

Collegio Carlo Alberto



## A Robustly Efficient Auction

Kyungmin Kim

Antonio Penta

**No. 248**

**April 2012**

# Carlo Alberto Notebooks

[www.carloalberto.org/research/working-papers](http://www.carloalberto.org/research/working-papers)

# A Robustly Efficient Auction\*

Kyungmin Kim<sup>†</sup>

*University of Iowa*

Antonio Penta<sup>‡</sup>

*University of Wisconsin-Madison*

April 5, 2012

## Abstract

We study the problem of efficient auction design in environments with interdependent values, under arbitrary common knowledge assumptions. We propose a simple mechanism and show that, under a rather mild condition, it ‘robustly’ achieves efficiency. Our mechanism consists in a standard Vickrey auction, preceded by one round of communication, where agents report their private signals and receive transfers from the designer. We interpret the transfers as the cost for the designer to robustly achieve efficiency. We introduce a notion of *robust informational size* and show that the transfers are small if agents are informationally small in our sense. Furthermore, the transfers are decreasing in the amount of information available to the designer and in the strength of the common knowledge assumptions. In other words, the more robust the efficient implementation result, the higher the cost of achieving efficiency. We thus formalize the intuitive idea of a trade-off between robustness and efficient implementation and analyze the determinants of the ‘*cost of robustness*’.

**Keywords:** Cost of robustness; efficient auctions; informational size; interdependent values; robust mechanism design

---

\*We are especially grateful for the helpful comments of Alia Gizatulina, Andrew Postlewaite, Marzena Rostek and Bill Sandholm. We also thank seminar audiences at the Toulouse School of Economics, Collegio Carlo Alberto, University of Bonn, University of Wisconsin-Milwaukee, University of Michigan, University of Arizona, University of Wisconsin-Madison and University of Toronto. Antonio Penta is grateful for the hospitality of the Collegio Carlo Alberto (Turin).

<sup>†</sup>E-mail: kyungmin-kim@uiowa.edu.

<sup>‡</sup>E-mail: apenta@ssc.wisc.edu

“Game theory has a great advantage in explicitly analyzing the consequences of trading rules that presumably are really common knowledge; it is deficient to the extent it assumes other features to be common knowledge, such as one agent’s probability assessment about another’s preferences or information. [...] I foresee the progress of game theory as depending on successive reduction in the base of common knowledge required to conduct useful analyses of practical problems. Only by repeated weakening of common knowledge assumptions will the theory approximate reality.” (Wilson, 1987)

“In the present state of the art, academic mechanism design theory relies on stark and exaggerated assumptions to reach theoretical conclusions that can sometimes be fragile. [...] Mechanisms that are optimized to perform well when the assumptions are exactly true, may still fail miserably in the much more frequent cases when the assumptions are untrue. Useful real-life mechanisms need to be robust. Those that are too fragile should be discarded, whereas a robust mechanism can sometimes be confidently adopted even if, in the corresponding mechanism design model, it is not provably optimal.” (Milgrom, 2004, p. 22-23).

## 1 Introduction

The so called *Wilson doctrine* holds that practical mechanisms should be simple and should not rely on the strong common knowledge assumptions often implicit in the classical theory of mechanism design. This view provides the basic motivation for the recent development of the theory of ‘robust’ mechanism design.<sup>1</sup>

Although the robustness of a mechanism is a central concern, it is only one of several desiderata in ‘practical’ mechanism design. A key difficulty in the ‘art’, as opposed to the ‘theory’, of mechanism design consists precisely in finding a balance among the various properties that a mechanism should have. A real-world designer faces the problem of essentially choosing *how robust* a mechanism should be, vis-à-vis the other desiderata dictated by the underlying economic problem. The existing literature, however, provides no guidance to inform this choice of the practical designer. This paper takes a first step towards coping with this limitation of the theory, in the context of efficient auction design with interdependent values.

Our main point of departure from the existing literature is that we allow arbitrary common knowledge restrictions on agents’ preferences and beliefs. The strength of these restrictions qualifies the extent of the robustness requirement: the weaker the maintained common

---

<sup>1</sup>For a comprehensive account of the literature, see Bergemann and Morris (2012) and references therein (see also footnote 8). In the context of game theory, recent work that can be ascribed to the more broadly defined *Wilson doctrine* includes Weinstein and Yildiz (2007, 2012) and Penta (2012a,b).

knowledge assumptions, the more demanding the robustness requirement. We propose a simple mechanism and show that, under a rather mild condition, it robustly achieves efficiency. We then examine the performance of this mechanism as we vary the robustness requirements. We provide an upper bound to the designer’s cost of achieving efficiency and show that this bound increases as the robustness requirements get more demanding. This bound thus captures a trade-off between robustness and the cost of achieving efficiency, enabling us to isolate the determinants of the ‘cost of robustness’.<sup>2</sup>

Specifically, we consider the classical ‘mineral rights’ auction environment. Each agent receives a private signal (e.g. privately observed outcomes of a probing test) about a common unobservable payoff-relevant state (e.g. the quantity of oil). This problem is of both practical and theoretical interest. It has served as a workhorse to study auctions of various natural resources, such as oil-drilling rights. Theoretically, it is the environment that most naturally gives rise to both interdependence among agents’ values and correlations among their signals.

In the classical approach, the joint distribution of payoff states and agents’ signals as well as agents’ valuation functions (their willingness to pay as a function of agents’ signals) are assumed common knowledge, and known to the designer. In this environment, Cremer and McLean (1985, 1988) show that efficiency can be achieved granting no informational rents to the agents.<sup>3</sup> This result obtains through a properly designed scheme of side lotteries, which exploits the designer’s information about the agents’ beliefs. This result has been subject to a number of criticisms.<sup>4</sup> The most compelling criticisms, however, point to the fact that the result relies too heavily on the stark and unrealistic common knowledge assumptions implicit in the standard model.<sup>5</sup>

We depart from the classical approach in a number of ways. On a conceptual level, we

---

<sup>2</sup>Two recent studies are closest to ours in spirit. Similar to our approach, Artemov, Kunimoto and Serrano (2011) also consider robustness maintaining some restrictions on agents’ beliefs, but focus on virtual implementation. Yamashita (2012) formalizes the idea of the cost of robustness, comparing a ‘robust’ environment with a Bayesian benchmark. However, he does not investigate the determinants of this cost in terms of the underlying robustness requirements, whereas it is our main focus.

<sup>3</sup>See also McAfee, McMillan and Reny (1989) and McAfee and Reny (1992).

<sup>4</sup>For instance, if correlations among agents’ signals are sufficiently small, then the payments involved in the side-lotteries can be arbitrarily large. This property is undesirable in the presence of limited-liability constraints or risk aversion (McLean and Postlewaite, 2004). Furthermore, the optimal mechanism is not ‘detail free’ in the sense that it depends on the agents’ valuation functions (Perry and Reny, 2002). Finally, the full-surplus extraction result hinges upon a ‘belief-determine-preference’ condition that may not be satisfied in ‘rich’ type spaces (Neeman, 2004).

<sup>5</sup>These criticisms can be found in the very papers that establish the result: “Although the paper develops tools for solving mechanism design problems with correlated information, the results (full rent extraction) cast doubt on the value of the current mechanism design paradigm as a model of institutional design.” (McAfee and Reny, 1992, p. 400). “[...] The assumption that a common knowledge probability distribution  $\pi$  exists is very strong. Though economic theorists have found this assumption convenient [...], little discussion has been devoted to its ramifications for ‘real life’ problems of mechanism design.” (Cremer and McLean, 1988, p. 1254).

model agents' preferences, beliefs and information (all conflated in the classical notion of 'type') separately. This way, we can be explicit about the meaning and the strength of the common knowledge assumptions that we make, which will be key in investigating the performance of our mechanism as the robustness requirements vary. On a more technical level, our model differs from the classical framework mainly in two respects. First, we do not assume that agents' valuation functions are common knowledge. Second, we assume that agents only share common knowledge of *a set* of priors over the signals and payoff states (as opposed to a singleton). Different agents may entertain different subjective priors within this set, and such beliefs are private information to each agent.

From an applied viewpoint, allowing a set of priors is important to accommodate a wide variety of situations that cannot be cast within the classical framework. For instance, in a standard mineral rights problem, it is often the case that the engineers that inform firms about the probing technology do not express their assessments as a single joint probability distribution. Rather, they provide confidence intervals on the reliability of each signal. Furthermore, even if the consultants express their assessment via a single probability distribution, it may be that different consultants disagree on the reliability of the test. If the opinion of a firm's consultant is private information to the firm, it may be difficult to accommodate this situation in the classical 'single prior' framework.<sup>6</sup> Alternatively, agents may be uncertain about the reliability of the sampling technology. This may be the case, for instance, if a new yet unknown technology is introduced and different firms disagree on its informativeness. Such uncertainty may not be represented by a single commonly known prior. Finally, suppose that different probing technologies are available to different competing firms and each firm has private information over its technology. The single prior model would be inadequate to represent a situation in which firms are uncertain about the sampling technologies of their competitors.

Our first result shows that, under a rather mild condition, a simple modification of the Vickrey (second-price) auction guarantees efficiency. This condition requires that the agents share 'sufficient agreement' about the relative quality of the different signals. We develop a simple measure that captures this idea, *adjusted variability of beliefs*, and show that whenever it is positive, efficiency is achievable.

Our mechanism consists in introducing, prior to the Vickrey auction, one round of communication in which bidders publicly announce their private signals (e.g., the outcomes of the probing test). Based on these reports, agents receive monetary transfers from the de-

---

<sup>6</sup>This is the case, unless one is willing to make the somewhat bold assumption that the firms agree on the probability distribution that matches each firm to her consultant, and each consultant's opinion is common knowledge.

signer. Intuitively, if agents' signals are revealed truthfully, then the problem reduces to one of private values, in which the Vickrey auction guarantees efficiency. Essentially the same mechanism is adopted by McLean and Postlewaite (2004), who extend the techniques of Cremer and McLean (1985, 1988) to the mineral rights auction environment. The novelty of our result is its *robustness*. Efficiency is guaranteed without the designer's having fine information or imposing strong common knowledge assumptions among the agents. We thus show that the techniques of Cremer and McLean can be extended to 'robust' settings, despite the fact that the designer does not possess precise information about agents' beliefs.

We introduce next a notion of *robust informational size*, a measure that summarizes the amount of agents' private information, taking into account the limited information that the designer has about agents' beliefs. We show (Proposition 3) that the transfers in the first stage of our mechanism, hence the agents' informational rents, are related to the agents' informational size in an intuitive manner: the transfers are small if agents are informationally small in our sense. To the best of our knowledge, this is the first measure of informational size in settings in which the prior distribution of agents' types is not common knowledge.<sup>7</sup>

The payments to the agents in the first round of our mechanism have one natural interpretation. They can be interpreted as the cost borne by the designer to 'robustly' achieve efficiency in the presence of interdependence in valuations. When agents' informational size is zero, no payments are necessary and efficiency is achieved by the standard second price auction. Payments increase as agents' informational size increases. Furthermore, informational size (weakly) increases as the robustness requirements become more demanding. This comparative statics is particularly interesting because it indirectly uncovers the incremental cost of increased robustness. We explore these issues in Section 6. Proposition 4 provides explicit upper bounds to the transfers in the first stage of our mechanism. We show that the bounds are monotonically decreasing in the amount of information available to the designer or in the amount of 'common knowledge' shared by the agents, both about preferences and about beliefs. Hence, the bounds do uncover a trade-off between the robustness of the implementation result and the designer's cost of achieving efficiency.

**Other Related Literature.** Our approach to robustness differs from the others that have been put forward. For instance, Dasgupta and Maskin (2000) and Perry and Reny (2002) present ex-post incentive compatible and 'detail-free' efficient auctions. Their auctions are

---

<sup>7</sup>In related work, Gizatulina and Hellwig (2010) show that, in an environment that does not satisfy the 'belief-determine-preference' (BDP) condition (Neeman, 2004), techniques analogous to Cremer and McLean's could still be used if agents are 'informationally small'. However, they do not provide a definition of informational size, pointing at the difficulty of doing it in environments that do not satisfy the BDP condition. Our environment does not satisfy that condition.

detail-free in the sense that they require neither the designer’s knowledge of agents’ beliefs nor of their valuation functions. However, their efficiency properties crucially rely on strong common knowledge assumptions *among the agents*. Our reading of the ‘Wilson doctrine’ is that not only is it desirable to weaken the assumptions on the information available to the designer, but also the reliance of the results on common knowledge among the agents. From this viewpoint, our notion of ‘robustness’ is more demanding than those of these authors.

On the other hand, our framework allows the designer to have *some* information about agents’ beliefs. This feature differentiates us from the recent literature on robust mechanism design, which has focused on the extreme case in which the designer has no information about such beliefs.<sup>8</sup> These studies (and, more generally, ex-post incentive compatible mechanisms) are often criticized on the ground that the notion of robustness is too restrictive.<sup>9</sup> Our model nests both the belief-free and the Bayesian settings as special cases, but it also accommodates intermediate situations in which some restrictions on agents’ beliefs are maintained. As a consequence, a weaker notion than ex-post incentive compatibility in general suffices for the results.

The main departure from the literature, however, regards the spirit of our results. A typical paper on auctions or mechanism design provides an analysis of a given model (hence, for *given* common knowledge assumptions), whereas our aim is to inform the modeling choices of the designer, particularly regarding the strength of the robustness requirement. The idea of a trade-off between robustness and the extent by which certain goals can be achieved is intuitive. Nonetheless, the existing literature is essentially silent on the issue. Exploring the nature of this trade-off inherently involves varying the common knowledge assumptions. To the best of our knowledge, this is the first paper to pursue this kind of exercise.

In seeking to gain some economic insights about the determinants of the cost of robustness, we adopt a non-optimal mechanism. In this regard, besides the opening quote by Milgrom (2004), our approach is close to a recent literature on ‘approximate mechanism design’, at the boundary between the economics and computer science literatures.<sup>10</sup> The typical result in this literature is that a simple ‘non-optimal’ mechanism guarantees a certain fraction of an upper bound to the natural objective of the designer (e.g., revenue maximization or efficiency). Our results can be interpreted in a similar way, except that

---

<sup>8</sup>Bergemann and Morris (2005, 2009a, 2009b, 2011) develop a “belief-free” approach. Mueller (2011) and Penta (2011) pursued it in dynamic settings. Alternative approaches have been proposed by Artemov, Kunimoto and Serrano (2011) and Chung and Ely (2007). Chung and Ely (2003), Aghion, Fudenberg, Holden, Kunimoto and Tercieux (2012) and Oury and Tercieux (2012) instead explore properties of ‘local robustness’, considering arbitrarily small perturbations of agents’ beliefs.

<sup>9</sup>See, e.g., Jehiel, Meyer-ter-Vehn, Moldovanu, and Zame (2006)

<sup>10</sup>See Hartline (2011) and references therein. For a similar ‘*approximately optimal*’ approach in the context of contract theory, see Chassang (2011).

we provide an upper bound on the absolute value of the designer’s loss (measured by the payments in the first round), as opposed to a fraction. In many cases, the absolute value is a better measure of performance than a fraction. This is especially the case if the benchmark is only a theoretical concept, not a known quantity. Furthermore, we relate this bound to the fundamentals of the economic environment, studying how the bound varies as structural features of the model change.

**Structure of the paper.** The rest of the paper is organized as follows: Section 2 introduces the economic environment; Section 3 presents the efficient design problem and the notion of ‘robust’ incentive compatibility (Proposition 1); Section 4 introduces our mechanism and the notion of *variability of beliefs*, and then provides the efficient implementation result (Proposition 2). Section 5 presents the notion of *informational size* and our main result (Proposition 3). Section 6 investigates the determinants of the trade-off between robustness and efficiency. Section 7 concludes.

## 2 The Model

### 2.1 Primitives of the Environment

We consider a standard auction problem in which a finite number of agents compete for the allocation of one unit of an indivisible object. Let  $N = \{1, \dots, n\}$  denote the set of agents. Agents’ valuations for the object depend on some physical characteristics of the good, referred to as a *payoff state* hereafter. Let  $\Theta_0 = \{\theta^1, \dots, \theta^m\} \subseteq \mathbb{R}$  denote the set of payoff states, and  $u_i : \Theta_0 \rightarrow \mathbb{R}$  represent agent  $i$ ’s preferences, where  $u_i(\theta_0)$  is agent  $i$ ’s willingness-to-pay for the object when the payoff state is  $\theta_0$ . The characteristics of the good are not directly observable by the agents, but they receive private noisy signals. Assume that the set of signals for agent  $i$  is given by  $S_i = \{s_{i,1}, s_{i,2}, \dots, s_{i,n_i}\}$ . Let  $S \equiv \times_{i \in N} S_i$  denote the set of signal profiles and  $S_{-i} \equiv \times_{j \neq i} S_j$  the set of signal profiles of agents other than  $i$ . These signals are generated according to some distribution  $p \in \Delta(\Theta_0 \times S)$ .<sup>11</sup> To the extent that each agent’s signal is informative about the underlying characteristics, this setting naturally gives rise to interdependence in agents’ valuations.

The prototypical example is the ‘mineral rights’ auction problem: wildcatters competing for drilling rights on an oil field. The quantity of oil in the field is the common unobservable component ( $\theta_0$ ), and signals ( $s_i$ ) are the outcomes of some probing technology,  $p$ . Firms’ willingness-to-pay may differ across firms ( $u_i \neq u_j$ ), due to heterogeneity in drilling costs, in

---

<sup>11</sup>Throughout the paper,  $\Delta(X)$  denotes the set of probability measures on  $X$ .

their positions in the market, etc.

## 2.2 Modeling Assumptions

In the classical approach, the distribution of the signals and agents' valuation functions are assumed common knowledge and known to the designer (see subsection 2.3). We depart from the traditional approach in both respects: agents do not know the exact prior that generates the signals, but do know that it belongs to a commonly known *set* of priors; agents' valuation functions are neither known to the designer nor common knowledge among the agents.

Let  $\Pi \subseteq \Delta(\Theta_0 \times S)$  be a set of priors, and for each  $i \in N$ , let  $U_i$  be a set of strictly increasing functions  $u_i : \Theta_0 \rightarrow \mathbb{R}$ . In our model, an *environment* is described by a tuple

$$\mathcal{E} = \langle N, \Theta_0, \Pi, (S_i, U_i)_{i \in N} \rangle,$$

and we assume that  $\mathcal{E}$  is common knowledge among the agents and known to the designer.<sup>12</sup>

The set  $\Pi$  represents agents' information about the signal generating process. We assume that it is common knowledge that the signals are generated according to some distribution that belongs to the set  $\Pi$ , but agents do not know exactly which distribution it is. Each agent  $i$  entertains subjective beliefs, denoted by  $p_i \in \Pi$ . Importantly, we make the natural assumption that agent  $i$ 's beliefs  $p_i$  are his private information.

The sets  $(U_i)_{i \in N}$  represent the information about agents' preferences,  $u_i : \Theta_0 \rightarrow \mathbb{R}$ . We assume that it is common knowledge that  $u_i$  belongs to  $U_i$ , but  $u_i$  is also private information to agent  $i$ . We also define the sets  $U \equiv \times_{i \in N} U_i$ ,  $U_{-i} \equiv \times_{j \neq i} U_j$ ,  $U^* = \bigcup_{i \in N} U_i$  and  $U_{-i}^* = \bigcup_{j \neq i} U_j$ .

We make the following technical assumptions: The elements of the set  $\Theta_0 = \{\theta^1, \dots, \theta^m\}$  are ordered so that  $\theta^k < \theta^{k+1}$  for each  $k$ .<sup>13</sup> We normalize the value of the agents' outside option to zero, and assume that there are always gains from trade:  $\min_{u_i \in U_i} u_i(\theta^1) > 0$  for each  $i$ . Finally, we assume that the sets  $\Pi$  and  $(U_i)_{i \in N}$  are compact.

This model can be interpreted as one in which agents have imprecise information about the signal-generating process. This formulation raises the question of what beliefs each agent would have and how to compute his *expected valuation*. In general, expected valuation functions might depend on agents' attitudes toward ambiguity.<sup>14</sup> Given our focus on robustness,

---

<sup>12</sup>In our baseline model, the tuple  $\mathcal{E}$  is the only common knowledge available to the agents. In Section 6 we consider alternative settings, in which agents may share more common knowledge.

<sup>13</sup>The finiteness of the set  $\Theta_0$  is assumed only for simplicity. It can be relaxed at the expense of additional technicalities. Also the one-dimensionality of the payoff state is not crucial. What is necessary is that agents agree about the relative desirability of any pair of states  $\theta_0, \theta'_0$ .

<sup>14</sup>See, e.g., Gilboa (2010) and references therein.

we abstract from the issues related to ambiguity and maintain the standard assumption that every agent  $i$  is an expected utility maximizer, with respect to his subjective beliefs  $p_i$ . This implies that given  $s_i$ , agent  $i$  believes that other agents' signals are  $s_{-i}$  with probability  $p_i(s_{-i}|s_i)$ . In addition, given  $s \in S$  and  $u_i \in U_i$ , agent  $i$ 's expected valuation is defined as

$$v_i(s, u_i, p_i) = \sum_{\theta_0 \in \Theta_0} u_i(\theta_0) \cdot p_i(\theta_0|s). \quad (1)$$

Each agent has three pieces of private information that are directly relevant to his valuation: a signal,  $s_i \in S_i$ , representing his information about the payoff state, his preferences  $u_i$ , and his beliefs  $p_i$ . We refer to the triple  $(s_i, u_i, p_i)$  as agent  $i$ 's *payoff type* and denote the set of  $i$ 's payoff types as  $\Theta_i = S_i \times U_i \times \Pi$ , with generic element  $\theta_i = (s_i, u_i, p_i)$ . Also, we let  $\Theta = \times_{i \in N} \Theta_i$  (note that  $\Theta$  does not include the unobservable component  $\Theta_0$ .)

We note that, unlike in the literature on robust mechanism design (e.g., Bergemann and Morris (2005)), our payoff types do entail some restrictions on  $i$ 's beliefs. However, payoff types do not coincide with the notion of *type* in a Bayesian game. This is because  $p_i$  does not describe  $i$ 's beliefs about the opponents' beliefs  $p_{-i}$ , nor about the opponents' preferences  $u_{-i}$ . The only restriction that our notion of environment imposes on agent  $i$ 's beliefs about  $(p_{-i}, u_{-i})$  is that they must be concentrated on  $(p_{-i}, u_{-i}) \in \Pi^{n-1} \times U_{-i}$ . Agents would have subjective beliefs about  $(p_{-i}, u_{-i})$  and the opponents' higher order beliefs, but they are unknown to the designer and are not part of the environment. We discuss this point further in the next section.

## 2.3 Discussion

**Bayesian and belief-free mechanism design.** Our analytical framework is non-standard in several ways: given our notion of *environment*, an auction mechanism induces neither a Bayesian game, as in classical mechanism design or auction theory, nor a belief-free game, as in the literature on robust mechanism design. A belief-free environment is obtained in the special case in which  $\Pi$  coincides with the set of all possible distributions over  $\Theta_0 \times S$ , while our environment reduces to a standard Bayesian one if both  $\Pi$  and all the  $U_i$ 's are singletons.<sup>15</sup> In all other cases, the set  $\Pi$  imposes some restrictions on agents' beliefs, but not to the point where their infinite hierarchies of beliefs are pinned down by their private signals. However, our model is consistent with a Bayesian approach in which players do have fully specified hierarchies of beliefs, though unknown to the designer (see subsection 3.1).

---

<sup>15</sup>Note that  $\Pi = \{p\}$  alone does not suffice to obtain a Bayesian environment, because  $p \in \Delta(\Theta_0 \times S)$  does not pin down agents' beliefs about  $U$ .

**Comparison to the classical approach.** The classical model abstracts from the details of the baseline environment and introduces some non-primitive objects. In particular, the classical model specifies *types*  $t_i \in T_i$  for each agent, and valuation functions  $w_i : T \rightarrow \mathbb{R}$ . Finally, the model is closed by specifying a prior  $\pi^* \in \Delta(T)$ . The tuple  $\langle N, \pi^*, (T_i, w_i)_{i \in N} \rangle$  is assumed common knowledge. This model can be interpreted in two ways.

According to one interpretation, types are ‘signals’ about the quantity of oil, i.e.  $T_i \approx S_i$ . Since both  $\pi^*$  and  $w_i$ ’s are assumed common knowledge, the classical model implicitly assumes that agents’ preferences are common knowledge as well (that is, for every  $i$ , there exists  $u_i^*$  such that  $w_i(s) = v_i(s; u_i^*, \pi^*)$ ). Hence, the following primitives are common knowledge:  $\langle N, \Theta_0, \pi^*, (S_i, u_i^*)_{i \in N} \rangle$ . The classical model thus involves common knowledge of both the sampling technology and of agents’ preferences. Clearly, the latter assumption is very strong, whereas the former postulates that agents fully agree about the reliability of the sampling technology. As discussed in the introduction, this assumption may be violated in various settings. Those situations cannot be cast within the classical framework, but can be easily accommodated by our setting, where the set  $\Pi$  describes the agents’ coarse agreement on the sampling technology.

The classical framework can accommodate lack of common knowledge of agents’ preferences, but only provided that ‘types’ are interpreted in a different way. Let  $T_i \approx S_i \times U_i$ , so that  $t_i = (s_i, u_i) \in T_i$  represents both the private signal and the preferences of agent  $i$ . Letting  $w_i(t_i, t_{-i}) = v_i(s_i, s_{-i}; u_i, \pi^*)$  for any  $t = (s, u)$ , the classical model entails common knowledge of the following objects:  $\langle N, \Theta_0, \pi^*, (S_i, U_i)_{i \in N} \rangle$ . In this case, however, the common prior  $\pi^*$  does not represent agents’ agreement on the sampling technology, but their agreement over the *joint* distribution of signals and preferences. This stark assumption may often be a useful simplification. But, unlike the case in which the prior represents information about the sampling technology, it is quite difficult to imagine plausible situations in which the joint distribution of preferences and information is really common knowledge. Furthermore, there are serious theoretical concerns about the usefulness of this simplification when modeling problems of institutional design (e.g., see Cremer and McLean (1988) and McAfee and Reny (1999), quoted in footnote 4.)

### 3 Efficiency, Incentive Compatibility and Robustness

We consider environments with transferable utility. A *selling mechanism* in these environments is a tuple  $((M_i, x_i)_{i \in N}, f)$  where  $M_i$  is the set of agent  $i$ ’s messages (or reports),  $(x_i)_{i \in N}$  is a *payment scheme* ( $x_i : M \rightarrow \mathbb{R}$  specifies the payment from agent  $i$ ), and  $f : M \rightarrow \Delta(N)$  is the *allocation function* which assigns an allocation (a lottery over the agents) to each

message profile. A selling mechanism is ‘ $\mathcal{E}$ -direct’ if  $M_i = \Theta_i$  for every  $i \in N$ . In that case, we omit the specification of the message space, and simply denote a selling mechanism by  $((x_i)_{i \in N}, q)$ , where  $q : \Theta \rightarrow \Delta(N)$  is the allocation function.

We consider  $\mathcal{E}$ -direct mechanisms and define ‘robust implementation’ in terms of a special notion of incentive compatibility. Proposition 1 in subsection 3.1 will show that this notion of robustness is consistent with a Bayesian foundation, and that the restriction to  $\mathcal{E}$ -direct mechanisms entails no loss of generality.

An *allocation rule* is a function mapping from the set of payoff types to lotteries over the agents,  $q : \Theta \rightarrow \Delta(N)$ . For each  $\theta$ , let  $q_i(\theta)$  denote the probability that agent  $i$  obtains the object. We say that an allocation rule is *efficient* if it always assigns the object to the agent with the highest willingness-to-pay for the object:

**Definition 1** *An allocation rule  $q$  is efficient if for every  $(s, u, p) \in \Theta$ ,  $q_i(s, u, p) > 0$  implies that  $i \in \arg \max_{j \in N} v_j(s, u_j, p_j)$ .*

Notice that this is an ‘ex-post’ concept in the sense that it is defined with respect to the pooled information available to the agents. It is not ‘ex-post’ in the stronger sense of conditioning on the realization of  $\theta_0$ . Since  $\theta_0$  is never observed in our setup, Definition 1 is the appropriate ex-post concept. The allocation rule clearly depends on agents’ signals,  $s$ , their preferences,  $u$ , as well as their beliefs,  $p$ .

We now introduce a notion of incentive compatibility that captures our concerns for robustness. Since the designer has only limited information about the agents’ beliefs, we will require a mechanism to achieve efficiency for all possible beliefs the agents may entertain.

Intuitively, the environment  $\mathcal{E}$  contains some information about agents’ beliefs about the payoff states and the opponents’ signals, namely, that for each  $s_i$ , agent  $i$ ’s beliefs about  $(\theta_0, s_{-i})$  would be equal to  $p_i(\theta_0, s_i | s_{-i})$  for some  $p_i \in \Pi$ . Not knowing the agent’s private signal  $s_i$  and prior  $p_i$ , we require incentive compatibility to hold for all beliefs  $(p_i(\theta_0, s_i | s_{-i}))_{p_i \in \Pi, s_i \in S_i}$ . In addition, the environment  $\mathcal{E}$  does not contain any information about agents’ beliefs about the opponents’ preferences and beliefs  $(u_{-i}, p_{-i})$ , except that they belong to the sets  $U_{-i} \times \Pi^{n-1}$ . Consequently, we also require  $i$ ’s incentive compatibility to be satisfied for all such  $(p_{-i}, u_{-i}) \in \Pi_{-i} \times U_{-i}$ . Proposition 1 in Section 3.1 proves that this is indeed the relevant notion of incentive compatibility in our setting.

**Definition 2** *An  $\mathcal{E}$ -direct selling mechanism  $((x_i)_{i \in N}, q)$  is ‘ $\mathcal{E}$ -incentive compatible’ ( $\mathcal{E}$ -IC)*

if for any  $i \in N$ ,  $\theta_i, \theta'_i \in \Theta_i$  and any  $(p_{-i}, u_{-i}) \in \Pi_{-i} \times U_{-i}$ ,

$$\begin{aligned} & \sum_{\theta_0} \sum_{s_{-i}} [q_i(\theta_i, s_{-i}, p_{-i}, u_{-i}) \cdot u_i(\theta_0) - x_i(\theta_i, s_{-i}, p_{-i}, u_{-i})] \cdot p_i(\theta_0, s_{-i} | s_i) \\ \geq & \sum_{\theta_0} \sum_{s_{-i}} [q_i(\theta'_i, s_{-i}, p_{-i}, u_{-i}) \cdot u_i(\theta_0) - x_i(\theta'_i, s_{-i}, p_{-i}, u_{-i})] \cdot p_i(\theta_0, s_{-i} | s_i) \end{aligned} \quad (2)$$

$\mathcal{E}$ -Incentive Compatibility is a hybrid of interim and ex-post incentive compatibility: while  $\mathcal{E}$ -IC is ‘ex-post’ with respect to  $(u_{-i}, p_{-i})$ , this definition is weaker than ex-post incentive compatibility, because (2) is not required to hold for all  $s_{-i} \in S_{-i}$ . The hybrid nature of  $\mathcal{E}$ -IC mirrors the information contained in our notion of environment, which is neither a classical Bayesian nor a belief-free environment. Notice, however, that  $\mathcal{E}$ -IC coincides with interim incentive compatibility in the case were  $\Pi$  and  $(U_i)_{i \in N}$  are singletons, while it coincides with ex-post incentive compatibility if  $\Pi = \Delta(\Theta_0 \times S)$ .

We finally define our notion of robustness:

**Definition 3** *An  $\mathcal{E}$ -direct selling mechanism  $((x_i)_{i \in N}, q)$  is ‘robustly efficient’ if it is  $\mathcal{E}$ -IC and the allocation rule  $q$  is efficient.*

We now show that Definition 3 is consistent with an alternative view of ‘robustness’, which requires the classical *interim incentive compatibility* to be satisfied for all the hierarchies of beliefs that are consistent with common knowledge of the environment  $\mathcal{E}$ .

### 3.1 A Bayesian Foundation

Unlike standard Bayesian settings, our environment does not include a complete description of agents’ higher-order beliefs. However, our model is consistent with a Bayesian view, in which agents have well-defined hierarchies of beliefs, but those beliefs are unknown to the designer whose only information is represented by the tuple  $\mathcal{E}$ .

Following Harsanyi (1967-68), we model agents’ hierarchies of beliefs implicitly, by means of type spaces. Since the environment  $\mathcal{E}$  imposes restrictions on agents’ hierarchies of beliefs, we should also make sure that the hierarchies represented by the type space are consistent with the common knowledge assumptions specified by the environment  $\mathcal{E}$ .

**Definition 4** *An  $\mathcal{E}$ -consistent type space is a tuple  $\mathcal{T} = (T_i, \hat{\theta}_i, \tau_i)_{i \in I}$  such that  $T_i$  is a compact set of player  $i$ ’s types;  $\hat{\theta}_i (\equiv (\hat{s}_i, \hat{u}_i, \hat{p}_i)) : T_i \rightarrow S_i \times U_i \times \Pi$  is a measurable function assigning to each type a payoff type, and  $\tau_i : T_i \rightarrow \Delta(\Theta_0 \times T_{-i})$  is a belief function such*

that, for every  $t_i \in T_i$  and every  $(\theta_0, s_{-i}) \in \Theta_0 \times T_{-i}$ ,

$$\tau_i(t_i) [\{(\theta_0, t_{-i}) \in \Theta_0 \times T_{-i} : \hat{s}_{-i}(t_{-i}) = s_{-i}\}] = \hat{p}_i(t_i) [\theta_0, s_{-i} | \hat{s}_i(t_i)]. \quad (3)$$

Equation (3) is a consistency condition, which requires that type  $t_i$ 's beliefs about the opponents' types and states of nature (as specified in the type space) be consistent with  $t_i$ 's prior beliefs over the set  $\Theta_0 \times S$ . It can be shown that any hierarchy of beliefs consistent with the restrictions implicit in the environment  $\mathcal{E}$  can find an implicit representation as a type in an  $\mathcal{E}$ -consistent type space.

The fact that  $\mathcal{E}$  does not contain information on agents' higher order beliefs does not automatically imply that adopting  $\mathcal{E}$ -direct mechanisms is without loss of generality. Although unspecified in the environment, agents do entertain subjective beliefs about the opponents'  $(u_{-i}, p_{-i})$  as well as about their opponents' beliefs, of any order. In principle, the designer could adopt mechanisms that elicit such hierarchies of beliefs. We show next that, for the problem at hand, there is no need to consider such richer mechanisms.

Appending an  $\mathcal{E}$ -consistent type space  $\mathcal{T}$  to  $\mathcal{E}$  delivers a standard Bayesian environment,  $(\mathcal{T}, \mathcal{E})$ . In these environments, we can apply to the (Bayesian) revelation principle and restrict attention to ' $\mathcal{T}$ -direct' mechanisms, i.e. such that  $M_i = T_i$  for every  $i \in N$ .

**Definition 5** A  $\mathcal{T}$ -direct mechanism  $(f, x)$  is interim incentive compatible on  $\mathcal{T}$  if for any  $i$ , and  $t_i, t'_i \in T_i$ ,

$$\begin{aligned} & \int_{\theta_0, t_{-i}} [f_i(t_i, t_{-i}) \cdot \hat{u}_i(t_i)(\theta_0) - x_i(t_i, t_{-i})] d\tau_i(t_i) \\ & \geq \int_{\theta_0, t_{-i}} [f_i(t'_i, t_{-i}) \cdot \hat{u}_i(t_i)(\theta_0) - x_i(t'_i, t_{-i})] d\tau_i(t_i) \end{aligned}$$

**Proposition 1** A trading mechanism is  $\mathcal{E}$ -incentive compatible if and only if it is interim incentive compatible for all the  $\mathcal{E}$ -consistent type spaces.

**Proof.** See the Appendix. ■

Proposition 1 can be interpreted as the counterpart, in our non-belief-free environment, to Propositions 1 and 2 in Bergemann and Morris (2005).<sup>16</sup> It implies that a trading mechanism is interim incentive compatible for all  $\mathcal{E}$ -consistent type spaces if and only if it is  $\mathcal{E}$ -incentive compatible. Hence, the two notions of robustness (definition 3 and 'interim incentive compatibility for all  $\mathcal{E}$ -consistent type spaces) coincide.

---

<sup>16</sup>It can be shown that, for general implementation problems, an analogue of Proposition 1 holds for social choice correspondences in *separable environments*. The allocation rules in environments with transferable utility of this section are a special case of separable environments.

## 4 An Efficient Auction

### 4.1 Augmented Vickrey Auction with Conditional Payments

We employ a simple mechanism, which is essentially a Vickrey auction augmented by conditional payments. The mechanism consists of two stages: in the first stage, agents report their signals  $\tilde{s}_i$  and receive payments conditional on the reported signal profile,  $\tilde{s}$ . For each  $i$ , let a function  $z_i : S \rightarrow \mathbb{R}_+$  represent the conditional payment to agent  $i$ , where  $z_i(\tilde{s})$  is the transfer to agent  $i$  when the reported signals are  $\tilde{s}$ . In the second stage, the reported signals are made public and the Vickrey auction takes place. In the Vickrey auction, ties are broken by a fair coin toss. This mechanism, which is also employed by McLean and Postlewaite (2004), fully utilizes the robustness properties of the Vickrey auction in private values environments, supplementing it with conditional payments designed to solve the interdependent values problem. In our environment, each bidder's signal has a common value, but if the mechanism designer can collect and disseminate bidders' signals, then the problem reduces to one of private values, where the standard Vickrey auction ensures efficiency. Of course, bidders must be provided with an incentive to truthfully report their signals.

Observe that this mechanism is equivalent to the following ( $\mathcal{E}$ -direct) mechanism in which each agent reports his payoff type  $\tilde{\theta}_i = (\tilde{s}_i, \tilde{u}_i, \tilde{p}_i)$  and the following allocation rule is employed: for each  $\tilde{\theta} = (\tilde{s}, \tilde{u}, \tilde{p})$ , let

$$I(\tilde{\theta}) = \left\{ i \in N \mid v_i(\tilde{\theta}) = \max_j v_j(\tilde{\theta}) \right\},$$

where

$$v_j(\tilde{\theta}) = \sum_{\theta_0} \tilde{u}_j(\theta_0) \tilde{p}_j(\theta_0 | \tilde{s}).$$

In addition, let

$$\bar{w}_i(\tilde{\theta}) = \max_{j \neq i} v_j(\tilde{\theta}).$$

Then, the winning probability of agent  $i$  is determined so that

$$q_i^*(\tilde{\theta}) = \begin{cases} \frac{1}{|I(\tilde{\theta})|}, & \text{if } i \in I(\tilde{\theta}), \\ 0, & \text{if } i \notin I(\tilde{\theta}). \end{cases}$$

The payments scheme,  $x^* : \Theta \rightarrow \mathbb{R}^n$ , can be written as  $x^*(\theta) = x(\theta) - z(s)$ : the first term,  $x : \Theta \rightarrow \mathbb{R}_+^n$  corresponds to agents' payments in the Vickrey auction (that is,  $x_i(\tilde{\theta}) = q_i^*(\tilde{\theta}) \bar{w}_i(\tilde{\theta})$  for each  $i$ ); the second term corresponds to the conditional payments to the agents,  $z : S \rightarrow \mathbb{R}_+^n$ .

## 4.2 Adjusted Variability of Beliefs

The efficient robust implementation problem boils down to designing conditional payments  $z : S \rightarrow \mathbb{R}_+^N$  in an  $\mathcal{E}$ -incentive compatible manner. We extend the technique, originally employed by Cremer and McLean (1988), of taking advantage of correlations among agents' signals to our robust implementation problem. The technique is typically very sensitive to the introduction of higher order uncertainty, as entailed by our robustness requirement.<sup>17</sup> Nonetheless, we illustrate how the scheme of McLean and Postlewaite can be adapted to our setting and provide a sufficient condition under which agents truthfully report their signals, independently of their beliefs (Proposition 2).

To illustrate our construction, consider the single-prior case first. That is, suppose  $\Pi$  is a singleton set, including only  $p$ . Under a common prior, the mechanism designer's ability to provide each agent with an incentive to reveal his private information through conditional payments depends on the magnitude of the difference between the conditional distributions on  $S_{-i}$  given different signals. Precisely, set  $z_i : S \rightarrow \mathbb{R}$  so that

$$z_i(\tilde{s}) = \kappa \cdot \frac{p(\tilde{s}_{-i}|\tilde{s}_i)}{\|p(\tilde{s}_{-i}|\tilde{s}_i)\|_2} \quad (4)$$

for some  $\kappa > 0$  ( $\|\cdot\|$  and  $\|\cdot\|_2$  denote 1-norm and 2-norm, respectively).<sup>18</sup> Given bidder  $i$ 's signal  $s_i$ , his beliefs over the others' signals  $S_{-i}$  are given by  $p(s_{-i}|s_i)$ . Therefore, bidder  $i$ 's expected reward when he receives  $s_i$  but reports  $\tilde{s}_i$  is

$$\kappa \cdot \sum_{s_{-i} \in S_{-i}} \frac{p(s_{-i}|\tilde{s}_i)}{\|p(s_{-i}|\tilde{s}_i)\|_2} p(s_{-i}|s_i).$$

This expression is always maximized at  $\tilde{s}_i = s_i$  and uniquely maximized if  $p(\cdot|s_i)$  is different from  $p(\cdot|s'_i)$  for every  $s'_i \neq s_i$ . By adjusting  $\kappa$ , we can give the agents sufficient incentives to truthfully report their signals.

Based on this idea, McLean and Postlewaite (2002, 2004) introduce the notion of variability of beliefs and show that agents' signals can be extracted whenever variability of beliefs

---

<sup>17</sup>Our environments do not satisfy the "Belief-Determine-Preferences" condition of Neeman (2004).

<sup>18</sup>The  $p$ -norm of vector  $\mathbf{x} \in \mathbb{R}^n$  is defined as

$$\|\mathbf{x}\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}$$

is positive. Their definition of variability of beliefs is

$$\Lambda_i^{MP} \equiv \min_{s_i, s'_i \in S_i, s_i \neq s'_i} \left\| \frac{p(\cdot | s_i)}{\|p(\cdot | s_i)\|_2} - \frac{p(\cdot | s'_i)}{\|p(\cdot | s'_i)\|_2} \right\|_2^2. \quad (5)$$

The same scheme obviously cannot be used with a set of priors, as  $p_i$  is private information to agent  $i$ . This prevents the design of a mechanism that fine-tunes agents' incentives based on their beliefs. Nonetheless, the information contained in the set  $\Pi$  may enable us to design a mechanism that provides sufficient incentives for the bidders.

Fix  $p^* \in \Delta(\Theta_0 \times S)$  and suppose the same payment scheme as in the single-prior case is used according to this particular prior. That is, if the reported signal profile is  $\tilde{s}$ , then agent  $i$  receives  $z_i(\tilde{s}) = \kappa \frac{p^*(\tilde{s}_{-i} | \tilde{s}_i)}{\|p^*(\tilde{s}_{-i} | \tilde{s}_i)\|_2}$ . Suppose agent  $i$ 's belief is given by  $p_i \in \Pi$ . Then, agent  $i$ 's (subjective) expected reward by reporting  $\tilde{s}_i$  while his true signal is  $s_i$  is

$$\kappa \cdot \sum_{s_{-i} \in S_{-i}} \frac{p^*(s_{-i} | \tilde{s}_i)}{\|p^*(s_{-i} | \tilde{s}_i)\|_2} p_i(s_{-i} | s_i).$$

Unless  $p_i$  happens to coincide with  $p^*$ , this expected reward is not necessarily maximized at  $\tilde{s}_i = s_i$ . However, if  $p_i$  is 'not so different' from  $p^*$  and there is enough variation in  $p_i$  and  $p^*$  (that is,  $p_i(\cdot | s_i)$  and  $p^*(\cdot | s_i)$  change significantly depending on  $s_i$ ), then the expected payoff would be maximized at  $\tilde{s}_i = s_i$ .

Clearly, the possibility of ensuring that  $p_i$  is 'not so different' from the  $p^*$  adopted by the mechanism depends on the designer's information about agents' beliefs. As the set  $\Pi$  gets larger, the designer has less precise information about the agents' beliefs and, therefore, it becomes harder to guarantee that a specific  $p^*$  would work. Intuitively, the variability of beliefs (which, if positive, suffices for the result in the setting of McLean and Postlewaite) needs to be penalized (or adjusted) in order to reflect the designer's imprecise information about agents' beliefs. But, as long as there is enough variability of beliefs, it should still be possible to design a scheme of payments.

The key difficulty in formalizing this idea is how to identify the relevant measure. In particular, which information in  $\Pi$  would suffice for the designer to design such a scheme of payments. We propose the following notion of *adjusted variability of beliefs*:

---

<sup>19</sup>McLean and Postlewaite use different versions of the definition in their papers, but they are interchangeable. The definition in (5) corresponds to the one in their 2002 paper.

**Definition 6** For each  $i$ ,  $p \in \Delta(\Theta_0 \times S)$ , let

$$\Lambda_i(p) = \min_{s_i, s'_i \in S_i, s_i \neq s'_i} \min_{p_i \in \Pi} \left\{ \left\| \frac{p(\cdot|s'_i)}{\|p(\cdot|s'_i)\|_2} - \frac{p_i(\cdot|s_i)}{\|p_i(\cdot|s_i)\|_2} \right\|_2^2 - \left\| \frac{p(\cdot|s_i)}{\|p(\cdot|s_i)\|_2} - \frac{p_i(\cdot|s_i)}{\|p_i(\cdot|s_i)\|_2} \right\|_2^2 \right\}. \quad (6)$$

Agent  $i$ 's 'adjusted variability of belief' is defined as

$$\Lambda_i = \max_{p \in \Delta(\Theta_0 \times S)} \Lambda_i(p). \quad (7)$$

The first term in (6) is analogous to McLean and Postlewaite's variability of beliefs (5). The second term is a penalty, measured by the distance between the conditional distributions induced by  $p$  and  $p_i$ . Consider a designer who uses prior  $p$  to design the payments in (4). The min-operator taken over  $p_i \in \Pi$  (also over  $s_i, s'_i \in S_i$ , as in (5)) represents the worst-case scenario from the designer's viewpoint. Hence,  $\Lambda_i(p)$  represents a measure of the worst-case scenario in case prior  $p$  is used to design the payments in (4). Observe that, if  $\Pi = \{p\}$ , then (5) coincides with (6).

Equation (7) informs the optimal choice of  $p$ , relative to the worst-case scenario  $\Lambda_i(p)$ . Notice that this choice is not restricted to  $p \in \Pi$ : the optimal  $p$  may lie outside of  $\Pi$ . This is because, while  $p_i$  represents agent  $i$ 's beliefs, which must be in  $\Pi$ ,  $p$  is the designer's instrument and, therefore, can be chosen arbitrarily. Hence, it may be that  $\Lambda_i > 0$  even if  $\Lambda_i(p) < 0$  for all  $p \in \Pi$ .

**Proposition 2** If  $\Lambda_i > 0$  for all  $i$ , then it is possible to design a robustly efficient augmented Vickrey auction  $\{q_i^*, x_i^* - z_i^*\}_{i \in N}$  using the side payments  $(z_i^*)_{i \in N}$  obtained from (4) using priors  $p_i^* \in \arg \max_{p \in \Delta(\Theta_0 \times S)} \Lambda_i(p)$ .

**Proof.** See the Appendix. ■

The condition  $\Lambda_i > 0$  requires that the possible disagreement in agents' priors, represented by the 'size' of  $\Pi$ , is small relative to the variability of beliefs. The following example shows that a very mild level of agreement among players might suffice.

**Example 1** Suppose three wildcatters are competing for the drilling right on an oil field. The quantity of oil  $\theta_0$  is either high or low,  $\theta_0 \in \{H, L\}$ , each state equally likely. Wildcatters observe private signals  $s_i$ , which may be  $h$  or  $l$ . Assume that conditional on the state, signals are conditionally independent across wildcatters, and let  $\rho$  denote the probability that agent  $i$  receives the 'correct' signal ( $\Pr(s_i = h|\theta_0 = H) = \Pr(s_i = l|\theta_0 = L) = \rho$ ). Then,  $\rho$  can be naturally be interpreted as a measure of the quality of the probing technology. If  $\rho$  were common knowledge among the agents, McLean and Postlewaite's variability of

beliefs condition would be satisfied for any  $\rho \neq 1/2$  (i.e., unless signals were completely uninformative).

Now, suppose that the quality of the technology is not common knowledge, and let  $\Pi$  denote the set of (conditionally independent) distributions where  $\rho \in [\underline{\rho}, \bar{\rho}]$ ,  $0 \leq \underline{\rho} \leq \bar{\rho} \leq 1$ . Then, it can be verified that, whenever  $\underline{\rho} > 1/2$ ,<sup>20</sup>

$$\Lambda_i = \frac{(\underline{\rho}^3 + (1 - \underline{\rho})^3 - \underline{\rho}(1 - \underline{\rho}))}{\sqrt{(\underline{\rho}^3 + (1 - \underline{\rho})^3)^2 + 3(\underline{\rho}(1 - \underline{\rho}))^2}} > 0.$$

That is, as long as players agree that  $s_i = H$  is a better signal than  $s_i = L$  (in the sense of being more indicative of  $\theta_0 = H$ ), the adjusted variability of beliefs is positive, and it is possible to design an  $\mathcal{E}$ -incentive compatible auction.

It is interesting that, in this example, the prior  $p^*$  that achieves  $\Lambda_i$  (i.e., such that  $\Lambda(p^*) = \Lambda_i$ ) corresponds to the case  $\rho = 1$ . That is, provided that players agree in the relative ranking of signals ( $\underline{\rho} > 1/2$ ), the optimal prior to be used in designing the conditional payments is the one that treats signals as if they were perfectly informative, which lies outside of  $\Pi$  whenever  $\bar{\rho} < 1$ .

## 5 Robust Informational Size and a Bound on Transfers

The mechanism constructed in the previous section may be too costly to the seller, as the conditional payments necessary to induce agents to truthfully report their signals could be arbitrarily large. If the seller has limited resources, then she may not be able to ensure efficiency. In this section, we address this issue and provide a sufficient condition under which efficiency obtains at a small cost.

### 5.1 Robust Informational Size

To bound the cost of inducing agents to truthfully reveal their signals, we need to understand agents' incentives to misreport. In our mechanism, the incentives stem from the possibility that they can manipulate the second-stage auction outcome. In particular, if an agent reports a lower signal, then the price he pays in case he wins the auction would be lower, because all other agents would bid lower prices. Therefore, the more an agent can change the opponents' posteriors, the stronger incentive to misreport would he have in the first stage.

The notion of 'informational size' captures precisely the agent's ability to manipulate

---

<sup>20</sup>One can get a similar condition whenever  $\bar{\rho} < 1/2$ .

the opponents' posteriors by changing his report. Clearly, from agent  $i$ 's viewpoint, his incentives to misreport depend on his beliefs about the opponents' prior beliefs and their private signals: if  $i$  thinks that the opponents' beliefs are unresponsive to changes in  $s_i$ , or that the opponents' signals  $s_{-i}$  are sufficiently informative about  $\theta_0$ , then he would have low incentives to misreports. In our setting, however, the designer does not know such beliefs of player  $i$ . We thus need a measure of informational size that works for all the beliefs that the agent may have, as long as they are consistent with the minimal information available to the designer,  $\Pi$ .

Suppose that agent  $i$  believes that agent  $j$ 's prior is  $p_j \in \Pi$ . Then, (agent  $i$  believes that) given a signal vector  $s = (s_i, s_{-i}) \in S$ , agent  $j$ 's posterior distribution on  $\Theta_0$  would be  $p_j(\cdot|s_i, s_{-i})$ . If agent  $i$  unilaterally deviates and reports  $s'_i$  instead of  $s_i$ , bidder  $j$ 's posterior distribution would change. Measure this change by  $\|p_j(\theta|s_i, s_{-i}) - p_j(\theta|s'_i, s_{-i})\|$ . For each  $\epsilon \geq 0$ , let  $A_i(s_i, s'_i, p_j; \epsilon)$  be the set that consists of those  $s_{-i}$  for which bidder  $i$  has at least “ $\epsilon$ -effect” on bidder  $j$ 's posterior, when bidder  $i$  believes that  $j$ 's prior is  $p_j$ . Precisely,

$$A_i(s_i, s'_i, p_j; \epsilon) \equiv \{s_{-i} \in S_{-i} : \|p_j(\theta_0|s_{-i}, s_i) - p_j(\theta_0|s_{-i}, s'_i)\| > \epsilon\}. \quad (8)$$

Recall that the mechanism designer does not know agent  $i$ 's beliefs about the other agents' beliefs, but we aim at a measure that works for all beliefs consistent with common knowledge of the environment,  $\mathcal{E}$ . We thus define

$$A_i(s_i, s'_i; \epsilon) \equiv \cup_{p_j \in \Pi} A_i(s_i, s'_i, p_j; \epsilon).$$

The set  $A_i(s_i, s'_i; \epsilon)$  consists of the  $s_{-i}$  for which, for some beliefs that  $i$  may have about opponents' priors, bidder  $i$  thinks that he has at least  $\epsilon$ -effect on their posteriors. Equivalently,

$$A_i(s_i, s'_i; \epsilon) = \left\{ s_{-i} \in S_{-i} : \max_{p_j \in \Pi} \|p_j(\theta_0|s_{-i}, s_i) - p_j(\theta_0|s_{-i}, s'_i)\| > \epsilon \right\}. \quad (9)$$

The maximum operator in (9) represents the worst-case scenario from the viewpoint of the designer. The maximum is achieved at player  $i$ 's most optimistic beliefs about the opponents' prior, in terms of being able to manipulate their posterior by reporting  $s'_i$  instead of  $s_i$ .

From the viewpoint of player  $i$ , the set  $A_i(s_i, s'_i; \epsilon)$  is an event. Its probability depends on  $i$ 's beliefs about the opponents' signals, that is  $p_i(s_{-i}|s_i)$ . For each  $s_i, s'_i$  and  $p_i$ , let

$$\mu_i(s_i, s'_i, p_i) = \min\{\epsilon \in [0, 1] : p_i(A(s_i, s'_i; \epsilon)|s_i) \leq \epsilon\}.$$

The value  $\mu_i(s_i, s'_i, p_i)$  is the minimum value of  $\epsilon$  such that, when bidder  $i$ 's prior is  $p_i$  and

his signal is  $s_i$ , the probability that  $i$  attaches to the other bidders receiving a signal profile on which reporting  $s'_i$  has at least an  $\epsilon$ -effect is less than or equal to  $\epsilon$ .<sup>21</sup> To accommodate the situation in which agent  $i$  has the strongest incentive to manipulate the opponents' posteriors, let

$$\mu_i(p_i) = \max_{s_i, s'_i \in S_i} \mu_i(s_i, s'_i, p_i).$$

Finally, we define the informational size of agent  $i$  considering once again the worst-case scenario from the designer's viewpoint. In this context, this means considering the agent  $i$ 's beliefs such that he is most optimistic about his ability to manipulate the opponents' posterior. This measure is obtained by taking the maximum of  $\mu_i(p_i)$  over  $p_i \in \Pi$ .

**Definition 7** *Agent  $i$ 's robust informational size is defined as:*

$$\mu_i = \max_{p_i \in \Pi} \mu_i(p_i). \quad (11)$$

It is straightforward that with a single common prior  $\Pi = \{p\}$ , the definition shrinks to the following, which coincides with that of McLean and Postlewaite:

$$\mu_i^*(p) = \max_{s_i, s'_i} \{ \min\{\epsilon \in [0, 1] : p(\{s_{-i} \in S_{-i} : \|p(\theta_0|s_{-i}, s_i) - p(\theta_0|s_{-i}, s'_i)\| > \epsilon\} | s_i) \leq \epsilon\} \}.$$

**Example 2** *Consider the same problem in Example 1, and suppose  $\underline{\rho} > \frac{1}{2}$ .<sup>22</sup> In this case, informational size delivers the following value:*

$$\mu_i = 2\underline{\rho}(1 - \underline{\rho}).$$

*In this particular example,  $\mu$  only depends on the lower bound,  $\underline{\rho}$ . This is natural because the worst-case scenario from the designer's perspective is when agent  $i$  is most optimistic about his ability to manipulate the opponents' posteriors, which is when the precision of the opponents' signals is smallest ( $\rho = \underline{\rho}$ ). In addition,  $\mu_i$  is strictly decreasing in  $\underline{\rho}$ . Hence, as the information of the designer improves (the set  $[\underline{\rho}, \bar{\rho}]$  shrinks),  $\mu_i$  decreases (strictly decreasing in  $\underline{\rho}$ , while constant in  $\bar{\rho}$ ).*

---

<sup>21</sup>To see that this value is always well-defined, let

$$F(\epsilon) = p_i \left( \left\{ s_{-i} \in S_{-i} : \max_{p_j \in \Pi} \|p_j(\theta|s_{-i}, s_i) - p_j(\theta|s_{-i}, s'_i)\| \leq \epsilon \right\} | s_i \right). \quad (10)$$

Hence  $\{\epsilon \in [0, 1] : 1 - F(\epsilon) \leq \epsilon\}$  is nonempty (because  $1 - F(1) \leq 1$ ), bounded and closed (since  $F$  is right continuous with left-hand limits, using the fact that  $\Pi$  is compact).

<sup>22</sup>If  $\underline{\rho} < 1/2$ , while  $\bar{\rho} > 1/2$ , then adjusted variability of beliefs is negative, and the augmented auction would not be  $\mathcal{E}$ -incentive compatible.

The next result proves that the definition in (11) indeed serves as a relevant measure of informational size in our ‘robust’ setting. It shows that the payments to the agents in the augmented auction can be made arbitrarily small, if agents are sufficiently informationally small in our sense, relative to their (adjusted) variability of beliefs:

**Proposition 3** *For every  $\epsilon > 0$  and for every  $i$ , there exists a  $\delta_i > 0$  such that, whenever  $\mu_i \leq \delta_i \cdot \Lambda_i$ , there exists an  $\mathcal{E}$ -incentive compatible efficient augmented Vickrey auction  $\{q_i^*, x_i^* - z_i\}_{i \in N}$  satisfying  $0 \leq z_i(s) \leq \epsilon$  for every  $i$  and  $s$ .*

**Proof.** See the Appendix. ■

## 5.2 An Explicit Upper Bound

Proposition 3 relates agents’ informational size to the cost of efficient implementation. The latter, however, does not depend only on agents’ informational size. Ultimately, an agent’s incentive to misreport his signal depends on the impact that he has on the opponent’s expected *valuations*, that is, the expected utility given the conditional probabilities. Informational size measures only the agent’s ability to manipulate the opponents’ posteriors. Given agents’ informational size, an agent’s impact on the opponents’ valuations would be larger if the opponents’ utilities vary faster with  $\theta_0$ . In our setting, the designer has only limited information about agents’ beliefs about their opponents’ preferences  $u_{-i}$ , namely that agent  $i$  believes that agent  $j$ ’s preferences lie in the set  $U_j$ . Consequently, the cost of robustly achieving efficiency would depend not only on the set of priors  $\Pi$  (which determines adjusted variability of beliefs  $\Lambda_i$  and informational size  $\mu_i$ ), but also on the maintained common knowledge assumptions on agents’ preferences  $(U_i)_{i \in N}$ . This intuition is formalized by the following proposition, which provides explicit upper bounds to the cost of robustly achieving efficiency.

**Proposition 4** *Fix an environment  $\mathcal{E}$ , and consider the corresponding measures  $\mu_i$  and  $\Lambda_i$  (for all  $i \in N$ ). Let  $\kappa = 2 \cdot \sqrt{\sum_{j \neq i} n_j}$ . If  $\Lambda_i > 0$  for all  $i$ , then the payments to agent  $i$  in the first stage are bounded above by*

$$\bar{z}_i = \kappa \cdot M_{i,1} (1 + M_{i,2}) \cdot \frac{\mu_i}{\Lambda_i}, \quad (12)$$

where  $M_{i,1} = \max_{u \in U_{-i}^*} [u(\theta^m) - u(\theta^1)]$ , and  $M_{i,2} = \frac{\max_{u \in U_{-i}^*} \sum_{\theta_0 \in \Theta_0} u(\theta_0)}{\min_{c \in C_{-i}^*} [u(\theta^m, c) - u(\theta^1, c)]}$ .

**Proof.** See the Appendix. ■

The upper bounds are characterized by three more variables other than agents' variability of beliefs and informational size. The parameter  $\kappa$  is just a constant that depends on the cardinality of the sets of signals. The other two terms  $M_{i,1}$  and  $M_{i,2}$  depend only on the restrictions on agents' preferences, as specified by the sets  $(U_i)_{i \in N}$ .<sup>23</sup> The constant  $M_{i,1}$  is a bound on the *responsiveness* of other agents' valuation functions to changes in  $\theta_0$ :  $M_{i,1}$  is small if the opponents' valuation functions are rather insensitive to increases in  $\theta_0$ . The constant  $M_{i,2}$  provides a measure of the *level* of the opponents' valuations for the good, normalized by a measure of its variability:  $M_{i,2}$  is small if other agents' valuations tend to be low. Intuitively, an agent's impact on the opponents' expected valuations would be larger if the opponents' valuations vary a lot with  $\theta_0$  (large  $M_{i,1}$ ). In addition, holding  $M_{i,1}$  constant, the impact would be larger if the valuation tends to be high ( $M_{i,2}$ ), due to a scaling effect.

We note that the bounds in Proposition 4 are not tight. That is, the designer may be able to achieve efficiency at a lower cost. Nonetheless, we think that the bounds are informative, and capture well the constraints faced by the designer. In particular, they can be used to provide a measure of the cost, from the designer's viewpoint, of achieving more robust implementation results. As we discuss in the next section, the 'cost of robustness' would be identified with the variation of  $\bar{z}_i$ , not with its level. The non-tightness is thus less important, provided that the variation of  $\bar{z}_i$  does capture the relevant economic insights.

## 6 The Determinants of the 'Cost of Robustness'

Consider the problem of a social planner, who desires to achieve certain objectives through the design of a mechanism. Not knowing the details of agents' beliefs about each other, she is concerned with the possible non-robustness of the mechanisms supported by the classical theory. In this situation, it seems plausible that the planner would compromise on her objectives, if it could deliver more robustness. But what is the cost of achieving a more robust implementation result?

The idea of a trade-off between robustness and the extent to which certain goals can be achieved is intuitive. Nonetheless, the existing literature does not provide much insight on the issue. In fact, it is not even clear whether that conjecture is correct at all. Proposition 4 indirectly provides an answer to this question, and offers some insights on the determinants of this trade-off.

To identify the determinants of the cost of robustness, we study the behavior of the bounds (12) as the robustness requirements vary. The exercise is essentially to compare

---

<sup>23</sup>Recall that agents' informational size  $\mu_i$  and variability of beliefs  $\Lambda_i$  depend only on  $\Pi$ .

the bounds  $\bar{z}$  and  $\bar{z}'$  associated with two environments that differ in the strength of the maintained common knowledge assumptions. The difference  $(\bar{z}' - \bar{z})$  provides a measure of the cost of extending the robustness of the implementation from one environment to the other, allowing us to identify the factors that affect the cost of robustness. We perform two kinds of thought experiments: In a given environment  $\mathcal{E}$ , we change the designer's information either on agents' preferences ( $U' \subset U$ ) or on their beliefs ( $\Pi' \subset \Pi$ ), and examine how the bound  $\bar{z}$  changes.

Note that an environment  $\mathcal{E}$  determines the information commonly known to the agents as well as the designer's information. Thus, by varying the sets  $U'$  or  $\Pi'$ , not only do we provide more information for the designer, but we also endow the agents with more common knowledge. This identification is common in the literature, but the two effects are conceptually distinct. For instance, agents may commonly know more than what is known to the designer. Varying the common knowledge assumptions, *given* information of the designer, is thus a robustness exercise of independent interest and allows us to disentangle the two effects. We first study how robust informational size, adjusted variability of beliefs and the upper bounds in (12) change as, holding the designer's information fixed, agents share more common knowledge about each other. We then examine how the measures would change as the designer is endowed with more information about agents' preferences or beliefs.

## 6.1 More common knowledge among the agents

**Preferences.** Suppose agents have better information about each other's preferences (while  $\mathcal{E}$  remains the only information available to the designer). Precisely, agents commonly know that their preferences belong to  $U' \subset U$ . In this case, neither the adjusted variability of beliefs (7) nor the measure of informational size (11) change, as these measures depend only on the commonly known restrictions on a commonly known set of priors,  $\Pi$ . The other two measures affecting the bounds  $\bar{z}$  also do not change. Agents' preferences do affect their incentives to misreport their signals, and therefore their information about each other's preferences potentially matters in determining the conditional payments. However, unless the designer possesses that information, she cannot exploit it to reduce the transfers to the agents. Hence, the bounds  $\bar{z}$  do not change.

**Beliefs.** Suppose agents share a common prior  $p^* \in \Pi$  (but the designer's information is still represented by  $\mathcal{E}$ ). The designer knows that agents agree on some prior in  $\Pi$ , but she does not know which one it is.

Parameters  $M_{i,1}$  and  $M_{i,2}$  are clearly not affected by this change, because they are concerned with agents' preferences. The same is true for variability of beliefs, but for a more

subtle reason. The discussion in Section 4.2 that led to the definition of  $\Lambda_i$  was based on the fact that the prior  $p$  chosen by the planner does not necessarily coincide with each agent's beliefs  $p_i$ . Whether or not agents share the same prior is irrelevant to that argument and, therefore, to the notion of variability of beliefs. Conceptually, an agent's variability of beliefs measures the sensitivity of that agent's beliefs about others' signals (not beliefs) to changes of his own signal. Therefore, it must be independent of the agent's beliefs about others' beliefs. The only way  $\Lambda_i$  might be affected is when the designer obtains more information about agents' priors.

On the contrary, informational size does change with the common prior assumption. The informational size of an agent measures the impact he has on the posterior distributions of other agents on  $\Theta_0$ . When the designer does not know that agents share a common prior, she must consider the possibility that an agent is overly optimistic about his impact on others' posterior distributions. Therefore, she needs to provide a sufficient incentive for each agent to truthfully report his signal no matter what beliefs he has. When the designer knows that agents share a common prior, she knows that each agent has a better estimate about his impact on others' posterior distributions, that is, she does not have to take into account the possibility that an agent believes that other agents' beliefs are such that he can have a huge impact. Therefore, the designer can induce agents to truthfully report their signals at a smaller cost.

Formally, for each  $p_i \in \Pi$ , let

$$\mu_i^*(p_i) = \max_{s_i, s'_i \in S_i} \min \{ \epsilon \in [0, 1] : p_i(\{s_{-i} \in S_{-i} : \|p_i(\theta_0|s_{-i}, s_i) - p_i(\theta_0|s_{-i}, s'_i)\| > \epsilon\} | s_i) \leq \epsilon \}.$$

This value coincides with McLean and Postlewaite's measure of informational size in the case of a single prior  $p_i$  and would be the relevant measure if it were to be known to the designer that agents share this prior. When the designer does not know what prior agents agree on, the relevant measure is

$$\mu_i^{CP} = \max_{p_i \in \Pi} \mu_i^*(p_i).$$

This measure looks similar to  $\mu_i$  in equation (11), but it is not the same. This difference lies in, given  $p_i$ , how to identify the set  $A_i(s_i, s'_i)$  (the set of others' signals given which agent  $i$  has more than  $\epsilon$ -effect on others' posterior distributions). When agents share a common prior, once  $p_i$  is fixed, then the designer can assume that all other agents adopt the same prior, and thus

$$A_i(s_i, s'_i) = A_i(s_i, s'_i, p_i) = \{s_{-i} \in S_{-i} : \|p_i(\theta|s_{-i}, s_i) - p_i(\theta|s_{-i}, s'_i)\| > \epsilon\}.$$

When agents do not necessarily share a common prior, agent  $i$  can have any beliefs about agent  $j$ 's prior  $p_j$ . The designer must take into account all possible beliefs of agent  $i$ , and thus

$$A_i(s_i, s'_i) = \bigcup_{p_j \in \Pi} A_i(s_i, s'_i, p_j).$$

Due to this difference,  $\mu_i \geq \mu_i^{CP}$  for any  $i$ , and thus knowing that agents share a common prior (more generally, any subset of  $\Pi$ ) reduces the designer's cost to extract agents' private information, even if the designer does not know what the prior is.

Overall, providing agents with more common knowledge of the prior reduces the bounds  $\bar{z}$ , via its impact on agents' informational size, even if the designer's information is held constant.

## 6.2 More information to the designer

Now we study how the relevant measures change as the designer is endowed with more information about the environment. Formally, we examine how the measures are affected as we change the set of agents' possible preferences or the set of priors. Of course, when we move from one environment to another, common knowledge among agents also changes. Therefore, the discussion below must be complemented with that in the previous subsection.

**Preferences.** Suppose the designer has better information about agents' preferences. Formally, now the designer knows that agents' preference parameters belong to  $U' \subset U$ . In this case, neither the measure of adjusted variability of beliefs (7) nor that of informational size (11) change, as these measures depend only on the commonly known restrictions on the set of priors,  $\Pi$ . However, the two measures on agents' preferences,  $M_{i,1}$  and  $M_{i,2}$ , become smaller. In particular, if

$$\begin{aligned} \operatorname{argmax}_{u \in U_{-i}^*} [u(\theta^m) - u(\theta^1)] &\notin \bigcup_{j \neq i} U'_j, \\ \operatorname{argmin}_{u \in U_{-i}^*} [u(\theta^m) - u(\theta^1)] &\notin \bigcup_{j \neq i} U'_j, \end{aligned}$$

or

$$\operatorname{argmax}_{u \in U_{-i}^*} \sum_{\theta_0 \in \Theta_0} u(\theta_0) \notin \bigcup_{j \neq i} U'_j,$$

then at least one of  $M_{i,1}$  and  $M_{i,2}$  becomes strictly smaller, and thus the designer's cost of implementing efficient allocation decreases. Consequently, the corresponding variation of the bounds  $\bar{z}$  captures the cost of incorporating the preferences in  $U \setminus U'$ .

**Beliefs.** Now suppose the designer has better information about agents' beliefs. That is,  $\Pi' \subset \Pi$  is given. The two measures on preferences  $M_{i,1}$  and  $M_{i,2}$  are obviously unaffected by this change.

Variability of beliefs increases. Mathematically, this is because the domain of the minimum operator in (6) shrinks. Economically, now the designer has better information about each agent's beliefs and, therefore, can better tailor the payment scheme to each agent.

Informational size decreases. There are two effects, one direct and the other indirect. First, now the designer has better information about each agent's conditional beliefs on others' signals,  $p_i(s_{-i}|s_i)$ . This directly decreases informational size, as apparent in (11). Second, now the designer knows that agents share beliefs  $\Pi'$ . As we discussed in the previous subsection, this also decreases agents' informational size.

**Example 3** *Let us illustrate the main points of this section with the example we have used throughout the paper. As the results regarding two parameters on preferences,  $M_{i,1}$  and  $M_{i,2}$ , are rather straightforward, we focus on our two measures on beliefs. Consider the problem of a designer facing the efficient implementation problem in the setting of examples 1 and 2. Not knowing the details of agents' beliefs about the quality of the sampling technology, she intends to design a robust auction that guarantees efficient allocation for all beliefs  $\rho \in [\underline{\rho}, \bar{\rho}]$ . The larger the set  $[\underline{\rho}, \bar{\rho}]$ , the more demanding the robustness requirement. Since varying the set  $[\underline{\rho}, \bar{\rho}]$  corresponds to a change in the set  $\Pi$ , the critical term in (12) involved in this change is the ratio  $(\mu_i/\Lambda_i)$ . Combining the expressions for  $\mu_i$  and  $\Lambda_i$  and arranging terms, we obtain*

$$\frac{\mu_i}{\Lambda_i} = \frac{2\underline{\rho}(1-\underline{\rho}) \cdot \sqrt{(\underline{\rho}^3 + (1-\underline{\rho})^3)^2 + 3(\underline{\rho}(1-\underline{\rho}))^2}}{(\underline{\rho}^3 + (1-\underline{\rho})^3 - \underline{\rho}(1-\underline{\rho}))}.$$

*First, notice that this ratio only depends on the value of  $\underline{\rho}$ . Hence, increasing the robustness including more precise sampling technologies ( $\bar{\rho} \rightarrow 1$ ) does not entail any cost increase. We can thus set  $\bar{\rho} = 1$  at no extra cost and focus on varying  $\underline{\rho}$ . We already know from Example 1 that if  $\underline{\rho} \leq 1/2$ , the mechanism would not work, because the adjusted variability of beliefs would be negative. Furthermore, it is straightforward to show that  $\mu_i/\Lambda_i \rightarrow \infty$  as  $\underline{\rho} \rightarrow 1/2$ . Hence, as we approach the non-informative distribution, robust implementation becomes arbitrarily costly. However, for sufficiently informative distributions, the costs are bounded and approach zero as  $\underline{\rho} \rightarrow 1$ . Figure 1 plots the behavior of  $\mu_i/\Lambda_i$  as a function of  $\underline{\rho}$ . For instance, if the designer is considering  $\rho \in [0.85, 1]$ , and wishes to increase the robustness to include precisions  $\rho \geq 0.75$ , the ratio  $\mu_i/\Lambda_i$  roughly doubles.*

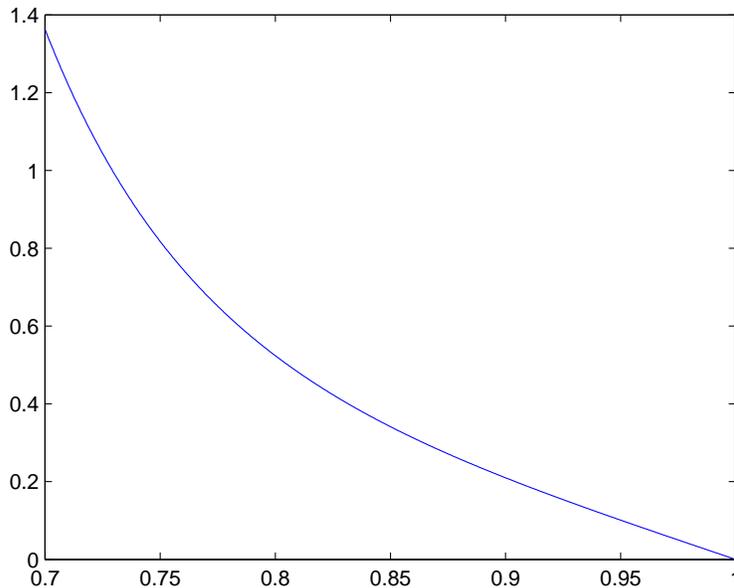


Figure 1:  $x$ -axis:  $\underline{\rho} \in [0.7, 0.99]$ ,  $y$ -axis:  $\frac{\mu_i}{\Lambda_i}$

## 7 Conclusion

We studied a classical efficient auction problem with interdependent values under *arbitrary* common knowledge assumptions. We constructed a simple mechanism and showed that it robustly achieves efficiency, albeit at some cost. We also showed that if agents are informationally small in an appropriate sense, then the cost of robustly achieving efficiency can be arbitrarily small.

From an applied viewpoint, these results are important because our model covers a variety of ‘real world’ situations that could not be cast within the classical models of auctions with interdependent values. From a more theoretical perspective, our analysis allows us to relate the size of the cost necessary for robust efficiency to the strength of the robustness requirement, both in terms of information available to the designer, and in terms of common knowledge among the agents. Although the idea of a trade-off between robustness and implementation is fairly intuitive, the literature is essentially silent on the issue, mainly because addressing the problem requires a number of conceptual innovations and intermediate results. By studying how the cost of achieving efficiency varies with the strength of the robustness requirement, we moved a first step towards understanding the determinants of the ‘cost of robustness’.

We departed from the existing literature in several ways. Our model differs from the

classical framework in that we impose less common knowledge assumptions, and assume that the designer has only imprecise information about agents’ beliefs. However, we also departed significantly from the recent literature on robust mechanism design, in that we maintain that the designer has *some* information about agents’ beliefs, and that agents share some common knowledge of their beliefs.<sup>24</sup> Another important innovation, relative to the literature on robust mechanism design (see footnote 8), is that our model accommodates situations in which the state of the world is not ‘distributed knowledge’ among the agent. Lack of distributed knowledge is particularly important in applied settings, and the inability to accommodate it is a serious limitation of the existing literature on robust mechanism design.

We approached the robust design problem by introducing a novel notion of incentive compatibility, which we called  $\mathcal{E}$ -incentive compatibility ( $\mathcal{E}$ -IC).  $\mathcal{E}$ -IC is in general stronger than interim incentive compatibility (normally adopted in classical mechanism design) but weaker than ex-post incentive compatibility (central in the literature on robust mechanism design.) We showed that imposing  $\mathcal{E}$ -IC is equivalent to imposing interim incentive compatibility for all beliefs consistent with the minimal information available to the designer.<sup>25</sup>

An important aspect of our results is that our mechanism involves a transfers scheme based on the ideas developed by Cremer and McLean (1985, 1988) to obtain the famous full-surplus extraction (FSE) result. These techniques are often criticized for the heavy reliance on the strong common knowledge assumptions implicit in the classical model (see Neeman (2004)). Although we do not focus on the FSE problem, our results show that these techniques can be extended to ‘robust’ settings, provided that minimal common knowledge assumptions are satisfied.<sup>26</sup>

In deriving a measure of the cost of robust efficiency, we developed a measure of *robust informational size*. This measure quantifies the value of private information held by an agent in settings in which agents’ beliefs (and in particular, the mapping from beliefs to preferences) are not common knowledge. This measure is thus of independent interest, and provides an answer in the affirmative to the open question of whether the original notion of informational size by McLean and Postlewaite (2002, 2004) could be extended to settings that do not satisfy the BDP condition (see Gizatulina and Hellwig (2011)).

---

<sup>24</sup>Among the works referenced in footnote 8, the paper by Artemov, Kunimoto and Serrano (2010) is closest in this respect, in that it maintains partial information on agents’ beliefs.

<sup>25</sup>Since our environments encompass both the classical Bayesian environments and the Belief-Free settings common in the literature on the mechanism design, and  $\mathcal{E}$ -IC coincides with ex-post IC in the latter, this result generalizes Propositions 1 and 2 from Bergemann and Morris (2005) to environments in which *some* restrictions on agents’ beliefs are maintained.

<sup>26</sup>In particular, our settings do not satisfy the ‘Beliefs-Determine-Preference’ (BDP) condition (Neeman, 2004): even once agent  $i$ ’s beliefs  $p_i$  are known, the only information about his preferences is that  $u_i \in U_i$ .

# Appendix: Omitted Proofs

## Proof of Proposition 1

Allocation rule  $q : \Theta \rightarrow A$  is  $\mathcal{E}$ -Incentive compatible if and only if it is robustly  $\mathcal{E}$ -implementable.

For the “only if” part, let  $x_i : \Theta \rightarrow \mathbb{R}$  ( $i = 1, \dots, n$ ) be the transfers that make  $q : \Theta \rightarrow X$  be the  $\mathcal{E}$ -IC. Fix a  $\mathcal{E}$ -consistent type space  $\mathcal{T}$ , and consider the direct mechanism  $(f^*(t), (x_i^*(t))_{i \in I})$  such that  $f^*(t) = q(\theta)$  whenever  $\hat{\theta}(t) = \theta$ . Now, consider type  $t_i \in T_i$ . Interim incentive compatibility requires that

$$\begin{aligned} t_i &\in \arg \max_{t'_i \in T_i} \int_{\theta_0, t_{-i}} [f_i^*(t'_i, t_{-i}) \cdot \hat{u}_i(t_i)(\theta_0) - x_i^*(t'_i, t_{-i})] \cdot d\tau_i(t_i) \\ &= \arg \max_{t'_i \in T_i} \int_{\theta_0, t_{-i}} \left[ f_i^*(\hat{\theta}_i(t'_i), \hat{\theta}_{-i}(t_{-i})) \cdot \hat{u}_i(t_i)(\theta_0) - x_i^*(\hat{\theta}_i(t'_i), \hat{\theta}_{-i}(t_{-i})) \right] \cdot d\tau_i(t_i) \\ &\quad \text{which implies} \\ \hat{\theta}_i(t_i) &\in \arg \max_{\theta'_i \in \Theta_i} \int_{\theta_0, t_{-i}} \left[ f_i^*(\theta'_i, \hat{\theta}_{-i}(t_{-i})) \cdot \hat{u}_i(t_i)(\theta_0) - x_i^*(\theta'_i, \hat{\theta}_{-i}(t_{-i})) \right] \cdot d\tau_i(t_i) \end{aligned}$$

By the  $\mathcal{E}$ -consistency requirement, for each  $(\theta_0, s_{-i})$ ,  $\int_{\theta_0, t_{-i}: \hat{s}_{-i}(t_{-i})=s_{-i}} d\tau_i(t_i) = \hat{p}_i(t_i)[\theta_0, s_{-i} | \hat{s}_i(t_i)]$ ; furthermore, the fact that  $q$  is  $\mathcal{E}$ -IC implies that

$$\theta_i \in \arg \max_{\theta'_i \in \Theta_i} \sum_{\theta_0} \sum_{s_{-i}} [q_i(\theta'_i, s_{-i}, \bar{p}_{-i}, \bar{u}_{-i}) \cdot \hat{u}_i(t_i)(\theta_0) - x_i(\theta'_i, s_{-i}, \bar{p}_{-i}, \bar{u}_{-i})] \cdot \hat{p}_i(t_i)[\theta_0, s_{-i} | \hat{s}_i(t_i)]$$

holds for all  $\bar{p}_{-i}, \bar{u}_{-i}$ , then interim IC on  $\mathcal{T}$  is trivially satisfied.

For the “if” part, suppose that  $q$  is interim implementable on all  $\mathcal{E}$ -consistent type spaces. Then, for every  $(\bar{p}_{-i}, \bar{u}_{-i}) \in \Pi_{-i} \times U_{-i}$ ,  $q$  is interim implementable on the following type space: for all  $j \neq i$ , let  $T_j = S_j \times \{\bar{p}_j\} \times \{\bar{u}_j\}$ ; functions  $\hat{\theta}_j$  (for all  $j \neq i$ ) in this type space are given by the natural projections; finally, let  $T_i, \tau_i$  be such that and for any  $t_i \in T_i$ ,  $\tau_i(t_i)[\theta_0, s_{-i}, \bar{p}_{-i}, \bar{u}_{-i}] = \hat{p}_i(t_i)[\theta_0, s_{-i} | \hat{s}_i(t_i)]$ . Hence, for each  $(\bar{p}_{-i}, \bar{u}_{-i}) \in \Pi_{-i} \times U_{-i}$ , there

exist

$$q^{i,(\bar{p}_{-i},\bar{u}_{-i})} : T_i \times S_{-i} \rightarrow A \text{ and } \left( x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})} \right)_{i \in N} : T_i \times S_{-i} \rightarrow \mathbb{R}^n$$

such that :

$$\begin{aligned} (1). & \quad q^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t_i, s_{-i}) = q\left(\hat{\theta}_i(t_i), (s_{-i}, \bar{p}_{-i}, \bar{u}_{-i})\right) \text{ for all } (t_i, s_{-i}), \text{ and} \\ (2). & \quad \text{for all } t'_i \in T_i, \\ & \quad \int_{\theta_0, t_{-i}} \left[ q_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t_i, s_{-i}) \cdot \hat{u}_i(t_i)(\theta_0) - x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t_i, s_{-i}) \right] d\tau_i(t_i) \\ & \geq \int_{\theta_0, t_{-i}} \left[ q_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t'_i, s_{-i}) \cdot \hat{u}_i(t_i)(\theta_0) - x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t'_i, s_{-i}) \right] d\tau_i(t_i) \end{aligned}$$

Because of (1), (2) can be written, for any  $(\bar{p}_{-i}, \bar{u}_{-i})$ , as:

$$\begin{aligned} & \int_{\theta_0, t_{-i}} \left[ q_i\left(\hat{\theta}_i(t_i), (s_{-i}, \bar{p}_{-i}, \bar{u}_{-i})\right) \cdot \hat{u}_i(t_i)(\theta_0) - x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t_i, s_{-i}) \right] d\tau_i(t_i) \\ \geq & \int_{\theta_0, t_{-i}} \left[ q_i\left(\hat{\theta}_i(t'_i), (s_{-i}, \bar{p}_{-i}, \bar{u}_{-i})\right) \cdot \hat{u}_i(t_i)(\theta_0) - x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t'_i, s_{-i}) \right] d\tau_i(t_i) \text{ for all } t'_i \\ & \text{or, equivalently (exploiting the } \mathcal{E}\text{-consistency restriction):} \\ & \sum_{\theta_0, s_{-i}} \left[ q_i\left(\hat{\theta}_i(t_i), (s_{-i}, \bar{p}_{-i}, \bar{u}_{-i})\right) \cdot \hat{u}_i(t_i)(\theta_0) - x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t_i, s_{-i}) \right] \\ & \cdot \hat{p}_i(t_i) [\theta_0, s_{-i} | \hat{s}_i(t_i)] \\ \geq & \sum_{\theta_0, s_{-i}} \left[ q_i\left(\hat{\theta}_i(t'_i), (s_{-i}, \bar{p}_{-i}, \bar{u}_{-i})\right) \cdot \hat{u}_i(t_i)(\theta_0) - x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t'_i, s_{-i}) \right] \\ & \cdot \hat{p}_i(t_i) [\theta_0, s_{-i} | \hat{s}_i(t_i)] \text{ for all } t'_i \end{aligned}$$

Also, this implies that for any  $t_i, t'_i \in T_i$  such that  $\hat{\theta}_i(t_i) = \hat{\theta}_i(t'_i)$ , it must be that  $x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t_i, s_{-i}) = x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t'_i, s_{-i})$ , otherwise  $t_i$  would have an incentive to report  $t'_i$  if  $x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t_i, s_{-i}) > x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t'_i, s_{-i})$ , and viceversa if  $x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t_i, s_{-i}) < x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t'_i, s_{-i})$  (this is because the allocation  $q$  would not change, and the two types also have the same beliefs over  $\theta_{-i}$ ). Hence, we can write  $x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(t_i, s_{-i})$  as  $x_i^{i,(\bar{p}_{-i},\bar{u}_{-i})}(\hat{\theta}_i(t_i), s_{-i})$  (that is, only payoff types affect the payments).

Now, robust  $\mathcal{E}$ -implementation can be achieved setting, for each  $\theta^* = (s_i^*, p_i^*, u_i^*)_{i \in N}$ , and for each  $i \in N$ ,  $x_i^*(\theta^*) = x_{i1}^{i,(p_{-i}^*, u_{-i}^*)}(\theta_i^*, s_{-i}^*)$ . Clearly, by construction,  $(q, x^*)$  is  $\mathcal{E}$ -IC.

## Proof of Proposition 2

Let  $p_i^*$  be the prior that achieves  $\Lambda_i$  in (7). Set  $z_i^*(\tilde{s}) = \kappa \frac{p_i^*(\tilde{s}_{-i} | \tilde{s}_i)}{\|p_i^*(\tilde{s}_{-i} | \tilde{s}_i)\|_2}$  for some  $\kappa > 0$ .

Suppose agent  $i$ 's beliefs are given by  $p_i$ . Then, his deviation loss by reporting  $s'_i$  when

his true signal is  $s_i$  is

$$\kappa \left( \frac{p_i^*(\cdot|s_i)}{\|p_i^*(\cdot|s_i)\|_2} - \frac{p_i^*(\cdot|s'_i)}{\|p_i^*(\cdot|s'_i)\|_2} \right) p_i(\cdot|s_i).$$

It suffices to show that this term is always larger than  $\Lambda_i$  times some constant: If so, whenever  $\Lambda_i > 0$  for each  $i$ , by making  $\kappa$  sufficiently large, agents can be induced to truthfully report their signals, irrespective of their beliefs. By the definition of  $\Lambda_i$ ,

$$\begin{aligned} \Lambda_i &\leq \left\| \frac{p_i^*(\cdot|s'_i)}{\|p_i^*(\cdot|s'_i)\|_2} - \frac{p_i(\cdot|s_i)}{\|p_i(\cdot|s_i)\|_2} \right\|_2^2 - \left\| \frac{p_i^*(\cdot|s_i)}{\|p_i^*(\cdot|s_i)\|_2} - \frac{p_i(\cdot|s_i)}{\|p_i(\cdot|s_i)\|_2} \right\|_2^2 \\ &= -\frac{2p_i^*(\cdot|s'_i) \cdot p_i(\cdot|s_i)}{\|p_i^*(\cdot|s'_i)\|_2 \cdot \|p_i(\cdot|s_i)\|_2} + \frac{2p_i^*(\cdot|s_i) \cdot p_i(\cdot|s_i)}{\|p_i^*(\cdot|s_i)\|_2 \cdot \|p_i(\cdot|s_i)\|_2} \\ &= 2 \left( \frac{p_i^*(\cdot|s_i)}{\|p_i^*(\cdot|s_i)\|_2} - \frac{p_i^*(\cdot|s'_i)}{\|p_i^*(\cdot|s'_i)\|_2} \right) \cdot \frac{p_i(\cdot|s_i)}{\|p_i(\cdot|s_i)\|_2} \end{aligned}$$

Therefore,

$$\left( \frac{p_i^*(\cdot|s_i)}{\|p_i^*(\cdot|s_i)\|_2} - \frac{p_i^*(\cdot|s'_i)}{\|p_i^*(\cdot|s'_i)\|_2} \right) p_i(\cdot|s_i) \geq \Lambda_i \frac{\|p_i(\cdot|s_i)\|_2}{2} \geq \frac{\Lambda_i}{2\sqrt{\sum_{j \neq i} n_j}}.$$

## Proof of Propositions 3 and 4

Let  $\phi_i(s_i, s'_i, u_i, p_i; \theta_{-i})$  be agent  $i$ 's deviation gains from the Vickrey auction  $\{q_i^*, x_i^*\}_{i \in N}$  when he reported  $s'_i$ , while his true signal is  $s_i$ , given other agents' types  $\theta_{-i}$ . We show that the deviation gains, when his prior is  $p_i$ , are bounded above by  $\mu_i(p_i) \equiv \max_{s_i, s'_i \in S_i} \mu_i(s_i, s'_i, p_i) (\leq \mu_i)$  times some constant.

First, observe that given  $\theta_{-i}$ ,

$$\phi_i(s_i, s'_i, u_i, p_i; \theta_{-i}) \leq \max\{\bar{w}_i(s_i, u_i, p_i, \theta_{-i}) - \bar{w}_i(s'_i, u_i, p_i, \theta_{-i}), 0\},$$

where

$$\bar{w}_i(s''_i, u_i, p_i, \theta_{-i}) = \max_{j \neq i} \sum_{\theta_0} u_j(\theta_0) p_j(\theta_0 | s''_i, s_{-i}).$$

This is because an agent can get positive gains by misreporting only when he could become the winner or decrease his payment in the Vickrey auction. In the latter case, the deviation gains are clearly weakly less than  $\bar{w}_i(s_i, u_i, p_i, \theta_{-i}) - \bar{w}_i(s'_i, u_i, p_i, \theta_{-i})$ . In the former case,  $v_i(s_i, s_{-i}) \leq \bar{w}_i(\theta_i, \theta_{-i})$  and, therefore, the gains are again bounded by  $\bar{w}_i(s_i, u_i, p_i, \theta_{-i}) - \bar{w}_i(s'_i, u_i, p_i, \theta_{-i})$ .

In addition, the fact that

$$\bar{w}_i(s'_i, u_i, p_i, \theta_{-i}) = \max_{j \neq i} \sum_{\theta_0} u_j(\theta_0) p_j(\theta_0 | s'_i, s_{-i}) \geq \sum_{\theta_0} u_l(\theta_0) p_l(\theta_0 | s'_i, s_{-i}), \forall l \neq i$$

implies

$$\begin{aligned} \phi_i(s_i, s'_i, u_i, p_i; \theta_{-i}) &\leq \max\{\bar{w}_i(s_i, u_i, p_i, \theta_{-i}) - \bar{w}_i(s'_i, u_i, p_i, \theta_{-i}), 0\} \\ &\leq \max\left\{\max_{j \neq i} \sum_{\theta_0} u_j(\theta_0) (p_j(\theta_0 | s_i, s_{-i}) - p_j(\theta_0 | (s'_i, s_{-i}))), 0\right\} \\ &\leq \max\left\{\max_{j \neq i, p_j \in \Pi, u_j \in U_j} \sum_{\theta_0} u_j(\theta_0) (p_j(\theta_0 | s_i, s_{-i}) - p_j(\theta_0 | (s'_i, s_{-i}))), 0\right\}. \end{aligned}$$

Observe that the final expression is independent of other agents' types  $\theta_{-i}$ .

Then, independently of agent  $i$ 's beliefs on  $p_{-i}$  and  $u_{-i}$ ,

$$\begin{aligned} &E_{p_i}[\phi_i(s_i, s'_i, u_i, p_i; \theta_{-i}) | s_i] \\ &\leq E_{p_i}\left[\max\left\{\max_{j \neq i, p_j \in \Pi, u_j \in U_j} \sum_{\theta_0} u_j(\theta_0) (p_j(\theta_0 | s) - p_j(\theta_0 | s'_i, s_{-i})) | s_i, 0\right\}\right] \\ &= \sum_{s_{-i} \in S_{-i}} p_i(s_{-i} | s_i) \max\left\{\max_{j \neq i, p_j \in \Pi, u_j \in U_j} \sum_{\theta_0} u_j(\theta_0) (p_j(\theta_0 | s) - p_j(\theta_0 | (s'_i, s_{-i}))), 0\right\} \\ &= \sum_{s_{-i} \in A(s_i, s'_i)} p_i(s_{-i} | s_i) \max\left\{\max_{j \neq i, p_j \in \Pi, u_j \in U_j} \sum_{\theta_0} u_j(\theta_0) (p_j(\theta_0 | s) - p_j(\theta_0 | (s'_i, s_{-i}))), 0\right\} \\ &+ \sum_{s_{-i} \in S_i \setminus A(s_i, s'_i)} p_i(s_{-i} | s_i) \max\left\{\max_{j \neq i, p_j \in \Pi, u_j \in U_j} \sum_{\theta_0} u_j(\theta_0) (p_j(\theta_0 | s) - p_j(\theta_0 | (s'_i, s_{-i}))), 0\right\}. \end{aligned}$$

We will find a bound for each term in the last expression.

For the first term, let  $M_{i,1} = \max_{j \neq i, u_j \in U_j} \{u_j(\theta^m) - u(\theta^1)\}$ . Then, by the definitions of  $\mu_i(s_i, s'_i, p_i)$  and  $\mu_i(p_i)$ ,

$$\begin{aligned} &\sum_{s_{-i} \in A(s_i, s'_i)} p_i(s_{-i} | s_i) \max\left\{\max_{j \neq i, p_j \in \Pi, u_j \in U_j} \sum_{\theta_0} u_j(\theta_0) (p_j(\theta_0 | s) - p_j(\theta_0 | (s'_i, s_{-i}))), 0\right\} \\ &\leq \sum_{s_{-i} \in A(s_i, s'_i)} p_i(s_{-i} | s_i) \max_{j \neq i, p_j \in \Pi, u_j \in U_j} \{u_j(\theta^m) - u(\theta^1)\} \\ &= M_{i,1} \sum_{A(s_i, s'_i)} p_i(s_{-i} | s_i) = M_{i,1} \cdot p_i(\{A(s_i, s'_i) | s_i\}) \leq M_{i,1} \cdot \mu_i(s_i, s'_i, p_i) \leq M_{i,1} \cdot \mu_i(p_i). \end{aligned}$$

For the second term, for any  $j \neq i$  and  $u_j \in U_j$ , let  $\kappa_j(u_j) = \min\{u_j(\theta^m) - u(\theta^1)\}$  (since  $u_j$  is strictly increasing,  $0 < \kappa_j(u_j) \leq M_{i,1}$ ) and  $M_{i,2} = \max_{j \neq i, u_j \in U_j} \left\{ \frac{1}{\kappa_j(u_j)} \sum_{\theta_0} u_j(\theta_0) \right\}$ . Fix  $s_{-i} \in S_{-i} \setminus A(s_i, s'_i)$ . Then for any  $p_j \in \Pi$ ,  $p_j(\theta_0 | s_i, s_{-i}) - p_j(\theta_0 | s'_i, s_{-i}) \leq \mu_i(s_i, s'_i, p_i)$ . Therefore,

$$\begin{aligned}
& \sum_{s_{-i} \in S_{-i} \setminus A(s_i, s'_i)} p_i(s_{-i} | s_i) \max \left\{ \max_{j \neq i, p_j \in \Pi, u_j \in U_j} \sum_{\theta_0} u_j(\theta_0) (p_j(\theta_0 | s) - p_j(\theta_0 | (s'_i, s_{-i}))), 0 \right\} \\
& \leq \sum_{s_{-i} \in S_{-i} \setminus A(s_i, s'_i)} p_i(s_{-i} | s_i) \max \left\{ \max_{j \neq i, p_j \in \Pi, u_j \in U_j} \frac{M_{i,1}}{\kappa_j(u_j)} \sum_{\theta_0} u_j(\theta_0) (p_j(\theta_0 | s) - p_j(\theta_0 | (s'_i, s_{-i}))), 0 \right\} \\
& \leq \mu_i(s_i, s'_i, p_i) \sum_{s_{-i} \in S_{-i} \setminus A(s_i, s'_i)} p_i(s_{-i} | s_i) \max_{j \neq i, p_j \in \Pi, u_j \in U_j} \left\{ \frac{M_{i,1}}{\kappa_j(u_j)} \sum_{\theta_0 \in \Theta_0} u_j(\theta_0) \right\} \\
& \leq \mu_i(s_i, s'_i, p_i) \cdot M_{i,1} \cdot M_{i,2} \leq \mu_i(p_i) \cdot M_{i,1} \cdot M_{i,2}.
\end{aligned}$$

Overall, we have shown that

$$E_{p_i}[\phi_i(s_i, s'_i, v_i; \theta_{-i}) | s_i] \leq M_{i,1} (1 + M_{i,2}) \cdot \mu_i(p_i).$$

From the proof of Proposition 2, we know that agent  $i$ 's deviation losses are bounded below by  $\Lambda_i / (2\sqrt{\sum_{j \neq i} n_j})$ . Proposition 4 thus follows, and Proposition 3 is obtained letting

$$\delta_i = \frac{\epsilon}{2\sqrt{\sum_{j \neq i} n_j} M_{i,1} (1 + M_{i,2})}.$$

## REFERENCES

1. Aghion, P., D. Fudenberg, R. Holden, T. Kunimoto and O. Tercieux (2012), "Subgame-Perfect Implementation Under Information Perturbations," *mimeo*, Harvard University.
2. Artemov, G., T. Kunimoto and R. Serrano (2011) "Robust Virtual Implementation with Incomplete Information: Towards a Reinterpretation of the Wilson Doctrine," *mimeo*.
3. Bergemann D. and S, Morris (2005), "Robust Mechanism Design," *Econometrica*, 73, 1521-1534.
4. Bergemann, D. and S. Morris (2009a) "Robust Implementation in Direct Mechanisms," *Review of Economic Studies*, 76, 1175-1204.
5. Bergemann, D. and S. Morris (2009b) "Robust Virtual Implementation," *Theoretical Economics*, 4(1), 45-88
6. Bergemann, D. and S. Morris (2011) "Robust Implementation in General Mechanisms," *Games and Economic Behavior*, 71(2), 261-281.
7. Bergemann D. and S, Morris (2012), *Robust Mechanism Design*. World Scientific Publishing, Singapore
8. Chassang, S. (2011), "Calibrated Incentive Contracts," *mimeo*, Princeton.
9. Cremer, J. and R.P. McLean (1985), "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent," *Econometrica*, 53(2), 345-361
10. Cremer, J. and R.P. McLean (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56(6), 1247-1257.
11. Dasgupta, P. and E, Maskin (2000), "Efficient Auctions," *Quarterly Journal of Economics*, 115(2), 341-388.
12. Hartline, J. (2011), "Approximate Economic Design," *mimeo*, Northwestern.
13. Heifetz, A. and Z. Neeman (2006), "On the Generic (Im)possibility of Full Surplus Extraction," *Econometrica*, 74, 213-233.

14. Jehiel, P., M. Meyer-ter-Vehn, B. Moldovanu, and W.R. Zame, "The Limits of Ex Post Implementation," *Econometrica*, 74(3), 585-610.
15. McAfee, P., J. McMillan, and P. Reny (1989), "Extracting the Surplus in the Common-Value Auction," *Econometrica*, 57(6), 1451-1459.
16. McAfee, P. and P. Reny (1989), "Correlated Information and Mechanism Design," *Econometrica*, 60(2), 395-421.
17. McLean, R. and A. Postlewaite (2002), "Informational Size and Incentive Compatibility," *Econometrica*, 70, 2421-2453.
18. McLean, R. and A. Postlewaite (2004), "Informational Size and Efficient Auctions," *Review of Economic Studies*, 71, 809-827.
19. Milgrom, P. (2004), *Putting auction theory to work*. Cambridge, U.K.: Cambridge University Press.
20. Mueller, C. (2011) "Robust Virtual Implementation under Common Strong Belief in Rationality" *mimeo*, Carnegie Mellon University.
21. Oury, M. and O. Tercieux, "Continuous Implementation," *Econometrica*, forthcoming.
22. Neeman, Z. (2004), "The Relevance of Private Information in Mechanism Design," *Journal of Economic Theory*, 117, 155-177.
23. Penta, A. (2011) "Robust Dynamic Mechanism Design," *mimeo*, University of Wisconsin-Madison.
24. Penta, A. (2012a) "Higher Order Uncertainty and Information: Static and Dynamic Games," *Econometrica*, 80(2), 631-660.
25. Penta, A. (2012b) "On the Structure of Rationalizability on Arbitrary Spaces of Uncertainty," *mimeo*, University of Wisconsin-Madison.
26. Perry, M. and P. J. Reny (2002), "An Efficient Auction," *Econometrica* 70, 1199-1212.
27. Postlewaite, A. and D. Schmeidler (1986), "Implementation in Differential Information Economies," *Journal of Economic Theory* 39, 14-33.
28. Schmeidler, D. (1989), "Subjective Probability and Expected Utility Without Additivity," *Econometrica*, 57, 571-587.

29. Weinstein, J. and M. Yildiz, (2007) “A Structure Theorem for Rationalizability With Application to Robust Predictions of Refinements,” *Econometrica*, 75, 365–400.
30. Weinstein, J. and M. Yildiz, (2012) “A Structure Theorem for Rationalizability in Infinite Horizon Games,” *mimeo*, MIT.
31. Yamashita, T. (2012), “A Necessary Condition on Robust Implementation: Theory and Applications,” *mimeo*, Toulouse School of Economics.