Uncertainty 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

November 2007

Abstract

We consider mechanism design problems in which agents perceive Knightian uncertainty. Uncertainty is formalized using incomplete preferences, as in Bewley (1986). Without completeness, individual decision making depends on a set of beliefs, and an action is preferred to another if and only if it has larger expected utility for all beliefs in this set. We consider two natural notions of incentive compatibility in this setting: maximal incentive compatibility requires that no strategy has larger expected utility than truth-telling for all beliefs, while optimal incentive compatibility requires that truth-telling has larger expected utility than all other strategies for all beliefs. In a model with a continuum of types, we show that optimal incentive compatibility is equivalent to ex-post incentive compatibility under fairly general conditions on beliefs. In a model with a discrete type space, we characterize full extraction of rents from private information. We show that full extraction requires sufficient disagreement in beliefs across types for optimal incentive compatible mechanisms. Contrary to the standard Bayesian case, we show that full extraction in optimal incentive compatible mechanisms is not generically possible.

JEL Codes: D0, D5, D8, G1

Keywords: Knightian uncertainty, mechanism design, auctions, incomplete preferences.

---

*We thank David Ahn, Dino Gerardi, Botond Koszegi, Paul Milgrom, Ben Polak, Debraj Ray, and Ennio Stacchetti for helpful discussions and comments. We also thank audiences at Decentralization 2004, RUD 2004, TARK 2005, CETC 2006, University of British Columbia, University of Iowa, Stanford, and NYU 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) suggests a distinction between uncertainty and risk, arguing that while risky events occur with measurable probabilities, the likelihood of uncertain events is more qualitative in nature and cannot be computed precisely. Ellsberg (1961) advanced a more precise definition of uncertainty, in which an event is uncertain or ambiguous if it has unknown probability. Choice behavior reflecting this difference is inconsistent with the expected utility model, and this observation has inspired a significant amount of recent research in economics and decision theory. The extent to which Knightian uncertainty affects economic institutions rests on whether the ambiguity individuals perceive about the environment translates into equilibrium effects. We develop a simple extension of the standard mechanism design framework, and use this framework to study how uncertainty influences economic outcomes in settings with private information.

Following Harsanyi (1967), we assume an agent's type encodes beliefs regarding the environment in which she acts. We also assume that these beliefs are described by a set of probability distributions to reflect the presence of Knightian uncertainty. This is in the spirit of the decision theory developed by Bewley (1986), in which preferences may be incomplete. Without completeness, individual decision making depends on a set of beliefs over the state space, and one state-contingent consumption bundle is preferred to another if and only if it yields larger expected utility for all probability measures in this set of beliefs. When uncertainty is absent, this set is a singleton and the model reduces to standard expected utility.

We consider a mechanism design setup with a single agent and a continuum of types. The agent's utility depends on her type, an uncertain state, and the outcome of the mechanism.[1] Because of Knightian uncertainty, the beliefs of each type are described by a set of probability distributions over the states, and we allow the preference order on state-contingent outcomes that underlies the agent's choices to be incomplete at the interim stage. The standard notion of interim incentive compatibility, which requires that truth-telling has expected utility at least as large as any other strategy, becomes ambiguous when beliefs are not singletons. We therefore develop two natural notions of incentive compatibility in this framework. A weak form of interim incentive compatibility requires that truth-telling be a maximal choice for the agent, that is, no other strategy gives higher expected utility for all beliefs; we call this *maximal incentive compatibility.* In a maximal incentive compatible mechanism, truth-telling is not dominated, but may be incomparable to other lies. A stronger notion of interim incentive compatibility requires instead that truth-telling be an optimal choice for the agent, that is, the expected utility from truth-telling is at least as large as that from any other strategy for all possible beliefs; we call this *optimal incentive compatibility.*

Our main theorem provides conditions under which optimal incentive compatibility is equivalent to ex post incentive compatibility. The conditions include standard regularity assumptions

[1]Since the state can include the types of other players, this framework can include mechanisms with many players.

often invoked in mechanism design, such as smoothness of the utility function, and novel restrictions regarding the richness of the agent's beliefs and the relationship between types and beliefs. In the standard framework, when only risk is allowed, interim and ex-post incentive compatible mechanisms are in general very different. The introduction of uncertainty, instead, results in sharp restrictions on feasible mechanisms, even when uncertainty is arbitrarily small. In this sense, the introduction of Knightian uncertainty generates a significant discontinuity with respect to standard Bayesian mechanism design.

We then proceed to study the nature of information rents in settings with Knightian uncertainty. While the seminal work of Akerlof (1970) shows that asymmetric information can have welfare consequences, a recent series of papers suggests that appropriately constructed mechanisms can extract all rents agents derive from private information. Crémer and McLean (1985, 1988) and McAfee and Reny (1992) have shown that, in the standard Harsanyi framework, correlation in beliefs across types allows the designer to extract all of the surplus in a wide array of settings. Since beliefs are generically correlated, these results suggest that private information typically has no value.

We provide sufficient and necessary conditions for full extraction of rents in our model. For these results we focus on settings with a discrete type space. With a continuum of types, our main results show that optimal incentive compatibility is equivalent to ex post incentive compatibility under fairly general conditions. In this case, well-known results imply that full extraction is not possible. In contrast, with a discrete type space, full extraction may still be possible with Knightian uncertainty. We show that with a maximal incentive compatible mechanism, full extraction follows from the familiar condition of correlation in beliefs across types, applied to some selection from agents' sets of beliefs. In contrast, full extraction with an optimal incentive compatible mechanism requires that this correlation condition holds uniformly across all beliefs. When uncertainty is sufficiently large, full extraction under optimal incentive constraints becomes impossible. As a consequence of these results, we show that full extraction is neither generically possible nor generically impossible with Knightian uncertainty.

Our paper is related to the recent and large body of work on "detail-free" mechanisms and robustness in mechanism design. Much of this work is motivated by relaxing aspects of standard mechanism design assumptions related to common knowledge and restrictions on higher order beliefs. For example, Bergemann and Morris (2005) model robustness to higher order beliefs by requiring implementation on the universal type space, and show that this is equivalent to ex-post implementation in many settings. This result can be viewed as a generalization of earlier work in Ledyard (1978, 1979) using the modern language of the universal type space. In Ledyard (1978, 1979), as in Bergemann and Morris (2005), valuations are fixed as beliefs vary from type to type. By taking the resulting union over all possible beliefs, their setup has the flavor of our model for the special case in which each type is fully uncertain, i.e. his belief set includes all probability measures, and optimal incentive compatibility is imposed. In this sense, our results can be interpreted as showing that ex-post incentive compatibility obtains even if belief sets are potentially significantly smaller than the full uncertainty case.

Chung and Ely (2007), instead, consider robustness of auction mechanisms with respect to the designer's beliefs about the agents' types. Focusing on optimal mechanisms as the designer's beliefs vary, Chung and Ely (2007) show that there exists at least one belief for which ex-post incentive compatible mechanisms are optimal. **?** also study the optimal auction design problem. In their case the bidders and the seller may be ambiguity averse in the sense of Gilboa and Schmeidler (1989). Their main result shows that for an ambiguity neutral seller, the optimal auction must provide full insurance to all types of all bidders provided the bidders's belief sets contain the seller's (unique) belief. In contrast, we do not model the designer's beliefs explicitly, and focus instead on robustness to models of individual agents' beliefs.

Our results on the existence of information rents and the potential for full extraction are also related to recent work emphasizing robustness. Neeman (2004) points out that the possibility of full extraction hinges critically on the assumption that types with different values have different beliefs. Heifetz and Neeman (2006) argue that this assumption is not satisfied generically in appropriate type spaces allowing for richer higher order beliefs. Our approach is similar in spirit, but instead considers robustness to the introduction of Knightian uncertainty in simple type spaces.[2] Our results show that the presence of Knightian uncertainty provides an alternative justification for the emergence of ex-post mechanisms, and for the impossibility of full rent extraction in private information settings.

The paper proceeds as follows. Section 2 describes the decision-theoretic framework that motivates our model. Section 3 introduces the setup. Section 4 develops the equivalence between optimal and ex-post incentive compatibility. In Section 5 we consider auction models. In Section 6 we study the problem of full extraction of information rents in Knightian mechanisms. Section 7 concludes.

[2] Ahn (2005) develops a theory of interactive beliefs allowing for agents to hold a compact set of beliefs at any level, and constructs a corresponding universal type space.

2 Preliminaries: Incomplete Preferences and Uncertainty

In this section we describe the basic model of decision making under uncertainty that underlies our work. Our model is in the spirit of Knightian decision theory axiomatized in Bewley (1986). The main result in Bewley (1986) shows that a strict preference relation that is not necessarily complete, but satisfies all other axioms of the standard Anscombe-Aumann framework, can be represented by a family of expected utility functions generated by a utility index and a set of probability distributions.[3] Incompleteness is thus reflected in the multiplicity of beliefs: the unique subjective probability distribution of the standard expected utility framework is replaced by a set of probability distributions.

To illustrate more formally, consider a finite state space S, and let $x = (x_1, \ldots, x_S)$ and $y = (y_1, \ldots, y_S)$ denote payoff vectors in $\mathbf{R}_+^S$ (with abuse of notation, S also denotes the cardinality of the state space). An agent's preference relation $\succ$ over state-contingent payoffs is represented by a closed and convex set Π of probability distributions on S and a continuous and 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 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 expectation with respect to the probability distribution $\pi = (\pi_1, \ldots, \pi_S)$. We say $\succ$ is *complete* if for all $x \in \mathbf{R}_+^S$, $\text{cl}\,\{y \in \mathbf{R}_+^S : x \succ y \text{ or } y \succ x\} = \mathbf{R}_+^S$. If $\succ$ is complete, the set Π reduces to a singleton and the usual expected utility representation obtains. Without completeness, comparisons between alternatives are carried out "one probability distribution at a time", with one bundle strictly preferred to another if and only if its expected utility is larger under every probability distribution in the set Π.[4]

Bewley (1986) suggests that the above representation captures the Knightian distinction between risk and uncertainty, where an event is risky if its probability is known, and uncertain otherwise. The decision maker perceives only risk when Π is a singleton, and uncertainty otherwise. Incompleteness and uncertainty are two sides of the same phenomenon in this framework; both the amount of uncertainty that the decision maker perceives and the degree of incomplete-

[3] Bewley's original paper has been published recently as Bewley (2002). Incompleteness in decision making in a von Neumann Morgenstern setting was first studied by Aumann (1962) and Aumann (1964). Recently, this work has been extended and clarified by Dubra, Maccheroni, and Ok (2004), Ok (2002) and Shapley and Baucells (1998). Preferences of this kind have also been studied by Ghirardato, Maccheroni, Marinacci, and Siniscalchi (2003), and axiomatized by Girotto and Holzer (2005) in an infinite state space.

[4] The natural notion of indifference defines two bundles to be indifferent if they have the same expected utility for each probability distribution in Π. We return later to explore some implications of this definition with sufficiently rich uncertainty.

ness of her preference order $\succ$ are measured by the size of the set Π.[5]

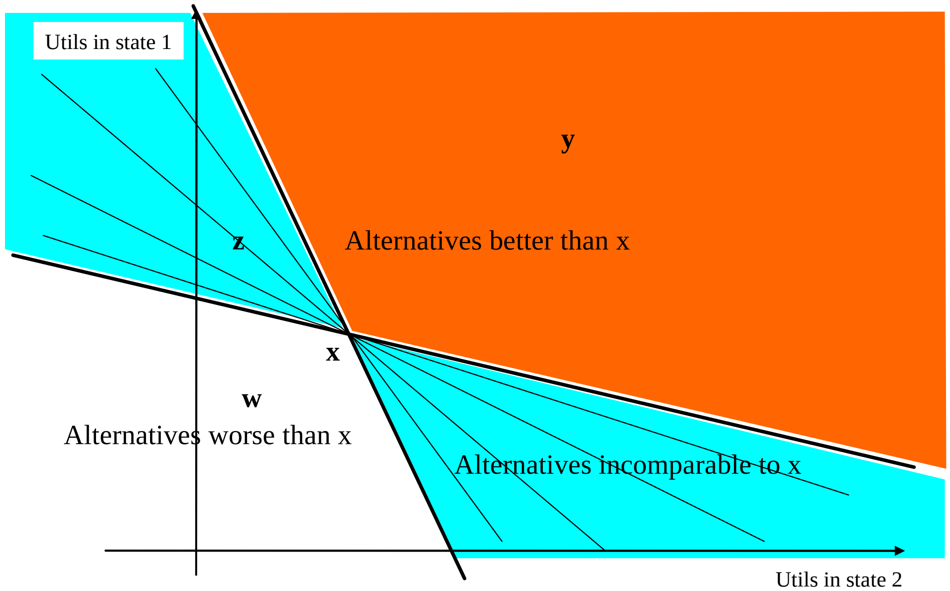

Figure 1: Incomplete Preferences

A picture may help to clarify. In Figure 1 the axes measure utils in each of two possible states. Any probability distribution over the states determines the slope of a standard indifference curve through the bundle x, as well as the set of other bundles having the same expected utility as x for that distribution. As this distribution changes, we obtain a family of indifference curves corresponding to different probabilities. The thick curves represent the most extreme elements of this family, while thin curves represent other possible elements.

A bundle like y is preferred to x since it lies above all indifference curves through x. Also, x is preferred to w since w lies below all indifference curves through x. Finally, z is not comparable to x since it lies above some indifference curves through x and below others. Thus each bundle x generates three regions: bundles preferred to x, worse than x, and incomparable to x. This last set is empty only if there is a unique probability distribution over the two states (i.e. the preferences are complete).

Usual revealed preference arguments may not apply when preferences are incomplete. If y is chosen when x is available, we cannot say y is revealed preferred to x; we can only say x is not revealed preferred to y. In other words, choice among incomparable alternatives cannot be linked directly to preferences. This observation will have important implications for modeling implementation in mechanism design problems since it bears directly on the idea of incentive compatibility.

[5]For precise results along these lines, see Ghirardato, Maccheroni, and Marinacci (2004) or Rigotti and Shannon (2005).

3 The Setup

In this section, we describe a simple mechanism design framework with Knightian uncertainty, and introduce two notions of incentive compatibility. Our setup is entirely standard, except for the possibility that agents' perceptions of the state space reflect Knightian uncertainty. We focus on a framework with a single agent so that we can strip down the model of higher order beliefs and abstract from issues of strategic uncertainty, aspects that would complicate the analysis without adding much substance.

Let S be the set of states, and O be the set of outcomes. We assume that S is a compact metric space. We let $\Delta(E)$ denote the set of all Borel probability measures on a set E, and endow $\Delta(E)$ with the weak* topology.

There is a single agent with privately known type $t \in T$. We restrict attention to direct mechanisms.[6] Any such mechanism is a function $\phi : T \times S \to O$ that specifies an outcome $\phi(\theta, s) \in O$ for any reported type $\theta \in T$ and any realized state $s \in S$. The agent's payoff function is

$$u : O \times T \times S \to \mathbf{R}.$$

Thus when the agent reports θ, while her true type is t and the realized state is s, her *ex post utility* is

$$u(\phi(\theta, s), t, s).$$

This basic framework can be easily modified to allow for many agents, by specifying that the type spaces of other agents be part of each agent's state space.[7]

Each type $t \in T$ has a closed, convex set of beliefs $\Pi(t) \subset \Delta(S)$. When the agent reports θ, while her true type is t, we denote her expected payoff according to any $\pi \in \Pi(t)$ by

$$E_\pi \left[u(\phi(\theta, s), t, s) \right].$$

In the standard Bayesian setup, $\Pi(t)$ is a singleton, and this expected value corresponds to the agent's *interim* expected utility when reporting θ. This case corresponds to the absence of Knightian uncertainty in our model.[8] In our analysis we will also refer occasionally to the opposite extreme case of "full ignorance," where the belief set of each type is the entire simplex, $\Pi(t) = \Delta(S)$ for all $t \in T$.

A *mixed strategy* is a function $\sigma : T \to \Delta(T)$ specifying a probability distribution $\sigma(t) \in \Delta(T)$

[6] In the appendix we verify that the revelation principle holds for the equilibrium notions we use: for any equilibrium outcome of any (indirect) mechanism there exists a direct mechanism in which truth-telling is an equilibrium inducing the same outcome.

[7] In Section 5 we consider an auction environment with multiple agents.

[8] The terms "ex post" and "interim" are in accordance with standard mechanism design terminology: *interim* refers to the stage at which the agent knows her type but has not observed the state, while *ex post* refers to the stage at which there is no uncertainty about the realized state.

for each $t \in T$.[9] The expected payoff generated by $\sigma(t) \in \Delta(T)$ is

$$E_{\sigma(t)}\left[u\left(\phi\left(\theta, s\right), t, s\right)\right],$$

and the interim expected utility of $\sigma(t)$ according to any $\pi \in \Pi(t)$ is

$$E_{\pi}\left[E_{\sigma(t)}[u(\phi(\theta, s), t, s)]\right].$$

We now define three notions of incentive compatibility. The first is the standard ex-post incentive compatibility notion. As usual, it requires that the agent prefers reporting her true type to reporting any other type, in each state.

Definition 1 *A mechanism $\phi : T \times S \to O$ is* ex post incentive compatible *for type t if for each $s \in S$*

$$u(\phi(t, s), t, s) \geq u(\phi(\theta, s), t, s) \qquad \forall \theta \in T.$$

The two notions of incentive compatibility pertain to the interim stage. They differ from each other due to the presence of Knightian uncertainty. Interim incentive compatibility, in setting with compete preferences, requires that truth-telling generates at least as much expected utility as any other strategy. This notion however becomes ambiguous when the belief set $\Pi(t)$ is not a singleton, i.e. when the preference order on state-contingent outcomes are incomplete at the interim stage. A weak notion of interim incentive compatibility in this case simply for each type $t \in T$ there is no strategy that generates a higher expected utility than truth-telling utility, for all beliefs in $\Pi(t)$. In this case we say that truth telling is a *maximal* strategy.

Definition 2 *A mechanism $\phi : T \times S \to O$ is* maximal incentive compatible *for type t if there exists no $\sigma(t) \in \Delta(T)$ such that*

$$E_{\pi}\left[E_{\sigma(t)}\left[u\left(\phi\left(\theta, s\right), t, s\right)\right]\right] > E_{\pi}\left[u\left(\phi\left(t, s\right), t, s\right)\right] \qquad \forall \pi \in \Pi(t).$$

A stronger notion of interim incentive compatibility requires instead that truth-telling generates at least as much expected utility as any other strategy for all beliefs in $\Pi(t)$, for each type $t \in T$. In this case we say that truth telling is an *optimal* strategy.[10]

Definition 3 *A mechanism $\phi : T \times S \to O$ is* optimal incentive compatible *for type t if for all $\sigma(t) \in \Delta(T)$*

$$E_{\pi}\left[u\left(\phi\left(t, s\right), t, s\right)\right] \geq E_{\pi}\left[E_{\sigma(t)}\left[u\left(\phi\left(\theta, s\right), t, s\right)\right]\right] \qquad \forall \pi \in \Pi(t).$$

[9] Note that $\sigma(t)$ specifies a unique distribution for each type; thus we assume that the agent views the randomness induced by his use of a mixed strategy as risk rather than uncertainty.

[10] Note that to verify optimal incentive compatibility, it suffices to check for deviations in pure strategies, while for maximal incentive compatibility deviations in mixed strategies must also be checked.

We say that a mechanism is *ex post, maximal, or optimal incentive compatible* if it is ex post, maximal, or optimal incentive compatible for all types.

It is obvious that optimal incentive compatibility implies maximal incentive compatibility. However any maximal incentive compatible mechanism such that there exists a type t and a mixed action $\sigma(t)$ such that

$$E_{\pi'}\left[u\left(\phi\left(t,s\right),t,s\right)\right] > E_{\pi'}\left[E_{\sigma(t)}\left[u\left(\phi\left(\theta,s\right),t,s\right)\right]\right] \text{ for some } \pi' \in \Pi\left(t\right)$$

and

$$E_{\pi''}\left[u\left(\phi\left(t,s\right),t,s\right)\right] < E_{\pi''}\left[E_{\sigma(t)}\left[u\left(\phi\left(\theta,s\right),t,s\right)\right]\right] \text{ for some } \pi'' \in \Pi\left(t\right)$$

fails to satisfy optimal incentive compatibility.

It is worth highlighting that the relation between maximal and optimal incentive compatibility is conceptually different from the familiar relation between best-response and strict best-response in the standard Bayesian framework, due to the fundamental difference between incomparability and indifference.[11] In particular, when preferences are strictly monotone, small changes in payoffs can always break ties due to indifference, but cannot in general eliminate incomparable alternatives.

The three notions just defined are clearly nested: ex post incentive compatibility implies optimal incentive compatibility, which in turn implies maximal incentive compatibility. Intuitively, as $\Pi(t)$ becomes larger, the maximal incentive constraints become less demanding while the optimal incentive constraints become tighter. In this sense, optimal incentive compatibility captures a notion of robustness to Knightian uncertainty.

In the special case where $\Pi(t)$ is a singleton, i.e. preferences are complete, both maximal and optimal incentive compatibility become identical to the standard notion of interim incentive compatibility. In the opposite special case of full ignorance, where the agent's belief set is the entire simplex $\Delta(S)$, optimal incentive compatibility becomes equivalent to ex post incentive compatibility, and maximal incentive compatibility becomes equivalent to the property that truth-telling is not strictly dominated.[12] We record these observations as Lemma 1, and provide the easy proof for the sake of completeness.

Lemma 1 *If $\Pi\left(t\right) = \Delta\left(S\right)$, then*

(i) a mechanism is optimal incentive compatible for type t if and only if it is ex post incentive compatible for type t;

[11] For example, indifference is transitive while incomparability need not be.

[12] Truth-telling is not strictly dominated for type t when there is no $\sigma\left(t\right) \in \Delta\left(T\right)$ such that

$$E_{\sigma(t)}\left[u\left(\phi\left(\theta,s\right),t,s\right)\right] > u\left(\phi\left(t,s\right),t,s\right) \qquad \forall s \in S.$$

(ii) a mechanism is maximal incentive compatible for type t if and only if truth-telling is not strictly dominated for type t.

Proof (i) Clearly ex post implies optimal incentive compatibility. For the converse, suppose that ϕ is not ex post incentive compatible for type t. Then there exists a state s and a report θ such that

$$u(\phi(t,s),t,s) < u(\phi(\theta,s),t,s).$$

Let π^s denote the measure assigning probability one to s. Since $\pi^s \in \Pi(t) = \Delta(S)$, truth-telling cannot be optimal incentive compatible.

(ii) Clearly if truth-telling is not strictly dominated, maximal incentive compatibility holds. For the converse, suppose that ϕ is maximal incentive compatible for type t. Then there exists no $\sigma(t) \in \Delta(T)$ such that

$$E_\pi\left[u\left(\phi\left(t,s\right),t,s\right)\right] < E_\pi\left[E_{\sigma(t)}\left[u\left(\phi\left(\theta,s\right),t,s\right)\right]\right] \qquad \forall \pi \in \Delta\left(S\right).$$

Since $\Pi(t) = \Delta(S)$, the point mass π^s is contained in $\Pi(t)$ for each state $s \in S$. For these measures, the above inequality becomes

$$u\left(\phi\left(t,s\right),t,s\right) < E_{\sigma(t)}\left[u\left(\phi\left(\theta,s\right),t,s\right)\right].$$

Since this applies to any $s \in S$, truth-telling is not strictly dominated. ■

4 Optimal and Ex Post Incentive Compatibility

In this section we show that under fairly mild conditions on the agent's payoff function and beliefs, optimal incentive compatibility is equivalent to ex post incentive compatibility. Since ex post incentive compatibility always implies optimal incentive compatibility, we concentrate on establishing the reverse implication. The result hinges on two sets of assumptions, imposed on the agents' payoff function and beliefs, respectively. First, the following set of regularity conditions is imposed on the payoff function to enable the use of the envelope theorem.

Assumption 1 *(a) The type space is $T = [0,1]$;*

(b) The payoff function $u : O \times T \times S \to \mathbf{R}$ is differentiable with respect to t, and $u_2 := \frac{\partial u}{\partial t}$ is non-negative and bounded.

Given the differentiability assumption in (1b), the non negativity of u_2 is without additional loss of generality, since types can always be reordered. Since u_2 can be equal to zero on some interval, distinct types t and t' can have the same payoff function $u(\cdot,t,\cdot) = u(\cdot,t',\cdot)$ but different belief sets $\Pi(t) \neq \Pi(t')$.

Under Assumption 1, ex post incentive compatibility can be easily characterized in terms of an envelope condition and a monotonicity condition.

Definition 4 *A mechanism* $\phi : T \times S \to O$ *satisfies the* ex post envelope condition *if for each* $s \in S$

$$u(\phi(t', s), t', s) - u(\phi(t, s), t, s) = \int_t^{t'} u_2(\phi(\tau, s), \tau, s) d\tau \qquad \forall t, t' \in T.$$

Definition 5 *A mechanism* $\phi : T \times S \to O$ *is* ex post monotone *if for each* $s \in S$

$$0 \leq t' < t \leq 1 \quad \textit{implies} \quad u(\phi(t, s), t, s) - u(\phi(t, s), t', s) \geq u(\phi(t', s), t, s) - u(\phi(t', s), t', s).$$

When specialized to the case of quasi-linear utility functions, ex post monotonicity is equivalent to weak monotonicity as defined in Bikhchandani, Chatterji, Lavi, Mu'alem, Nisan, and Sen (2006). The next lemma extends well known results about the equivalence between ex post incentive compatibility and the ex post envelope condition plus ex post monotonicity (see, for example, Krishna (2002)).

Lemma 2 *Suppose Assumption 1 holds. Then a mechanism is ex post incentive compatible for all types* $t \in T$ *if and only if it satisfies the ex post envelope condition and is ex post monotone.*

Proof Fix $s \in S$, and set $w(\theta, t) := u(\phi(\theta, s), t, s)$. Using this notation, the ex post envelope condition, ex post monotonicity, and ex post incentive compatibility can be rewritten as follows:

$$w(t', t') - w(t, t) = \int_t^{t'} w_2(\tau, \tau) d\tau \qquad \forall t, t' \in T, \tag{EPEC}$$

$$w(t, t) - w(t, t') \geq w(t', t) - w(t', t') \quad \forall\ t, t' \in T, \tag{EPM}$$

and

$$w(t, t) \geq w(\theta, t) \qquad \forall t, \theta \in T. \tag{EPIC}$$

First we show that EPEC and EPM imply EPIC. Choosing $t > t'$ and dividing both sides in EPM by $t - t'$ yields

$$\frac{w(t, t) - w(t, t')}{t - t'} \geq \frac{w(t', t) - w(t', t')}{t - t'}.$$

Since w is differentiable in its second argument, we have

$$\lim_{t' \to t} \frac{w(t, t) - w(t, t')}{t - t'} \geq \lim_{t \to t'} \frac{w(t', t) - w(t', t')}{t - t'},$$

i.e.

$$w_2(t, t) \geq w_2(t', t) \text{ for all } t > t',$$

which in turn implies

$$\int_\theta^t w_2(\tau, \tau) d\tau \geq \int_\theta^t w_2(\theta, \tau) d\tau \quad \forall\ t, \theta \in T.$$

Using EPEC on the left hand side and the fundamental theorem of calculus on the right hand side yields

$$w(t,t) - w(\theta,\theta) \geq w(\theta,t) - w(\theta,\theta),$$

or, after simplification,

$$w(t,t) \geq w(\theta,t).$$

Since t and θ were arbitrary, we conclude that EPIC holds.

Now suppose that EPIC holds. Using Assumption 1, EPEC follows immediately from the envelope theorem as in Theorem 2 in Milgrom and Segal (2002). To see that EPM must hold, consider the two EPIC constraints that involve types t and t':

$$w(t,t) \geq w(t',t),$$

$$w(t',t') \geq w(t,t').$$

Adding and rearranging yields EPM. ■

The second set of assumptions pertains to the agent's beliefs and is tied to our model of Knightian uncertainty. Loosely speaking, it requires that the agent's beliefs are sufficiently rich and locally overlapping. We formalize these ideas as follows.

Definition 6 *A set* $\Pi \subset \Delta(S)$ *has* full dimension *if, given any continuous function* $g : S \to \mathbf{R}$*,*

$$\int_S g(s)\, d\pi = 0 \quad \forall \pi \in \Pi \qquad \textit{implies} \quad g = 0.$$

Definition 7 *The agent's beliefs are* fully overlapping *if for each* $t \in T$ *there exists a neighborhood* $N(t) \subset T$ *such that* $\bigcap_{t' \in N(t)} \Pi(t')$ *has full dimension.*

We are now ready to state and prove the equivalence result.

Theorem 1 *Suppose that Assumption 1 holds and that the agent's beliefs are fully overlapping. Then any mechanism that is both optimal incentive compatible and ex post monotone is also ex post incentive compatible.*

Lemma 2 has already established that ex post monotonicity is necessary for ex post incentive compatibility. The main step in the proof consists in using the envelope theorem in integral form (Milgrom and Segal (2002)) and the assumption of fully overlapping beliefs to arrive at the ex post envelope condition of Definition 4. The result then follows immediately from Lemma 2.

Proof Let ϕ be a mechanism that satisfies optimal incentive compatibility; that is, for all $t, \theta \in T$,

$$\int_S u(\phi(t,s),t,s)d\pi - \int_S u(\phi(\theta,s),t,s)d\pi \geq 0 \qquad \forall \pi \in \Pi(t). \tag{OIC}$$

Fix $t_0 \in T$ and, using fully overlapping beliefs, choose a neighborhood $N(t_0) \subset T$ of t_0 such that $\Pi^N(t_0) := \cap_{t' \in N(t_0)} \Pi(t')$ has full dimension. For each $\pi \in \Pi^N(t_0)$ and each $t' \in N(t_0)$, OIC together with an envelope theorem (e.g., Theorem 2 in Milgrom and Segal (2002)) ensures that

$$\int_S u(\phi(t_0,s),t_0,s)d\pi - \int_S u(\phi(t',s),t',s)d\pi = \int_{t'}^{t_0} \left(\int_S u_2(\phi(\tau,s),\tau,s)d\pi \right) d\tau \tag{1}$$

or equivalently

$$\int_S \left[u(\phi(t_0,s),t_0,s) - u(\phi(t',s),t',s) - \int_{t'}^{t_0} u_2(\phi(\tau,s),\tau,s)d\tau \right] d\pi = 0 \quad \forall \pi \in \Pi^N(t_0).$$

Since $\Pi^N(t_0)$ has full dimension, for each $t' \in N(t_0)$,

$$u(\phi(t_0,s),t_0,s) - u(\phi(t',s),t',s) = \int_{t'}^{t_0} u_2(\phi(\tau,s),\tau,s)d\tau \qquad \forall s \in S. \tag{2}$$

This in turn implies that for almost all $t' \in N(t_0)$,

$$\frac{d}{dt} u(\phi(t',s),t',s) = u_2(\phi(t',s),t',s) \quad \forall s \in S.$$

Since t_0 was arbitrary, we conclude that equation (2) holds for all $t, t' \in T$. Thus ϕ satisfies the ex post envelope condition. By assumption, ϕ is also ex post monotone, so Lemma 2 yields the desired conclusion that ϕ is ex post incentive compatible. ■

Assumption 1 is indispensable for the equivalence result. In particular, if T is finite, even if all types have the same belief set of full dimension, a large class of mechanisms will be ex post monotone and optimal incentive compatible but not ex post incentive compatible.

The condition of fully overlapping beliefs is also critical. For example, suppose that there are two states, $S = \{1, 2\}$, and the agent has quasilinear utility with value t for an object. An outcome is determined by the pair (q, m), where q denotes the probability that the agent is awarded the object and m denotes her payment. A mechanism specifies the probability $q(\theta, s)$ that the agent is awarded the object and her payment $m(\theta, s)$, for any pair $(\theta, s) \in [0, 1] \times \{1, 2\}$. Thus $u(\phi(\theta, s), t, s) = t\ q(\theta, s) - m(\theta, s)$. Any belief $\pi \in \Delta(S)$ is identified by the number $\pi_1 := \Pr[s = 1]$, and any belief set $\Pi(t)$ corresponds to an interval $[\underline{\pi}_1(t), \overline{\pi}_1(t)] \subset [0, 1]$. Let $\varepsilon \in (0, \frac{1}{6}]$, and

$$[\underline{\pi}_1(t), \overline{\pi}_1(t)] = \begin{cases} [\frac{1}{3} - 2\varepsilon, \frac{1}{3} - \varepsilon], & 0 \le t \le \frac{1}{2}; \\ [\frac{2}{3} + \varepsilon, \frac{2}{3} + 2\varepsilon], & \frac{1}{2} < t \le 1. \end{cases}$$

Consider the mechanism (q, m), where $q(\theta, s) = 1$ for all (θ, s), and

$$m(\theta, s) = \begin{cases} \frac{1}{2} g, & \text{if } \theta > \frac{1}{2} \text{ and } s = 1; \\ -g, & \text{if } \theta > \frac{1}{2} \text{ and } s = 2; \\ 0, & \text{if } \theta < \frac{1}{2}. \end{cases}$$

Since $\pi(t-\frac{1}{2}g)+(1-\pi)(t+g)=t+g(1-\frac{3}{2}\pi)$ the interim expected utility function

$$U\left(\phi\left(\theta,s\right),t\right)=\begin{cases} t+g(1-\frac{3}{2}\pi) & \text{if } \theta>\frac{1}{2}, \\ t & \text{if } \theta\leq\frac{1}{2}, \end{cases}$$

is maximized by any $\theta>\frac{1}{2}$ if $\pi<\frac{2}{3}$, and by any $\theta\leq\frac{1}{2}$ if $\pi>\frac{2}{3}$. Since $\pi<\frac{2}{3}$ for all $\pi\in\Pi(t)$ when $t\leq\frac{1}{2}$, and $\pi>\frac{2}{3}$ for all $\pi\in\Pi(t)$ when $t>\frac{1}{2}$, the mechanism satisfies optimal incentive compatibility. However is it easy to see that ϕ is not ex post incentive compatible: since $g>0$ any type $t>\frac{1}{2}$ prefers to report $\theta\leq\frac{1}{2}$ when $s=2$.

Lemma 2 and Theorem 1 can be easily generalized to multi-dimensional type spaces. For example, suppose that (i) the outcome space consists of all probability distribution over a finite set, i.e. $O=\Delta\left(\{1,\ldots,k\}\right)$, so that any mechanism can be represented by $\phi\left(\theta,s\right)=\left(\phi_1\left(\theta,s\right),\ldots,\phi_k\left(\theta,s\right)\right)$, where $\phi_j\left(\theta,s\right)$ denotes the probability of choosing outcome j; (ii) T is a smoothly connected subset of R^k, i.e. for any two points $t,t'\in T$ there exists a differentiable function $f:[0,1]\rightarrow T$ such that $f\left(0\right)=t$ and $f\left(1\right)=t'$; and (iii) $u\left(\phi\left(\theta,s\right),s,t\right)=\sum_{j=1}^{k}t_j\phi_j\left(\theta,s\right)$. In this case, the function $\widetilde{u}\left(j,s,t\right)=f(t_j(s))$ satisfies Assumption 1, and the proof of Theorem 1 follows with the requisite change that the integrals in equations (1) and (2) are interpreted as path integrals along any path connecting t and t'.

Theorem 1 shows that the presence of uncertainty, together with the notion of robustness imposed by optimal incentive compatibility, can severely limit the set of feasible mechanisms. The designer can only choose among ex post incentive compatible mechanisms, despite the fact that we maintain the standard common knowledge assumptions that usually make the designer's problem easier to solve. In particular, the designer does not need to elicit beliefs since there is a one-to-one relationship between beliefs and types. Nonetheless, this informational advantage is not sufficient to utilize mechanisms that are not ex post incentive compatible.

Viewed from this angle, our results are complementary to recent work on robust mechanism design, which has focused on relaxing common knowledge assumptions and allowing mechanisms to depend on higher order beliefs. In particular, our conclusions are similar to those of Bergemann and Morris (2005). Bergemann and Morris (2005) consider the problem of implementing a given social choice correspondence in an environment with finitely many agents. In their setting, agent i has payoff function $u_i:Y\times\Theta_i\rightarrow\mathbf{R}$, where Y denotes the set of feasible outcomes and Θ_i denotes the set of all possible "payoff types". The agent's privately known type $t_i\in T_i$ determines both her payoff type θ_i, via a function $\widehat{\theta}_i:T_i\rightarrow\Theta_i$, and her belief about the other agents' types $\widehat{\pi}_i\left(\cdot|t_i\right)\in\Delta\left(T_{-i}\right)$. A type space in this environment is a collection $(T_i,\widehat{\theta}_i,\widehat{\pi}_i)_{i=1}^{I}$, and a *payoff type space* is a type space in which for each i, $T_i=\Theta_i$ and $\widehat{\theta}_i$ is the identity map. Our setup can be interpreted as a payoff type space with a single agent ($\widehat{\theta}_1$ is the identity map), in which the state s includes the type profile of other agents whose behavior is not modeled explicitly.

In separable environments (e.g. with quasi-linear payoff functions and unrestricted payments), Bergemann and Morris (2005) show that ex post incentive compatibility is equivalent to interim incentive compatibility in all payoff type spaces in which the agents' beliefs are consistent with a common prior. The key observation, which can be viewed as embedding earlier results of Ledyard (1978, 1979) in the modern framework of type spaces, is the following. If a mechanism ϕ satisfies interim incentive compatibility for all distributions, then it must also satisfy interim incentive compatibility in the type space where all agents but i have a fixed payoff type profile θ_{-i}. Since this is true for any profile $\theta_{-i} \in \Theta_{-i}$, the mechanism ϕ must be ex post incentive compatible. This reasoning does not rely on the cardinality of the type space and can be used without essential alterations to establish the equivalence between optimal and ex post incentive compatibility in our setting for the case of full ignorance, in which $\Pi(t) = \Delta(S)$ for all $t \in T$. We have applied a similar reasoning in the proof of Lemma 1. In contrast, Theorem 1 shows that with a continuum of types, the equivalence between optimal incentive compatibility and ex post incentive compatibility may hold even when the belief sets of all types are significantly restricted.

We close this section by examining some sufficient conditions for beliefs to be fully overlapping that also illustrate the extent to which this condition differs from an assumption of full ignorance.

One sufficient condition comes from the observation that a convex set $\Pi \subset \Delta(S)$ has full dimension whenever its algebraic interior in $\Delta(S)$ is non-empty, where the algebraic interior of Π is given by

$$\text{alg-int } \Pi := \{\pi \in \Pi : \forall \tilde{\pi} \in \Delta(S) \text{ there exists } \delta \in (0,1] \text{ such that } (1-\delta)\pi + \delta\tilde{\pi} \in \Pi\}.$$

A stronger sufficient condition, equivalent when S is finite, is that Π has non-empty relative interior. Using this observation, a simple example of such a set with full dimension arises from the common ε-contamination model of ambiguity, in which for a fixed $\varepsilon \in (0,1)$ and $\bar{\pi} \in \Delta(S)$,

$$\Pi_\varepsilon := \{\pi \in \Delta(S) : \pi = (1-\varepsilon)\bar{\pi} + \varepsilon\tilde{\pi} \text{ for some } \tilde{\pi} \in \Delta(S)\}.$$

This set Π_ε has full dimension for any $\varepsilon \in (0,1)$.

Beliefs are fully overlapping under a mild form of continuity of the correspondence describing the agent's beliefs as her type varies. This point is formalized by the following result.

Theorem 2 *If the correspondence $\Pi : T \to 2^{\Delta(S)}$ is lower hemi-continuous and $\Pi(t)$ has non-empty relative interior for each $t \in T$, then the beliefs $\{\Pi(t) : t \in T\}$ are fully overlapping.*

Proof For each t, let rint $\Pi(t)$ denote the relative interior of $\Pi(t)$, which is non-empty by assumption. The correspondence $t \mapsto$ rint $\Pi(t)$ is lower hemi-continuous, so has a continuous selection $\pi(t) \in$ rint $\Pi(t)$ for each t. Fix t, and choose $\epsilon > 0$ and a neighborhood $N(t)$ such that $B_\epsilon(\pi(t')) \subset$ rint $\Pi(t')$ for each $t' \in N(t)$. Next, using the continuity of $t \mapsto \pi(t)$, choose a neighborhood $\hat{N}(t) \subset N(t)$ such that for all $t' \in \hat{N}(t)$, $\pi(t') \in B_{\epsilon/2}(\pi(t))$. Then by construction, $B_{\epsilon/2}(\pi(t)) \subset B_\epsilon(\pi(t')) \subset$ rint $\Pi(t')$ for each $t' \in \hat{N}(t)$, hence $B_{\epsilon/2}(\pi(t)) \subset \cap_{t' \in \hat{N}(t)}$ rint $\Pi(t')$. ■

Returning to the ε-contamination model for an example, if $t \mapsto (\bar{\pi}(t), \varepsilon(t)) \in \Delta(S) \times (0,1)$ is continuous, where $\{\bar{\pi}(t) : t \in T\}$ are mutually absolutely continuous, and $\Pi(t) := \Pi_{\varepsilon(t)}(\bar{\pi}(t))$, then beliefs are fully overlapping.

5 Optimal Incentive Compatible Auctions and Revenue Equivalence

In this section, we focus on auction environments. This helps to illustrate the significance of the equivalence result of the previous section in relation to the current literature, and sheds further light on the role played by various assumptions in a setting with quasi-linear utilities and multiple agents. In this setting, we also establish a strong equivalence result: the ex-post payment functions of two mechanisms that satisfy optimal incentive compatibility and allocate the object identically can differ at most by a constant.

To keep the notation as simple as possible we focus on the standard case where the owner of an indivisible object faces a set $N = \{1, ..., n\}$ of potential buyers, with interdependent values (see e.g. Krishna (2002)). Each buyer has type t_i drawn from $T_i = [0, 1]$. Buyer i's ex-post utility function is quasi-linear, and takes the form

$$q_i \; v_i(t_i, t_{-i}) - m_i$$

where q_i denotes the probability that buyer i receives the good, m_i denotes his payment, and $v_i : T_i \times T_{-i} \to \mathbf{R}$ denotes his value for the good.[13] Each function v_i is non-decreasing and continuously differentiable, with $\frac{\partial v_i}{\partial t_i}(\tau, t_{-i}) > 0$ for all $\tau \in T_i$, $t_{-i} \in T_{-i}$.

Each type t_i of each buyer i has a closed, convex set of beliefs over her opponents' types given by $\Pi_i(t_i) \subset \Delta(T_{-i})$. This model collapses to the standard auction model with interdependent values if all belief set are singletons and consistent with a common prior. Thus our setting departs from the benchmark model in a minimal way: we simply assume that each buyer has a set of beliefs over her opponents' types that is not necessarily a singleton.[14]

A direct mechanism consists of $2n$ functions $(q, m) = (q_i, m_i)_{i \in N}$ where

$$q_i : T_i \times T_{-i} \to \Delta(N \cup \{0\}),$$

with $q_0(t_i, t_{-i}) = 1 - \sum_{i=1}^{n} q_i(t_i, t_{-i})$ denoting the probability that the object is not sold, and $m_i : T_i \times T_{-i} \to \mathbf{R}$ specifying the payment that buyer i makes to the seller. For any profiles of true types $(t_i, t_{-i}) \in T_i \times T_{-i}$ and reported types $(\theta_i, \theta_{-i}) \in T_i \times T_{-i}$, the ex post utility of buyer i is

$$q_i(\theta_i, \theta_{-i}) v_i(t_i, t_{-i}) - m_i(\theta_i, \theta_{-i}).$$

Adapting definitions from the previous section in the natural way, we say that a mechanism (q, m) satisfies *optimal incentive compatibility* if for each $t_i, \theta_i \in T_i$, and for each $i \in N$

$$E_\pi \left[q_i\left(t_i, t_{-i}\right) v_i\left(t_i, t_{-i}\right) - m_i\left(t_i, t_{-i}\right) \right] \geq E_\pi \left[q_i\left(\theta_i, t_{-i}\right) v_i\left(t_i, t_{-i}\right) - m_i\left(\theta_i, t_{-i}\right) \right] \qquad \forall \pi \in \Pi\left(t_i\right); \tag{3}$$

[13] Following usual conventions, the subscript $-i$ is used for variables that pertain to all players except i.

[14] In particular, we do not introduce heterogeneity of higher order beliefs.

and it satisfies *ex-post incentive compatibility* if for each $t_i, \theta_i \in T_i$, and for each $i \in N$

$$q_i\left(t_i, t_{-i}\right) v_i\left(t_i, t_{-i}\right) - m_i\left(t_i, t_{-i}\right) \geq q_i\left(\theta_i, t_{-i}\right) v_i\left(t_i, t_{-i}\right) - m_i\left(\theta_i, t_{-i}\right) \qquad \forall t_{-i} \in T_{-i}. \quad (4)$$

The next lemma provides the foundation for the two theorems of this section.

Lemma 3 *If the beliefs of each buyer are fully overlapping, then any mechanism (q, m) that satisfies optimal incentive compatibility also satisfies the following ex-post envelope condition*

$$w_i\left(t_i', t_{-i}\right) - w_i\left(t_i, t_{-i}\right) = \int_{t_i}^{t_i'} q_i\left(\tau, t_{-i}\right) \frac{\partial v_i\left(\tau, t_{-i}\right)}{\partial t_i} d\tau \qquad \forall t_{-i} \in T_{-i}, \quad t_i, t_i' \in T_i, \quad i \in N, \quad (5)$$

where $w_i(t_i, t_{-i}) := q_i\left(t_i, t_{-i}\right) v_i\left(t_i, t_{-i}\right) - m_i\left(t_i, t_{-i}\right).$

Proof Fix $i \in N$. By taking $\theta_i = t_i'$, the optimal incentive compatibility constraints in (3) can be written as

$$E_\pi\left[w_i\left(t_i, t_{-i}\right)\right] - E_\pi\left[w_i\left(t_i', t_{-i}\right)\right] \geq E_\pi\left[q_i\left(t_i', t_{-i}\right)\left[v_i\left(t_i, t_{-i}\right) - v_i\left(t_i', t_{-i}\right)\right]\right] \qquad \forall \pi \in \Pi_i\left(t_i\right).$$

Combining this inequality with the one that results from switching the roles of t_i and t_i' yields

$$\begin{aligned} E_\pi\left[q_i\left(t_i, t_{-i}\right)\left[v_i\left(t_i, t_{-i}\right) - v_i\left(t_i', t_{-i}\right)\right]\right] &\geq E_\pi\left[w_i\left(t_i, t_{-i}\right)\right] - E_\pi\left[w_i\left(t_i', t_{-i}\right)\right] \\ &\geq E_\pi\left[q_i\left(t_i', t_{-i}\right)\left[v_i\left(t_i, t_{-i}\right) - v_i\left(t_i', t_{-i}\right)\right]\right] \quad \forall \pi \in \Pi_i(t_i) \cap \Pi_i(t_i') \end{aligned}$$

Taking $t_i' > t_i$, dividing all terms by $t_i' - t_i$, letting $N(t_i)$ be a neighborhood of t_i on which beliefs are fully overlapping, and taking the limit as $t_i' \to t_i$ yields

$$\frac{\partial\left[E_\pi\left[w_i\left(t_i, t_{-i}\right)\right]\right]}{\partial t_i} = E_\pi\left[q_i\left(t_i, t_{-i}\right) \frac{\partial v_i\left(t_i, t_{-i}\right)}{\partial t_i}\right] \quad \text{for all } \pi \in \bigcap_{t_i' \in N(t_i)} \Pi_i\left(t_i'\right).$$

Integrating both sides from t_i to any $t_i' \in N\left(t_i\right)$ yields

$$E_\pi\left[w_i\left(t_i', t_{-i}\right)\right] - E_\pi\left[w_i\left(t_i, t_{-i}\right)\right] = \int_{t_i}^{t_i'} E_\pi\left[q_i\left(\tau, t_{-i}\right) \frac{\partial v_i\left(\tau, t_{-i}\right)}{\partial t_i}\right] d\tau,$$

or equivalently

$$E_\pi\left[w_i\left(t_i', t_{-i}\right) - w_i\left(t_i, t_{-i}\right)\right] = E_\pi\left[\int_{t_i}^{t_i'} q_i\left(\tau, y\right) \frac{\partial v_i\left(\tau, t_{-i}\right)}{\partial t_i} d\tau\right] \quad \text{for all } \pi \in \bigcap_{\tau \in N(t_i)} \Pi_i\left(\tau\right).$$

By fully overlapping beliefs, this implies

$$w_i(t_i', t_{-i}) - w_i(t_i, t_{-i}) = \int_{t_i}^{t_i'} q_i\left(\tau, t_{-i}\right) \frac{\partial v_i\left(\tau, t_{-i}\right)}{\partial t_i} d\tau.$$

Differentiating with respect to t_i yields

$$\frac{\partial w_i\left(t_i, t_{-i}\right)}{\partial t_i}=q_i\left(t_i, t_{-i}\right) \frac{\partial v_i\left(t_i, t_{-i}\right)}{\partial t_i} \quad \textit{a.e.};$$

and integrating again yields (5), proving the claim. ■

The next theorem reestablishes the equivalence between optimal and ex post incentive compatibility in the auction context. Most of the proof involves linking the ex post monotonicity condition to the monotonicity of the allocation function q_i in buyer i's type t_i.

Theorem 3 *Suppose that the beliefs of all buyers are fully overlapping. If (q, m) satisfies optimal incentive compatibility and each q_i is non-decreasing in its first argument, then (q, m) satisfies ex-post incentive compatibility.*

Proof We first show that $u_i\left(q(\theta_i, \theta_{-i}), m(\theta_i, \theta_{-i}), t_i, t_{-i}\right)$ satisfies ex-post monotonicity if and only if q_i is non-decreasing in its first argument. To that end, fix $i \in N$ and fix $t_{-i} \in T_{-i}$. For any $t_i, t_i' \in T_i$, ex-post monotonicity can be written as

$$\begin{aligned} & \left[q_i(t_i, t_{-i}) v_i(t_i, t_{-i})-m_i(t_i, t_{-i})\right]-\left[q_i(t_i, t_{-i}) v_i(t_i', t_{-i})-m_i(t_i, t_{-i})\right] \\ & \quad \geq \left[q_i(t_i', t_{-i}) v_i(t_i, t_{-i})-m_i(t_i', t_{-i})\right]-\left[q_i(t_i', t_{-i}) v_i(t_i', t_{-i})-m_i(t_i', t_{-i})\right] \end{aligned}$$

or, after simplification

$$q_i(t_i, t_{-i})\left[v_i(t_i, t_{-i})-v_i(t_i', t_{-i})\right] \geq q_i(t_i', t_{-i})\left[v_i(t_i, t_{-i})-v_i(t_i', t_{-i})\right]. \tag{EPM$_i$}$$

Since v_i is non-decreasing in t_i, if EPM$_i$ holds then $q_i(t_i, t_{-i}) \geq q_i(t_i', t_{-i})$ whenever $t_i > t_i'$. Conversely, if q_i is non-decreasing in its first argument, EPM$_i$ is satisfied.

Now fix $i \in N$ and set $S := T_{-i}$. By Theorem 1, (q, m) must satisfy ex-post incentive compatibility for i. Repeating this argument for each $i \in N$ allows us to conclude that (q, m) is ex-post incentive compatible for all $i \in N$. ■

It is well known that there are payment functions m_i such that the pair (q_i, m_i) satisfies ex post incentive compatibility if and only if q_i is non-decreasing in t_i. Any such mechanism (q, m) in which q_i is fails to be non-decreasing in t_i cannot satisfy the ex post incentive constraints for buyer i.

We conclude this section with an ex post revenue equivalence result. Consider two mechanisms (q, m) and $(q, \widetilde{m})$ that satisfy optimal incentive compatibility and allocate the object in the same way. If the beliefs of buyer i are fully overlapping, and there is a type of buyer i who is indifferent between participating in either mechanism, then the ex post payment functions m_i and $\widetilde{m}_i$ must be identical.

Theorem 4 *Suppose that the beliefs of all buyers are fully overlapping. If* (q, m) *and* $(q, \widetilde{m})$ *are mechanisms satisfying optimal incentive compatibility, then for each* $i \in N$ *there exists a function* $k_i : T_{-i} \to R$ *such that*

$$k_i(t_{-i}) = m_i(t_i, t_{-i}) - \widetilde{m}_i(t_i, t_{-i}) \qquad \forall t_i \in T_i \quad \forall t_{-i} \in T_{-i}.$$

Moreover, if there exists a type t_i^* *that is indifferent between participating in either mechanism, then* $k_i \equiv 0$.

Proof By Lemma 3, optimal incentive compatibility implies the ex-post envelope condition (5). Rearranging this equation yields $\forall i \in N$:

$$m_i(t_i, t_{-i}) - m_i(0, t_{-i}) = q_i(t_i, t_{-i}) v_i(t_i, t_{-i}) - q_i(0, t_{-i}) v_i(0, t_{-i}) - \int_0^{t_i} q_i(\tau, t_{-i}) \frac{\partial v_i(\tau, t_{-i})}{\partial t_i} d\tau.$$

Since both (q, m) and $(q, \widetilde{m})$ must satisfy the previous equality, we have

$$m_i(t_i, t_{-i}) - m_i(0, t_{-i}) = \widetilde{m}_i(t_i, t_{-i}) - \widetilde{m}_i(0, t_{-i}),$$

or

$$m_i(t_i, t_{-i}) - \widetilde{m}_i(t_i, t_{-i}) = m_i(0, t_{-i}) - \widetilde{m}_i(0, t_{-i}) \tag{6}$$

This proves the first claim.

If type t_i^* is indifferent between participating in either mechanism then

$$E_\pi[q_i(t_i^*, t_{-i}) v_i(t_i^*, t_{-i}) - m_i(t_i^*, t_{-i})] = E_\pi[q_i(t_i^*, t_{-i}) v_i(t_i^*, t_{-i}) - \widetilde{m}_i(t_i^*, t_{-i})] \qquad \forall \pi \in \Pi_i(t_i^*);$$

or, after simplifying,

$$E_\pi[m_i(t_i^*, t_{-i})] = E_\pi[\widetilde{m}_i(t_i^*, t_{-i})] \qquad \forall \pi \in \Pi_i(t_i^*).$$

Because $\Pi(t_i^*)$ has full dimension, $m_i(t_i^*, t_{-i}) = \widetilde{m}_i(t_i^*, t_{-i})$ for all $t_{-i} \in T_{-i}$. Substituting into (6) yields

$$m_i(0, t_{-i}) = \widetilde{m}_i(0, t_{-i}),$$

hence $m_i(t_i, t_{-i}) = \widetilde{m}_i(t_i, t_{-i})$ for all $t_i \in T_i$. ■

Theorem 4 implies that the introduction of uncertainty can have a significant impact on the set of mechanisms that satisfy optimal incentive compatibility. For example, in the standard auction model with independent types, in the absence of Knightian uncertainty the set of all interim incentive compatible mechanisms contains all equilibria of all standard auction formats, e.g. first-price, second-price, all pay, etc.. With the introduction of a small amount of (full-dimensional) uncertainty, only ex post equilibria satisfy optimal incentive compatibility. In particular, there exists at most one mechanism (q, m) for any given assignment rule q that satisfies optimal incentive compatibility and leaves at least one type of each buyer indifferent between participating and earning zero ex post utility.

6 Full Extraction in Knightian Mechanisms

In this section we examine the extent to which private information can generate rents in the presence of Knightian uncertainty. The answer to this question will depend on the notion of incentive compatibility used and on how information rents are defined in the presence of Knightian uncertainty.

Our single agent setup is similar to the one studied in McAfee and Reny (1992), except for the introduction of Knightian uncertainty. The agent can participate in a game that will leave him with ex post rents. These rents may depend on the agent's private information, summarized by a set of types T, and on publicly observed information, summarized by a set of states S. The precise nature of the game e.g. an auction, a bargaining game, etc., is irrelevant for our purposes; we take the agent's payoff function $r : T \times S \to \mathbf{R}_+$ as given, with $r(t) : S \to \mathbf{R}_+$ denoting the rents for type t as function of the publicly observed state s. The designer, who might be a seller, a mediator, etc., can charge the agent for participating in the game, and the agent can choose not to participate, in which case his payoff is zero.

We focus on the case where both the type space T and the state space S are finite; and, with slight abuse of notation, we use S and T to denote both the sets and their cardinality.[15] The agent knows his type t privately, and may perceive Knightian uncertainty about the state variable s. For each $t \in T$ we assume that there is a closed, convex set $\Pi(t) \subset \Delta(S)$ describing the beliefs of type t.

Since the realization of the state s is publicly observable, the participation fee charged by the designer can be made contingent on s. Let $z \in \mathbf{R}^S$ denote a *participation fee schedule*. We assume that the designer can offer a finite menu $Z \subset \mathbf{R}^S$ of participation fee schedules, from which the agent must select one.

The extent to which the designer can construct the menu Z to extract the agent's rent will then depend both on the nature of the belief correspondence $t \mapsto \Pi(t)$, and on the notion of incentive compatibility invoked. We explore characterizations of full rent extraction in this setup, first under maximal incentive compatibility, and then under the more restrictive notion of optimal incentive compatibility. In the last subsection we will discuss the robustness of the necessary and sufficient conditions for full extraction that we identify.

6.1 Surplus extraction with maximal choices

We begin by thinking of incentive compatibility in terms of maximal choices. For a given menu Z, this means that the agent's choice from Z should be "minimal" in terms of expected payments.

[15] We already know from Theorem 1 that, with a continuum of types, optimal incentive compatibility is equivalent to ex post incentive compatibility under fairly general conditions on beliefs. Therefore whenever this result applies, full extraction with ex post incentive compatible and ex post individually rational mechanisms is not feasible. This motivates our choice of focusing on the case with finitely many types in this section.

Definition 8 *A participation fee schedule* $z^m \in Z$ *is* minimal for type t in a menu Z *if there is no mixed strategy* $\sigma \in \Delta(Z)$ *such that*

$$\sum_{z \in Z} \sigma_z [\pi \cdot z] < \pi \cdot z^m \qquad \forall \pi \in \Pi(t).$$

When the agent's choices are minimal, the corresponding notion of surplus for type $t \in T$, relative to a given menu Z, is

$$S^m(t) := \{\pi \cdot [r(t) - z] : z \text{ is minimal for type } t \text{ in } Z, \quad \pi \in \Pi(t)\}.$$

Note that S^m also depends on the rent function r and the menu Z, although we have suppressed this dependence.

Corresponding to the notion of minimal choices, we have the following definition of full rent extraction.

Definition 9 *A menu of participation fee schedules* Z *achieves* maximal full extraction of the rent function r *if for each* $t \in T$ *there exists* $z(t) \in Z$ *that is minimal for type* t *in* Z *and for which*

$$\pi \cdot [r(t) - z(t)] = 0 \text{ for some } \pi \in \Pi(t).$$

The designer can achieve maximal full extraction *if for every* $r \in \mathbf{R}_+^{S \times T}$ *there exists a menu* Z *that achieves maximal full extraction of* r*.*

Using the notation defined above, maximal rent extraction of a given rent function r can also be described as finding a menu Z for which $0 \in S^m(t)$ for each $t \in T$.

Under maximal incentive compatibility, the presence of ambiguity weakens the incentive constraints. Maximal full rent extraction is feasible under conditions on beliefs that mimic those in the standard case: if there exists a selection from the correspondence $\{\Pi(t) : t \in T\}$ that satisfies the correlation condition familiar from the work of Crémer and McLean, maximal full rent extraction is achievable.

Theorem 5 *The designer can achieve maximal full rent extraction if there exists a selection* $\{\pi(t) : \pi(t) \in \Pi(t) \text{ for each } t \in T\}$ *such that*

$$\forall\, t \in T: \qquad \pi(t) \notin co\ \{\{\pi(t')\}_{t' \neq t}\} \tag{MFE}$$

where co A denotes the convex hull of the set A.

Proof Fix a rent function $r \in \mathbf{R}_+^{T \times S}$. Suppose that (MFE) holds, and choose a selection $\{\pi(t) : \pi(t) \in \Pi(t) \text{ for each } t \in T\}$ such that

$$\forall t \in T: \qquad \pi(t) \notin \text{co}\ \{\{\pi(t')\}_{t' \neq t}\}.$$

For each t, using the Separating Hyperplane Theorem, choose $\tilde{z}_t \in \mathbf{R}^S$ such that

$$\begin{aligned} \pi(t) \cdot \tilde{z}_t &= 0 \\ \pi(t') \cdot \tilde{z}_t &> 0 \ \ \forall t' \neq t, \end{aligned}$$

and define

$$c_t := \pi(t) \cdot r(t)$$

Now by construction, for every $\alpha > 0$,

$$\pi(t) \cdot r(t) - c_t - \alpha\pi(t) \cdot \tilde{z}_t = 0$$

Then choose scaling factors $\alpha_t > 0$ for each t so that $\forall t' \neq t$,

$$\pi(t) \cdot r(t) - c_t - \alpha_t \pi(t) \cdot \tilde{z}_t \geq \pi(t) \cdot r(t) - c_{t'} - \alpha_{t'} \pi(t) \cdot \tilde{z}_{t'}$$

Because $\pi(t) \cdot \tilde{z}_{t'} > 0$ while $\pi(t) \cdot \tilde{z}_t = 0$, such a collection $\{\alpha_t\}_{t \in T}$ exists.

Also, for each t, set

$$z_t := c_t + \alpha_t \tilde{z}_t$$

and define the menu $Z := \{z_t : t \in T\}$. Now by construction, for all $t \in T$ we have

$$\pi(t) \cdot [r(t) - z_t] \geq \pi(t) \cdot [r(t) - z_{t'}] \quad \forall t' \neq t,$$

which guarantees that z_t is minimal for type t in the menu Z. (Note that it suffices to consider pure strategies because all comparisons are made with the same distribution $\pi(t)$); and by construction, for each $t \in T$

$$\pi(t) \cdot [r(t) - z_t] = 0 \in S^m(t)$$

Thus the menu Z achieves maximal full rent extraction of the rent function r. ■

6.2 Surplus extraction with optimal choices

When incentive compatibility is defined in terms of optimal choices, the presence of ambiguity strengthens the incentive constraints, and full rent extraction becomes more demanding than in the standard case with no ambiguity. In this case, we require the agent's choice from the proposed menu of participation fee schedules Z to be "pessimal" with respect to expected payments.

Definition 10 *A participation fee schedule z^p is* pessimal for type t in a menu Z *if for all mixed strategies $\sigma \in \Delta(Z)$*

$$\pi \cdot z^p \leq \sum_{z \in Z} \sigma_z [\pi \cdot z] \qquad \forall \pi \in \Pi(t).$$

In analogy with the case of minimal choices, we define the corresponding measure of surplus

$$S^o(t) := \{\pi \cdot [r(t) - z] : z \text{ is pessimal for type } t \text{ in } Z,\ \pi \in \Pi(t)\}$$

and the corresponding notion of full rent extraction.

Definition 11 *The menu Z achieves* optimal full rent extraction of the rent function r *if for each $t \in T$:*

$$\pi \cdot [r(t) - z(t)] \geq 0 \text{ for all } z(t) \text{ pessimal for } t \text{ in the menu } Z \text{ and for all } \pi \in \Pi(t)$$

with equality for at least one such $z(t)$ and $\pi \in \Pi(t)$.

The designer can achieve optimal full rent extraction *if for every rent function $r \in \mathbf{R}_+^{S\times T}$ there exists a menu Z that achieves optimal full rent extraction of r.*

Optimal full rent extraction of a given rent function r can alternatively be characterized via the surplus measure for each $t \in T$ as the requirement:

$$0 \in S^o(t) \subset [0, a(t)], \text{ where } a(t) \geq 0$$

Next we provide necessary and sufficient conditions for optimal full rent extraction. The sufficient condition is a uniform version of the selection condition that ensures maximal full extraction, and requires that beliefs be sufficiently different across types. The necessary condition we give is weaker, and instead requires that beliefs not be too similar across types.[16] We formalize this below.

Theorem 6 *The designer can achieve optimal full rent extraction if*

$$\forall t \in T: \qquad \Pi(t) \cap co\ \left\{\cup_{t' \neq t} \Pi(t')\right\} = \emptyset \tag{OFE}$$

Moreover, the designer can achieve optimal full rent extraction only if

$$\forall t \in T: \qquad \Pi(t) \not\supset co\ \left\{\cup_{t' \neq t} \Pi(t')\right\} \tag{NOFE}$$

Proof First, we claim that the condition (OFE) is sufficient for optimal full rent extraction. To see this, fix $r \in \mathbf{R}_+^{S\times T}$, and suppose that (OFE) holds. For each t, using the Separating Hyperplane Theorem, choose $\tilde{z}_t \in \mathbf{R}^S$ such that

$$\begin{aligned} \pi(t) \cdot \tilde{z}_t &\leq 0 \ \ \forall \pi(t) \in \Pi(t) \\ \pi(t') \cdot \tilde{z}_t &> 0 \ \ \forall \pi(t') \in \Pi(t'), \forall t' \neq t \end{aligned}$$

[16] The gap between the necessary and sufficient conditions in the optimal case arises due to the slack that is allowed in the incentive compatibility constraints for some beliefs.

Adjusting $\tilde{z}_t$ if necessary, $\tilde{z}_t$ can be chosen such that in addition

$$\max_{\pi(t)\in\Pi(t)} \pi(t)\cdot\tilde{z}_t = 0$$

For each t, set

$$\begin{aligned} c_t := \quad & \min_{\pi(t)\in\Pi(t)} \pi(t)\cdot r(t) \\ & \text{s.t. } \pi(t)\cdot\tilde{z}_t = 0 \end{aligned}$$

Now by construction, for every $\alpha > 0$,

$$\begin{aligned} \pi(t)\cdot r(t) - c_t - \alpha\pi(t)\cdot\tilde{z}_t &\geq 0 \quad \forall\pi(t)\in\Pi(t) \\ \pi(t)\cdot r(t) - c_t - \alpha\pi(t)\cdot\tilde{z}_t &= 0 \quad \text{for some } \pi(t)\in\Pi(t) \end{aligned}$$

Then for each t, choose scaling factors $\alpha_t > 0$ so that $\forall t' \neq t$,

$$\pi(t)\cdot r(t) - c_t - \alpha_t\pi(t)\cdot\tilde{z}_t \geq \pi(t)\cdot r(t) - c_{t'} - \alpha_{t'}\pi(t)\cdot\tilde{z}_{t'} \quad \forall\pi(t)\in\Pi(t)$$

Because $\pi(t)\cdot\tilde{z}_{t'} > 0\ \forall\pi(t)\in\Pi(t)$ while $\pi(t)\cdot\tilde{z}_t \leq 0\ \forall\pi(t)\in\Pi(t)$, this is possible.

For each t, set

$$z_t := c_t + \alpha_t\tilde{z}_t$$

and define the menu $Z := \{z_t : t \in T\}$. Now by construction, for each $t \in T$:

$$\pi\cdot[r(t) - z_t] \geq 0 \quad \forall\pi\in\Pi(t)$$

with equality for some $\pi(t)\in\Pi(t)$. Thus the menu Z achieves optimal full rent extraction of r.

To see that (NOFE) is necessary for optimal full rent extraction, fix $t_0 \in T$ and consider the rent function $r(t) = (t - t_0)^2$. Suppose Z is a menu that achieves optimal extraction of the rents given by r.

For each $t \neq t_0$, choose $z(t) \in Z$ pessimal for t in Z and $\pi(t)\in\Pi(t)$ such that

$$\pi(t)\cdot[r(t) - z(t)] = 0$$

Similarly, let $z(t_0) \in Z$ be pessimal for t_0 in Z.

Suppose there exists $\mu \in \Delta(T \setminus \{t_0\})$ such that

$$\pi(t_0) := \sum_{t\neq t_0} \mu_t\pi(t) \in \Pi(t_0)$$

Then

$$\begin{aligned}
0 = \pi(t_0) \cdot r(t_0) &\geq \pi(t_0) \cdot z(t_0) \\
&= \sum_{t \neq t_0} \mu_t \pi(t) \cdot z(t_0) \\
&\geq \sum_{t \neq t_0} \mu_t \pi(t) \cdot z(t) \\
&= \sum_{t \neq t_0} \mu_t \pi(t) \cdot r(t) \\
&= \sum_{t \neq t_0} \mu_t (t - t_0)^2 \\
&> 0
\end{aligned}$$

But this is impossible. ■

6.3 Genericity of optimal full extraction

We conclude this section with results investigating the robustness of the necessary and sufficient conditions for full extraction identified above. The work of Crémer and McLean and others uncovered the connection between correlated values and full extraction in standard Bayesian mechanism design. Perhaps the most powerful and negative aspect of this work was showing that these conditions are generic in an appropriate sense. Related work showed that these generic conditions lead to full extraction in a wide array of settings with private information. We seek a similar measure of the extent of full rent extraction and the existence of information rents in the presence of Knightian uncertainty.

To formalize this discussion, recall that in the standard Bayesian setting, each type $t \in T$ is associated with a unique conditional distribution $\pi(t) \in \Delta(S)$, and full rent extraction is possible if and only if the correlation condition (MFE) holds for the collection $\{\pi(t) : t \in T\}$. It is straightforward to see that the subset of $\Delta(S)^T$ on which this condition is satisfied is an open set of full Lebesgue measure. In a setting with Knightian uncertainty, each type t holds a set of distributions $\Pi(t)$ drawn from

$$\mathcal{C} := \{\Pi \subset \Delta(S) : \Pi \text{ is closed and convex}\}$$

and full extraction is characterized in terms of conditions on the collection $\{\Pi(t) : t \in T\} \in \mathcal{C}^T$. To gauge how widespread the absence of information rents is in this setting, we seek to measure the size of the subset of $\mathcal{C}^T$ corresponding to the various conditions we have identified above.

To make this precise, we endow $\mathcal{C}$ with the Hausdorff topology, and $\mathcal{C}^T$ with the product topology. Let $\mathcal{M}$ denote the subset of $\mathcal{C}^T$ satisfying (MFE), $\mathcal{O}$ denote the subset of $\mathcal{C}^T$ satisfying (OFE), and $\mathcal{N}$ denote the subset of $\mathcal{C}^T$ violating (NOFE). Thus $\mathcal{M}$ is a set on which maximal full extraction is always possible. Similarly, $\mathcal{O}$ is a set of beliefs for which optimal full extraction is always possible, and $\mathcal{N}$ is a set for which optimal full extraction is never possible.

The set $\mathcal{C}^T$ is infinite-dimensional, which means the issue of measuring the sizes of these sets is no longer straightforward due to the absence of a natural analogue of Lebesgue measure in infinite-dimensional spaces. Genericity in infinite-dimensional spaces is typically defined either using topological notions such as open and dense or residual, or using measure-theoretic notions such as prevalence. Prevalence and its complement shyness, developed by Christensen (1974) and Hunt, Sauer, and Yorke (1992), and made relative by Anderson and Zame (2001), are analogues of Lebesgue measure 0 and full Lebesgue measure that more closely mimic properties of Lebesgue measure in many problems.[17] We first give formal definitions, and then discuss some important properties shared by these notions of genericity.

Because we are interested in the relative size of subsets of $\mathcal{C}^T$, we use the relative notions of prevalence and shyness developed by Anderson and Zame (2001) for use in a convex subset which may be a shy subset of the ambient space. The formal definitions are given below.

Definition 12 *Let Z be a topological vector space and let $C \subset Z$ be a convex Borel subset of Z which is completely metrizable in the relative topology. Let $c \in C$. A universally measurable subset $E \subset Z$ is* shy in C at c *if for each $\delta > 0$ and each neighborhood W of 0 in Z, there is a regular Borel probability measure μ on Z with compact support such that* $\operatorname{supp} \mu \subset (\delta(C-c)+c) \cap (W+c)$ *and $\mu(E+z) = 0$ for every $z \in Z$.*[18] *The set E is* shy in C *if it is shy at each point $c \in C$. A (not necessarily universally measurable) subset $F \subset C$ is* shy in C *if it is contained in a shy universally measurable set. A subset $K \subset C$ is* prevalent in C *if its complement $C \setminus K$ is shy in C.*

Like Lebesgue measure 0, relative shyness and prevalence have many properties desirable for measure-theoretic notions of "smallness" and "largeness": relative shyness is translation invariant, preserved under countable unions, and coincides with Lebesgue measure 0 in $\mathbf{R}^n$, and no relatively open set is relatively shy.

We note a simple but important property common to both residual and relative prevalence as notions of genericity with respect to subsets of $\mathcal{C}^T$.

Lemma 4 *Let $X \subset \mathcal{C}^T$ be universally measurable. If $X^c = \mathcal{C}^T \setminus X$ has a non-empty relative interior, then X is neither residual nor relatively prevalent in $\mathcal{C}^T$.*

Proof For relative prevalence the result is immediate from the definitions and the fact that no relatively open set is relatively shy. To see that X is not residual in $\mathcal{C}^T$, note that $\mathcal{C}^T$ is a

[17] Well-known problems with interpreting topological notions of genericity are illustrated by simple examples of open and dense sets in $\mathbf{R}^n$ having arbitrarily small Lebesgue measure, and residual sets of Lebesgue measure 0.

[18] A set $E \subset Y$ is universally measurable if for every Borel measure η on Y, E belongs to the completion with respect to η of the sigma algebra of Borel sets.

compact metric space, hence a Baire space. The conclusion then follows immediately from the Baire Category Theorem. ■

From this simple observation, we conclude that optimal full extraction is neither generically possible nor generically impossible.

Theorem 7 *Neither $\mathcal{O}$ nor $\mathcal{N}$ is residual in $\mathcal{C}^T$. Neither $\mathcal{O}$ nor $\mathcal{N}$ is relatively prevalent in $\mathcal{C}^T$.*

Proof Both $\mathcal{O}$ and $\mathcal{N}$ are Borel sets, hence are universally measurable. By definition, $\mathcal{O} \cap \mathcal{N} = \emptyset$, so $\mathcal{O} \subset \mathcal{N}^c$ and $\mathcal{N} \subset \mathcal{O}^c$. The results will all follow provided both $\mathcal{O}$ and $\mathcal{N}$ have non-empty interior. In both cases, we will establish this by constructing interior points.

First consider $\mathcal{O}$. Choose $\{\pi(t) \in \Delta : t \in T\}$ such that $\pi(t) \notin \text{co}\ \{\pi(t') : t \neq t'\}$ for each t, i.e., such that $\{\{\pi(t)\} : t \in T\}$ satisfies (OFE). Choose $\epsilon > 0$ such that

$$\forall t \in T: \quad B_\epsilon(\pi(t)) \cap \text{co}\ \left\{\cup_{t' \neq t} B_\epsilon(\pi(t'))\right\} = \emptyset$$

that is, such that $\{B_\epsilon(\pi(t)) : t \in T\}$ also satisfies (OFE). Now if $\Pi \in \mathcal{C}$ and $d(\Pi, \{\pi(t)\}) < \epsilon$, $\Pi \subset B_\epsilon(\pi(t))$.[19] Thus any collection $\{\Pi(t) \in \mathcal{C} : t \in T\}$ such that $d(\Pi(t), \{\pi(t)\}) < \epsilon$ for each t must satisfy (OFE) as well. From this we conclude that $\{\pi(t) \in \Delta : t \in T\}$ is an interior point of $\mathcal{O}$.

Next, consider $\mathcal{N}$. Fix $t_0 \in T$. Choose $\Pi(t_0) \in \mathcal{C}$ such that $\text{rint}\ \Pi(t_0) \neq \emptyset$. Fix $\epsilon > 0$ and choose $\bar{\pi}(t_0) \in \text{rint}\ \Pi(t_0)$ such that $B_\epsilon(\bar{\pi}(t_0)) \subset \Pi(t_0)$. Now choose $\{\Pi(t) \subset \Delta^c(S) : t \in T \setminus \{t_0\}\}$ such that

$$\text{co}\ \{\cup_{t \neq t_0} \Pi(t)\} \subset B_{\epsilon/4}(\bar{\pi}(t_0))$$

In particular then, $\{\Pi(t) : t \in T\}$ violates (NOFE), so $\{\Pi(t) : t \in T\} \in \mathcal{N}$.

If $\{\tilde{\Pi}(t) \in \mathcal{C} : t \in T \setminus \{t_0\}\}$ is any collection such that $d(\tilde{\Pi}(t), \Pi(t)) < \epsilon/4T$ for each $t \neq t_0$, then

$$\text{co}\ \{\cup_{t \neq t_0} \tilde{\Pi}(t)\} \subset B_{\epsilon/2}(\bar{\pi}(t_0))$$

Finally, if $\Pi \in \mathcal{C}$ and $d(\Pi, \Pi(t_0)) < \epsilon/2$, then $B_{\epsilon/2}(\bar{\pi}(t_0)) \subset \Pi$. Putting these observations together, any collection $\{\tilde{\Pi}(t) \in \mathcal{C} : t \in T\}$ such that $d(\tilde{\Pi}(t_0), \Pi(t_0)) < \epsilon/2$ and $d(\tilde{\Pi}(t), \Pi(t)) < \epsilon/4T$ for each $t \neq t_0$ will also violate (NOFE) and hence will belong to $\mathcal{N}$. ■

Taken together, Theorems 6 and 7 show that the conditions for optimal full extraction are more stringent in a world with Knightian uncertainty than the generic conditions for full rent extraction in Bayesian mechanisms. As argued above, optimality makes incentive constraints

[19] Here $d(A, B)$ denotes the Hausdorff distance between $A, B \in \mathcal{C}$, defined by

$$d(A, B) = \max\left\{\sup_{x \in A} \text{dist}(x, B), \sup_{y \in B} \text{dist}(y, A)\right\}.$$

harder to satisfy when Knightian uncertainty is introduced, so it is natural that full extraction would correspondingly become more difficult. These results also suggest that concerns about the possibility full extraction and the existence of information rents translate naturally to a choice between methods of resolving incentive compatibility with Knightian uncertainty.

7 Conclusion

We have developed a framework for mechanism design in the presence of Knightian uncertainty. In this setting, the distinction between maximal and optimal implementation is crucial for the resulting mechanisms. Under maximal incentive compatibility, uncertainty weakens incentive constraints and provides more scope for a designer, resulting in predictions that more closely resemble standard Bayesian mechanism design. The interpretation of implementation that relies on particular choices in the presence of incomparable alternatives is problematic, however. Mechanisms that are robust to uncertainty instead require optimal incentive compatibility.

Robustness to uncertainty can then result in significant discontinuities in standard Bayesian mechanism design predictions. Without uncertainty, the sets of mechanisms that satisfy interim and ex-post incentive compatibility are significantly different. Instead, the addition of uncertainty implies that only ex-post incentive compatible mechanisms may be robust to small perturbations in standard models. Similarly, with uncertainty, full extraction of information rents in optimal incentive compatible mechanisms imposes a nontrivial and non-generic constraint on the differences in beliefs across types, while in the absence of uncertainty, full extraction of these rents is possible even when beliefs are arbitrarily close.

These results contrast with recent work in the robust mechanism design literature while reaching qualitatively similar conclusions. We maintain standard assumptions regarding common knowledge for all participants, while allowing for Knightian uncertainty. In contrast, following Wilson (1987), much recent work has taken up the charge of the oft-quoted "Wilson doctrine" to consider robustness to weakening common knowledge assumptions. Our results indicate that even maintaining standard assumptions regarding common knowledge, simple ex-post mechanisms may emerge as the only feasible mechanisms robust to Knightian uncertainty.

8 Appendix: Direct Mechanisms and the Revelation Principle

In this appendix, we define the general class of mechanisms and verify that the revelation principle holds for both maximal and optimal strategies so that our focus on direct mechanisms is justified.

For simplicity we focus on the single agent case. The notation and basic setup are as in Section 3. The outcome space is O, the state space is S, the type space is T, and the agent has type-dependent ex post payoff

$$u : O \times T \times S \to \mathbf{R}.$$

A mechanism in this environment consists of a message space B and a function $g : B \times S \to O$. If the agent chooses to participate in the mechanism, she chooses a message $b \in B$, and the mechanism specifies a state-dependent outcome $g(b, s)$. Thus moral hazard is ruled out: by assumption the agent always obeys the mechanism.

A mixed strategy is a function $\sigma : T \to \Delta(B)$. For each type $t \in T$ there is a closed, convex set of beliefs $\Pi(t) \subset \Delta(S)$ such that

$$\sigma(t) \succeq_t \sigma'(t) \text{ if and only if } E_\pi\left[E_{\sigma(t)}\left[u\left(g\left(b,s\right),t,s\right)\right]\right] \geq E_\pi\left[E_{\sigma'(t)}\left[u\left(g\left(b,s\right),t,s\right)\right]\right] \text{ for all } \pi \in \Pi(t).$$

We consider two notions of best response: a maximal strategy corresponds to the requirement of "no incentive to deviate", while an optimal strategy must be better than any other feasible strategy.[20]

Definition 13 *A strategy σ is* maximal *if there is no other strategy $\sigma' : T \to \Delta(B)$ such that*

$$E_\pi\left[E_{\sigma'(t)}\left[u\left(g\left(b,s\right),t,s\right)\right]\right] > E_\pi\left[E_{\sigma(t)}\left[u\left(g\left(b,s\right),t,s\right)\right]\right] \qquad \textit{for all } \pi \in \Pi(t).$$

Definition 14 *A strategy σ is* optimal *if for each $\sigma' : T \to \Delta(B)$:*

$$E_\pi\left[E_{\sigma(t)}\left[u\left(g\left(b,s\right),t,s\right)\right]\right] \geq E_\pi\left[E_{\sigma'(t)}\left[u\left(g\left(b,s\right),t,s\right)\right]\right] \qquad \textit{for all } \pi \in \Pi(t).$$

A social choice function $\psi : T \times S \to O$ specifies a feasible outcome for any pair (t, s).

Definition 15 *A mechanism g* implements the social choice function ψ in maximal (optimal) strategies *if there exists a maximal (optimal) pure strategy $\beta : T \to B$ such that $g(\beta(t), s) = \psi(t, s)$ for all $t \in T$ and for all $s \in S$.*

Our analysis will focus on truth-telling in direct mechanisms. At truth-telling, the ex-post utility of the agent is:

$$u\left(g\left(t,s\right),t,s\right)$$

With this in mind, we can define mechanisms that implement truth-telling as follows.

[20] The notion of maximal best response in games with incomplete preferences, and the corresponding notion of Nash equilibrium, was introduced by Shapley (1959) and Aumann (1962).

Definition 16 *A social choice function ψ is* truthfully implementable in maximal strategies *if there exists a mechanism g such that truth-telling is a maximal strategy in g and $g(t,s) = \psi(t,s)$ for each $t \in T$ and all $s \in S$.*

Equivalently, ψ is truthfully implementable in maximal strategies if for each $t \in T$ there exists no $\sigma'(t) \in \Delta(B)$ such that

$$E_\pi \left[E_{\sigma'(t)} \left[u\left(g\left(b,s\right),t,s\right)\right]\right] > E_\pi \left[u\left(g\left(t,s\right),t,s\right)\right] \qquad \text{for all } \pi \in \Pi\left(t\right) \tag{7}$$

Definition 17 *A social choice function ψ is* truthfully implementable in optimal strategies *if there exists a mechanism g such that truth-telling is an optimal strategy in g and $g(t,s) = \psi(t,s)$ for each $t \in T$ and all $s \in S$.*

Thus if ψ is truthfully implementable in optimal strategies, then for each $t \in T$ and for all $\sigma'(t) \in \Delta(B)$

$$E_\pi \left[u\left(g\left(t,s\right),t,s\right)\right] \geq E_\pi \left[E_{\sigma'(t)} \left[u\left(g\left(b,s\right),t,s\right)\right]\right] \qquad \text{for all } \pi \in \Pi\left(t\right). \tag{8}$$

With these formalities in place, a version of the revelation principle follows.

Proposition 1 (The Revelation Principle) *If a social choice function ψ can be implemented in maximal (optimal) strategies by a mechanism g, then ψ is also truthfully implementable in maximal (optimal) strategies.*

Proof We prove the result for the case of maximal strategies; for optimal strategies the argument is analogous. By assumption, there exists a maximal pure strategy β such that $g(\beta(t), s) = \psi(t,s)$. In particular, for each $t \in T$ there is no $\sigma'(t) \in \Delta(B)$ such that

$$E_\pi \left[E_{\sigma'(t)} \left[u\left(g\left(b,s\right),t,s\right)\right]\right] > E_\pi \left[u\left(g\left(\beta(t),s\right),t,s\right)\right] \quad \text{for all } \pi \in \Pi\left(t\right).$$

But $u(g(\beta(t),s),t,s) = u(\psi(t,s),t)$, which implies by definition that ψ is truthfully implementable. ■

References

Ahn, D. (2005): "Hierarchies of Ambiguous Beliefs," Discussion paper, University of California, Berkeley.

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

Anderson, R. M., and W. R. Zame (2001): "Genericity with Infinitely Many Parameters," *Advances in Theoretical Economics*, 1, Article 1.

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.

Bergemann, D., and S. Morris (2005): "Robust Mechanism Design," *Econometrica*, 73, 1771–1813.

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

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

Bikhchandani, S., S. Chatterji, R. Lavi, A. Mu'alem, N. Nisan, and A. Sen (2006): "Weak Monotonicity Characterizes Deterministic Dominant-Strategy Implementation," *Econometrica*, 74, 1109–1132.

Christensen, J. P. R. (1974): *Topology and Borel Structure*. Amsterdam: North Holland.

Chung, K.-S., and J. Ely (2007): "Foundations of Dominant Strategy Mechanisms," *Review of Economic Studies*, 74, 447–476.

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 (2004): "Expected Utility Theory without the Completeness Axiom," *Journal of Economic Theory*, 115, 118–133.

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

GHIRARDATO, P., F. MACCHERONI, AND M. MARINACCI (2004): "Differentiating Ambiguity and Ambiguity Attitude," *Journal of Economic Theory*, 118, 133–173.

GHIRARDATO, P., F. MACCHERONI, M. MARINACCI, AND M. SINISCALCHI (2003): "A Subjective Spin on Roulette Wheels," *Econometrica*, 71, 1897–1908.

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

GIROTTO, B., AND S. HOLZER (2005): "Representation of Subjective Preferences Under Ambiguity," *Journal of Mathematical Psychology*, 49, 372–382.

HARSANYI, J. C. (1967): "Games with Incomplete Information Played by 'Bayesian' Players," *Management Science*, 14, 159–182 320–334 486–502.

HEIFETZ, A., AND Z. NEEMAN (2006): "On the Generic Impossibility of Full Surplus Extraction in Mechanism Design," *Econometrica*, 74, 213–233.

HUNT, B., T. SAUER, AND J. YORKE (1992): "Prevalence: A Translation Invariant 'Almost Every' on Infinite Dimensional Spaces," *Bulletin (New Series) of the American Mathematical Society*, 27, 217–238.

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

KRISHNA, V. (2002): *Auction Theory*. Academic Press.

LEDYARD, J. O. (1978): "Incentive compatibility and Incomplete Information," *Journal of Economic Theory*, 18, 171–189.

——— (1979): "Dominant Strategy Mechanisms and Incomplete Information," in *Aggregation and Revelation of Preferences,*, ed. by J.-J. Laffont. Amsterdam: North-Holland, chap. 1979.

LOPOMO, G. (1998): "The English Auction is Optimal among Simple Sequential Auctions," *Journal of Economic Theory*, 82, 144–166.

——— (2000): "Optimality and Robustness of the English Auction," *Games and Economic Behavior*, 36, 219–240.

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

MILGROM, P., AND I. SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

NEEMAN, Z. (2004): "The Relevance of Private Information in Mechanism Design," *Journal of Economic Theory*, 117, 55–77.

OK, E. A. (2002): "Utility Representation of an Incomplete Preference Relation," *Journal of Economic Theory*, 104, 429–449.

Rigotti, L., and C. Shannon (2005): "Uncertainty and Risk in Financial Markets," *Econometrica*, 73, 203–243.

Shapley, L. S. (1959): "Equilibrium Points in Games with Vector Payoffs," *Naval Research Logistics Quarterly*, 6, 57–61.

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

Wilson, R. (1987): "Game-Theoretic Analyses of Trading Processes," in *Advances in Economic Theory: Fifth World Congress*, ed. by T. Bewley. Cambridge: Cambridge University Press.