,\ldots,T}$ are measures over product subsets of the strategy subspace $A_{-i} \in \times_{j\neq i} A_j$, $i \in \mathcal{I}$. Strategies $\{\mathbf{f}_t\}_{t=1,\ldots,T}$ and beliefs $\boldsymbol{\lambda} = \{\boldsymbol{\lambda}_\mathbf{t}\}_{t=1,\ldots,T}$ induce a distribution on outcome subsets, $\boldsymbol{\mu} = \{\boldsymbol{\mu}_\mathbf{t}\}_{t=1,\ldots,T} \in \{\Delta_t\left(A_{-i}\right)\}_{t=1,\ldots,T}$. Condition (ii) implies that $\boldsymbol{\mu}$ is absolutely continuous with respect to $\boldsymbol{\lambda}$. The assertion is then implied by the theorem in Blackwell and Dubins (1962). ■

[72] Because of the lack of a consistency condition that ties observed reasoning depths to players conjectures, the Cognitive Hierarchy model as such, cannot be embedded in a consistent way in a dynamic game with beliefs updating. This can be done by dropping some auxiliary assumptions (e.g., *a priori* precluding a player to believe that others reason as far as she does herself) and introducing a consistency condition between the conjectures about others' depth of reasonign and the realized outcomes.

Remark 6 *(1) The result is stronger than stated. The convergence holds, regardless of whether the players observe their opponents' payoffs.*[73] *Imperfect monitoring is thus allowed, which might be of interest for applications without an explicit focus on depth of rationality.*

(2) An interesting connection of the reasoning and dynamics presented here is to subjective games *by Kalai and Lehrer (1995). Our structural assumption that players hold perceptions about others' depth of reasoning can be re-cast as what they call "an environment response function," an exogenously given subjective belief about what outcomes will occur. Our formulation is a set-theoretic extension of subjective games.*

[73] Or, the static model need not have the literal interpretation regarding the depths of reasoning. It suffices for all results that players reason about the set of actions that the oponents might take for some beliefs (or, the range of $\mathbf{f}_{-i}$). In particular, the players need not reason about strategies *per se*. Consequently, if each player $i \in \mathcal{I}$ reasons directly about action subsets, A_{-i}, the prediction for a given game Γ can be found as $\bigcap_{i \in \mathcal{I}} A_{-i} \cap \mathcal{R}^{\mathbf{k}}(\Gamma)$.

Appendix 1: Axioms of Savage (1954), Machina and Schmeidler (1992), Grant (1995), Chew and Sagi (2006)

We state the set of axioms developed by Savage (1954) that characterize Subjective Utility Maximization.[74]

AXIOM P1 (**ORDERING**): *The relation $\succsim$ is complete, transitive and reflexive.*

AXIOM P2 (**SURE-THING PRINCIPLE**): *For all events E and acts f, f^*, g, h,*

$$\begin{bmatrix} f^*(s) \text{ if } s \in E \\ g(s) \text{ if } s \notin E \end{bmatrix} \succsim \begin{bmatrix} f(s) \text{ if } s \in E \\ g(s) \text{ if } s \notin E \end{bmatrix} \Rightarrow \begin{bmatrix} f^*(s) \text{ if } s \in E \\ h(s) \text{ if } s \notin E \end{bmatrix} \succsim \begin{bmatrix} f(s) \text{ if } s \in E \\ h(s) \text{ if } s \notin E \end{bmatrix}. \tag{57}$$

AXIOM P3 (**EVENTWISE MONOTONICITY**): *For all outcomes x* and *y, non-null events E* and acts *g,*

$$\begin{bmatrix} x \text{ if } E \\ g \text{ if } E^c \end{bmatrix} \succsim \begin{bmatrix} y \text{ if } E \\ g \text{ if } E^c \end{bmatrix} \Leftrightarrow x \succsim y. \tag{58}$$

AXIOM P4 (**WEAK COMPARATIVE PROBABILITY**): *For all events E, F and outcomes $x^* \succ x$ and $y^* \succ y$,*

$$\begin{bmatrix} x^* \text{ if } E \\ x \text{ if } E^c \end{bmatrix} \succsim \begin{bmatrix} x^* \text{ if } F \\ x \text{ if } F^c \end{bmatrix} \Rightarrow \begin{bmatrix} y^* \text{ if } E \\ y \text{ if } E^c \end{bmatrix} \succsim \begin{bmatrix} y^* \text{ if } F \\ y \text{ if } F^c \end{bmatrix}. \tag{59}$$

AXIOM P5 (**NONDEGENERACY**): *There exist outcomes x and y such that $x \succ y$.*

AXIOM P6 (**SMALL-EVENT CONTINUITY**): *For any acts $f \succ g$ and outcome x, there exists a finite set of events $\{E_1, ..., E_N\}$ forming a partition of $\mathcal{S}$ such that*

$$f \succ \begin{bmatrix} x \text{ if } E_n \\ g \text{ if } E_n^c \end{bmatrix} \text{ and } \begin{bmatrix} x \text{ if } E_m \\ f \text{ if } E_m^c \end{bmatrix} \succ g \tag{60}$$

[74] We present axioms for the setting with finite-outcome acts and therefore we omit **P7** (which holds for τ-maximization).

for all $m, n = 1, ..., N$.

Machina and Schmeidler (1992) characterize probabilistically sophisticated preferences of a non-expected utility maximizer by **P1**, **P3**, **P4***, **P5**, and **P6**, where

AXIOM P4* (**STRONG COMPARATIVE PROBABILITY**): *For all pairs of outcomes* $x^* \succ x$, $y^* \succ y$, *events* E, F, $E \cap F \neq \varnothing$, *and acts* g, h,

$$\begin{bmatrix} x^* \text{ if } s \in E \\ x \text{ if } s \in F \\ g \text{ if } s \notin E \cup F \end{bmatrix} \succsim \begin{bmatrix} x^* \text{ if } s \in F \\ x \text{ if } s \in E \\ g \text{ if } s \notin E \cup F \end{bmatrix} \Rightarrow \begin{bmatrix} y^* \text{ if } s \in E \\ y \text{ if } s \in F \\ h \text{ if } s \notin E \cup F \end{bmatrix} \succsim \begin{bmatrix} y^* \text{ if } s \in F \\ y \text{ if } s \in E \\ h \text{ if } s \notin E \cup F \end{bmatrix} \tag{61}$$

Grant (1995) shows that conditions **P1**, **P4**^{CE}, **P5**, **P6**†, and one of **P3**^{CU} or **P3**^{CL} imply probabilistic sophistication, as defined in ((24) and (25)), where

AXIOM P3^{CU} (**CONDITIONAL UPPER EVENTWISE MONOTONICITY**): *For all pairs of non-null, disjoint events* E *and* F, *all outcomes* x *and* y, *and all acts* g,

$$\left(\begin{bmatrix} x \text{ if } E \cup F \\ g \text{ if } (E \cup F)^c \end{bmatrix} \succ (\succsim) \begin{bmatrix} y \text{ if } E \cup F \\ g \text{ if } (E \cup F)^c \end{bmatrix} \right) \Rightarrow \begin{bmatrix} x \text{ if } E \\ y \text{ if } F \\ g \text{ if } (E \cup F)^c \end{bmatrix} \succ (\succsim) \begin{bmatrix} y \text{ if } E \cup F \\ g \text{ if } (E \cup F)^c \end{bmatrix}. \tag{62}$$

AXIOM P3^{CL} (**CONDITIONAL LOWER EVENTWISE MONOTONICITY**): *For all pairs of non-null, disjoint events* E *and* F, *all outcomes* x *and* y, *and all acts* g,

$$\left(\begin{bmatrix} x \text{ if } E \cup F \\ g \text{ if } (E \cup F)^c \end{bmatrix} \succ (\succsim) \begin{bmatrix} y \text{ if } E \cup F \\ g \text{ if } (E \cup F)^c \end{bmatrix} \right) \Rightarrow \begin{bmatrix} x \text{ if } E \cup F \\ g \text{ if } (E \cup F)^c \end{bmatrix} \succ (\succsim) \begin{bmatrix} x \text{ if } E \\ y \text{ if } F \\ g \text{ if } (E \cup F)^c \end{bmatrix}. \tag{63}$$

AXIOM P4^{CE} (STRONG CONDITIONAL EQUIVALENT PROBABILITY): *For all pairs of disjoint events E and F, and outcomes w, x, y, and z, and acts g and h,*

$$\begin{bmatrix} x \text{ if } E \cup F \\ g \text{ if } (E \cup F)^c \end{bmatrix} \succ \begin{bmatrix} x \text{ if } E \\ y \text{ if } F \\ g \text{ if } (E \cup F)^c \end{bmatrix} \sim \begin{bmatrix} y \text{ if } E \\ x \text{ if } F \\ g \text{ if } (E \cup F)^c \end{bmatrix} \succ \begin{bmatrix} y \text{ if } E \cup F \\ g \text{ if } (E \cup F)^c \end{bmatrix} \tag{64}$$

implies

$$\begin{bmatrix} w \text{ if } E \\ z \text{ if } F \\ h \text{ if } (E \cup F)^c \end{bmatrix} \sim \begin{bmatrix} z \text{ if } E \\ w \text{ if } F \\ h \text{ if } (E \cup F)^c \end{bmatrix}. \tag{65}$$

AXIOM P6† (SMALL-EVENT CONTINUITY): *For any acts $f \succ g$ and outcome x, there exists a finite set of events $\{E_1, ..., E_N\}$ forming a partition of $\mathcal{S}$, such that for all $n = 1, ..., N$, and all $F_n \subseteq E_n$,*

$$f \succ \begin{bmatrix} x \text{ if } s \in F_n \\ g \text{ if } s \notin F_n \end{bmatrix} \text{ and } \begin{bmatrix} x \text{ if } s \in F_n \\ f \text{ if } s \notin F_n \end{bmatrix} \succ g \tag{66}$$

Chew and Sagi (2006) demonstrate that axioms **P1**, **P5**, **N**, **C** and **A** are equivalent to there being a unique, solvable, and finitely additive agreeing[75] probability measure, π, for $\succsim^C$. In addition, π is either atomless or purely and uniformly atomic,[76] any two events have the same measure if and only if they are exchangeable, and the decision maker is indifferent between any two acts that induce the same lottery with respect to π.

AXIOM N (EVENT NON-SATIATION): *For any pairwise disjoint events $E, F, G \in \mathcal{E}$, if E is exchangeable with F and G is non-null, then no subevent of F is exchangeable with*

[75] π is an *agreeing* probability measure for a likelihood relation $\succsim_l$ if it is a probability measure over $\mathcal{E}$ and for every $A, B \in \mathcal{E}$, $A \succsim_l B \Leftrightarrow \pi(A) \geq \pi(B)$.

π is *solvable* if for every $A, B \in \mathcal{E}$, if $\pi(A) \geq \pi(B)$ then there exists a subevent $C \subseteq A$ with $\pi(C) = \pi(B)$.

[76] π is *purely* and *uniformly atomic* if the union of all atoms has unit measure and all atoms have equal measure.

$E \cup G$.

Axiom C (**Completeness of** $\succsim^C$): *Given any disjoint pair of events, one of the two must contain a subevent exchangeable with the other.*

Axiom A (**Event Archimedean Property**): *Any sequence of pairwise disjoint and non-null events, $\{E_n\}_{n=0}^{N} \subseteq \mathcal{E}$, such that $E_n \approx E_{n+1}$ for every $n = 0, \ldots$ is necessarily finite.*

Appendix 2: Proof of Theorem 1

In part 2A, we establish several auxiliary results that will be frequently used in the main part of the proof. Part 2B presents the proof for the non-extreme preferences (cf. definition 2). Although the general line 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 are employed. In order to highlight the differences, we present the proof for the extreme preferences separately in part 2C. Part 2D contains the proofs of Lemmas 2 and 3.

2A. Auxiliary results

The lemmas in this section serve to characterize the relations over events in $\mathcal{E}$ used in the paper.[77]

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 any mapping such that $g_{x^*+}(\mathcal{S}) \succ x^*$, $g_{x^*-,x+}$ as any mapping such that $x \prec g_{x^*-,x+}(\mathcal{S}) \prec x^*$, and g_{x-} as any mapping such that $g_{x-}(\mathcal{S}) \prec x$.

Lemma 6 *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-},*

$$\begin{bmatrix} g_{x^*+} \text{ if } s \in G_1 \\ x^* \text{ if } s \in E \\ g_{x^*-,x+} \text{ if } s \in G_2 \\ x \text{ if } s \in F \\ g_{x-} \text{ if } s \in G_3 \end{bmatrix} \succ \begin{bmatrix} g_{x^*+} \text{ if } s \in G_1 \\ x^* \text{ if } s \in F \\ g_{x^*-,x+} \text{ if } s \in G_2 \\ x \text{ if } s \in E \\ g_{x-} \text{ if } s \in G_3 \end{bmatrix} \Rightarrow \begin{bmatrix} h_{y^*+} \text{ if } s \in G_1 \\ y^* \text{ if } s \in E \\ h_{y^*-,y+} \text{ if } s \in G_2 \\ y \text{ if } s \in F \\ h_{y-} \text{ if } s \in G_3 \end{bmatrix} \succ \begin{bmatrix} h_{y^*+} \text{ if } s \in G_1 \\ y^* \text{ if } s \in F \\ h_{y^*-,y+} \text{ if } s \in G_2 \\ y \text{ if } s \in E \\ h_{y-} \text{ if } s \in G_3 \end{bmatrix}. \tag{67}$$

[77] **P5'** ensures that the relations $\succ_*$, $\succ^*$, and $\succ_{**}$ are nontrivial.

Proof. Let $x^* \succ x$ and $y^* \succ y$. Assume

$$f = \begin{bmatrix} g_{x^*+} \text{ if } s \in G_1 \\ x^* \text{ if } s \in E \\ g_{x^*-,x+} \text{ if } s \in G_2 \\ x \text{ if } s \in F \\ g_{x-} \text{ if } s \in G_3 \end{bmatrix} \succ \begin{bmatrix} g_{x^*+} \text{ if } s \in G_1 \\ x^* \text{ if } s \in F \\ g_{x^*-,x+} \text{ if } s \in G_2 \\ x \text{ if } s \in E \\ g_{x-} \text{ if } s \in G_3 \end{bmatrix} = f'. \tag{68}$$

Then, the pivotal events for acts in (68) are E and G_2, or G_2 and E. Suppose

$$f'' = \begin{bmatrix} h_{y^*+} \text{ if } s \in G_1 \\ y^* \text{ if } s \in E \\ h_{y^*-,y+} \text{ if } s \in G_2 \\ y \text{ if } s \in F \\ h_{y-} \text{ if } s \in G_3 \end{bmatrix} \not\succ \begin{bmatrix} h_{y^*+} \text{ if } s \in G_1 \\ y^* \text{ if } s \in F \\ h_{y^*-,y+} \text{ if } s \in G_2 \\ y \text{ if } s \in E \\ h_{y-} \text{ if } s \in G_3 \end{bmatrix} = f'''. \tag{69}$$

Consider the case when the acts f'' and f''' in (69) are indifferent. Then, the pivotal events are G_1 and G_1, or E and F, or G_2 and G_2, F and E, or G_3 and G_3, respectively for f'' and f'''. For the pair, E and F, using $\mathbf{P3}^Q$ replace the outcome on the pivotal event in f''', F, with x^*, and the outcomes on G_1, G_2, E, and G_3 with g_{x^*+}, $g_{x^*-,x+}$, x, and g_{x-} , respectively. Then, by $\mathbf{P3}^Q$, the new act is indifferent to x^*. Applying $\mathbf{P3}^Q$ again to act f' and using **P1**, a contradiction obtains. An analogous argument can be used for the remaining pairs of pivotal events, which are possible when $f'' \sim f'''$ or when $f'' \prec f'''$. ■

One implication of Lemma 6 is that if event $(E \cup F)^c$ is null in $\mathbf{P4}^Q$, then a strict (rather than only a weak) implication can be assured.

Lemma 7 $E \succ^* (\succ_*) F \Leftrightarrow \mathcal{S} \sim^* (\sim_*) E$ *and* $F \sim^* (\sim_*) \varnothing$.

Proof. Since the arguments for $\succsim^*$ and $\succsim_*$ are very similar, we only prove the assertion for $\succsim_*$. Let $E \succ_* F$, that is by (22) and **P5**,

$$f = \begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix} = g, \ x \succ y. \tag{70}$$

Then, it must be that event E is pivotal for act f and F^c - for g. By $\mathbf{P3}^Q$,

$$f \sim y \sim \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix} \text{ and } g \sim x \sim \begin{bmatrix} x \text{ if } s \in \mathcal{S} \\ y \text{ if } s \notin \mathcal{S} \end{bmatrix}. \tag{71}$$

The definition of $\succsim_*$, **P1** and $\mathbf{P4}^Q$ yield $\mathcal{S} \sim_* E$ and $F \sim_* \varnothing$.

For the converse, assume $\mathcal{S} \sim^* (\sim_*) E, \ F \sim^* (\sim_*)\varnothing$. Using definition of $\succ_*$ in (22) and **P5**, for all $x \succ y$

$$\begin{array}{ccc} \begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} & & \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix} \\ \wr & & \wr \\ \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix} & & \begin{bmatrix} x \text{ if } s \in \mathcal{S} \\ y \text{ if } s \notin \mathcal{S} \end{bmatrix} \\ \wr & & \wr \\ y & \prec & x \end{array} \tag{72}$$

Then, by **P1**, $\mathbf{P4}^Q$ and (22), $E \succ_* F$, as desired. ■

Lemma 1 in Section 1.4 is a direct corollary:

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

Lemma 8 $E \succ^* (\succ_*)F \Rightarrow E \not\prec_* (\not\prec^*)F$.

Proof. This follows by $\mathbf{P4}^Q$ and Lemma 6. ■

Lemma 9 *If $F \subset E$, then $F \precsim_* (\precsim^*) E$.*

Proof. Let $F \subset E$ and define $H = E \backslash F$. Suppose $F \succ_* E$. Using definition of $\succ_*$,

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in H \\ y \text{ if } s \in F \end{bmatrix} \sim (by\ \mathbf{P3}^Q) \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} = \begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in H \\ y \text{ if } s \in F \end{bmatrix}. \tag{73}$$

Contradiction to **P1**. An analogous construction can be used for the relation $\succ^*$. ■

Lemma 10 *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 $E \cup G' \prec_* F \cup G'$.*

Proof. The assertion follows by $\mathbf{P4}^Q$. ■

Lemma 11 *If $E \prec_* (\prec^*) \mathcal{S}$, then E^c is non-null.*

Proof. Let $E \prec_* \mathcal{S}$ and suppose E^c is null. Then,

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \sim y, \tag{74}$$

which contradicts $E \prec_* \mathcal{S}$, using **P1**, $\mathbf{P4}^Q$, **P5**, Lemma 6 and the definition of $\succ_*$,

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix}, \text{ for any } x \succ y. \tag{75}$$

The argument for $\prec^*$ is analogous. ■

2B. Proof of Theorem 1 for non-extreme preferences

Proof. Assume that preferences are not extreme (cf. definition 2).

The proof consists of a series of steps. Step 1 demonstrates that $\succ_*$ and $\succ^*$ are weak orders. Step 2 characterizes the set of 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'** 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 $\mathcal{E}_{**}$, π. Next, Step 6 constructs a likelihood relation which is complete on the entire set of events, $\mathcal{E}$, and shows that measure π extends there. Finally, Step 7 establishes that $\succ$ is probabilistically sophisticated w.r.t. π. We will repeatedly invoke Lemma 6 without mentioning; it assures that the likelihood relations used in the axiomatization can be defined as revealed from preferences over acts.

Step 1 ($\succ_*$ **AND** $\succ^*$ **ARE WEAK ORDERS**):

1. We prove that $\succ_*$ is a weak order. Asymmetry is implied by the definition of $\succ_*$ (22), **P1** and **P5**. To show negative transitivity, suppose $E \not\succ_* F$ and $F \not\succ_* G$. Using $x \succ y$ (**P5**), (22) and **P4**^{Q}

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \not\prec \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix} \text{ and } \begin{bmatrix} x \text{ if } s \notin F \\ y \text{ if } s \in F \end{bmatrix} \not\prec \begin{bmatrix} x \text{ if } s \notin G \\ y \text{ if } s \in G \end{bmatrix}. \quad (76)$$

Then, **P1** yields

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \not\prec \begin{bmatrix} x \text{ if } s \notin G \\ y \text{ if } s \in G \end{bmatrix}, \quad (77)$$

hence $E \not\succ_* G$. An analogous argument proves that $\succ^*$ is a weak order.

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

Step 2 (**CHARACTERIZATION OF EQUIVALENCE CLASSES OF** $\mathcal{E}$ **UNDER** $\sim_*$ **AND** $\sim^*$):

1. Since $\succ_*$ on $\mathcal{E}$ is a weak order, $\sim_*$ is an equivalence relation. By Lemma 7, 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 $\mathcal{E}|_{\sim^*\mathcal{S}}$ =$\{F \in \mathcal{E}|F \sim^* \mathcal{S}\}$.

2. That the sets $\mathcal{E}|_{\sim_*\varnothing}$, $\mathcal{E}|_{\sim_*\mathcal{S}}$ and $\mathcal{E}|_{\sim^*\varnothing}$, $\mathcal{E}|_{\sim^*\mathcal{S}}$ are nondegenerate, follows from the assumption of non-extreme preferences. (See Step 3.3.)

Step 3 (**CONSTRUCTION OF** $\mathcal{E}_{**}$):

1. Define a binary relation over events, $\succ_{**}$, as in Definition 3.

2. Define $\mathcal{E}_{**} = \{\bar{E} \in \mathcal{E}|\bar{E} \prec_* E^c\}$. Fix $E \in \mathcal{E}_{**}$. Then, using $\mathbf{P3}^Q$, for any $F \in \mathcal{E}|_{\sim_*\varnothing}$ disjoint (w.l.o.g.) with E, there exists $G \subseteq (E \cup F)^c$ such that $F \cup G \succ_* \varnothing$. Hence, for any $E, F \sim \mathcal{E}_{**}$, 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 $E \cup G \sim_* F \cup G \sim_* \mathcal{S}$.

3. To verify that the relation $\succ_{**}$ is nondegenerate, observe that by the assumption of non-extreme preferences, there is a non-null event F such that $F^c \sim_* \mathcal{S}$ and F is not pivotal. If $F \succ_* \varnothing$, applying $\mathbf{P6}^{Q_*}$ to F and $\varnothing$, there is a non-null subevent of F, $\tilde{F}$, such that $\tilde{F} \sim_* \varnothing$ and $\tilde{F} \prec_* \tilde{F}^c$, where we used the assumption of non-extreme preferences and $\mathbf{P6}^{Q^*}$.

Applying $\mathbf{P6}^{Q_*}$ to $\tilde{F}^c$ and $\varnothing$, and again using the assumption of non-extreme preferences, yields an event $H \sim_* \varnothing$, $H \cap \tilde{F} = \varnothing$ in $\tilde{F}^c$ such that $H \cup G \succ_* \varnothing$ for some $G \subseteq \tilde{F}^c$.

4. We prove that either $\mathcal{E}_{**} = \mathcal{E}|_{\sim_*\varnothing}$ or $\mathcal{E}_{**} = \mathcal{E}|_{\sim^*\varnothing}$. Assume first that there exists $E \in \mathcal{E}|_{\sim_*\varnothing}$ such that $E \sim_* E^c \sim_* \varnothing$. We will show that $F \sim_* \varnothing$. Consider an event $F \in \mathcal{E}|_{\sim^*\varnothing}$. By the definitions of $\succ_*$ and $\succ^*$ and Lemma 1, $F^c \succ_* \varnothing$. Using the definitions of $\succ_*$ and $\succ^*$ again, $F \prec^* E$ and $F \prec^* E^c$. Then, by Lemma 8, $F \precsim_* E$ and $F \precsim_* E^c$.

Since $\succ_*$ is a weak order (Step 1), it follows from Lemma 7 that $F^c \succ_* F \sim_* \varnothing$. 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 $\mathcal{E}_{**} = \mathcal{E}|_{\sim_*\varnothing}$.

Step 4 (AXIOMS A1, A3, A4, A5' HOLD ON $\mathcal{E}_{**}$):

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

(**A1**) Consider $E \sim_* \varnothing$. By an argument as in Lemma 9, there cannot exist an event G such that

$$f = \begin{bmatrix} x \text{ if } s \notin E \cup G \\ y \text{ if } s \in E \cup G \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin G \\ y \text{ if } s \in G \end{bmatrix} = g. \tag{78}$$

Hence $E \not\succ_{**} \varnothing$. If E is null, then for all G

$$\begin{bmatrix} x \text{ if } s \notin E \cup G \\ y \text{ if } s \in E \cup G \end{bmatrix} \sim \begin{bmatrix} x \text{ if } s \notin G \\ y \text{ if } s \in G \end{bmatrix}. \tag{79}$$

Again, $E \not\succ_{**} \varnothing$.

(**A3**) That $\succ_{**}$ is asymmetric is implied by Lemma 10. Condition (i) in negative transitivity follows from the transitivity of $\sim_*$ ($E, F, H \in \mathcal{E}_{**}$). To prove that condition (ii) holds, suppose that there does not exist G non-null such that $E \cup G \succ_* F \cup G$ and there does not exist G non-null such that $F \cup G \succ_* H \cup G$. We need to show that for no G non-null, $E \cup G \succ_* H \cup G$. Since for all $G' \subseteq F^c \cap H^c$

$$\begin{bmatrix} x \text{ if } s \notin F \cup G' \\ y \text{ if } s \in F \cup G' \end{bmatrix} \succsim \begin{bmatrix} x \text{ if } s \notin H \cup G' \\ y \text{ if } s \in H \cup G' \end{bmatrix}, \tag{80}$$

by $\mathbf{P3}^Q$, (80) holds for any $G' \subseteq H^c$. Suppose that for some $G'' \subseteq (E \cup F)^c$

$$\begin{bmatrix} x \text{ if } s \notin E \cup G'' \\ y \text{ if } s \in E \cup G'' \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin F \cup G'' \\ y \text{ if } s \in F \cup G'' \end{bmatrix}. \tag{81}$$

Then, by Lemma 10, for any $G \subseteq (E \cup F)^c$

$$\begin{bmatrix} x \text{ if } s \notin E \cup G \\ y \text{ if } s \in E \cup G \end{bmatrix} \succsim \begin{bmatrix} x \text{ if } s \notin F \cup G \\ y \text{ if } s \in F \cup G \end{bmatrix}. \tag{82}$$

By $\mathbf{P3}^Q$, (82) is also satisfied for $G \subseteq E^c$. Suppose now that for all $G'' \subseteq (E \cup F)^c$

$$\begin{bmatrix} x \text{ if } s \notin E \cup G'' \\ y \text{ if } s \in E \cup G'' \end{bmatrix} \sim \begin{bmatrix} x \text{ if } s \notin F \cup G'' \\ y \text{ if } s \in F \cup G'' \end{bmatrix}. \tag{83}$$

Then, using the definition of $\mathcal{E}_{**}$ and $\mathbf{P3}^Q$, there cannot exist $G''' \subseteq E^c$ such that

$$\begin{bmatrix} x \text{ if } s \notin E \cup G''' \\ y \text{ if } s \in E \cup G''' \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin F \cup G''' \\ y \text{ if } s \in F \cup G''' \end{bmatrix}. \tag{84}$$

Using **P1**, it follows that there is no event $G \subseteq E^c \cap H^c$ for which

$$\begin{bmatrix} x \text{ if } s \notin E \cup G \\ y \text{ if } s \in E \cup G \end{bmatrix} \prec \begin{bmatrix} x \text{ if } s \notin H \cup G \\ y \text{ if } s \in H \cup G \end{bmatrix}. \tag{85}$$

That is, there does not exist $G \subseteq E^c \cap H^c$ such that $E \cup G \succ_* H \cup G$.

(**A4**) Assume ($E \cap H = F \cap H = \varnothing$) and $x \succ y$ (possible by **P5**).

($\Leftarrow$) Suppose (i) $E \cup H \sim_* F \cup H$ and (ii) $\exists G \subseteq (E \cup H)^c \cap (F \cup H)^c : E \cup H \cup G \succ_* F \cup H \cup G$. Since $E, F \in \mathcal{E}_{**}$, by Step 1.1, $E \sim_* F \sim_* \varnothing$. Taking $G' = G \cup H$ in (ii) directly gives $\exists G'$ non-null: $E \cup G' \succ_* F \cup G'$.

($\Rightarrow$) Suppose now that (i) $E \sim_* F$ and (ii) $\exists G'' : E \cup G'' \succ_* F \cup G''$. Given that $E \cup H,\ F \cup H \in \mathcal{E}_{**}$, by Step 1.1 we have $E \cup H \sim_* F \cup H$. By Lemma 10, $E \cup H \cup G''' \succsim_* F \cup H \cup G'''$. Suppose for all $G''' \subseteq (E \cup H)^c \cap (F \cup H)^c$, $E \cup H \cup G''' \sim_* F \cup H \cup G'''$; then, a contradiction obtains, using part ($\Leftarrow$).

(**A5'**) Suppose $E \sim_* F$ and $E \cup G \succ_* F \cup G$ for some G. Applying **P6**^{Q_*} to the latter, there exists a finite partition $\{H_1, ..., H_N\}$ such that for all $n = 1, ..., N$, $E \cup G \succ_* F \cup G \cup H_n$.

Step 5 (**DERIVATION OF** π **ON** $\mathcal{E}_{**}$):

The axioms **A1**, **A3**, **A4**, **A5'** hold for all subsets of $\mathcal{E}_{**}$. Therefore, there exists a unique, finitely additive, convex-ranged (and hence also nonatomic) probability measure π that represents $\succ_{**}$ on $\mathcal{E}_{**}$ (see, for example, Fishburn [1970, Ch.14]).

Step 6 (**EXTENDING** π **TO** $\mathcal{E}$):

From relation $\succ_{**}$ on $\mathcal{E}_{**}$, we derive a binary relation $\succ_{***}$ which is complete on $\mathcal{E}$.

1. We show that all events in $\mathcal{E} \backslash \mathcal{E}_{**}$ can be partitioned into finitely many events in $\mathcal{E}_{**}$. To this end, we find a partition of $\mathcal{S}$, each element 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 $\mathcal{S}$, $\{G_1, ..., G_N\}$ such that $E \succ_* G_n$, for all $n = 1, ..., N$. By Lemma 7, $G_n \sim_* \varnothing$, for all $n = 1, ..., N$, and hence $G_n^N \in \mathcal{E}_{**}$. Write $E = \left(\bigcup\limits_{\tilde{n}=1}^{m} G_{\tilde{n}}^N \right) \cup \left(E \backslash \bigcup\limits_{\tilde{n}} G_{\tilde{n}}^N \right)$. By construction, for all $\tilde{n} = 1, ..., m$, $G_{\tilde{n}}^N \in \mathcal{E}_{**}$, while $F \backslash \bigcup\limits_{\tilde{n}} G_{\tilde{n}}^N \subset G_{m+1}^N$ and hence $F \backslash \bigcup\limits_{\tilde{n}} G_{\tilde{n}}^N \in \mathcal{E}_{**}$. Thus events in the set $\mathcal{E} \backslash \mathcal{E}_{**}$ can be decomposed into events from the set $\mathcal{E}_{**}$.

Next, assume that $\mathcal{E}_{**} = \mathcal{E}|_{\sim^* \varnothing}$. By **P6**$^{Q^*}$, the state space $\mathcal{S}$ can be partitioned into finitely many events $\{H_1, ..., H_M\}$ such that $\mathcal{S} \succ^* H_m$ (and hence, by Lemma 7, $H_m \in \mathcal{E}_{**}$) for every $m = 1, ..., M$. Using the argument as for $\mathcal{E}_{**} = \mathcal{E}|_{\sim_* \varnothing}$, each event in $\mathcal{E}|_{\sim^* \mathcal{S}}$ can be partitioned into finitely many events in $\mathcal{E}|_{\sim_* \varnothing}$.

2. Define a binary relation over events in $\mathcal{E}$: $E \succ_{***} F$ if there exists N-partitions of

E and F such that for all $n = 1, ..., N$, $E_n \succ_{**} F_n$. Existence of such a partition follows from convex-rangedness of π: Consider $E, F \notin \mathcal{E}_{**}$ and let $\{E_1, ..., E_N\}$ and $\{F_1, ..., F_N\}$ be partitions of E and F, respectively, into elements in $\mathcal{E}_{**}$. By convex-rangedness of π, those partitions can be made equi-numbered and such that if $\sum_{n=1,...,N} \pi(E_n) > \sum_{n=1,...,N} \pi(F_n)$, then for each $n = 1, ..., N$, $\pi(E_n) > \pi(F_n)$.

3. We show that the probability measure π can be uniquely extended to a finitely additive and convex-ranged probability measure on the whole set $\mathcal{E}$, as follows: for each $E \in \mathcal{E} \backslash \mathcal{E}_{**}$ and its finite partition $\{E_1, ..., E_N\}$ let $\tilde{\pi} = \sum_{n=1,...,N} \pi(E_n)$.

Consider an event $E \in \mathcal{E} \backslash \mathcal{E}_{**}$ and its partitions $\{E_1, ..., E_N\}$, $\{F_1, ..., F_M\}$. Let $\{H_1, ..., H_L\}$ be the coarsest common refinement of those partitions. Uniqueness of summations follows immediately and since each event is finitely decomposed, $\tilde{\pi}$ is finitely additive on $\mathcal{E}$. To see that it is convex-ranged, for any $\rho \in [0, 1]$, take $\rho \cdot \tilde{\pi}(E) = \rho \cdot \sum_{n=1,...,N} \pi(E_n) = \sum_{n=1,...,N} \rho \cdot \pi(E_n) = \sum_{n=1,...,N} \pi(G_n)$, where we used that for each $n = 1, ..., N$, there is a $G_n \subseteq E_n$ such that $\pi(G_n) = \rho \cdot \pi(E_n)$.

This also shows that if there exists an N-partition of E and F such that for all $n \leq N$: $E_n \succ_{**} F_n$, then it cannot hold for any N'-partition that for all $n' \leq N'$: $E_{n'} \prec_{**} F_{n'}$.

Step 7 ($\succ$ **IS PROBABILISTICALLY SOPHISTICATED W.R.T.** π):

1. The proof of (23) 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 2A.

2. 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 $\pi \circ f^{-1} = P$. Therefore, using in addition that $\succ$ is a weak order, the stronger version of probabilistic sophistication, (2), from Section 1.3.2 is also satisfied. ■

2C. Proof of Theorem 1 for extreme preferences

Proof. Note: to aid in contrasting the arguments with those in the proof for the non-extreme preferences, each step is numbered with the same number as its counterpart step in that proof. Some steps are left out as no longer relevant.

Assume that preferences are extreme (definition 2).

Step 1:

As above.

Step 2:

1. Given that $\succsim_*^*$ on $\mathcal{E}$ is a weak order, Lemma 7 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 $\mathcal{E}|_{\sim_*^*\mathcal{S}}$ $=\{F \in \mathcal{E}|F \sim_*^* \mathcal{S}\}$.

2. We show that under (**H**), the equivalence classes $\mathcal{E}|_{\sim_*^*\varnothing}$ and $\mathcal{E}|_{\sim_*^*\mathcal{S}}$ are degenerate in that they contain events that differ from $\varnothing$ and $\mathcal{S}$, respectively, only on a null subevent:

$$\begin{aligned}\mathcal{E}|_{\sim_*^*\varnothing} &= \{E \in \mathcal{E}|E \text{ is null}\} \\ \mathcal{E}|_{\sim_*^*\mathcal{S}} &= \{F \in \mathcal{E}|F = \mathcal{S}\backslash H,\ H \text{ null}\}.\end{aligned} \tag{86}$$

Consider a non-null event E such that $\mathcal{S}\backslash E$ is non-null (possible by $\mathbf{P6}^{Q_*}$ and Lemma 11). Then, given the assumption (**H**),

$$\begin{bmatrix} x \text{ if } s \notin E \\ y \text{ if } s \in E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix} \text{ and } \begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \notin E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix},\ x \succ y. \tag{87}$$

Using $\mathbf{P4}^Q$ and definitions of $\succ_*$ and $\succ^*$, it follows accordingly that $E \prec_* \mathcal{S}$ and $E \succ^* \varnothing$. Thus by the definition of $\succsim_*^*$, the equivalence classes $\varnothing \prec_*^* E \prec_*^* \mathcal{S}$ and $\mathcal{E}|_{\sim_*^*\varnothing}$, $\mathcal{E}|_{\sim_*^*\mathcal{S}}$ are as defined in (86). The proof for the case (**L**) is analogous.

4. By Lemma 2B, all non-null events are ranked as equally likely by $\sim_*^*$ and are thus contained in $\mathcal{E}|_{\succ_*^*\varnothing}^{\prec_*^*\mathcal{S}}$.

Step 3:

Under (**H**) or (**L**), the subset of events in $\mathcal{E}|_{\succ_*^*\varnothing}^{\prec_*^*\mathcal{S}}$ that can be compared through relation $\succ_{**}$ only contains nested events that differ on non-null subevents. For example, under (**H**), consider two events $E_1, E_2 \in \mathcal{E}|_{\succ_*^*\varnothing}^{\prec_*^*\mathcal{S}}$ 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 $G = \mathcal{S} \backslash E_2$:

$$\begin{bmatrix} x \text{ if } s \notin E_1 \cup G \\ y \text{ if } s \in E_1 \cup G \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \notin \mathcal{S} \\ y \text{ if } s \in \mathcal{S} \end{bmatrix} = \begin{bmatrix} x \text{ if } s \notin E_2 \cup G \\ y \text{ if } s \in E_2 \cup G \end{bmatrix} \tag{88}$$

and hence by $\mathbf{P4}^Q$ and definition of $\succ_{**}$, $E_1 \prec_* E_2$.

However, the strict relation cannot be extended to non-nested events that differ on a non-null subset. What fails is condition (ii). This is because under (**H**) or (**L**), the events that could satisfy (ii) can be strictly ranked by the relation $\succ_*^*$ only with the events in $\mathcal{E}|_{\sim_*^*\mathcal{S}}$, and hence by Step 2 there cannot exist a non-null event in the common complement of the non-nested events for which (ii) would hold. Therefore, the relation $\succsim_{**}$ cannot rank events as indifferent and therefore, the strict ranking cannot be extended to non-nested events. In other words, there are no non-null events E such that $E \prec_*^* E^c$.

Step 4: By Step 3, there are no events in $\mathcal{E}|_{\succ_*^*\varnothing}^{\prec_*^*\mathcal{S}}$ satisfying **A4**. The remaining axioms hold.

Step 5:

1. Take an 2^N-uniform partition of $\mathcal{S}$. Let $E \subseteq S$ and $\underline{k}_E(2^N)$ be the *largest* integer

$m \in \mathbb{N}_+$ such that $\bigcup_{n_l=1}^{m} E_n^N \precsim_*^* E$, where $\bigcup_{n_l=1}^{m} E_n^N$ is a union of any m elements of some 2^N-partition of $\mathcal{S}$. Let $\bar{k}_E(2^N)$ be the *smallest* integer $m \in \mathbb{N}_+$ such that $\bigcup_{n_l=1}^{m} E_n^N \succsim_*^* E$. Define

$$
\begin{aligned}
\pi(E) &= \sup\left\{ \frac{\underline{k}_E(2^N)}{2^N} \middle| N \in \mathbb{N}_+ \right\} \text{ if } \inf \frac{\bar{k}_E(2^N)}{2^N} = \sup \frac{\underline{k}_E(2^N)}{2^N} && (89)\\
\pi(E) &\in \lim_N \left[\inf \frac{\bar{k}_E(2^N)}{2^N}, \sup \frac{\underline{k}_E(2^N)}{2^N}\right] \text{ otherwise.} && (90)
\end{aligned}
$$

By Lemma 2B, for each $E \in \mathcal{E}|_{\succ_*^*\varnothing}^{\prec_*^*\mathcal{S}}$, $\pi(E) \in \lim_N[\frac{1}{2^N}, \frac{2^N-1}{2^N}] = (0,1)$. (89) gives $\pi(\varnothing) = 0$, $\pi(\mathcal{S}) = 1$ and $\pi(E) \geq 0$, for all $E \subseteq \mathcal{S}$.

2. π is *totally finitely additive*: if $\bigcup_{n=1,\ldots,N} E_n = \mathcal{S}$, then, since $\underline{k}_{\bigcup_{n=1,\ldots,N} E_n}(2^N) = \underline{k}_{\mathcal{S}}(2^N) = 2^N$,

$$\pi\left(\bigcup_{n=1,\ldots,N} E_n\right) = 1. \quad (91)$$

3. By (89), (90) and Step 2, $E \succsim_*^* F \Rightarrow \pi(E) \geq \pi(F)$.

4. By Steps 1-4, any totally additive probability measure that

(i) satisfies $\pi(\varnothing) = 0$, $\pi(\mathcal{S}) = 1$;

(ii) assigns $0 < \pi(E) < 1$ to $E \in \mathcal{E}|_{\succ_*^*\varnothing}^{\prec_*^*\mathcal{S}}$;

(iii) is monotonic by respecting non-null differences on nested events: if $F \subset E$ and $E \backslash F$ non-null, then, $\pi(F) < \pi(E)$;

represents $\succsim_*^*$ under (**H**) or (**L**): $E \succ_*^* F \Leftrightarrow \pi(E) > \pi(F)$. Call the set of all measures that represent $\succsim_*^*$ under (**H**) $\Pi(\mathcal{E})^H$, and under (**L**) $\Pi(\mathcal{E})^L$.

5. Each measure $\pi \in \Pi(\mathcal{E})^H$ and $\pi \in \Pi(\mathcal{E})^L$ is nonatomic. We will prove this for (**H**). Fix $\pi \in \Pi(\mathcal{E})^H$ and consider an event $E \in \mathcal{E}|_{\succ_*^*\varnothing}^{\prec_*^*\mathcal{S}}$. By Steps 2 and 5.1, $0 < \pi(E)$ and

$\pi(E^c) > 0$. Using **P5** construct acts

$$\begin{bmatrix} x \text{ if } E \\ y \text{ if } E^c \end{bmatrix} \succ \begin{bmatrix} x \text{ if } \varnothing \\ y \text{ if } \mathcal{S} \end{bmatrix}, \tag{92}$$

that is, using the definition of $\succ_*$, $E^c \prec_* \mathcal{S}$. It follows by $\mathbf{P6}^{Q_*}$ that 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 $\mathbf{P6}^{Q_*}$. By Step 3 and Step 5.5, $\pi(E) > \pi(F)$ and $\pi(E) > \pi(E \backslash F)$. This completes the proof.[78] ■

Proof. Steps 6,7: Do not apply. ■

2D. Proofs of Lemmas 2 and 3

Lemma 2A. *If preferences over acts are not extreme, an individual is indifferent to exchanging outcomes on events equally likely according to* $\sim_{**}$.

Proof. Assume that preferences $\succ$ are not extreme. Let $E, F \in \mathcal{E}_{**}$ be a pair of disjoint events satisfying $\pi(E) = \pi(F)$. Using the definition of $\sim_{**}$,

$$E \sim_{**} F \text{ if for any } x \succ y, \begin{bmatrix} x \text{ if } s \notin E \cup G \\ y \text{ if } s \in E \cup G \end{bmatrix} \sim \begin{bmatrix} x \text{ if } s \notin F \cup G \\ y \text{ if } s \in F \cup G \end{bmatrix}, \text{ for any } G \subseteq (E^c \cap F^c). \tag{93}$$

[78] We sketch an argument to construct a nonatomic but not convex-range measure which satisfies all the required properties, derived in Step 5. Consider a measure $\tilde{\pi}$ and assume that all the properties from Step 5 hold. Take two nested events $E \supset F$ so that by Steps 3 and 5.4 $\tilde{\pi}(E) > \tilde{\pi}(F)$, and a sequence of strictly nested subsets of E, all of which are strict supersets of $F : \{H_n\}_{n \in \mathbb{N}_{++}}$, $F \subset H_1 \subset H_2 \subset ... \subset H_n \subset ... \subset E$. A measure $\tilde{\pi}$ that assigns $\tilde{\pi}(H_n) = 0.3 \cdot (\tilde{\pi}(E) - \tilde{\pi}(F)) \cdot (0.5)^{\frac{1}{n}}$ is clearly not convex-range.

Consider two events E, F satisfying (93). Take acts

$$f = \begin{bmatrix} h_{x+} \text{ if } s \in H \\ x \text{ if } s \in F \\ y \text{ if } s \in E \\ h_{y-} \text{ if } s \in I \end{bmatrix} \text{ and } g = \begin{bmatrix} h_{x+} \text{ if } s \in H \\ x \text{ if } s \in E \\ y \text{ if } s \in F \\ h_{y-} \text{ if } s \in I \end{bmatrix}. \tag{94}$$

If neither E nor F is pivotal, then the result follows by $\mathbf{P3}^Q$. Assume that E or F is pivotal.[79] Using $\mathbf{P4}^Q$, we have

$$\begin{array}{cccc} f = & \begin{bmatrix} h_{+} \text{ if } s \in H \\ x \text{ if } s \in F \\ y \text{ if } s \in E \\ h_{-} \text{ if } s \in I \end{bmatrix} & \begin{bmatrix} h_{+} \text{ if } s \in H \\ x \text{ if } s \in E \\ y \text{ if } s \in F \\ h_{-} \text{ if } s \in I \end{bmatrix} & = g \\ (\text{by } \mathbf{P3}^Q) & \wr & \wr & \\ & \begin{bmatrix} x \text{ if } s \in F \cup H \\ y \text{ if } s \in E \cup I \end{bmatrix} & \begin{bmatrix} x \text{ if } s \in E \cup H \\ y \text{ if } s \in F \cup I \end{bmatrix} & \\ & \shortparallel & \shortparallel & \\ (\text{by } (93)) & \begin{bmatrix} x \text{ if } s \notin E \cup I \\ y \text{ if } s \in E \cup I \end{bmatrix} \sim & \begin{bmatrix} x \text{ if } s \notin F \cup I \\ y \text{ if } s \in F \cup I \end{bmatrix} & \end{array} \tag{95}$$

■

Lemma 2B. *If preferences over acts are extreme, an individual is indifferent to exchanging outcomes on disjoint non-null events.*

Proof. Fix two non-null events E' and F', $E' \cap F' \neq \varnothing$ (disjointedness is w.l.o.g.) and

[79] Notice that we do not use here that the pivotal event is unique.

assume (**H**). Consider the following acts

$$f = \begin{bmatrix} x \text{ if } s \in F' \\ y \text{ if } s \in E' \\ h_{y-} \text{ if } s \notin E' \cup F' \end{bmatrix} \text{ and } g = \begin{bmatrix} x \text{ if } s \in E' \\ y \text{ if } s \in F' \\ h_{y-} \text{ if } s \notin E' \cup F' \end{bmatrix}, \; x \sim y. \tag{96}$$

By assumption, F' is pivotal in f and E' in g. It follows that $f \sim g$. The argument is exactly analogous under (**L**), and therefore it is omitted. ■

Lemma 3

(i) Given the partition of $\mathcal{S}$ induced by act $f \in \mathcal{F}$, there is a unique pivotal event.

(ii) Let acts $f, g \in \mathcal{F}$ be such that for any $s, s' \in \mathcal{S}$, $f(s') \succ f(s) \Rightarrow g(s') \succsim g(s)$. Then, $E_f \Delta E_g$ cannot be pivotal ($E_f \cap E_g \neq \varnothing$).

Proof. We prove the lemma in three steps: the property of being pivotal is (1) state-independent; the pivotal event is (2) unique to an act; and it is (3) unique between acts characterized in (ii). Having two non-indifferent outcomes suffices in each step of the proof. (**P5** guarantees that they exist.)

Step 1: We first state the key assertion, implied by Lemma 2A,B: the property of being pivotal is state-independent.

For the next two steps, the nontrivial case involves non-extreme preferences, which we assume for the remainder of the proof.

Step 2: (Part (i)) Suppose that disjoint events E_f and E_f^c are both pivotal to act f, and map to non-indifferent outcomes (**P5**). Applying $\mathbf{P3}^Q$ twice to f and using **P1** yields a contradiction:

$$x \sim \begin{bmatrix} x \text{ if } s \in E_f^{c\prime} \\ x \text{ if } s \in E_f \end{bmatrix} \sim_{\text{Pivotal } E_f'} \begin{bmatrix} x \text{ if } s \in E_f^c \\ y \text{ if } s \in E_f \end{bmatrix} \sim_{\text{Pivotal } E_f} \begin{bmatrix} y \text{ if } s \in E_f^c \\ y \text{ if } s \in E_f \end{bmatrix} \sim y. \tag{97}$$

Step 3: (Part (ii)) Take any pair of acts $f, g \in \mathcal{F}$. Consider the coarsest measurable partitions induced by f and g, and take their coarsest common refinement. Using convex-rangedness of measure, Lemma 2A, Step 2, and the construction analogous to the one from Machina and Schmeidler (1992, Theorem 1, Step 5), any pair of acts $f, g \in \mathcal{F}$ can be transformed, without affecting preferences $\succ$, in a finite sequence of steps to be comonotonic; that is to satisfy the following property: for all $s, s' \in S$, $f(s') \succ f(s) \Rightarrow g(s') \succsim g(s)$.

Let E_f be pivotal for act f and E_g - for act g. Assume that pivotal events map into indifferent outcomes (letting them be the same, x),

$$f = \begin{bmatrix} f_+ \text{ if } s \in E_{f+} \\ x \text{ if } s \in E_f \\ f_- \text{ if } s \in E_{f-} \end{bmatrix}, g = \begin{bmatrix} g_+ \text{ if } s \in E_{g+} \\ x \text{ if } s \in E_g \\ g_- \text{ if } s \in E_{g-} \end{bmatrix}. \tag{98}$$

We need to show that $E_f \cap E_g \neq \varnothing$ and $E_f \Delta E_g$ is not pivotal in any act.

By way of contradiction, suppose $E_f \cap E_g = \varnothing$; specifically, assume $E_g \subset E_{f-}$. (An analogous argument follows for $E_g \subset E_{f+}$.) Then, applying $\mathbf{P3}^Q$ to E_g,

$$x = \begin{bmatrix} x \text{ if } s \in E_{g+} \\ x \text{ if } s \in E_g \\ x \text{ if } s \in E_{g-} \end{bmatrix} \sim g \succ \begin{bmatrix} x \text{ if } s \in E_{f+} \\ x \text{ if } s \in (E_f \cup E_{f-}) \backslash (E_g \cup E_{g-}) \\ y \text{ if } s \in E_{g-} \cup E_g \end{bmatrix}. \tag{99}$$

But by $\mathbf{P3}^Q$, act f is indifferent to the last act in (99). Then, applying $\mathbf{P3}^Q$ to E_f, we obtain $f \sim x$, which, using $\mathbf{P1}$, contradicts (99).

If the disjoint pivotal events E_f, E_g map into non-indifferent outcomes, respectively x and y, $x \succ y$, then by $\mathbf{P1}$ and $\mathbf{P3}^Q$,

$$f \sim x \succ y \sim g \tag{100}$$

and further (still for $E_g \subset E_{f-}$; if $E_g \subset E_{f+}$, then construct an act that maps E_f to y and E_g to x),

$$f = \begin{bmatrix} f_+ \text{ if } s \in E_{f+} \\ x \text{ if } s \in E_f \\ f_- \text{ if } s \in E_{f-} \end{bmatrix} \sim \begin{bmatrix} x \text{ if } s \in E_{f+} \\ x \text{ if } s \in (E_f \cup E_{f-}) \backslash (E_g \cup E_{g-}) \\ y \text{ if } s \in E_g \cup E_{g-} \end{bmatrix} \tag{101}$$

$$g = \begin{bmatrix} g_+ \text{ if } s \in E_{g+} \\ y \text{ if } s \in E_g \\ g_- \text{ if } s \in E_{g-} \end{bmatrix} \sim \begin{bmatrix} x \text{ if } s \in E_{g+} \\ y \text{ if } s \in E_g \\ y \text{ if } s \in E_{g-} \end{bmatrix}. \tag{102}$$

But the last acts in (101) and (102) are the same, which given (100) contradicts the result established in Step 2). Hence $E_f \Delta E_g$ cannot be pivotal and $E_f \cap E_g \neq \varnothing$. ■

Appendix 3: Proof of Theorem 2

Proof. **((1) $\Rightarrow$ (2))**

Step 1 of the necessity part establishes the existence and uniqueness of a quantile τ. Step 2 constructs a preference functional over acts that represents $\succ$ on $\mathcal{F}$. Step 3 proves that the quantile is left-continuous.

Step 1 (**EXISTENCE AND UNIQUENESS OF** τ):

Assume that preferences are not extreme.

1. We will use repeatedly that: if $F \succ_{***} \varnothing$,[80] then for any $N \in \mathbb{N}_{++}$, there exists a 2^N-partition of F, $\{F_1^{2^N}, ..., F_{2^N}^{2^N}\}$ such that $F_1^{2^N} \sim_{***} ... \sim_{***} F_n^{2^N} \sim_{***} ... \sim_{***} F_{2^N}^{2^N}$ (given that the axioms **A1 - A5'** hold on the set $\mathcal{E}$, the argument in Fishburn [1970, Ch.14.2] can be applied).[81] Such a partition will be called a uniform 2^N-partition of F.

2. Consider a sequence of 2^N -uniform partitions. For a fixed N and $x \succ y$ (**P5**), consider acts

$$\begin{bmatrix} x \text{ if } \bigcup\limits_{l=n+1,...,2^N} F_l^{2^N} \\ y \text{ if } \bigcup\limits_{l=1,...,n} F_l^{2^N} \end{bmatrix}, \ n = 1, ..., 2^N. \tag{103}$$

By Lemma 3, for each n, only one of the subsets in an act $\bigcup\limits_{l=1,...,n} F_l^{2^N}$ and $\bigcup\limits_{l=n+1,...,2^N} F_l^{2^N}$ can be pivotal. Define event $F_p^{2^N} = \bigcup\limits_{l=1,...,n^p} F_l^{2^N}$ where n^p is such that

$$\begin{bmatrix} x \text{ if } \mathcal{S} \\ y \text{ if } \varnothing \end{bmatrix} \sim ... \sim \begin{bmatrix} x \text{ if } \left(\bigcup\limits_{l=1,...,n^p-1} F_l^{2^N}\right)^c \\ y \text{ if } \bigcup\limits_{l=1,...,n^p-1} F_l^{2^N} \end{bmatrix} \succ \begin{bmatrix} x \text{ if } \left(\bigcup\limits_{l=1,...,n^p} F_l^{2^N}\right)^c \\ y \text{ if } \bigcup\limits_{l=1,...,n^p} F_l^{2^N} \end{bmatrix} \sim ... \sim \begin{bmatrix} x \text{ if } \varnothing \\ y \text{ if } \mathcal{S} \end{bmatrix}. \tag{104}$$

By construction, $F_p^{2^N}$ is a pivotal event. Now, take the sequence of events, $\{F_p^{2^N}\}_{N \in \mathbb{N}_{++}}$. By Lemma 3, the events $F_p^{2^N}$ are nested and weakly decreasing. Define $\tau = \lim_{N\to\infty} \pi\left(F_p^{2^N}\right) =$

[80] The relation $\succ_{***}$ is defined in step 6.2 in the proof of Theorem 1.

[81] Step 7.2 could be used instead.

$\bigcap_N \pi\left(F_p^{2^N}\right)$. $\left\{\pi(F_p^{2^N})\right\}_{N\in\mathbb{N}_{++}}$ converges as a bounded monotonic sequence. Hence, τ exists and is unique.

3. Fix τ from Step 1.2. Using the definition of $\succ_P$, **P1** and **P3**^{Q}, it is straightforward to show that acts that imply indifferent τ^{th} outcomes $x \sim y$ are indifferent.

Now, assume that preferences are extreme.

4. Apply the reasoning from the proof for non-extreme preferences to uniformly bounded N-partitions, generated by **P6**^{Q_*} for (**H**) and **P6**$^{Q^*}$ for (**L**).

Under (**H**): Fix any measure $\pi \in \Pi(\mathcal{E})^H$ and define $\tau \equiv \lim_{k\to\infty} \pi(G_{p-}^{N^k} \cup G_p^{N^k}) = \pi(\mathcal{S}) = 1$, for any $\pi \in \Pi(\mathcal{E})^H$, where $G_p^{N^k}$ is the pivotal event in the collections of N^k-partitions constructed as in (104) and

$$G_{p-}^{N^k} = \bigcup_{l=1,\dots,n^p} G_l^{N^k}. \tag{105}$$

Under (**L**): Fix any measure $\pi \in \Pi(\mathcal{E})^L$ and define $\tau \equiv \lim_{k\to\infty}(1 - \pi(G_{p+}^{N^k} \cup G_p^{N^k})) = 1 - \pi(\mathcal{S}) = 0$, for any $\pi \in \Pi(\mathcal{E})^L$, where

$$G_{p+}^{N^0} = \bigcup_{l=n^p+1,\dots,N^0} G_l^{N^0};\ G_{p+}^{N^k} = G_{p+}^{N^{k-1}} \cup \left(\bigcup_{l=n^p+1,\dots,N^k} G_l^{N^k}\right). \tag{106}$$

Step 2 (**CONSTRUCTION OF REPRESENTATION FUNCTIONAL**): Suppose $\succ$ satisfies the axioms **P1**, **P3**^{Q}, **P4**^{Q}, **P5**, **P6**^{Q}. Fix probability measure(s) π and a quantile $\tau \in [0,1]$ derived in Theorem 1. Using the derived probabilities, map each act f a probability distribution $P \in \mathcal{P}_0(\mathcal{X})$ through $\pi \circ f^{-1} = P$. Let

$$\Pi_f^{-1}(\tau) = \begin{cases} \inf\{x \in \mathcal{X} | \pi[f(s) \precsim x] \geq \tau\} \text{ if } \tau \in (0,1] \\ \sup\{x \in \mathcal{X} | \pi[f(s) \precsim x] \leq 0\} \text{ if } \tau = 0 \end{cases}. \tag{107}$$

For the non-extreme preferences, probabilistic sophistication (Step 7 of Theorem 1), convex-rangedness of π and **P1** imply that $\succ$ induces relation $\succ_P$ over probability distributions in $\mathcal{P}_0(\mathcal{X})$, which is asymmetric and negatively transitive, and such that

$$\big(P \succsim_P Q,\ \pi \circ f^{-1} = P,\ \pi \circ g^{-1} = Q\big) \Leftrightarrow f \succsim g. \tag{108}$$

Moreover, by Step 1, all acts in the set $\mathcal{F}(\tau, x^*|\pi) = \{f \in \mathcal{F}|\Pi_\pi^{-1}(\tau) \sim x^*,\ x^* \in f(\mathcal{S})\}$ are indifferent, and $P \succsim_P Q \Leftrightarrow f \succsim g,\ \ \pi \circ f^{-1} = P,\ \pi \circ g^{-1} = Q,\ f \in \mathcal{F}(\tau, x^*|\pi),\ g \in \mathcal{F}(\tau, y^*|\pi)$ for some $x^* \succsim y^*$.

For the extreme preferences, (108) holds for probability distributions induced by $\pi \circ f^{-1} = P$, $\pi \in \Pi(\mathcal{E}) = \Pi(\mathcal{E})^H, \Pi(\mathcal{E})^L$, for some x, y and

$$\left(\min_{\{x \in f(\mathcal{S})|\pi \circ f^{-1}(x) > 0, \pi \in \Pi(\mathcal{E})\}} \{f(\mathcal{S})\} = x \succsim_x \min_{\{x \in f(\mathcal{S})|\pi \circ g^{-1}(x) > 0, \pi \in \Pi(\mathcal{E})\}} \{g(\mathcal{S})\} = y\right) \Leftrightarrow f \succsim g. \tag{109}$$

By Step 1, all acts in the set $\mathcal{F}(\tau, x^*|\pi \in \Pi(\mathcal{E})) = \{f \in \mathcal{F}|\Pi_f^{-1}(\tau) \sim x^*,\ x^* \in f(\mathcal{S})$, for all $\pi \in \Pi(\mathcal{E})\}$ are indifferent.

Hence given (108), $\Pi_f^{-1}(\tau)$ defines a preference functional that represents $\succsim$:

$$f \succsim g \Leftrightarrow P \succsim_P Q \Leftrightarrow \Pi_f^{-1}(\tau) \succsim_P \Pi_g^{-1}(\tau). \tag{110}$$

For the preference functional on $\mathcal{F}$ to be real-valued, it suffices to ensure that there exists a real-valued utility function on certain outcomes, $u : \mathcal{X} \to \mathbb{R}$, which is a representation for $\succ_x$. 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 $\succ_x$. For any $u \in \mathcal{U}^O$,

$$\mathcal{V}(f) = \hat{\mathcal{V}}(P) = u \circ \Pi_f^{-1}(\tau) \geq u \circ \Pi_g^{-1}(\tau) = \hat{\mathcal{V}}(Q) = \mathcal{V}(g) \tag{111}$$

Step 3 (**LEFT-CONTINUITY OF QUANTILE REPRESENTATION**):

1. Assume that preferences are not extreme. For $\{G_{p-}^{N^k}\}_{k\in\mathbb{N}_{++}}$ and $\{G_p^{N^k}\}_{k\in\mathbb{N}_{++}}$ as defined in Step 7, take the sequence $\{\pi(G_{p-}^{N^k}\cup G_p^{N^k})\}_{k\in\mathbb{N}_{++}}$. Using Step 6, define the limit event $\hat{G}$ as $\pi(\hat{G}) = \pi(\lim_{k\to\infty}(G_{p-}^{N^k}\cup G_p^{N^k})) = \lim_{k\to\infty}\pi(G_{p-}^{N^k}\cup G_p^{N^k}) = \lim_{k\to\infty}\Big(\pi(G_{p-}^{N^k}) + \pi(G_p^{N^k})\Big) = \lim_{k\to\infty}\pi(G_{p-}^{N^k})+\lim_{k\to\infty}\pi(G_p^{N^k}) = \pi(\lim_{k\to\infty}G_{p-}^{N^k})+\pi(\lim_{k\to\infty}G_p^{N^k})$. By $\mathbf{P3}^Q$, for a fixed $x \succ y$ and acts

$$\begin{bmatrix} x \text{ if } s \notin G_{p-}^{N^k} \cup G_p^{N^k} \\ y \text{ if } s \in G_{p-}^{N^k} \cup G_p^{N^k} \end{bmatrix} \text{ and } \begin{bmatrix} x \text{ if } s \in \mathcal{S} \\ y \text{ if } s \notin \mathcal{S} \end{bmatrix}, \tag{112}$$

for all $k \in \mathbb{N}_{++}$, $G_{p-}^{N^k} \cup G_p^{N^k} \succ_* \varnothing$. $\hat{G}$ must belong to the sequence, that is $\hat{G} \succ_* \varnothing$. For otherwise, one can construct an act:

$$\begin{bmatrix} x \text{ if } \hat{G}^c \\ y \text{ if } \hat{G} \end{bmatrix} \succ \begin{bmatrix} x \text{ if } \hat{G}\cup\{s\}^c \\ y \text{ if } \{s\} \\ y \text{ if } \hat{G} \end{bmatrix}. \tag{113}$$

But then $\hat{G}\cup\{s\} \succ_* \hat{G}$ and there does not exist a *finite* N-partition of $\mathcal{S}$ such that for all $n = 1, ..., N$, $\hat{G}\cup\{s\} \succ_* \hat{G}\cup G_n$, a violation of $\mathbf{P6}^{Q_*}$. Fix $\{\mathcal{S}, \mathcal{X}, \mathcal{E}, \mathcal{F}, \succ\}$ and $\tau \in [0, 1]$, π (or $\Pi(\mathcal{E})$) from Theorem 1.

2. Assume that preferences are extreme.

Under (**H**): An argument from the Step 9 for $\tau \in (0, 1)$ can be employed.

Under (**L**): For a given act f, the 0^{th}-quantile outcome is equal to the infimum of the outcomes in the image of f that are mapped from non-null events.

((2) ⇒ (1)) Fix $\tau \in [0, 1]$ and π.

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

P3^Q (**PIVOTAL MONOTONICITY**)

(*only if*) Suppose $x \succsim y$, $F_{f,\pi}(\tau) = x$, and $F_{g,\pi}(\tau) = y$. Then, existence of the pivotal event E_f such that $f^{-1}(x) = E_f$ for some $x \in f(\mathcal{S})$, and $[g_{x+} \text{ if } E_{f+}; x \text{ if } E_f;\ g_{x-} \text{ if } E_{f-}] \succsim$

$[g_{y+}$ if $E_{f+}; y$ if E_f; g_{y-} if $E_{f-}]$ for any subacts g_{x+}, g_{x-}, g_{y+}, and g_{y-}, follows from τ-FOSD.

(*if*) Implied by τ-FOSD.

P4^Q (**COMPARATIVE PROBABILITY**) Let $x^* \succ x$, and

$$\begin{bmatrix} x^* \text{ if } s \in E \\ x \text{ if } s \in F \\ g \text{ if } s \notin E \cup F \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \in E \\ x^* \text{ if } s \in F \\ g \text{ if } s \notin E \cup F \end{bmatrix}. \tag{114}$$

Define $(E \cup F)_{gx^*+} = \{s \in \mathcal{S} | g(s) \succ x^*\}$ and $(E \cup F)_{gx-} = \{s \in \mathcal{S} | g(s) \prec x\}$. Let the function g_{x^*+} be any mapping $g_{x^*+} : (E \cup F)_{gx^*+} \to \mathcal{X}$ with $g_{x^*+}(s) \succsim x^*$, for all $s \in (E \cup F)_{gx^*+}$ and let g_{x-} be any map $g_{x-} : (E \cup F)_{gx-} \to \mathcal{X}$ with $g_{x-}(s) \precsim x$, for all $s \in (E \cup F)_{gx-}$. τ-FOSD and implies that $\pi((E \cup F)_{gx-}) + \pi(F) < \tau \leq \pi((E \cup F)_{gx-}) + \pi(E)$. Hence $\pi(E) > \pi(F)$. Given that, a similar argument forces

$$\begin{bmatrix} y^* \text{ if } s \in E \\ y \text{ if } s \in F \\ h \text{ if } s \notin E \cup F \end{bmatrix} \succsim \begin{bmatrix} y \text{ if } s \in E \\ y^* \text{ if } s \in F \\ h \text{ if } s \notin E \cup F \end{bmatrix}. \tag{115}$$

P5 (**NONDEGENERACY**) This follows, since the functional $V : \mathcal{F} \to \mathbb{R}$ is nonconstant.

P6^Q (**SMALL-EVENT CONTINUITY OF** $\succsim_l$) Let $\tau \in (0, 1)$. Suppose that for any $x \succ y$,

$$\begin{bmatrix} x \text{ if } s \in E \\ y \text{ if } s \notin E \end{bmatrix} \succ \begin{bmatrix} x \text{ if } s \in F \\ y \text{ if } s \notin F \end{bmatrix} \tag{116}$$

Using the definition of $\succ^*$, $E \succ^* F$. Then, given the measure π, $\pi(E) > \pi(F)$. The difference $\pi(E) - \pi(F) > 0$ defines an event G such that $\pi(G) \equiv \pi(E) - \pi(F)$. Since measure π is convex-ranged, we can partition set G into N events $G_1, ..., G_N$. By convex rangedness, again, for a sufficiently large N we have that $\pi(E) > \pi(F) + \pi(G_n)$, $n = 1, ..., N$.

Using definition of $\succ^*$, $E \succ_* F \cup G_n$, $n = 1, ..., N$.

An analogous argument can be used for the remaining cases. ∎

Appendix 4: Other results in Chapter 1

4A. Properties of lottery preferences (cont. from Section 3.4)

Substitution. Grant, Kajii and Polak (1992) introduced the following substitution axiom.

Axiom of Degenerate Independence (**ADI**): *For all simple lotteries* $P \in \mathcal{P}_0(\mathcal{X})$*, outcomes* x*,* y and $\alpha \in (0,1)$,

$$x \succsim_x y \Leftrightarrow \alpha\delta_x + (1-\alpha)P \succsim_P \ \alpha\delta_y + (1-\alpha)P$$

This axiom states that moving a probability mass from one outcome to another is weakly preferred (according to the induced lottery preferences) if and only if the second outcome is preferred to the first. Quantile Maximization requires that this holds only for outcomes on pivotal events. We call this condition *Axiom of Pivotal Independence*, **API**. For a fixed lottery $P \in \mathcal{P}_0(\mathcal{X})$ and outcome x such that $x \in$*supp*$\{P\}$, let

$$P_{x+} = \sum_{\{n|x_n \succ_P x\}} p_n, \ P_{x-} = \sum_{\{n|x_n \prec_P x\}} p_n \tag{117}$$

and let Q_{x+} and Q_{x-} be any sublotteries on P_{x+} and P_{x-} with supports such that *supp*$\{Q_{x+}\} \succsim_P x$ and *supp*$\{Q_{x+}\} \precsim_P x$, respectively.

Axiom of Pivotal Independence (API): *For any simple lottery* $P \in \mathcal{P}_0(\mathcal{X})$*, there is an outcome* $x \in$*supp*$\{P\}$ *such that for all outcomes* x, y*, and* $\lambda \in (0,1]$*, there exists a* $\gamma \in [0,1]$ *such that,*

$$x \succsim_x y \Leftrightarrow \gamma(1-\lambda)Q_{x-} + \lambda\delta_x + (1-\gamma)(1-\lambda)Q_{x+} \succsim \gamma(1-\lambda)Q_{y-} + \lambda\delta_y + (1-\gamma)(1-\lambda)Q_{y+}. \tag{118}$$

The following equivalence obtains.

Proposition 7 *If (i) $\succ$ is a weak order, and (ii) is probabilistically sophisticated with respect to π, then*

$$\succ \text{ satisfies } \mathbf{P3}^Q \text{ if and only if } \succ_P \text{ exhibits } \mathbf{API}. \tag{119}$$

Proof. (**API**$\Rightarrow$**P3**^{Q}) Fix an act $f \in \mathcal{F}$ and its pivotal event E_f such that $f(E_f) = x$ for some $x \in f(\mathcal{S})$ and for any outcomes $x, y,$ and subacts $g_{x+},\ g_{x-},\ g_{y+},\ g_{y-}$:

$$\begin{bmatrix} g_{x+}\ E_{f+} \\ x \text{ if } E_f \\ g_{x-} \text{ if } E_{f-} \end{bmatrix} \succsim \begin{bmatrix} g_{y+}\ E_{f+} \\ y \text{ if } E_f \\ g_{y-} \text{ if } E_{f-} \end{bmatrix} \Leftrightarrow x \succsim y. \tag{120}$$

Since the pivotal event E_f is non-null, i.e., $\pi(E_f) > 0$. Probabilistic sophistication implies condition (24). Given that $\succ$ is a weak order and π is convex-ranged, condition (25) is also satisfied. Using these conditions,

$$x \succsim y \Leftrightarrow \delta_x \succsim_P \delta_y \tag{121}$$

$$f = \begin{bmatrix} g_{x+}\ E_{f+} \\ x \text{ if } E_f \\ g_{x-} \text{ if } E_{f-} \end{bmatrix} \succsim \begin{bmatrix} g_{y+}\ E_{f+} \\ y \text{ if } E_f \\ g_{y-} \text{ if } E_{f-} \end{bmatrix} = f' \Leftrightarrow \tag{122}$$

$$\Leftrightarrow \quad \begin{array}{ccc} \pi \circ f^{-1} & \succsim_P & \pi \circ f'^{-1} \\ \shortparallel & & \shortparallel \\ \pi(E_{f-})G_{x-} + \pi(E_f)\delta_x + \pi(E_{f+})G_{x+} & & \pi(E_{f-})G_{y-} + \pi(E_f)\delta_y + \pi(E_{f+})G_{y+} \end{array}$$

Using **API** for $\gamma(1-\lambda) = \pi(E_{f-})$ completes the proof.

(**P3**^{Q} $\Rightarrow$**API**) Pick a simple lottery $P = [G_{x-}, P_{x-}, ..., x, p_x, G_{x+}, P_{x+}]$. By convex-rangedness of π, there exist events E_{f-}, E and E_{f+} such that $\pi(E_{f-}) = P_{x-}$, $\pi(E) = p_x$

and $\pi(E_{f+}) = P_{x+}$. The assertion follows from (121) and (122). ■

Monotonicity. It is clear that **API** implies that $\succsim_P$ satisfies weak first-order stochastic dominance, and that the implication is one-way only. A natural notion of monotonicity under Quantile Maximization is the following local version of first-order stochastic dominance:

Definition 9 *Given a complete preorder over outcomes $\succsim_x$, $P = (x_1, p_1; ...; x_N, p_N)$* τ-first-order stochastically dominates *(τ-FOSD) $Q = (y_1, q_1; ...; y_M, q_M)$ with respect to $\succsim_x$ if*

$$F_P(\tau) \succ_x F_Q(\tau) \tag{123}$$

where F_R is the cumulative probability distribution corresponding to lottery $R \in \mathcal{P}_0(\mathcal{X})$.

$\succsim_P$ is said to satisfy τ-first order stochastic dominance if $P \succ_P Q$ whenever P τ-FOSD Q with respect to $\succsim_x$. It turns out that a strengthening of **API** yields equivalence with τ-first-order stochastic dominance.

AXIOM OF PIVOTAL INDEPENDENCE (API'): *For any simple lottery $P \in \mathcal{P}_0(\mathcal{X})$, there is an outcome $x \in$supp$\{P\}$ such that for all outcomes x, y, and $\lambda \in (0, 1]$, there exists an interval $[\gamma', \gamma'') \subseteq [0, 1]$ such that for any $\gamma \in [\gamma', \gamma'')$,*

$$x \succsim_x y \Leftrightarrow \gamma(1-\lambda)Q_{x-} + \lambda\delta_x + (1-\gamma)(1-\lambda)Q_{x+} \succsim_P \gamma(1-\lambda)Q_{y-} + \lambda\delta_y + (1-\gamma)(1-\lambda)Q_{y+}. \tag{124}$$

Under the same assumptions as in Proposition 7, this property of risk preferences is equivalent to $\succ$ satisfying $\mathbf{P3}^Q$ and $\mathbf{P6}^Q$.

Proposition 8 *If (i) $\succ$ is a weak order, and (ii) is probabilistically sophisticated with re-*

spect to π, then

$$\succ \text{ satisfies } \boldsymbol{P3}^Q,\ \boldsymbol{P6}^Q \text{ if and only if } \succ_P \text{ exhibits } \boldsymbol{API'}.$$

Proof. Analogous to the proof of **Proposition** 7. ■

4B. Other proofs

Proposition 1

Proof. ((2)⇒(1)) Using Debreu's (1954) theorem to derive utility representation for outcomes u, as in the proof of Theorem 2, we can assume without loss of generality that $\mathcal{X} = \mathbb{R}$. Then monotonicity with respect to FOSD implies the following property: letting $\mathbf{x}_R$ be the ranked outcome vector of a lottery $R \in \mathcal{P}_0(\mathcal{X})$, for $P, Q \in \mathcal{P}_0(\mathcal{X})$ if $\mathbf{x}_P \geq \mathbf{x}_Q$, then $\mathbf{x}_P \succsim_P \mathbf{x}_Q$. Then, the proof of the assertion is an application of the result in Gevers (1979), with the anonymity condition implied by monotonicity with respect to FOSD.

((1)⇒(2)) Since rankwise monotonicity implies that for each distribution P there is an outcome x_k such that for any Q_{x-} and Q_{x+}:

$$P \sim (Q_{x-}, P_{x-},;\ x, p_x;\ Q_{x+}, P_{x+}), \tag{125}$$

in particular, $P \sim x_k$. Hence, it is straightforward that rankwise monotonicity implies monotonicity with respect to FOSD. Ordinal Invariance can be proved applying an argument analogous to the one used in the proof of Lemma 6. ■

Appendix 5: Proofs of results in Chapter 2

Proposition 2[82]

Proof. Fix utility over outcomes u. For a given decision problem F with K events, let $Q_F^*(0,c)$ be the set of all payoffs that are anticipated under optimal choices for beliefs with cardinality of support equal to c, $c = 1, ..., K$. Then,

$$Q_F^*(0) = \bigcup_{c=1,...,K} Q_F^*(0,c). \tag{126}$$

The proof proceeds in two steps:

Step 1 (Lemma 12 below): We show that for a given cardinality c, if $u(x) \in Q_F^*(0,c)$ for some τ' then $u(x) \in Q_F^*(0,c)$ for any $\tau < \tau'$.

Step 2 (Lemma 4): We prove that the difference sets $Q_F^*(0,c) \backslash Q_F^*(0,c-1)$ contain outcomes with lower ranks than those in $Q_F^*(0,c-1)$. ■

Lemma 12 *Fix decision problem F, utility u and cardinality of beliefs c. If $u(x) \in Q_F^*(0,c)$ for some τ' then $u(x) \in Q_F^*(0,c)$ for any $\tau < \tau'$.*

Proof. Throughout, we hold cardinality of beliefs π and π' as given. Fix a decision problem F, utility on outcomes u and act $f \in F$. For $f \in \mathcal{R}_F(\tau)$, define the set of[83] anticipated τ^{th}-quantile payoff corresponding to the choice, f, optimal for some beliefs $\tilde{\pi}$,

$$Q_{F,f}^*(\tau, \tilde{\pi}) = \{z \in \mathbb{R} | z = Q^\tau(\tilde{\Pi}_f),\ f \in \mathcal{R}_F(\tau)\}. \tag{127}$$

Let $\tau' > \tau$, fix $f \in \mathcal{R}_F(\tau')$ and consider x such that $u(x) \in Q_{F,f}^*(\tau', \pi')$ for some π'. We will show that for any $\tau < \tau'$, there exist beliefs π for which $u(x) \in Q_{F,f}^*(\tau, \pi)$. We need

[82] In fact, the result is stronger than stated: for the considered action the same payoffs can be anticipated under τ' and τ.

[83] Generically in payoffs, $Q_{F,f}^*(\tau, \tilde{\pi})$ is a singleton.

to prove that for the new beliefs π, the payoff $u(x)$ is still anticipated for act f, and that f remains optimal under π. Consider three cases: the rank of outcome x in act f, $r_f(x)$, is equal to 1, K and k, $1 < k < K$.

For $r_f(x) = 1$, the result follows for the same vector of beliefs $\pi = \pi'$ ($Q^*_{F,f}(\tau', \pi') = Q^*_{F,f}(\tau, \pi')$ and $Q^*_{F,\hat{f}}(\tau, \pi') \leq Q^*_{F,\hat{f}}(\tau, \pi')$, for all $\hat{f} \neq f$, $\hat{f} \in F$).

For $r_f(x) = K$, the assertion holds for $\pi_K = 1$.

Finally, let $r_f(x) = k$, $1 < k < K$. First, notice that it is w.l.o.g. to restrict attention to the following decomposition of the vector of ranked probabilities corresponding to an act f given beliefs π' : $\boldsymbol{\pi}'_f = (\boldsymbol{\pi}'_{f,k-}, \pi'_{f,k}, \boldsymbol{\pi}'_{f,k+})$, where $\boldsymbol{\pi}'_{f,k-} = (\pi'_{f,1}, ..., \pi'_{f,k-1})$ and $\boldsymbol{\pi}'_{f,k+} = (\pi'_{f,k+1}, ..., \pi'_{f,K})$. Suppose $\Delta\tau = \tau' - \tau$ is large enough so that $Q^{\tau}(\Pi'_f) < Q^{\tau'}(\Pi'_f)$ (otherwise, the result trivially holds for π'_f), that is

$$\tau \leq \sum_{\{\pi'_l | \pi'_l \in \boldsymbol{\pi}'_{f,k-}\}} \pi'_{f,l} < \tau' \leq \sum_{\{\pi'_l | \pi'_l \in \boldsymbol{\pi}'_{f,k-}\}} \pi'_{f,l} + \pi'_{f,k}. \tag{128}$$

For f to be optimal with the same payoff being anticipated under some new beliefs π ($u(x) \in Q^*_F(\tau, \pi)$), π must satisfy

$$\sum_{\{\pi_l | \pi_{f,l} \in \boldsymbol{\pi}_{f,k\_}\}} \pi_{f,l} < \tau \leq \sum_{\{\pi_l | \pi_{f,l} \in \boldsymbol{\pi}_{f,k\_}\}} \pi_{f,l} + \pi_{f,k}. \tag{129}$$

Conditions (128) and (129) imply $\sum_{\{\pi_l | \pi_{f,l} \in \boldsymbol{\pi}_{f,k\_}\}} \pi_{f,l} < \sum_{\{\pi'_l | \pi'_{f,l} \in \boldsymbol{\pi}'_{f,k\_}\}} \pi'_{f,l}$. To obtain that, we classify cases as summarized in Matrix M5. Fix $\tau < \tau'$.

Matrix M5

	${\pi'}_{k-}$	${\pi'}_{k}$	${\pi'}_{k+}$	
f	k_-	$\boxed{k}$	k_+	
	$\hat{k}_-$	$\boxed{\hat{k}}$		1
	$\hat{k}_+$	$\boxed{\hat{k}}$		2
$\hat{f}$	$\boxed{\hat{k}}$	$\hat{k}_-$		3
	$\boxed{\hat{k}}$	$\hat{k}_+$		4
	$\hat{k}_-$	$\hat{k}_+$	$\boxed{\hat{k}}$	5
	$\hat{k}_+$	$\hat{k}_-$	$\boxed{\hat{k}}$	6

Notes: W.l.o.g. the matrix restricts attention to generic outcomes. The first two rows contain act f for beliefs π'. The bottom six rows correspond to cases for a given act $\hat{f} \neq f$ and a new vector of beliefs π. $\hat{k}$= quantile $(Q^\tau(\Pi_{\hat{f}}))$, $\hat{k}_-$=outcome strictly less preferred to quantile ($x \in \hat{f}(S)$ such that $x \prec Q^\tau(\Pi_{\hat{f}})$), $\hat{k}_+$=outcome strictly preferred to quantile ($x \in \hat{f}(S)$ such that $x \succ Q^\tau(\Pi_{\hat{f}})$). In cells left blank, the argument will not be affected for any rank of outcome. The cases not included in the matrix are trivial.

Consider the signed vector difference

$$sign\{\pi_{\hat{f},k} - {\pi'}_{\hat{f},k}\} = \begin{pmatrix} sign\left\{ \sum_{\{\pi_l | \pi_{\hat{f},l} \in \boldsymbol{\pi}_{\hat{f},k_-}\}} \pi_{\hat{f},l} - \sum_{\{\pi'_l | {\pi'}_{\hat{f},l} \in \boldsymbol{\pi'}_{\hat{f},k_-}\}} {\pi'}_{\hat{f},l} \right\}, sign\{\pi_{\hat{f},k} - {\pi'}_{\hat{f},k}\}, \\ sign\left\{ \sum_{\{\pi_l | \pi_{\hat{f},l} \in \boldsymbol{\pi}_{\hat{f},k_+}\}} \pi_{\hat{f},l} - \sum_{\{\pi'_l | {\pi'}_{\hat{f},l} \in \boldsymbol{\pi'}_{\hat{f},k_+}\}} {\pi'}_{\hat{f},l} \right\} \end{pmatrix}. \tag{130}$$

The possible values are: $(-1, 1, 0)$, $(1, -1, 0)$, $(0, 1, -1)$, $(1, 0, -1)$, $(0, -1, 1)$, $(-1, 0, 1)$.

In the cases 1, 2, 3 and 6 in Matrix M5, for the probability mass not to shift upward

$(Q^{\tau'}(\Pi'_{\hat{f}}) \geq Q^{\tau}(\Pi_{\hat{f}}))$, let

$$\sum_{\{\pi'_l | \pi'_l \in \boldsymbol{\pi}'_{f,k-}\}} \pi'_{f,l} - \sum_{\{\pi_l | \pi_l \in \boldsymbol{\pi}_{f,k-}\}} \pi_{f,l} = \Delta\tau. \tag{131}$$

Since the change in the vector π' to π affects the vectors of ranked probabilities corresponding to acts other than f, we need to show that f is optimal for π, which will follow if $Q^{\tau}(\Pi'_{\hat{f}}) \geq Q^{\tau}(\Pi_{\hat{f}})$, for any $\hat{f} \in F$ such that $\hat{f} \neq f$. This obtains by taking, for a fixed act $\hat{f} \neq f$,

$$\pi_k = \pi'_k + \Delta\tau. \tag{132}$$

Then, (129) and (132), $Q^{\tau'}(\Pi'_f) = Q^{\tau}(\Pi_f)$ and $sign\{\pi_{\hat{f},k} - \pi'_{\hat{f},k}\}$ becomes $(-1, 1, 0)$, $(0, 1, -1)$, $(1, -1, 0)$, $(1, 0, -1)$, for the cases 1, 2, 3 and 6, respectively. With (131) and (132), in the cases 4 and 5 $sign\{\pi_{\hat{f},k} - \pi'_{\hat{f},k}\}$ takes values $(0, -1, 1)$ and $(-1, 0, 1)$, respectively. There, to ensure $Q^{\tau'}(\Pi'_{\hat{f}}) \geq Q^{\tau}(\Pi_{\hat{f}})$, we set

$$\sum_{\{\pi_l | \pi_{f,l} \in \boldsymbol{\pi}_{f,k_+}\}} \pi_{f,l} = \sum_{\{\pi'_l | \pi'_{f,l} \in \boldsymbol{\pi}'_{f,k_+}\}} \pi'_{f,l} + \Delta\tau, \tag{133}$$

$$\sum_{\{\pi_l | \pi_l \in \boldsymbol{\pi}_{f,k-}\}} \pi_{f,l} = \sum_{\{\pi'_l | \pi'_l \in \boldsymbol{\pi}'_{f,k-}\}} \pi'_{f,l} - \Delta\tau. \tag{134}$$

Then, $sign\{\pi_{\hat{f},k} - \pi'_{\hat{f},k}\}$ is $(1, -1, 0)$ and $(-1, 1, 0)$. Hence, for new belief vector π, f is the optimal choice and $Q^{\tau}(\Pi_f) = Q^{\tau'}(\Pi'_f)$, as desired.

The argument is true for any $f \in F$ and therefore

$$\bigcup_f \bigcup_\pi Q^*_{F,f}(\tau', \pi) = Q^*_F(\tau') \subseteq Q^*_F(\tau) = \bigcup_f \bigcup_\pi Q^*_{F,f}(\tau, \pi). \tag{135}$$

It is not difficult to show that the converse does not hold and we omit the proof. ∎

Lemma 4

Proof. We proceed by induction. Clearly, $Q_F^*(0, c-1) \subseteq Q_F^*(0, c)$ for any $c = 2, ..., K$, and $Q_F^*(0, 1)$ is nonempty.

($c = 2$) Pick any outcome x such that $u(x) \in Q_F^*(0, 1)$ for some act f ($\underline{c}(x) = 1$). Fix that act and consider $y \in Q_F^*(0, 2) \backslash Q_F^*(0, 1)$. Since $y \notin Q_F^*(0, 1)$, for any pair of events, there must exist an outcome $w > y$. (See Matrix M6.)

Matrix M6

$\cdots$	y	$\cdots$	x
	$\vdots$		$\vdots$
$\cdots$	w	$\cdots$	
	$\vdots$		$\vdots$

It must be that $y < x$ for some $x \in f(\mathcal{S})$. Otherwise, if $y = x$ for some $x \in f(\mathcal{S}) \cap Q_F^*(0)$, $x \neq y$, then $y \in Q_F^*(0, 1)$; and if $y > x$ for some $x \in f(\mathcal{S}) \cap Q_F^*(0)$, $x \neq y$, then $y \notin Q_F^*(0, 2)$. Hence $x > y$.

($c > 2$) Assume that for any $d < c$, $Q_F^*(0, d) \backslash Q_F^*(0, d-1)$ contain outcomes with lower ranks than those in $Q_F^*(0, d-1)$. Take an outcome x' with $u(x') \in Q_F^*(0, c-1) \backslash Q_F^*(0, c-2)$ for some act f' ($\underline{c}(x') = c-1$), and consider y' with $u(y') \in Q_F^*(0, c) \backslash Q_F^*(0, c-1)$. For any beliefs with the cardinality of support equal to $d-1$, there is an outcome w' such that $w' > y'$. It must be that $y' < x'$ for some $x' \in f'(\mathcal{S})$. By way of contradiction, suppose $x' \leq y'$. An argument analogous as above (note that it does not depend on y and w occurring in the same event) excludes $x' = y'$. If $x' < y'$ for some $x' \in f'(\mathcal{S}) \cap Q_F^*(0, c-2)$, $x' \neq y'$, then using the converse of the hypothesis, a contradiction obtains. We conclude that $x' > y'$. Using that by hypothesis outcomes in sets $Q_F^*(0, d)$, $d < c-2$, have higher ranks than those in $Q_F^*(0, d+1) \backslash Q_F^*(0, d)$, this proves the assertion. ■

Lemma 5

Proof. Fix a data set with one observation, $\{F, f^*\}$. Consider an outcome $x \in f^*(\mathcal{S})$ such that $u(x) \in Q^*_F(0)$ with the minimal cardinality $\underline{c}(x_{f^*})$, as defined in Section 2.2, equal to $c \in \mathbb{N}_{++}$. Using Lemma 4, the maximal τ for which x_{f^*} may be anticipated is equal to

$$\max_{\pi \in \Delta(\mathcal{S}):\ \Sigma_{k=1}^{\bar{K}} \pi_k = 1} \min\{\pi_1, ..., \pi_c\} = \frac{1}{c}. \tag{136}$$

It follows that the maximal τ for which there exists a belief such that f^* is chosen is equal to

$$\max_{x_{f^*} \in \{x \in f^*(\mathcal{S}) | u(x) \in Q^*_F(0)\}} \left\{ \frac{1}{\underline{c}(x_{f^*})} \right\}. \tag{137}$$

■

References

[1] Abdellaoui, M. and P. P. Wakker (2005): "The Likelihood Method for Decision under Uncertainty," *Theory and Decision*, 58, 1: 3-76.

[2] Abadie, A. and J. Angrist, and G. Imbens (2002): "Instrumental Variables Estimation of Quantile Treatment Effects," *Econometrica*, 70, 1: 91-117.

[3] Anscombe F. J. and R. J. Aumann (1963): "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, 34: 199-205.

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

[5] Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999): "Coherent Measures of Risk," *Mathematical Finance*, 9, 3: 203-29.

[6] Aumann, R. (1974): "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics*, 1: 67-96.

[7] Aumann, R. (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1: 1-18.

[8] Aumann, R. (1998): "Common Priors: A Reply to Gul", *Econometrica* 66, 4: 929-38.

[9] Ballestero, E. (2002): "Strict Uncertainty: A Criterion for Moderately Pessimistic Decision Makers," *Decision Sciences*, 33, 1: 87-98.

[10] Barbera, S. and M. Jackson (1988): "Maximin, Leximin and the Projective Criterion: Characterizations and Comparisons," *Journal of Economic Theory*, 46: 34-44.

[11] Barelli, P. (2006): "On Epistemic Conditions of Equilibiria," University of Rochester, working paper.

[12] Basu, K. and J. W. Weibull (1991): "Strategy subsets closed under rational behavior," *Economics Letters*, 36, 2: 141-46.

[13] Battigalli, P. (1987): "Comportamento Razionale ed Equilibrio nei Giochi e nelle Situazioni Sociali," unpublished dissertation, Universita' Bocconi, Milano.

[14] Battigalli, P. and D. Guaitoli (1997): "Conjectural Equilibria and Rationalizability in a Game with Incomplete Information," in *Decision, Games and Markets*, P. Battigalli, A. Montesano and F. Panunzi, eds., Dordrecht: Kluwer.

[15] Bhaskara Rao, K. P. S. and M. Bhaskara Rao (1983): *Theory of Charges. A Study of Finitely Additive Measures*, London: Academic Press.

[16] Blackwell, D. and L. Dubins (1962): "Merging of Opinions with Increasing Information," *The Annals of Mathematical Statistics*, 33, 3: 882-86.

[17] Blume, L., A. Brandenburger and E. Dekel (1991): "Lexicographic probabilities and choice under uncertainty," *Econometrica*, 59, 1 :61-9.

[18] Bosch-Domènech, A., J. G. Montalvo, R. Nagel and A. Satorra (2002): "One, Two, (Three), Infinity, ... : Newspaper and Lab Beauty-Contest Experiments," *American Economic Review*, 92: 1687-1701.

[19] Boutilier, C. (1994): "Towards a Logic for Qualitative Decision Theory", in: *Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning (KR'94, Bonn, Germany, May 24-27)*: 75-86.

[20] Börgers, T. (1993): "Pure Strategy Dominance," *Econometrica*, 61: 2, 423-30.

[21] Brandenburger, A. and E. Dekel (1987): "Rationalizability and Correlated Equilibria," *Econometrica* 55, 6: 1391-1402.

[22] Buchinsky, M. (1995): "Recent Advances in Quantile Regression Models," *Journal of Human Resources*, 33, 1: 88-126.

[23] Camerer, C. F., T. H. Ho and K. Chong (2004): "A Cognitive Hierarchy Model of Behavior in Games," *Quarterly Journal of Economics*, 119, 3: 861-98.

[24] Costa-Gomes, M. and V. Crawford (2006): "Cognition and Behavior in Two Person Guessing Games: An Experimental Study," *American Economic Review*, 96, in press.

[25] Costa-Gomes, M., V. Crawford and B. Broseta (2001): "Experimental Studies of Strategic Sophistication and Cognition in Normal-Form Games," *Econometrica*, 69: 1193-1235.

[26] Crawford, V. P., and N. Iriberri (2006a): "Fatal Attraction: Focality, Naivete, and Sophistication in Experimental 'Hide-and-Seek' Games," UCSD Discussion Paper 2004-12R.

[27] Crawford, V. P., and N. Iriberri (2006b): "Level-k Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?," working paper, UCSD.

[28] Chakraborty, A. and R. Harbaugh (2005): "Comparative Cheap Talk," forthcoming in *Journal of Economic Theory*.

[29] Chakraborty, A., N. Gupta and R. Harbaugh (2002): "Ordinal Cheap Talk in Common Value Auctions," working paper.

[30] Chamberlain, G. (1994): "Quantile Regression, Censoring and the Structure of Wages," in: *Advances in Econometrics*, Christopher Sims, ed., New York: Elsevier: 179-209.

[31] Chambers, C. P. (2005): "Quantiles and Medians," California Institute of Technology, working paper.

[32] Chateauneuf, A., J. Eichberger and S. Grant (2002): "Choice under Uncertainty with the Best and Worst in Mind: Neo-additive Capacities," Rice University Working paper 2002-10.

[33] Chew, S. H. and J. S. Sagi (2006): "Event Exchangeability: Probabilistic Sophistication without Continuity or Monotonicity," *Econometrica*, 74, 3: 771-86.

[34] Cohen, M. (1992): "Security Level, Potential Level, Expected Utility: a Three-criteria Decision Model under Risk," *Theory and Decision*, 33: 101-34.

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

[36] Debreu, G. (1964): "Continuity Properties of Paretian Utility," *International Economic Review*, 5, 3: 285-93.

[37] Dekel, E., D. Fudenberg and S. Morris (2006): "Interim Correlated Rationalizability," working paper.

[38] Dekel, E. and F. Gul (1997), "Rationality and knowledge in game theory," in: Kreps, D. M. and K. F. Wallis, eds., *Advances in economics and econometrics: theory and applications*, Vol. I, Econometric Society Monographs, No.26, Cambridge: Cambridge University Press.

[39] Denneberg, D. (1994): *Non-Additive Measure and Integral*, Dordrecht, Boston: Kluwer.

[40] Dubois, D., H. Fargier, H. Prade and P. Perny (2002): "Qualitative Decision Theory: From Savage's Axioms to Nonmonotonic Reasoning," *Journal of the ACM*, 49, 4: 455-95.

[41] Dubois, D., M. M. Grabisch, F. Modave and H. Prade (2000): "Relating Decision under Uncertainty and Multicriteria Decision Making Models," *International Journal of Intelligent Systems*, 15, 10: 967-79.

[42] Duffie, D. and J. Pan (1997): "An Overview of Value-at-Risk," *Journal of Derivatives*, 4: 7-49, reprinted in: *Options Markets*, G. Constantinides and A. G. Malliaris, eds., London: Edward Elgar, 2000.

[43] Embrechts, P., A. J. McNeil and R. Frey, (2005): *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton: Princeton University Press.

[44] Epstein, L. (1997): "Preference, rationalizability and equilibrium," *Journal of Economic Theory*, 73, 1-29.

[45] Epstein L. G. and M. Le Breton (1993): "Dynamically Consistent Beliefs must be Bayesian," *Journal of Economic Theory*, 61: 1-22.

[46] Epstein, L. G. J. Zhang (2001): "Subjective Probabilities on Subjectively Unambiguous Events," *Econometrica*, 69, 2: 265-306.

[47] Eyster, E. and M. Rabin (2005): "Cursed Equilibrium," *Econometrica*, 73: 1623-72.

[48] Fishburn, P.C. (1979): *Utility Theory for Decision Making*, New York: Wiley, 1970, reprinted in Huntington, NY: Robert E. Kreiger Publishing Company.

[49] Fudenberg, D. and D. K. Levine (1993): "Self-Confirming Equilibrium," *Econometrica*, 61, 3, 523-45.

[50] Gevers, L. (1979): "On Interpersonal Comparability and Social Welfare Orderings," *Econometrica*, 47: 75-89.

[51] Ghirardato, P. (2001): "Coping with Ignorance: Unforeseen Contingencies and Non-additive Uncertainty," *Economic Theory*, 17, 2: 247-76.

[52] Ghirardato, P., F. Maccheroni and M. Marinacci (2004): "Differentiating Ambiguity and Ambiguity Attitude," *Journal of Economic Theory*, 118: 133-73.

[53] Gilboa, I. (1988): "A Combination of Expected Utility and Maxmin Decision Criteria," *Journal of Mathematical Psychology*, 32: 405-20.

[54] Gilboa, I. and D. Schmeidler (1989): "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics*, 18: 141-53.

[55] Gilli, M. (1987): "On Non-Nash Equilibria," *Games and Economic Behavior*, 27, 184-203.

[56] Gollier, C. (2001): "Should we beware of the Precautionary Principle?," *Economic Policy*, 16: 303-327.

[57] Goeree J. K. and C. E. Holt (2001): "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions," *American Economic Review*, 91, 5: 1402-1422.

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

[59] Grant, S., A. Kajii and B. Polak (1992): “Many Good Choice Axioms: When Can Many-Good Lotteries be Treated as Money Lotteries?,” *Journal of Economic Theory*, 56: 313-337.

[60] Grant, S. and B. Polak (2006): “Bayesian Beliefs with Stochastic Monotonicity: An Extension of Machina and Schmeidler,” *Journal of Economic Theory*, 130, 1: 264-82.

[61] Green, J. R. and B. Jullien (1988): “Ordinal Independence in Nonlinear Utility Theory,” *Journal of Risk and Uncertainty*, 1: 355-387.

[62] Gul, F. (1998): ”A Comment on Aumann’s Bayesian View,” *Econometrica* 66, 4: 923-37.

[63] Hansen, L. P. and T. Sargent (2000): “Wanting Robustness in Macroeconomics,” working paper.

[64] Hansen, L. P. and T. J. Sargent (2004): *Robust Control and Economic Model Uncertainty*, Princeton, NJ: Princeton University Press.

[65] Hofbauer, J. and W. H. Sandholm (2006): “Survival of Dominated Strategies under Evolutionary Dynamics,” working paper.

[66] Hurwicz, L. (1951): “Optimality Criteria for Decision Making under Ignorance,” Cowles Commission Discussion Paper, 370.

[67] Jaffray, J.-Y. (1988): “Choice under Risk and the Security Factor: An Axiomatic Model,” *Theory and Decision*, 24: 169-200.

[68] Jehiel P. and F. Koessler (2006): “Revisiting Games of Incomplete Information with Analogy-based Expectations,” working paper.

[69] Kalai, E. and E. Lehrer (1993): "Subjective Equilibrium in Repeated Games," *Econometrica*, 61, 5: 1231-1240.

[70] Kalai, E. and E. Lehrer (1995): "Subjective Games and Equilibria," *Games and Economic Behavior*, 8, 1: 123-163.

[71] Karni, E. (1996): "Probabilities and Beliefs," *Journal of Risk and Uncertainty*, 13: 249-262.

[72] Karni, E. and D. Schmeidler (1993): "On the Uniqueness of Subjective Probabilities," *Economic Theory*, 3, 2: 267-77.

[73] Klibanoff, P., M. Marinacci and S. Mukerji (2005): "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, 73, 6: 1849-92.

[74] Koenker, R. and G. Bassett (1978a): "Regression Quantiles,“ *Econometrica*, 46, 1: 33-50.

[75] Koenker, R. and G. Bassett (1978b): "The Asymptotic Distribution of the Least Absolute Error Estimator,“ *Journal of the American Statistical Association*, 73: 618-22.

[76] Koenker, R. (2005): *Quantile Regression*, Econometric Society Monograph Series, Cambridge University Press.

[77] Kreps, D. M. (1988): *Notes on the Theory of Choice*, Boulder, London: Westwiev.

[78] Levy, H. (1992): "Stochastic Dominance and Expected Utility: Survey and Analyses," *Management Science*, 38, 4: 555-93.

[79] Maccheroni, F. (2002): "Maxmin under Risk," *Economic Theory*, 19: 823-31.

[80] Machina, M. J. (1989): "Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty," *Journal of Economic Literature*, 27, 4: 1622-68.

[81] Machina, M. and D. Schmeidler (1992): “A More Robust Definition of Subjective Probability,” *Econometrica*, 60, 4: 745-80.

[82] Manski, C. F. (1988): “Ordinal Utility Models of Decision Making Under Uncertainty,” *Theory and Decision*, 25, 1: 79-104.

[83] Manski, C. F. (2004): “Measuring Expectations,” *Econometrica*, 72, 5: 1329-76.

[84] Maskin, E. (1979): “Decision-Making under Ignorance with Implications for Social Choice,” *Theory and Decision*, 11: 319-37.

[85] Milnor, J. (1954): “Games against Nature,” in: *Decision Processes*, R. M. Thrall, C. H. Coombs, and R. L. Davis, eds., Wiley: New York.

[86] Morris, S. (1995): “The Common Prior Assumption in Economic Theory,” *Economics and Philosophy*, 11: 227-53.

[87] Mukerji, S. (1997): “A Theory of Play for Games in Strategic Form when Rationality is not Common Knowledge,” University of Southampton, Department of Economics Discussion Paper.

[88] Nagel, R. (1995): “Unravelling in Guessing Games: An Experimental Study,” *American Economic Review*, 85: 1313-26.

[89] O’Neill, B. (2001): “Risk Aversion in International Relations Theory,” *International Studies Quarterly*, 45: 617-40.

[90] O’Neill, B. (1987): “A Non-metric Test of the Minimax Theory of Two-Person Zerosum Games,” *Proceedings of the National Academy of Sciences*, 84: 2106-9.

[91] Pearce, D. G. (1984): “Rationalizable Strategic Behavior and the Problem of Perfection,” *Econometrica*, 52, 5, 1029-50.

[92] Ramsey, F. P. (1931): "Truth and Probability," in: *Foundations of Mathematics and Other Logical Essays*, London: K. Paul, Trench, Trubner, and Co., reprinted in Ramsey (1978), Kyburg and Smokler (1980), and Gärdenfors and Sahlin (1988).

[93] Rawls, J. (1971): *A Theory of Justice*, Cambridge, MA: Harvard Press.

[94] Roberts, K. W. S. (1980): "Possibility Theorems with Interpersonally Comparable Welfare Levels," *Review of Economic Studies*, 47: 409-20.

[95] Rogachev, A. (2002): "Dynamic Value-at-Risk", University of St. Gallen.

[96] Rostek, M. J. (2004): "p-Dominant Sets," working paper, Yale University.

[97] Rostek, M. J. (2006): "Robust Decision Making," working paper, Yale University.

[98] Roy, A. D. (1952): "Safety First and the Holding of Assets," *Econometrica*, 20, 3: 431-49.

[99] Rubinstein, A. (1996): "Why are Certain Properties of Binary Relations Relatively More Common in Natural Language?," *Econometrica*, 64: 343-56.

[100] Rubinstein, A. and A. Wolinsky (1994): "Rationalizable Conjectural Equilibrium: Between Nash and Rationalizability," *Games and Economic Behavior*, 6, 2: 299-311.

[101] Savage, L. J. (1954): *The Foundations of Statistics*, New York: Wiley, revised and enlarged, New York: Dover Publications, 1972.

[102] Schachter, B. (1997): "An Irreverent Guide to Value-at-Risk," *Financial Engineering News*, 1, 1, reprinted in: *Risks and Rewards*, 1998: 17-8.

[103] Segal, U. (1989): "Anticipated Utility Theory: A Measure Representation Approach," *Annals of Operations Research*, 19: 359-73.

[104] Segal, U. and J. Sobel (2002): "Min, Max, and Sum," *Journal of Economic Theory*, 106, 1: 126-50.

[105] Stahl, D. O. and P. Wilson (1994): "Experimental Evidence on Players' Models of Other Players," *Journal of Economic Behavior and Organization*, 25: 309-27.

[106] Stahl, D. O. and P. Wilson (1995): "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10: 218-54.

[107] Svensson L. E. O. (2000): "Robust Control Made Simple," Princeton University.

[108] Tercieux, O. (2004): "p-Best Response Set," forthcoming in *Journal of Economic Theory*.

[109] Viscusi, W. K. (1997): "Alarmist Decisions with Divergent Risk Information," *Economic Journal*, 107, 445: 1657-70.

[110] Wald, A. (1950): *Statistical Decision Functions*, New York: Wiley.

[111] Walsh, C. E. (2004): "Precautionary Policies," FRBSF Economic Letter, 2004-05.

[112] Wakker, P. P. (1989): *Additive Representations of Preferences. A New Foundation of Decision Analysis*, Dordrecht, Boston, London: Kluwer.

[113] Weymark, J. A. (2005): "Measurement Theory and the Foundations of Utilitarianism," *Social Choice and Welfare*, 25, 2-3: 527-55.

[114] Yildiz, M. (2004): "Wishful Thinking in Strategic Environments," forthcoming in *Review of Economic Studies*.

[115] Zambrano, E. (2005): "Epistemic Conditions for Rationalizability," forthcoming in *Games and Economic Behavior*.