Uncertainity in Mechanism Design *

Giuseppe Lopomo†
Fuqua School of Business
Duke University

Luca Rigotti‡
Fuqua School of Business
Duke University

Chris Shannon§
Department of Economics
UC Berkeley

May 2004

VERY PRELIMINARY

Abstract

We consider a simple mechanism design problem with private values in which each agent perceives Knightian uncertainty over his opponents' types. Uncertainty is formalized using incomplete preferences as in Bewley (1986). We show that the seller can extract all gains from trade with a direct mechanism in which truth-telling is a Nash equilibrium, in the sense that no buyer has a unilateral incentive to misrepresent his type. In a Nash equilibrium, however, truth-telling and mirepresenting can be incomparable alternatives. Thus we also consider an equilibrium concept in which truth-telling is optimal, that is, (weakly) preferred to other alternatives. In this case the full extraction of all gains from trade is feasible only if there is sufficient disagreement in beliefs. Otherwise, Vickrey mechanisms maximize the seller's expected revenue.

JEL Codes: D0, D5, D8, G1

Keywords: Knightian Uncertainty, Mechanism Design, Auction Theory, Incomplete Preferences.

*We thank Dino Gerardi for helpful discussions. We also thank the audiences at Decentralization 2004 and University of Iowa for many useful comments

†The Fuqua School of Business, Duke University; glopomo@duke.edu

‡The Fuqua School of Business, Duke University; rigotti@duke.edu

§Department of Economics, UC Berkeley; cshannon@econ.berkeley.edu

1 Introduction

In his classic work, Knight (1921) argues that there is an important difference between uncertainty and risk, where risk is characterized by randomness that can be measured precisely. Ellsberg (1961) suggests a more precise definition of uncertainty, in which an event is uncertain or ambiguous if it has unknown probability. In particular, Ellsberg's paradox illustrates important consequences of this distinction by showing that individuals may prefer gambles with precise probabilities to gambles with unknown odds. Uncertainty and risk are distinct characteristics of random environments, and they can also affect individuals' behavior very differently. Such behavior is inconsistent with the expected utility model, and this observation has inspired a significant amount of recent research in economics. Since uncertainty can exert a significant influence on individual behavior, it could also be a significant determinant of equilibrium outcomes.

In this paper, we introduce uncertainty in incomplete information games, and then study how this can affect mechanism design problems. In the mind of each player, the source of uncertainty is what the other players know. Following Harsanyi (1967), this knowledge is condensed into the definition of a player's type. Therefore, the set of all possible types for the other players determines a player's state space. We consider games in which the distinction between uncertainty and risk is formalized by assuming agents have incomplete preferences over state-contingent amounts of money, as in Bewley (1986). Without completeness, individual decision-making depends on a *set* of probability distributions over the state space. A state-contingent sum of money is preferred to another if and only if it has larger expected utility for all probabilities in this set. When preferences are complete this set is a singleton, and the model reduces to standard expected utility. Since incompleteness is reflected by multiple probabilities, this approach provides one way to formalize the distinction between risk and uncertainty.

In Bayesian games, each player assigns a probability distribution to the possible types of the other players. In our setup, this probability distribution is not unique. Hence, we say a *Knightian game* is a game in which each player assigns a set of probability distributions to the possible types of the other players. These games are a generalization of standard games and they can be used to ask whether conclusions one obtains in standard game theoretic frameworks are robust to the presence of uncertainty. One of such conclusions is whether the presence of asymmetric information has welfare consequences. Since Akerlof (1970) seminal lemons model, the idea that asymmetric information can have profound welfare consequences has become well-accepted in economics. More recently, a sequence of papers has cast doubts on that conclusion in a mechanism design setting. Crémer and McLean (1985), Crémer and McLean (1988), and McAfee and Reny (1992) have shown that when private information is correlated the designer of a mechanism can extract all rents from the mechanism's participants. If this is the case, private information has no value. Here, we show that when uncertainty is present this conclusion is sometimes overturned.

In a Knightian game, an alternative is preferred to another if and only if it has higher expected utility for every probability distribution. Hence, an action constitutes a profitable deviation only if it always yields higher expected utility. Similarly, incomparable alternatives are not necessarily profitable deviations. For a given strategy of the opponents, one can define two notions of best-response. In the first, called *maximal best-response*, a strategy is a best-response even if no other strategy is preferred to it. That is, other strategies yield worse or incomparable outcomes. In the second, called *optimal best-response*, a strategy is a best-response only if it is preferred to the other available strategies. That is, other strategies yield worse outcomes (which are therefore comparable). Corresponding to each notion of best-response, one can define equilibrium concepts like Nash equilibrium or equilibrium in dominant strategies.

Notice that, in a given Knightian game, the set of beliefs of each player determines how hard it is for some strategies to be maximal or optimal. Suppose there are only two possible strategies, x and y and a player's beliefs are represented by the entire simplex. Then, x is an optimal only if it yields more than y in every state of the world. In the same case, however, x is maximal unless it yields less than y in every state. Loosely, as uncertainty increases, it becomes easier for strategies to be maximal and difficult for them to be optimal. This difference implies that the equilibrium concepts corresponding these two notions of best-response can be very different. Moreover, optimality is a very stringent requirement that is more difficult to satisfy as uncertainty increases.

We consider a simple private values mechanism design problem in which players perceive uncertainty over the types of other players. We show that if one considers Nash equilibria and dominant strategy equilibria of mechanisms in which players choose maximal strategies, full extraction of informational rents is possible. This conclusion does not necessarily hold when one considers optimal actions. In a Nash equilibrium in optimal strategies, full extraction is possible only if belief sets are not too similar. In a dominant strategy equilibrium in optimal strategies, full extraction is incompatible with sufficiently rich uncertainty. These results build on the idea that uncertainty makes it harder for mechanisms to be incentive compatible in optimal best-responses.

The recent debate on the robustness of optimal mechanisms can then be recast is terms of equilibrium concepts. Full extraction mechanisms are not robust to the presence of uncertainty if the designer selects only among mechanisms that (in equilibrium) leave no incomparable choices to players. If, on the other hand, the designer selects among mechanisms that (in equilibrium) do leave incomparable choices then full extraction may be possible. The latter class of equilibria, obviously, can be characterized as having less predictive power. This unpredictability is deeper than the one which arises from indifference. For example, choice among indifferent alternatives can be "influenced" in the right direction with infinitesimal sums of money while choice among incomparable alternatives cannot.

We characterize the optimal mechanism when full extraction is not possible. Under standard conditions, we show that a Vickrey-like mechanism, like the second price auction, is optimal for the seller.[1] This mechanism is, in some sense, less subject to uncertainty since it is ex-post

[1] The optimality of the all standard auctions has been established by Myerson (1981) and Riley and Samuelson (1981). Optimal selling mechanisms with risk averse buyers and independent values have been characterized by Matthews (1987) and Maskin and Riley (1984). 'Full extraction' results with risk neutral buyers have been

incentive compatible. Hence, our results contribute to the discussion on the conditions simple auctions are optimal.[2]

2 Preliminaries: Incomplete Preferences and Uncertainty

In this section we briefly describe individual behavior under uncertainty when preferences are not necessarily complete. Incompleteness in decision making under uncertainty was first studied by Aumann (1962).[3] In a series of papers, Bewley (1986), (1987), and (1989), further developed this model, which he called Knightian decision theory.[4] The basic result of Bewley's approach is to modify the standard expected utility framework by replacing the unique subjective probability distribution used in expected utility with a *set* of probability distributions. When an individual's preferences satisfy the completeness axiom, she can compare any two state-contingent consumption bundles; she decides which one is preferred based on their respective expected utilities. If an individual's preferences are not complete, she is not necessarily able to compare every pair of consumption bundles. Because incompleteness is reflected by multiplicity of beliefs, she computes many expected utilities for each consumption bundle, and these might not be ranked uniformly.

To formalize this discussion, suppose the state space Ω is finite, and index the states by $s = 1, \ldots, S$. Let $x = (x_1, \ldots, x_S)$ and $y = (y_1, \ldots, y_S)$ be two consumption vectors in $\mathbf{R}_+^S$. We assume an individual's preference relation $\succ$ for consumption bundles is represented by a unique closed, convex set of probability distributions Π and a continuous, concave function $u : \mathbf{R}_+ \to \mathbf{R}$, unique up to positive affine transformations, such that

$$x \succ y \qquad \text{if and only if} \qquad \sum_{s=1}^{S} \pi_s u(x_s) > \sum_{s=1}^{S} \pi_s u(y_s) \text{ for all } \pi \in \Pi.$$

Abusing notation slightly, we can rewrite this as

$$x \succ y \qquad \text{if and only if} \qquad E_\pi [u(x)] > E_\pi [u(y)] \text{ for all } \pi \in \Pi$$

where $E_\pi [\cdot]$ denotes the expected value with respect to the probability distribution π, and $u(x)$ denotes the vector $(u(x_1), \ldots, u(x_S))$. Preferences of this kind have been characterized by Bewley (1986) in the Anscombe-Aumann framework, and by Ghirardato, Maccheroni, Marinacci, and Siniscalchi (2001) in a Savage setting.[5] Following Bewley, we say $\succ$ is *complete* if for all $x \in \mathbf{R}_+^S$,

established by Crémer and McLean (1988) for the case of discrete probability distributions. With continuous distributions McAfee and Reny (1992) have provided a nearly-full extraction result.

[2] This questions has been raised by McAfee and McMillan (1984) who write "A reasonable question for the mechanism design literature is how to capture the importance of robustness. Specifically, we think the answer to questions like 'under what circumstances are English auctions used?' has to do with the need for an institution perform well in a variety of circumstances." See also Milgrom (1985) and (1987).

[3] Recently, Aumann's work has been extended and clarified by Dubra, Maccheroni, and Ok (2001) and Shapley and Baucells (1998).

[4] Bewley's original paper has been published recently as Bewley (2002).

[5] Similar models have also been studied in statistical decision theory. See Nau (1992), Nau (2003), and Seidenfeld, Schervish, and Kadane (1995).

cl $\left\{y \in \mathbf{R}_+^S : x \succ y \text{ or } y \succ x\right\} = \mathbf{R}_+^S$. The set of probabilities Π reduces to a singleton whenever the preference ordering $\succ$ is complete, in which case the usual expected utility representation obtains. Without completeness, comparisons between alternatives are carried out "one probability distribution at a time", with one bundle preferred to another if and only if it is preferred under every probability distribution considered by the agent.[6]

Bewley (1986) notes that the above representation captures the distinction Knight (1921) draws between risk and uncertainty. An event is risky when the probability is known, and uncertain otherwise. Similarly, the decision maker perceives only risk when Π is a singleton, and uncertainty otherwise. Incompleteness and uncertainty are equivalent measures of the same phenomenon in this framework. That is, the amount of uncertainty the decision maker perceives is equivalently reflected by the size of the set of priors Π and the degree of incompleteness of the preference order $\succ$.

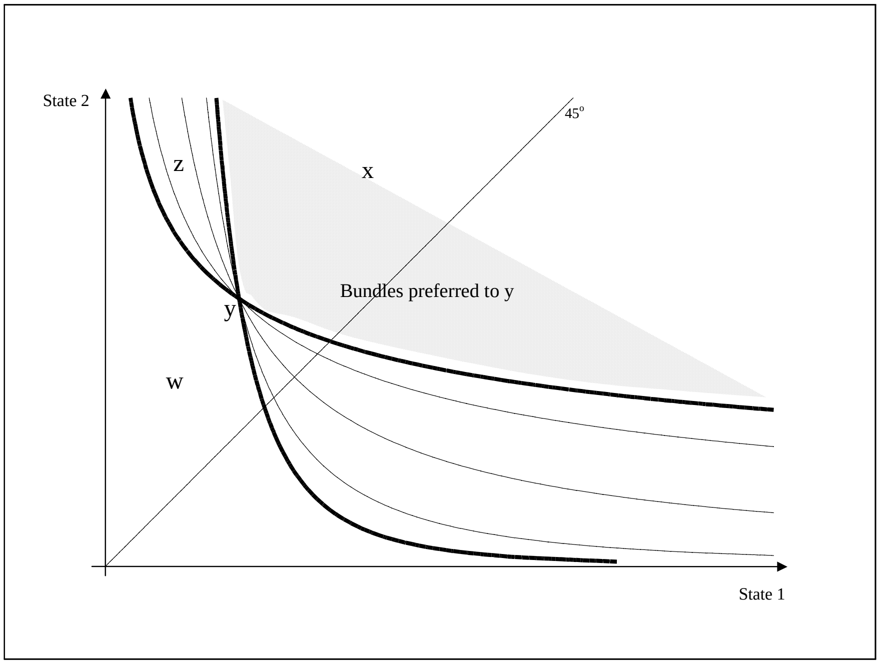


Figure 1: Incomplete Preferences

A graph may help clarify how Bewley's representation works. In Figure 1 the axes measure consumption in each of the two possible states. Given a probability distribution over the two states, a standard indifference curve through the bundle y represents all the bundles that have the same expected utility as y according to this distribution. As the probability distribution changes, we obtain a family of these indifference curves representing different expected utilities

[6]The natural notion of indifference in this setting says two bundles are indifferent whenever they have the same expected utility for each probability distribution in Π.

according to different probabilities. The thick curves represent the most extreme elements of this family, while thin curves represent other possible elements.

A bundle like x is preferred to y since it lies above all of the indifference curves corresponding to some expected utility of y. Also, y is preferred to w since w lies below all of the indifference curves through y. Finally, z is not comparable to y since it lies above some indifference curves through y and below others. Incompleteness induces three regions: bundles preferred to y, dominated by y, and incomparable to y. This last area is empty only if there is a unique probability distribution over the two states and the preferences are complete. Therefore, for any bundle y, the better-than$y$ set has a kink at y whenever there is uncertainty. This kink is a direct consequence of the multiplicity of probability distributions in Π, and vanishes only when Π is a singleton.

When preferences are not complete, the usual revealed preference arguments do not apply. If x is chosen when y is available, we cannot say x is revealed preferred to y, we can only say y is not revealed preferred to x. In other words, one cannot explain choice among incomparable alternatives. This observation has important implications for the definition of equilibria in games.

3 Knightian Games and Mechanisms

The mechanisms we study in this paper are particular examples of Knightian games. A *Knightian game* is a Bayesian game with Knightian uncertainty; that is, a game of incomplete information in which players' types are perceived uncertain in the sense of Knight (1921). In this section we first define Knightian games formally and discuss various solution concepts in general Knightian games, and then consider the formulation of mechanism design problems.

3.1 Knightian Games

Formally, in a Knightian game each player i has a set of actions A_i, a set of possible types Θ_i, and a payoff function $u_i : \times_i A_i \times_i \Theta_i \to \mathbf{R}$. A pure strategy for player i is a function $s_i : \Theta_i \to A_i$, giving the player's strategy $s_i(\theta_i)$ for each possible realization of his type $\theta_i \in \Theta_i$. A player's strategy set is thus $S_i := A_i^{\Theta_i}$. Following usual conventions, $-i$ denotes variables pertaining to all buyers except i, and no subscript for variables pertaining to all buyers, e.g. Θ, and $\theta \in \Theta$. Also, we write Δ^n to denote the simplex in $\mathbf{R}^n$, and $\Delta(E)$ to denote the set of all probability distributions defined over any finite set E. In contrast with standard Bayesian games, each player i has a set of priors $\Pi_i(\theta_i) \subset \Delta(\Theta_{-i})$ over his opponents' types for each possible realization $\theta_i \in \Theta_i$ of his own type. This set of priors corresponds to a strict preference order over strategy profiles $s \in \times_i S_i$ that is not necessarily complete, and under which $s \succ_i s'$ if and only if for each $\theta_i \in \Theta_i$,

$$E_{\pi_i(\theta_i)}[u_i(s_i(\theta_i), s_{-i}(\theta_{-i}), \theta)] > E_{\pi_i(\theta_i)}[u_i(s_i'(\theta_i), s_{-i}(\theta_{-i}), \theta)] \quad \text{for all } \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

The most natural notion of equilibrium in a Knightian game corresponds to a Bayesian Nash equilibrium, in which no player has a unilateral incentive to change his behavior. The addition of incomplete preferences makes the notion of best response somewhat ambiguous, however, due to the difference between maximal and optimal choices. A strategy is maximal if no other strategy does better, while it is optimal of all other strategies do worse. This distinction is captured by the following definitions.

Definition Given a strategy profile $s_{-i} \in S_{-i}$, a strategy s_i is a *maximal response* to s_{-i} if there is no other strategy s'_i such that $(s'_i, s_{-i}) \succ_i (s_i, s_{-i})$.

Note that s_i can be a maximal response and not comparable to s'_i.

Definition A strategy s_i is an *optimal response* to s_{-i} if for each $s'_i \in S_i$ and for each $\theta_i \in \Theta_i$, either

$$E_{\pi_i(\theta_i)}[u_i(s_i(\theta_i), s_{-i}(\theta_{-i}), \theta)] > E_{\pi_i(\theta_i)}[u_i(s'_i(\theta_i), s_{-i}(\theta_{-i}), \theta)] \quad \text{for all } \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

or

$$E_{\pi_i(\theta_i)}[u_i(s_i(\theta_i), s_{-i}(\theta_{-i}), \theta)] = E_{\pi_i(\theta_i)}[u_i(s'_i(\theta_i), s_{-i}(\theta_{-i}), \theta)] \quad \text{for all } \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

If s_i is optimal response to a given s_{-i}, there are no strategies that are not comparable to it.

Clearly every optimal response is also a maximal response, but typically an agent may have many maximal responses that are incomparable, and hence not optimal. These two notions of best response lead naturally to two nested notions of equilibrium. The first, which we call Knightian Nash equilibrium, is based on the idea that agents do not choose dominated strategies.

Definition A strategy profile s in a Knightian game is a *Knightian Nash equilibrium* (KNE) if for each i, s_i is a maximal response to s_{-i}.

There is a simple connection between equilibria in this setting and equilibria in the game in which everyone's beliefs are singletons. In particular, suppose a strategy profile is a Bayesian Nash equilibrium of a game in which there is no uncertainty about players' types. Then, consider the Knightian games in which all players have sets of beliefs which contain the precise beliefs of the Bayesian game. Then one can show that the Bayesian Nash equilibrium is also a KNE (players' strategies must be maximal).

Corresponding to the stronger notion of optimal response we have the following stronger equilibrium condition.

Definition A strategy profile s in a Knightian game is a *Knightian Nash equilibrium* in *optimal strategies* (KNEOS) if for each i, s_i is an optimal response to s_{-i}.

The difference between a KNE and an KNEOS is particularly relevant for mechanism design problems. In a KNE, the possibility of multiple incomparable equilibrium strategies compounds the usual problem of multi-valued best-responses arising due to indifference.

The distinction between maximal and optimal also applies to dominant strategies.

Definition A strategy s_i is a *maximal dominant strategy* for player i if there exists no $s_i' \in S_i$ such that for some $s_{-i} \in S_{-i}$, for each $\theta_i \in \Theta_i$

$$E_{\pi_i(\theta_i)}[u_i(s_i'(\theta_i), s_{-i}(\theta_{-i}), \theta)] > E_{\pi_i(\theta_i)}[u_i(s_i(\theta_i), s_{-i}(\theta_{-i}), \theta)] \quad \text{for all } \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

Here s_i can be a dominant strategy and yet be not comparable to some other strategies.

Definition A strategy $s_i \in S_i$ is an *optimal dominant strategy* for player i if for every $s_i' \in S_i$ and for every $s_{-i} \in S_{-i}$, for each $\theta_i \in \Theta_i$, either

$$E_{\pi_i(\theta_i)}\left[u_i\left(s_i(\theta_i), s_{-i}(\theta_{-i}), \theta\right)\right] > E_{\pi_i(\theta_i)}\left[u_i\left(s_i'(\theta_i), s_{-i}(\theta_{-i}), \theta\right)\right] \quad \text{for all } \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

or

$$E_{\pi_i(\theta_i)}\left[u_i\left(s_i(\theta_i), s_{-i}(\theta_{-i}), \theta\right)\right] = E_{\pi_i(\theta_i)}\left[u_i\left(s_i'(\theta_i), s_{-i}(\theta_{-i}), \theta\right)\right] \quad \text{for all } \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

The difference between maximal and optimal dominant strategies is conceptually different from the difference between weak and strict dominance. The latter only takes care of indifference, while the former takes care of non comparability. One could define strict optimality to rule out indifference among optimal strategies.

One can define two notions of equilibria in dominant strategies according to the two notions of dominance.

Definition A strategy profile s in a Knightian game is a *dominant strategy Knightian Nash equilibrium* if for each i, s_i is a dominant strategy for player i.

Definition A strategy profile s in a Knightian game is a *dominant strategy Knightian Nash equilibrium* in *optimal strategies* if for each i, s_i is an optimal dominant strategy for player i.

As for the Knightian Nash equilibrium, also a dominant strategy equilibrium appears particularly vague when it admits maximal rather than optimal strategies. This vagueness, as before, is not the product of indifference, it stems from incomparability.

3.2 Knightian Mechanisms

Although our focus in this paper is on direct mechanisms, we first define a general mechanism under Knightian uncertainty, and then verify that a version of the revelation principle holds. In standard mechanism theory, preferences are defined over *outcomes* only. In Bewley (1986), as in Savage and Ascombe-Aumann, the decision maker's preferences are defined over R^S, where S is the set of all conceivable states. In our setup, a state is a type profile θ, but buyer i has some (private) information about the state, i.e. he knows the component θ_i. In order to apply Bewley's model we need to show how buyer i's preferences, initially defined over social choice functions, determine his preferences over $R^{\Theta_{-i}}$.

A general mechanism G is a Knightian game consisting of a n-tuple of arbitrary message spaces $B_1, \ldots, B_n$, one for each buyer $i \in N$, and an outcome function $g : B_1 \times \cdots \times B_n \to A_1 \times \cdots \times A_n$, where A_i is player i's strategy space. Note that players only choose messages, while the actions are chosen by the mechanism and moral hazard is ruled out.

A pure strategy for buyer i is a function $\sigma_i : \Theta_i \to B_i$; hence the set of feasible strategies for buyer i is $S_i := B_i^{\Theta_i}$.

Then a strategy profile $(s_1^*, \ldots, s_n^*) \in \times_i S_i$ of a mechanism $G = (B_1, \ldots, B_n, g)$ is Knightian Nash equilibrium of G if for each $i \in N$, s_i^* is a maximal response to s_{-i}^*. By definition this requires for each $i \in N$, $\forall \theta_i \in \Theta_i$, $\nexists \mu \in \Delta(B_i)$ such that

$$\sum_{\theta_{-i} \in \Theta_{-i}} \pi_i(\theta_i) u_i \left(g\left(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})\right), \theta\right) < \sum_{\theta_{-i} \in \Theta_{-i}} \pi_i(\theta_i) \sum_{b \in B_i} \mu(b) u_i \left(g\left(b, s_{-i}^*(\theta_{-i})\right), \theta\right). \quad (1)$$

$\forall \pi_i(\theta_i) \in \Pi_i(\theta_i)$.

Similarly, a strategy profile $(s_1^*, \ldots, s_n^*) \in \times_i S_i$ of a mechanism $G = (B_1, \ldots, B_n, g)$ is an KNEOS of G if for each $i \in N$, s_i^* is an optimal response to s_{-i}^*. Again, by definition this requires that for each $i \in N$,

$$\begin{array}{l} \forall \theta_i \in \Theta_i, \text{ and } \forall b \in B_i \\ \text{either} \\ \qquad \displaystyle\sum_{\theta_{-i} \in \Theta_{-i}} \pi_i(\theta_i) u_i \left(g\left(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})\right), \theta\right) \\ \qquad\qquad = \\ \displaystyle\sum_{\theta_{-i} \in \Theta_{-i}} \pi_i(\theta_i) u_i \left(g\left(b, s_{-i}^*(\theta_{-i})\right), \theta\right) \qquad \forall \pi_i(\theta_i) \in \Pi_i(\theta_i), \\ \text{or} \\ \qquad \displaystyle\sum_{\theta_{-i} \in \Theta_{-i}} \pi_i(\theta_i) u_i \left(g\left(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})\right), \theta\right) \\ \qquad\qquad > \\ \displaystyle\sum_{\theta_{-i} \in \Theta_{-i}} \pi_i(\theta_i) u_i \left(g\left(b, s_{-i}^*(\theta_{-i})\right), \theta\right) \qquad \forall \pi_i(\theta_i) \in \Pi_i(\theta_i). \end{array} \quad (2)$$

Definition A mechanism G *implements* the social choice function ϕ in Knightian Nash (respectively optimal strategy Knightian Nash) equilibrium if there exists a KNE (respectively KNEOS) s^* of G such that $g(s^*(\theta)) = \phi(\theta)$ for all $\theta \in \Theta$.

Our analysis will focus on the truth-telling equilibria of direct mechanisms. With a slight abuse of notation, from now on we will write

$$u_i(\theta_i, \theta_{-i}, \theta)$$

for

$$u_i\left(g\left(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})\right), \theta\right)$$

Withi this in mind, we can define mechanism that implement truth-telling as follows.

Definition A social choice function ϕ is *truthfully implementable in KNE*, if truth-telling is a KNE of ϕ; that is, for each $i \in N$,

$$\begin{array}{l} \forall \theta_i \in \Theta_i, \ \nexists \mu \in \Delta(\Theta_i) \text{ such that} \\ \sum\limits_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\theta_i) u_i\left(\theta_i, \hat{\theta}_{-i}, \theta\right) < \sum\limits_{\hat{\theta}_i \in \Theta_i} \mu(\hat{\theta}_i) \sum\limits_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\theta_i) u_i\left(\hat{\theta}_i, \hat{\theta}_{-i}, \theta\right) \\ \hfill \text{for all } \pi_i(\theta_i) \in \Pi_i(\theta_i) \end{array} \tag{3}$$

Definition A social choice function ϕ is *truthfully implementable in KNEOS*, if truth-telling is a KNEOS of ϕ; that is, for each $i \in N$ and for all $\theta_i, \hat{\theta}_i \in \Theta_i$

$$\begin{array}{l} \text{either} \\ \sum\limits_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\theta_i) u_i\left(\theta_i, \hat{\theta}_{-i}, \theta\right) > \sum\limits_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\theta_i) u_i\left(\hat{\theta}_i, \hat{\theta}_{-i}, \theta\right) \ \forall \pi_i(\theta_i) \in \Pi_i(\theta_i), \\ \text{or} \\ \sum\limits_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\theta_i) u_i\left(\theta_i, \hat{\theta}_{-i}; \theta_i\right) = \sum\limits_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\theta_i) u_i\left(\hat{\theta}_i, \hat{\theta}_{-i}, \theta\right) \ \forall \pi_i(\theta_i) \in \Pi_i(\theta_i). \end{array} \tag{4}$$

One can easily establish a version of the revelation principle in Knightian Nash equilibria.

Proposition 1 (The Revelation Principle for KNE and KNEOS) *If a social choice function ϕ can be implemented in KNE (respectively KNEOS) by a mechanism $G = (B_1, \ldots, B_n, g)$, then ϕ is also truthfully implementable in KNE (respectively KNEOS).*

Proof: We prove the result for the case of KNE; for KNEOS the argument is analogous. By assumption, there exists a strategy profile $s^* = (s_1^*, \ldots, s_n^*) \in \times_i S_i$ which satisfies (1) such that $g(s^*(\theta)) \equiv \phi(\theta)$. In particular, (1) implies that $\forall i \in N$:

$$\begin{array}{l} \forall \theta_i \in \Theta_i \ \nexists \mu \in \Delta(\Theta_i) \text{ such that } \forall \pi_i(\theta_i) \in \Pi_i(\theta_i) \\ \sum\limits_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) u_i(g(s_i^*(\theta_i), s_{-i}^*(\hat{\theta}_{-i})), \theta) < \sum\limits_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \sum\limits_{\hat{\theta}_i \in \Theta_i} \mu(\hat{\theta}_i) u_i(g(s_i^*(\hat{\theta}_i), s_{-i}^*(\hat{\theta}_{-i})), \theta). \end{array} \tag{5}$$

But

$$u_i\left(g\left(s_i^*\left(\hat{\theta}_i\right), s_{-i}^*\left(\hat{\theta}_{-i}\right)\right), \theta\right) = u_i\left(\phi\left(\hat{\theta}_i, \hat{\theta}_{-i}\right), \theta_i\right) = u_i\left(\hat{\theta}_i, \hat{\theta}_{-i}, \theta\right),$$

hence (5) is equivalent to (3). The proof for the KNEOS case is essentially identical. ■

When truth-telling in a direct mechanism is a Knightian Nash equilibrium, then for each player, no feasible strategy is strictly preferred to truth-telling. With incomplete preferences, there may be reports which induce outcomes which are not comparable with truth-telling. In a sense, this is similar to what happens when best-responses are multivalued. In that case too, one cannot guarantee that a player will follow his part of the equilibrium strategy. In a mechanism design problem this can have undesirable consequences for the principal. With complete preferences, the "indifference problem" is not too troublesome since the indifference can be broken with small amounts of money. In our case, this argument does not necessarily apply since indifference is not the issue. This motivates restricting attention to the class of all mechanisms where truth-telling is not just maximal, but also optimal for each type of each buyer.

In an KNEOS, each player's strategy is *optimal* (not just maximal), given what his opponents are doing. When truth-telling is a KNEOS of a direct mechanism, then each type of each player prefers the outcome induced by when he reports his true type to the outcome induced by any other report, under the assumption that his opponents' are reporting their true types. The predictive power of the KNEOS notion is much stronger than that of KNE. If a particular strategy profile is an KNEOS, one can definitely say what strategy each player chooses when all other players conform to their equilibrium behavior.

Loosely speaking, if the designer opts for a mechanism in which truth-telling is a maximal strategy for each player, she can not really predict how the mechanism will actually be played. When faced with incomparable strategies, players will have choose what to report somehow, but they or the designer do not know much about that choice. If, however, the designer chooses a mechanism in which truth-telling is an optimal strategy, she can predict behavior much more confidently. Note that this distinction is independent of the distinction between dominant and non-dominant strategies. As noted above, a strategy can be dominant and maximal, in which case the issue of choice among incomparable actions is not eliminated.

4 The Auction Model

The owner of a single indivisible object faces a set $N = \{1, ..., n\}$ of potential buyers. The welfare of each buyer $i \in N$ depends on the (objective) probability of receiving the object, on his monetary payment to the seller, and on a privately known type θ_i which determines his willingness to pay for the object. All other agents only know that θ_i is an element of a finite set $\Theta_i := \{1, ..., |\Theta_i|\}$. The seller has no private information, and her welfare depends on the buyers' total payment to her, as well as on the monetary value that she attaches to the object.

In this environment, a *social choice function* $\phi = (q, m) : \Theta \to \Delta^{n+1} \times \mathbf{R}^n$ consists of an *assignment function*

$$q = (q_S, q_1, ..., q_n) \ : \ \Theta \to \Delta^{n+1},$$

and n *payment functions*

$$m = (m_1, ..., m_n) : \Theta \to \mathbf{R}^n,$$

assigning an *outcome* $\phi(\theta) = (q_0(\theta), q_1(\theta), ..., q_n(\theta), m_1(\theta), ..., m_n(\theta)) \in \Delta^{n+1} \times \mathbf{R}^n$ to each type profile $\theta \in \Theta \equiv \times_{i \in N} \Theta_i$. Here Δ^{n+1} denotes the standard simplex in $\mathbf{R}^{n+1}$, $q_S(\theta)$ denotes the probability that the seller retains the object and, for each $i \in N$, $q_i(\theta)$ denotes the probability that buyer i receives the object, and $m_i(\theta)$ his payment to the seller.[7]

Any social choice function $\phi = (q, m)$ can also be interpreted as the outcome function of a direct revelation mechanism, i.e. a game in which a *pure strategy* for buyer $i \in N$ is a function $\delta_i : \Theta_i \to \Theta_i$ specifying a report $\hat{\theta}_i = \delta_i(\theta_i)$ for each type $\theta_i \in \Theta_i$; and a *mixed strategy* is a function $\rho_i : \Theta_i \to \Delta(\Theta_i)$ specifying a probability distribution $\rho_i(\cdot|\theta_i)$ over the set Θ_i of feasible reports, for each type $\theta_i \in \Theta_i$; that is

$$\rho_i = \{\rho_i(\cdot|\theta_i) \in \Delta(\Theta_i) : \theta_i \in \Theta_i\}.$$

In any direct mechanism $\phi = (q, \ m)$, given a profile of opponents' strategies denoted $\rho_{-i} := (\rho_1, ..., \rho_{i-1}, \rho_{i+1}, ..., \rho_n)$, any report $\hat{\theta}_i \in \Theta_i$ induces a set of $|\Theta_{-i}|$ (objective) probability distributions over the set of feasible outcomes $\Delta^{n+1} \times \mathbf{R}^n$, one for each realization of his opponents' types θ_{-i} assigning probability $\rho_{-i}(\hat{\theta}_{-i}|\theta_{-i})$ to the outcome $(q(\hat{\theta}_i, \hat{\theta}_{-i}), m(\hat{\theta}_i, \hat{\theta}_{-i}))$.

Each buyer's preferences, defined over the set of all feasible social choice functions are not necessarily complete, as in Bewley (1986), and also linear in money. More precisely, from we maintain the following assumptions.

Assumption For each $i \in N$ and for each $\theta_i \in \Theta_i$, there exists

1. a continuous "utility index" (or "surplus function") $u_i(\cdot;\theta_i) : [0,1] \times \mathbf{R} \times \Theta_i \to \mathbf{R}$ of the form

   $$u_i(q_i, m_i; \theta_i) = v_i(\theta_i) \ q_i - m_i,$$

   where the function $v_i : \Theta_i \to \mathbf{R}_+$ can be assumed to be strictly increasing without loss of generality, $q_i \in [0,1]$ denotes the probability of being awarded the object, and $m_i \in \mathbf{R}$ denotes his payment;[8] and

2. a closed and convex set of probability distributions $\Pi_i(\theta_i) \subset \Delta(\Theta_{-i})$, with generic element $\pi_i(\theta_i)$ whose elements represent the possible beliefs that buyer i has over his opponents' types, conditional on his type being θ_i.

[7] We are restricting attention to deterministic payment functions. Since we will assume that all agents are risk neutral, this is without loss of generality.

[8] With slight abuse of notation, the same letters q and m denote both components of outcomes and outcome functions.

As pointed out in Bewley (1986), this form of incompleteness formalizes the Knightian distinction between risk and uncertainty.[9] In our setting, each buyer perceives uncertainty over his opponents' types, but regards any randomness built in the allocation mechanism or in opponents' mixed strategies as risk.

In any direct mechanism $\phi = (q, m)$, the *ex-post* surplus of type θ_i of buyer i, when he reports $\hat{\theta}_i \in \Theta_i$ and his opponents report $\hat{\theta}_{-i} \in \Theta_{-i}$, is

$$\widehat{u}_i \left(\hat{\theta}_i, \hat{\theta}_{-i}; \theta_i\right) := v_i (\theta_i) \ q_i \left(\hat{\theta}_i, \hat{\theta}_{-i}\right) - m_i \left(\hat{\theta}_i, \hat{\theta}_{-i}\right), \quad \theta_i, \hat{\theta}_i \in \Theta_i, \ \ \hat{\theta}_{-i} \in \Theta_{-i}.$$

Given a profile of opponents' strategies $\rho_{-i} := \left(\rho_1, ..., \rho_{i-1}, \rho_{i+1}, ..., \rho_n\right)$, choosing among the $|\Theta_i|$ feasible reports in Θ_i amounts to choosing among the $|\Theta_i|$ surplus vectors

$$\overline{\mathbf{u}}_i \left(\hat{\theta}_i, \theta_i | \rho_{-i}\right) := \left[\overline{u}_i \left(\hat{\theta}_i, 1; \theta_i | \rho_{-i}\right), ..., \overline{u}_i \left(\hat{\theta}_i, \theta_{-i}; \theta_i | \rho_{-i}\right), ..., \overline{u}_i \left(\hat{\theta}_i, |\Theta_{-i}|; \theta_i | \rho_{-i}\right)\right] \in \mathbf{R}^{|\Theta_{-i}|}$$

for $\hat{\theta}_i \in \Theta_i$, where each coordinate is the expected surplus computed according to the opponents' mixed strategy profile ρ_{-i}, conditional on the realization the opponents' type profile θ_{-i}, that is

$$\overline{u}_i \left(\hat{\theta}_i; \theta_{-i}, \theta_i | \rho_{-i}\right) := \sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \rho_{-i} \left(\hat{\theta}_{-i}, |\theta_{-i}\right) \widehat{u}_i \left(\hat{\theta}_i, \hat{\theta}_{-i}; \theta_i\right), \quad \forall \theta_{-i} \in \Theta_{-i}.$$

By randomizing over the set Θ_i of feasible reports, type θ_i of buyer i can choose any point in the convex hull

$$\mathcal{U}_i \left(\theta_i | \rho_{-i}\right) := co \left\{ \overline{\mathbf{u}}_i \left(\hat{\theta}_i, \theta_i | \rho_{-i}\right) : \hat{\theta}_i \in \Theta_i \right\} \subset \mathbf{R}^{|\Theta_{-i}|}.$$

Note that here we assume mixed strategies are perceived as risk by all players.

The seller's preferences are also incomplete, and represented by the pair (u_S, Π_S), where

$$u_S (q, m) = v_S + \sum_{i \in N} (m_i - v_S q_i),$$

where $v_S > 0$ is the seller's valuation of the item, and Π_S is a closed and convex set of "prior" beliefs over the profile of all buyers' types, i.e. $\Pi_S \subset \Delta(\Theta)$. For each $i \in N$, and for any $\theta_i \in \Theta_i$, the set of all seller's beliefs on θ_{-i}, conditional on θ_i, is given by

$$\Pi_{S_{-i}} (\theta_i) := \left\{ \pi_S (\theta_{-i} | \theta_i) \in \Delta(\Theta_{-i}) \colon \pi_S (\theta_{-i} | \theta_i) = \frac{\pi_S (\theta)}{\sum_{\theta_{-i} \in \Theta_{-i}} \pi_S (\theta_i, \theta_{-i})} \text{ for some } \pi_S \in \Pi_S \right\}.$$

For any buyers' strategy profile ρ, the seller's ex-post expected surplus is

$$\overline{u}_S (\theta | \rho) \equiv v_S + \left[\sum_{\theta' \in \Theta} \sum_{i \in N} \rho_i (\theta'_i | \theta_i)\right] \sum_{i \in N} \left[m_i (\theta') - v_S q_i (\theta')\right].$$

[9] A random variable represents risk if its probability distribution is known and uncertainty otherwise.

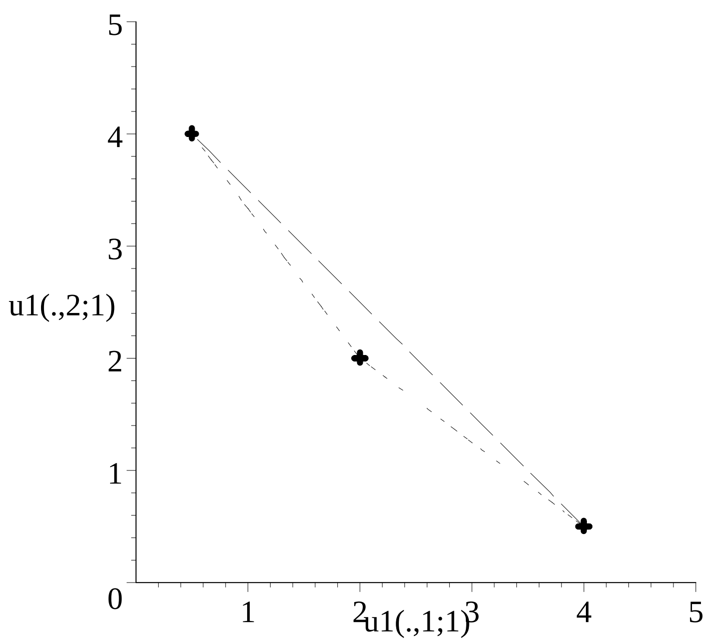
Example 1 *Let $n = 2$, $\Theta_1 = \{1, 2, 3\}$ and $\Theta_2 = \{1, 2\}$. Consider any mechanism (q, m) such that the ex-post surplus function of type 1 of buyer 1*

$$\widehat{u}_1\left(\hat{\theta}_1, \theta_2\,; 1\right) = v_1(1)\, q_1\left(\hat{\theta}_1, \theta_2\right) - m_1\left(\hat{\theta}_1, \theta_2\right)$$

is as in the following table

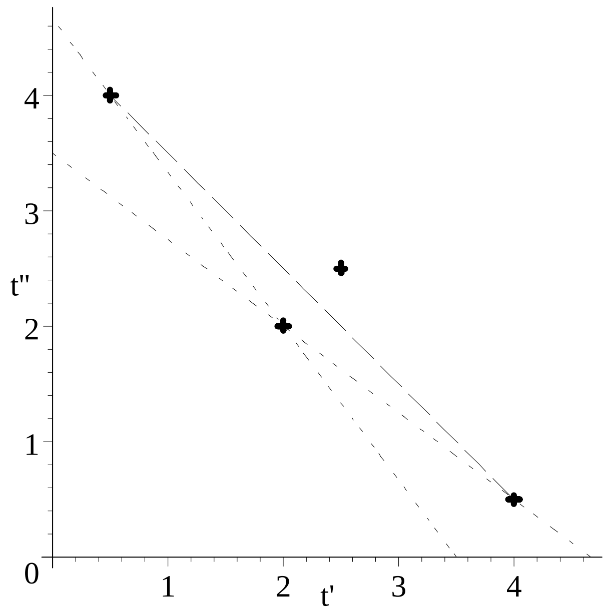
	$\theta_2 = 1$	$\theta_2 = 2$
$\hat{\theta}_1 = 1$	2	2
$\hat{\theta}_1 = 2$	$\frac{1}{2}$	4
$\hat{\theta}_1 = 3$	4	$\frac{1}{2}$

and suppose that buyer 2 reports her true type, i.e. $\rho_2(1|1) = \rho_2(2|2) = 1$, so that $\overline{u}_1\left(\hat{\theta}_1, \theta_2; \theta_1|\rho_2\right) \equiv \widehat{u}_1\left(\hat{\theta}_1, \theta_2; \theta_1\right)$. In this case the set $\mathcal{U}_1(1|\rho_2)$ is the triangle illustrated in the next picture.



Notice that it suffices to check for deviations in pure strategies to identify optimal responses, but one has to verify that there are no profitable deviations in mixed *strategies to establish that a strategy is maximal. This difference occurs because when a player's preferences are incomplete, it is possible that, given his opponents' strategies, he cannot compare outcome a with outcomes b or c, but strictly prefers a convex combination of b and c to a.*
To establish that a report is maximal, one has to verify that there are no profitable deviations in mixed *strategies. In the example given above, let $\widehat{u}_2\left(\hat{\theta}_2, \theta_1; \theta_2\right) = 1$ for all $\hat{\theta}_2, \theta_2 \in \Theta_2$,*

$\theta_1 \in \Theta_1$ so that it is optimal for buyer 2 to report his true type. Since any of buyer 1's belief can be represented by the number $\pi = \Pr[\theta_2 = 1] \in [0,1]$, the belief set $\Pi_1(1)$ corresponds to an interval $[\underline{\pi}, \overline{\pi}]$. It is easy to see that, if $\underline{\pi} < \frac{3}{7}$ and $\frac{4}{7} < \overline{\pi}$, the outcome induced by truth-telling is not comparable with any of the other two outcomes. However, mixing with equal probabilities between reporting type 2 and type 3 yields the surplus vector $(2.5, 2.5)$ which is strictly preferred to truth-telling. This is illustrated in the figure below



5 Full Extraction in Knightian Games

In this section we examine the extent to which the seller can extract all of the surplus in a Knightian mechanism. Clearly, the answer to this question depends on which equilibrium concepts we use to define equilibrium play in a given mechanism.

5.1 Full Extraction in Knightian Nash Equilibrium

In this section we examine the extent to which the seller can extract all of the surplus in a Knightian mechanism. We show that under a convex hull condition for sets we can replicate the full extraction result of Crémer and McLean (1988) in Knightian Nash equilibria.

A profile of surplus vectors $\{V_i(\theta_i) : \theta_i \in \Theta_i\}_{i \in N}$ is *individually rational* if for each $i \in N$, $V_i(\theta_i) \geq 0$ for each $\theta_i \in \Theta_i$.

Given a direct mechanism $\phi = (q, m)$, note that the restriction $V_i(\theta_i) \in \mathbf{R}_+$ captures individual rationality for each type of each player, as this says

$$V_i(\theta_i) = \sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \left[v_i(\theta_i) q_i(\theta) - m_i(\theta) \right] \geq 0 \qquad \text{for some } \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

Now we say that a seller can achieve full extraction if, given any individually rational surplus profile $\{V_i(\theta_i) : \theta_i \in \Theta_i\}_{i \in N}$, there exists a direct mechanism $\phi = (q, m)$ that implements truth-telling as a KNE and in which

$$\sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \left[v_i(\theta_i) q_i(\theta) - m_i(\theta) \right] = V_i(\theta_i) \qquad \text{for some } \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

Theorem 1 *The following are equivalent:*

(i) (the seller can achieve full extraction) for any individually rational surplus profile $\{V_i(\theta_i) : \theta_i \in \Theta_i\}_{i \in N}$ and any procedure to allocate the object q, there exists a mechanism ϕ such that at a truth-telling KNE, for each $i \in N$ and for each $\theta_i \in \Theta_i$,

$$\sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \left[v_i(\theta_i) q_i(\theta) - m_i(\theta) \right] = V_i(\theta_i) \qquad \textit{for some } \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

(ii) for each $i \in N$ there exists $\{\pi_i(\theta_i) \in \Pi_i(\theta_i) : \theta_i \in \Theta_i\}$ such that $\pi_i(\theta_i) \notin co\{\pi_i(\theta_i')\}_{\theta_i' \neq \theta_i}$ for each $\theta_i \in \Theta_i$ (where $co\{\cdot\}$ denotes the convex hull).

Proof: For each player, fix $\{\pi_i(\theta_i) \in \Pi_i(\theta_i) : \theta_i \in \Theta_i\}$ satisfying (ii). Consider the Bayesian game that corresponds to these (unique) probability distributions for every player. In this game, there is no Knightian uncertainty. Let $\{V_i(\theta_i) : \theta_i \in \Theta_i\}_{i \in N}$ be an individually rational surplus profile.

Fix $i \in N$. Since $\pi_i(\theta_{-i}; \theta_i) \notin co\{\pi_i(\theta_{-i}; \hat{\theta}_i)\}_{\hat{\theta}_i \neq \theta_i}$ for each $\theta_i \in \Theta_i$, by the separating hyperplane theorem, for each $\theta_i \in \Theta_i$ there exists an $x_{\theta_i} \in \mathbf{R}^{|\Theta_{-i}|}$ such that

$$\pi_i(\theta_i) \cdot x_{\theta_i} = 0 \qquad \text{and } \pi_i(\hat{\theta}_i) \cdot x_{\theta_i} > 0 \text{ for all } \hat{\theta}_i \neq \theta_i.$$

Define a family of payments $m_i : \Theta_i \times \Theta_{-i} \to \mathbf{R}^{|\Theta_i|}$ as follows:

$$m_i(\hat{\theta}_i, \theta_{-i}) := \left[\pi_i(\hat{\theta}_i) \cdot v_i(\hat{\theta}_i) q_i(\hat{\theta}_i, \theta_{-i}) - V_i(\hat{\theta}_i) \right] \mathbf{1} + K x_{\hat{\theta}_i}$$

where $\mathbf{1}$ is the $|\Theta_i|$-dimensional vector $(1, 1, \ldots, 1)$, $K > 0$ will be defined below, and $x_{\hat{\theta}_i}$ is the separating hyperplane corresponding to type $\hat{\theta}_i$.

The expected value of this payment with respect to $\pi_i(\theta_i)$ is

$$\pi_i(\theta_i) \cdot m_i(\hat{\theta}_i, \theta_{-i}) = \pi_i(\hat{\theta}_i) \cdot v_i(\hat{\theta}_i) q_i(\hat{\theta}_i, \theta_{-i}) - V_i(\hat{\theta}_i) + K\pi_i(\theta_i) \cdot x_{\hat{\theta}_i}.$$

Therefore, for the player of type θ_i, the expected surplus (computed with respect to $\pi_i(\theta_i)$) of announcing type $\hat{\theta}_i$ given the mechanism $\phi = (q, m)$ is

$$\begin{aligned}\pi_i(\theta_i) \cdot v_i(\theta_i) q_i(\hat{\theta}_i, \theta_{-i}) - \pi_i(\theta_i) \cdot m_i(\hat{\theta}_i, \theta_{-i}) &= \pi_i(\theta_i) \cdot v_i(\theta_i) q_i(\hat{\theta}_i, \theta_{-i}) \\ &\quad - \left[\pi_i(\hat{\theta}_i) \cdot v_i(\hat{\theta}_i) q_i(\hat{\theta}_i, \theta_{-i}) - V_i(\hat{\theta}_i) + K\pi_i(\theta_i) \cdot x_{\hat{\theta}_i}\right].\end{aligned}$$

Then there exists $K > 0$ sufficiently large so that for any $\hat{\theta}_i \neq \theta_i$ the right hand side of this expression is negative. Choose such a K. For $\hat{\theta}_i = \theta_i$, on the other hand the expression reduces to

$$\begin{aligned}v_i(\theta_i) \pi_i(\theta_i) \cdot q_i(\theta_i, \theta_{-i}) - \pi_i(\theta_i) \cdot m_i(\theta_i, \theta_{-i}) &= V_i(\theta_i) - K\pi_i(\theta_i) \cdot x_{\theta_i} \\ &= V_i(\theta_i).\end{aligned}$$

Therefore, the mechanism $\phi = (q, m)$ satisfies incentive compatibility.

This shows that ϕ implements truth-telling and induces the surplus profile $V_i(\theta_i)$ for player i. Since we can repeat this procedure for all players in $i \in N$, ϕ implements truth-telling as a KNE. Moreover, for each player $i \in N$, the interim expected surplus of type θ_i with respect to $\pi_i(\theta_i) \in \Pi_i(\theta_i)$ is just $V_i(\theta_i)$. ■

5.2 Full Extraction in Dominant Strategy Equilibrium

Here, following the same intuition of the previous section, we show that under a spanning condition for sets we can always replicate the full extraction result of Crémer and McLean (1988) in dominant strategies.

Theorem 2 *Let $\{V_i(\theta_i) : \theta_i \in \Theta_i\}_{i \in N}$ be an individually rational surplus profile. Suppose that for each $i \in N$ there exists $\{\pi_i(\theta_i) \in \Pi_i(\theta_i) : \theta_i \in \Theta_i\}$ such that*

$$\pi_i(\theta_i) \notin \textit{span}\left\{\pi_i(\theta_i')\right\}_{\theta_i' \in \Theta_i \setminus \{\theta_i\}} \quad \forall \theta_i \in \Theta_i. \qquad \text{(SPAN)}$$

Then there exists a mechanism (q, m) that implements truth-telling as a dominant strategy equilibrium and in which

$$\sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \left[v_i(\theta_i) q_i(\theta) - m_i(\theta)\right] = V_i(\theta_i) \qquad \textit{for some } \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

Proof: Let (q^*, m^*) denote a Vickrey mechanism. Let $\{\pi_i(\theta_i) \in \Pi_i(\theta_i) : \theta_i \in \Theta_i\}$ satisfy SPAN and let $V_i(\theta_i)$ denote the expected surplus with respect to $\pi_i(\theta_i)$ that type $\theta_i \in \Theta_i$ of buyer i receives from participating in this mechanism.

Fix a buyer i and let $A_{\theta_i} := span\left\{\pi_i\left(\theta'\right)\right\}_{\theta' \in \Theta_i \setminus \{\theta_i\}}$. By the Separating Hyperplane Theorem, for each $\theta_i \in \Theta_i$ there exists a vector $x_i\left(\theta_i\right) \in \mathbf{R}^{|\Theta_{-i}|}$ such that

$$\pi_i\left(\theta_i\right) \cdot x_i\left(\theta_i\right) > 0 \text{ and } p \cdot x_i\left(\theta_i\right) \leq 0 \ \ \forall p \in A_{\theta_i}.$$

However, since $p \in A_{\theta_i}$ implies $-p \in A_{\theta_i}$, it must be the case that

$$p \cdot x_i\left(\theta_i\right) = 0 \ \ \forall p \in A_{\theta_i}$$

as if $\hat{p} \cdot x_i\left(\theta_i\right) < 0$ for some $\hat{p} \in A_{\theta_i}$, then $\left(-\hat{p}\right) \cdot x_i\left(\theta_i\right) > 0$, violating separation. Now, normalize $x_i\left(\theta_i\right)$ if needed so that $\pi_i\left(\theta_i\right) \cdot x_i\left(\theta_i\right) = V_i\left(\theta_i\right)$.

Thus we have

$$\begin{aligned} \pi_i\left(\theta_i\right) \cdot x_i\left(\theta_i\right) &= V_i\left(\theta_i\right) \\ \pi_i\left(\hat{\theta}_i\right) \cdot x_i\left(\theta_i\right) &= 0, \quad \text{all } \hat{\theta}_i \neq \theta_i. \end{aligned}$$

Finally define $x_i = \sum_{\theta_i \in \Theta_i} x_i\left(\theta_i\right)$. This yields

$$\pi_i(\theta_i) \cdot x_i = V_i\left(\theta_i\right) \ \ \forall \theta_i \in \Theta_i.$$

The mechanism $\left(q^*, m\right)$, where $m_i\left(\tau, y\right) = x_i\left(\tau, y\right) + m_i^*\left(\tau, y\right)$ implements truth-telling as a dominant strategy equilibrium. For each $i \in N$, this yields the expected surplus for type θ_i with respect to $\pi_i(\theta_i)$ given by $V_i(\theta_i)$. ■

5.3 Full Extraction in Optimal Strategies

When we ask for full extraction in equilibrium involving truth-telling as an optimal strategy, we get two immediate analogues of the previous two theorems.

Theorem 3 *The following are equivalent:*

(i) (the seller can achieve full extraction) for any individually rational surplus profile $\{V_i(\theta_i) : \theta_i \in \Theta_i\}_{i \in N}$ and any procedure to allocate the object q, there exists a mechanism ϕ such that at a truth-telling KNEOS, for each $i \in N$ and for each $\theta_i \in \Theta_i$,

$$\sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i)\left[v_i\left(\theta_i\right) q_i\left(\theta\right) - m_i\left(\theta\right)\right] = V_i\left(\theta_i\right) \qquad \textit{for some } \pi_i\left(\theta_i\right) \in \Pi_i\left(\theta_i\right)$$

(ii) for each $i \in N$ for all $\theta_i \in \Theta_i$:

$$\Pi_i(\theta_i) \cap \left(co\left\{\cup_{\theta_i' \neq \theta_i} \Pi_i(\theta_i')\right\}\right) = \emptyset$$

This condition is obviously violated if for at least one player there exists two types whose beliefs have a non empty intersection. Therefore, it becomes more and more difficult to satisfy as one increases uncertainty.

Theorem 4 *Let* $\{V_i(\theta_i) : \theta_i \in \Theta_i\}_{i \in N}$ *be an individually rational surplus profile. Suppose that for each* $i \in N$*: for all* $\theta_i \in \Theta_i$

$$\Pi_i(\theta_i) \cap \left(span \left\{ \cup_{\theta'_i \neq \theta_i} \Pi_i(\theta'_i) \right\} \right) = \emptyset$$

Then there exists a mechanism (q, m) *that implements truth-telling as an optimal dominant strategy equilibrium and in which*

$$\sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \left[v_i(\theta_i) q_i(\theta) - m_i(\theta) \right] = V_i(\theta_i) \qquad \textit{for some } \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

6 Maximality and Optimality of Vickrey Mechanisms

In this section we find conditions under which the English auction is maximal or optimal among all social choice functions which are truthfully implementable in KNEOS and individually rational.

Formally, since we seek implementation in optimal strategies, the set of all feasible social choice functions is determined by the following constraints:

Optimal Incentive Compatibility (OIC): for each $i \in N$, for all $\theta_i, \theta'_i \in \Theta_i$,
either

$$\sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \left[v_i(\theta_i) q_i(\theta_i, \hat{\theta}_{-i}) - m_i(\theta_i, \hat{\theta}_{-i}) \right] = \sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}, \theta_i) \left[v_i(\theta_i) q_i(\theta'_i, \hat{\theta}_{-i}) - m_i(\theta'_i, \hat{\theta}_{-i}) \right]$$
$$\forall \pi_i(\theta_i) \in \Pi_i(\theta_i),$$

or

$$\sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \left[v_i(\theta_i) q_i(\theta_i, \hat{\theta}_{-i}) - m_i(\theta_i, \hat{\theta}_{-i}) \right] > \sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}, \theta_i) \left[v_i(\theta_i) q_i(\theta'_i, \hat{\theta}_{-i}) - m_i(\theta'_i, \hat{\theta}_{-i}) \right]$$
$$\forall \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

and
Optimal Individual Rationality (OIR): for each $i \in N$, for each $\theta_i \in \Theta_i$,
either

$$\sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \left[v_i(\tau) q_i(\theta_i, \hat{\theta}_{-i}) - m_i(\theta_i, \hat{\theta}_{-i}) \right] = 0 \qquad \forall \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

or

$$\sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \left[v_i(\tau) q_i(\theta_i, \hat{\theta}_{-i}) - m_i(\theta_i, \hat{\theta}_{-i}) \right] > 0 \qquad \forall \pi_i(\theta_i) \in \Pi_i(\theta_i)$$

These inequalities require that each type of each buyer prefers truth-telling to both not participating and to reporting any type different from his own (if these options yield different surplus vectors). Thus any mechanism in which some types of some buyers are asked to choose between incomparable alternatives is ruled out by these constraints.

6.1 Maximality and Optimality of Vickrey payments

Here we give conditions under which a q-Vickrey mechanism is maximal or optimal for the seller.

To this end, first we define the general class of "q-Vickrey" mechanisms as follows.

We introduce a convenient way of writing any payment function m as the sum of a "Vickrey" component r and a "bonus" b. For any assignment function q, the "q-Vickrey payment functions" $r^q = \{r_i^q\}_{i \in N}$ are defined recursively by

$$r_i^q\left(1, \hat{\theta}_{-i}\right) = v_i\left(1\right)\, q_i\left(1, \hat{\theta}_{-i}\right), \quad \hat{\theta}_{-i} \in \Theta_{-i}, \tag{6}$$

and

$$v_i\left(\theta_i\right)\, q_i(\theta_i, \hat{\theta}_{-i}) - r_i^q(\theta_i, \hat{\theta}_{-i}) = v_i(\theta_i)\, q_i(\theta_i - 1, \hat{\theta}_{-i}) - r_i^q(\theta_i - 1, \hat{\theta}_{-i}), \quad \hat{\theta}_{-i} \in \Theta_{-i}, \; \theta_i = 2, \ldots, |\Theta_i|. \tag{7}$$

By construction, in any "q-Vickrey mechanism" $\{q,\ r^q\}$, each type is indifferent *ex-post* — i.e. for any realization of his opponents' types $\hat{\theta}_{-i}$ — between reporting his true type θ_i and reporting his "downward-adjacent" type $\theta_i - 1$.

Since we can represent any payment function m_i as the sum of its q-Vickrey part r_i^q and the remaining "bonus" $b_i : \Theta \to \mathbf{R}$ defined as the difference

$$b_i^q\left(\theta_i, \hat{\theta}_{-i}\right) = r_i^q\left(\theta_i, \hat{\theta}_{-i}\right) - m_i\left(\theta_i, \hat{\theta}_{-i}\right), \qquad \hat{\theta}_{-i} \in \Theta_{-i}, \; \theta_i = 2, ..., |\Theta_i|$$

we can write the interim-expected surplus of type θ_i of buyer i, for any $\pi_i(\theta_i) \in \Pi_i\left(\theta_i\right)$, when he reports $\hat{\theta}_i$ as

$$\sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \left[v_i\left(\hat{\theta}_i\right) q_i\left(\hat{\theta}_i, \hat{\theta}_{-i}\right) - r_i^q\left(\hat{\theta}_i, \hat{\theta}_{-i}\right) + b_i^q\left(\hat{\theta}_i, \hat{\theta}_{-i}\right)\right].$$

Defining $u_i\left(\hat{\theta}_i, \hat{\theta}_{-i}\right) := \widehat{u}_i\left(\hat{\theta}_i, \hat{\theta}_{-i}; \theta_i\right)$ we have

$$\sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \left[u_i\left(\hat{\theta}_i, \hat{\theta}_{-i}\right) - u_i\left(\hat{\theta}_i - 1, \hat{\theta}_{-i}\right)\right] = \sum_{\hat{\theta}_{-i} \in \Theta_{-i}} \pi_i(\hat{\theta}_{-i}; \theta_i) \left[b_i\left(\hat{\theta}_i, \hat{\theta}_{-i}\right) - b_i\left(\hat{\theta}_i - 1, \hat{\theta}_{-i}\right)\right].$$

6.1.1 Maximality of Vickrey payments

The formal result is:

Theorem 5 *Assume that for each $i \in N$*

$$\bigcap_{\theta_i \in \Theta_i} \left(\Pi_S(\theta_i) \cap \Pi_i(\theta_i)\right) \neq \emptyset.$$

Then, for any q, the q-Vickrey mechanism $\phi^q = \{q_i, r_i^q\}_{i \in N}$ is maximal for the seller among all mechanisms with the same allocation function q.

Proof: We will show that the q-Vickrey mechanism is maximal among all mechanisms which satisfy the "downward adjacent" incentive constraints. Formally, these constraints are: for each $i \in N$: for each $\theta_i \in \Theta_i$,

$$\begin{array}{l} \text{either} \\ \sum_{\hat\theta_{-i} \in \Theta_{-i}} \pi_i(\hat\theta_{-i}; \theta_i) \left[v_i(\theta_i) q_i\left(\theta_i, \hat\theta_{-i}\right) - m_i\left(\theta_i, \hat\theta_{-i}\right)\right] \\ = \\ \sum_{\hat\theta_{-i} \in \Theta_{-i}} \pi_i(\hat\theta_{-i}; \theta_i) \left[v_i(\theta_i) q_i\left(\theta_i - 1, \hat\theta_{-i}\right) - m_i\left(\theta_i - 1, \hat\theta_{-i}\right)\right] \quad \forall \pi_i(\theta_i) \in \Pi_i(\theta_i), \\ \text{or} \\ \sum_{\hat\theta_{-i} \in \Theta_{-i}} \pi_i(\hat\theta_{-i}; \theta_i) \left[v_i(\theta_i) q_i\left(\theta_i, \hat\theta_{-i}\right) - m_i\left(\theta_i, \hat\theta_{-i}\right)\right] \\ > \\ \sum_{\hat\theta_{-i} \in \Theta_{-i}} \pi_i(\hat\theta_{-i}; \theta_i) \left[v_i(\theta_i) q_i\left(\theta_i - 1, \hat\theta_{-i}\right) - m_i\left(\theta_i - 1, \hat\theta_{-i}\right)\right] \quad \forall \pi_i(\theta_i) \in \Pi_i(\theta_i). \end{array} \tag{DAIC}$$

and the individual rationality constraint of the lowest type*:* for each $i \in N$,

$$\begin{array}{l} \text{either} \\ \sum_{\hat\theta_{-i} \in \Theta_{-i}} \pi_i(\hat\theta_{-i}; 1) \left[v_i(1) q_i\left(1, \hat\theta_{-i}\right) - m_i\left(1, \hat\theta_{-i}\right)\right] = 0 \quad \forall \pi_i(1) \in \Pi_i(1), \\ \text{or} \\ \sum_{\hat\theta_{-i} \in \Theta_{-i}} \pi_i(\hat\theta_{-i}; 1) \left[v_i(1) q_i\left(1, \hat\theta_{-i}\right) - m_i\left(1, \hat\theta_{-i}\right)\right] > 0 \quad \forall \pi_i(1) \in \Pi_i(1). \end{array} \tag{IR$_1$}$$

Writing each payment as $m_i\left(\theta_i, \hat\theta_{-i}\right) = r_i^q\left(\theta_i, \hat\theta_{-i}\right) + b_i^q\left(\theta_i, \hat\theta_{-i}\right)$, and recalling that, by (6) and (7), the q-Vickrey payment functions r_i^q are defined so that each type $\theta_i \in \Theta_i$ is indifferent *ex-post* (for any realization of the opponents' types θ_{-i}) between reporting his true type and reporting its downward-adjacent type $\theta_i - 1$, the previous inequalities simplify to

$$\begin{array}{l} \text{either} \\ \displaystyle\sum_{\hat\theta_{-i} \in \Theta_{-i}} \pi_i(\hat\theta_{-i}; \theta_i) b_i^q\left(\theta_i, \hat\theta_{-i}\right) = \sum_{\hat\theta_{-i} \in \Theta_{-i}} \pi_i(\hat\theta_{-i}; \theta_i) b_i^q\left(\theta_i - 1, \hat\theta_{-i}\right) \quad \forall \pi_i(\theta_i) \in \Pi_i(\theta_i), \\ \text{or} \\ \displaystyle\sum_{\hat\theta_{-i} \in \Theta_{-i}} \pi_i(\hat\theta_{-i}; \theta_i) b_i^q\left(\theta_i, \hat\theta_{-i}\right) > \sum_{\hat\theta_{-i} \in \Theta_{-i}} \pi_i(\hat\theta_{-i}; \theta_i) b_i^q\left(\theta_i - 1, \hat\theta_{-i}\right) \quad \forall \pi_i(\theta_i) \in \Pi_i(\theta_i). \end{array}$$

for each $\theta_i = 2, ..., |\Theta_i|$, and the individual rationality constraint of the lowest type becomes

$$\begin{array}{l} \text{either} \\ \sum_{\hat{\theta}_{-i}\in\Theta_{-i}} \pi_i(\hat{\theta}_{-i};1) b_i^q \left(1,\hat{\theta}_{-i}\right) = 0 \quad \forall \pi_i(1) \in \Pi_i(1), \\ \text{or} \\ \sum_{\hat{\theta}_{-i}\in\Theta_{-i}} \pi_i(\hat{\theta}_{-i};1) b_i^q \left(1,\hat{\theta}_{-i}\right) > 0 \quad \forall \pi_i(1) \in \Pi_i(1). \end{array}$$

Now we can compute the difference in expected payoff to the seller between the q-Vickrey mechanism and any other mechanism with the same assignment function q as follows.

$$\begin{aligned} \sum_{\theta\in\Theta} \pi_S(\theta) \left[\sum_{i\in N} m_i(\theta)\right] - \sum_{\theta\in\Theta} \pi_S(\theta) \left[\sum_{i\in N} m_i^q(\theta)\right] &= \sum_{\theta\in\Theta} \pi_S(\theta) \left[\sum_{i\in N} r_i^q(\theta_i,\theta_{-i}) - b_i(\theta)\right] \\ &\quad - \sum_{\theta\in\Theta} \pi_S(\theta) \left[\sum_{i\in N} r_i^q(\theta_i,\theta_{-i})\right] \\ &= \sum_{\theta\in\Theta} \pi_S(\theta) \sum_{i\in N} \left[-b_i(\theta)\right] \end{aligned}$$

Therefore, for a fixed q, a mechanism is preferred to the q-Vickrey mechanism if and only if

$$\sum_{\theta\in\Theta} \pi_S(\theta) \sum_{i\in N} b_i(\theta_i,\theta_{-i}) < 0 \;\; \forall \pi_S \in \Pi_S.$$

Using the notation developed in Section 4, a mechanism does better than a q-Vickrey mechanism if and only if

$$\sum_{\theta\in\Theta} \pi_S(\theta) \left[\sum_{i\in N} b_i(\theta_i,\theta_{-i})\right] = \sum_{\theta\in\Theta} \lambda_S(\pi_S(\cdot|\theta_i)) \pi_S(\theta_{-i}|\theta_i) \left[\sum_{i\in N} b_i(\theta_i,\theta_{-i})\right] < 0 \;\; \forall \pi_S \in \Pi_S(\theta_i).$$

By assumption, $\exists\ \pi \in \cap_{\theta_i\in\Theta_i} (\Pi_S(\theta_i) \cap \Pi_i(\theta_i))$ such that for each $\theta_i \in \Theta_i$:

$$\begin{aligned} \sum_{\hat{\theta}_{-i}\in\Theta_{-i}} \pi\left(\hat{\theta}_{-i}\right) b_i\left(\theta_i,\hat{\theta}_{-i}\right) &> \sum_{\hat{\theta}_{-i}\in\Theta_{-i}} \pi\left(\hat{\theta}_{-i}\right) b_i\left(\theta_i - 1,\hat{\theta}_{-i}\right) \\ \sum_{\hat{\theta}_{-i}\in\Theta_{-i}} \pi\left(\hat{\theta}_{-i}\right) b_i\left(\theta_1,\hat{\theta}_{-i}\right) &> 0 \end{aligned}$$

and such that

$$\sum_{\theta\in\Theta} \lambda_S(\pi(\theta_{-i})) \pi(\theta_{-i}) \left[\sum_{i\in N} b_i(\theta_i,\theta_{-i})\right] < 0$$

which is clearly impossible. ■

6.1.2 Optimality of Vickrey payments

The formal result is:

Theorem 6 *Suppose that the seller's beliefs are such that* $\Pi_i^S(\tau) \subset \Pi_i(\tau)$, *for all* $\tau \in \Theta_i$, $i \in N$. *Then, for any assignment function* q *the seller strictly prefers the* q*-Vickrey mechanism* $\{q_i, r_i^q\}_{i\in N}$ *to any mechanism* $\{q_i, m_i\}_{i\in N}$ *which satisfied all "downward adjacent incentive constraints" (DAIC) constraints. That is*

$$\{q_i, r_i^q\}_{i\in N} \succ_S \{q_i, m_i\}_{i\in N} .$$

Proof to be added.

7 Conclusion

To be written.

References

Akerlof, G. A. (1970): "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 84, 488–500.

Aumann, R. J. (1962): "Utility Theory without the Completeness Axiom," *Econometrica*, 30, 445–462.

——— (1964): "Utility Theory without the Completeness Axiom: A Correction," *Econometrica*, 32, 210–212.

Bewley, T. F. (1986): "Knightian Decision Theory: Part I," Discussion paper, Cowles Foundation.

——— (1987): "Knightian Decision Theory: Part II," Discussion paper, Cowles Foundation.

——— (1989): "Market Innovation and Entrepreneurship: A Knightian View," Discussion paper, Cowles Foundation.

——— (2002): "Knightian Decision Theory: Part I," *Decisions in Economics and Finance*, 2, 79–110.

Crémer, J.-J., and R. McLean (1985): "Optimal Selling Strategies under Uncertainty for a Discriminatory Monopolist when Demands Are Interdependent," *Econometrica*, 53, 345–61.

——— (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247–57.

Dubra, J., F. Maccheroni, and E. A. Ok (2001): "Expected Utility Theory without the Completeness Axiom," Discussion paper, New York University.

Ellsberg, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 643–669.

Ghirardato, P., F. Maccheroni, M. Marinacci, and M. Siniscalchi (2001): "A Subjective Spin on Roulette Wheels," Discussion paper, Caltech.

Harsanyi, J. C. (1967): "Games with Incomplete Information Played by 'Bayesian' Players," .

Knight, F. H. (1921): *Uncertainty and Profit*. Boston: Houghton Mifflin.

Maskin, E., and J. Riley (1984): "Optimal Auctions with Risk Averse Buyers," *Econometrica*, 52, 1473–518.

Matthews, S. (1987): "Comparing Auctions for Risk Averse Buyers: A Buyer's Point of View," *Econometrica*, 55, 633–646.

McAfee, P. R., and J. McMillan (1984): "Auctions and Bidding," *Journal of Economic Literature*, 25, 699–738.

McAfee, P. R., and P. J. Reny (1992): "Correlated Information and Mechanism Design," *Econometrica*, 60, 395–421.

Milgrom, P. (1985): "The Economics of Competitive Bidding: A Selective Survey," in *Social Goals and Social Organization. Essays in Memory of Elisha Pazner*, ed. by D. P. Jacobs, E. Kalai, and M. I. Kamien.

——— (1987): "Auction Theory," in *Advances in Economic Theory: Fifth World Congress*, ed. by T. F. Bewley.

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

Nau, R. F. (1992): "Indeterminate Probabilities on Finite Sets," *Annals of Statistics*, 20, 1737–1767.

——— (2003): "The Shape of incomplete Prefrences," Discussion paper, Fuqua School of Business, Duke University.

Riley, J., and W. Samuelson (1981): "Optimal Auctions," *American Economic Review*, 71, 381–92.

Seidenfeld, T., M. J. Schervish, and J. B. Kadane (1995): "A Representation of Partially Ordered Preferences," *Annals of Statistics*, 23, 2168–2217.

Shapley, L. S., and M. Baucells (1998): "Multiperson Utility," Discussion paper, University of California, Los Angeles.