Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Economic Theory 117 (2004) 55–77

JOURNAL OF Economic Theory

http://www.elsevier.com/locate/jet

# The relevance of private information in mechanism design

Zvika Neeman[a,b,*]

[a] *Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA*
[b] *Center for Rationality and Department of Economics, Hebrew University of Jerusalem, Jerusalem 91904, Israel*

Received 11 July 2001; final version received 17 September 2003

**Abstract**

Recent results in mechanism design show that as long as agents have correlated private information and are sufficiently risk neutral, it is possible to design mechanisms that leave agents with arbitrarily small information rents. We show that these full-rent-extraction results hinge on the implicit assumption that the agents' beliefs uniquely determine their preferences. We present an example of the voluntary provision of a public good in which this assumption is relaxed, and consequently, even in highly correlated environments, if agents' beliefs do not uniquely determine their preferences, then the extraction of the agents' entire information rents is impossible.

*JEL classification:* D44; D82; H41

*Keywords:* Surplus extraction; Information rents; Mechanism design; Private information; Correlated information; Public goods

## 1. Introduction

Casual observation as well as economic intuition suggest that possession of relevant private information confers a positive rent. This basic insight is at odds with

[*]Correspondence address: Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA. Fax: +1-617-353-4449.
*E-mail address:* zvika@bu.edu.
*URL:* http://www.people.bu.edu/zvika/.

0022-0531/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jet.2003.09.001

a number of recent results in mechanism design literature.[1] These results show that for sufficiently risk neutral agents, as long as each agent's private information (type) is not entirely uninformative about other agents' private information, then in every mechanism design environment of interest, including auctions, optimal taxation, regulation, monopolistic price discrimination, trade, or bargaining under asymmetric information, and public good provision, it is possible to design mechanisms that leave agents with arbitrarily small information rents. In a Bayesian context, this implies that if sufficiently risk neutral agents have correlated private information (types), then their information rents can be almost fully extracted.

These full rent extraction results imply that, at least as far as normative economic analysis is concerned, the agents' private information is irrelevant because it is possible to implement any outcome as if the agents' private information were commonly known. This implication has made several economists uncomfortable. McAfee and Reny [14, p. 400], for example, write that these results "cast doubt on the value of the current mechanism design paradigm as a model of institutional design."

Perhaps in response to these concerns, several authors have argued that the conditions that are imposed in order to obtain these full rent extraction results, while standard in many applications, are nevertheless very strong. Crémer and McLean [7] suggested that full rent extraction is not robust to the introduction of risk aversion or limited liability constraints, and emphasized the dependence of these results on the common prior assumption. Following their suggestion, Robert [21] showed that for any given auction mechanism, when agents are risk averse or face limited liability constraints, the function that relates the common prior to the seller's profit and to total surplus (and hence also to the sum of information rents captured by the agents) is continuous in the prior. Since it is known that agents do obtain positive information rents in independent environments, Robert concluded that full information rent extraction also fails in "nearly independent" environments with risk averse agents or agents that face limited liability constraints. More recently, Laffont and Martimort [12] have established the continuity of the mechanism's outcome function also for environments with risk-neutral agents who are not constrained by limited liability, but who may form collusive coalitions. Intuitively, the reason that full rent extraction fails under these circumstances is that the auction mechanisms that extract the full buyers' rent rely on lotteries whose variance increases to infinity at independence. Thus, in nearly independent environments, mechanisms that rely on such lotteries violate the buyers' limited liability or participation constraints. Because these lotteries also prescribe payments to and from agents that strongly depend on the actions of other agents, mechanisms that rely on such lotteries are highly susceptible to collusion among the agents, and fail in nearly independent environments where these payments are large.

In this paper we offer a different explanation for the apparent inconsistency between full rent extraction results and the intuition about the value of exclusive

[1] See [6,7,10,14,8, Section 7.6.1], and the references therein. For a general formulation of this result, which allows for a continuum of multidimensional, mutually payoff relevant, agents' types, see [11].

private information. We show that full rent extraction hinges on an implicit assumption about the nature of the relationship between agents' beliefs and agents' preferences, or more specifically, on the assumption that the agents' beliefs uniquely "determine" their preferences. More specifically, full rent extraction implicitly hinges on the assumption that the (common) prior assigns probability zero to the possibility that two different types of the same agent hold identical beliefs but two different preferences (for simplicity, we focus our attention on the "private values" case). We show that if this assumption fails to be satisfied, then a subtle channel opens up through which some independence among preferences and beliefs creeps into models with otherwise correlated types. Consequently, when this assumption is relaxed, risk-neutral agents who are not subject to limited liability constraints and who do not collude, do in fact obtain positive information rents when they have exclusive access to relevant private information, even though their information may be correlated.[2]

In order to appreciate more fully the claim that is made in this paper, it is useful to briefly describe the general context in which it is made. Most applications of the mechanism design paradigm include the assumption that each agent's private information is independent of the private information of other agents. This assumption implies that different types of the same agent necessarily hold the same identical belief about the state of the world, or in other words, that agents' beliefs are fixed and therefore also commonly known.[3] As mentioned above, it can be shown that if agents' types are independent, then they can generally obtain a positive information rent.

An important insight of Crémer and McLean [6,7] is that it is much more plausible to assume that agents' private information *is* informative about other agents' private information and that agents with different privately known preferences *are* likely to hold different privately known beliefs, and vice versa. For example, in the context of public good provision, it is likely that agents with high willingness to pay for the public good would be inclined to believe that other agents are likely to have high willingness to pay too. In the model studied by Crémer and McLean [7], for example, it is assumed that agents' "characteristics" are generated from some commonly known prior, agents' willingness to pay are a function of their respective characteristics, and agents' beliefs are obtained by conditioning the common prior

[2] The impossibility (of full rent extraction) result in this paper is thus different in nature from the impossibility result in [9] who assumed that agents' private information is independent. See also [16].

[3] Intuitively, if we denote the prior by Pr, the random variables that describe $i$'s and $j$'s types by $\tilde{t}^i$ and $\tilde{t}^j$, and their realizations by $t^i$ and $t^j$, respectively, then independence implies that

$$\Pr(\tilde{t}^i = t^i, \tilde{t}^j = t^j) = \Pr(\tilde{t}^i = t^i)\Pr(\tilde{t}^j = t^j).$$

Type $t^i$'s beliefs about $j$'s type are therefore given by

$$\begin{aligned}\Pr(\tilde{t}^j = t^j | \tilde{t}^i = t^i) &= \frac{\Pr(\tilde{t}^i = t^i, \tilde{t}^j = t^j)}{\Pr(\tilde{t}^i = t^i)}\\ &= \Pr(\tilde{t}^j = t^j),\end{aligned}$$

which is independent of $i$'s type.

on the agents' own characteristics. Consequently, in their model (as well as in all the other models where full rent extraction is possible; see footnote 1), agents with different willingness to pay "generically" hold different beliefs and agents' beliefs generically uniquely "determine" their willingness to pay. It thus appears that when the less plausible assumption that agents' types are independent is replaced with the more plausible assumption that agents' types are correlated, it becomes generically possible to extract the agents' almost entire information rents.[4]

Note however that whereas to a third party the uncertainty about an agent's type is in fact "two dimensional"—including both uncertainty about the agent's willingness to pay and uncertainty about the agent's belief—models in which beliefs uniquely determine willingness to pay and consequently full information rent extraction is generically possible, only allow for "one-dimensional" uncertainty. In contrast, in this paper we take the position that while it is certainly plausible that a certain belief of agent *i* would make a certain willingness to pay, say $v^i$, more likely, it is conceivable that *i* could possibly have a different willingness to pay $v^{i\prime} \neq v^i$, too. As we show, this implies that agents do retain positive information rents.

Although the ideas presented here apply to any mechanism design problem, for simplicity we consider the example of the voluntary provision of a public good.[5] We show that even in highly correlated environments, if agents' beliefs do not determine their preferences, then the extraction of the agents' entire informational rents is impossible. Free-riding cannot be prevented and as a consequence the final outcome may be inefficient regardless of the number of agents involved. Furthermore, as the number of agents involved increases, the probability that the public good can be voluntarily provided decreases (inefficiently) to zero.

The rest of the paper proceeds as follows. In Section 2 we describe a simple public goods example with two agents which illustrates the differences between the assumptions made and the results obtained in the full rent extraction literature, in the rest of mechanism design literature, and in this paper. A more general model is presented in Section 3 where we show that the impossibility of extracting the full rent implies that the probability of providing the public good decreases to zero as the number of agents involved increases. Section 4 concludes. All proofs are relegated to the appendix.

## 2. Example

Consider the following situation. Two risk-neutral agents need to decide about the probability of providing a certain indivisible public good. The cost of providing the

[4] The first to suggest that the presence of correlation among agents' types may facilitate the extraction of the agents' entire information rents was Myerson [19].

[5] Auriol and Laffont [3] consider another example with correlated types where full information rents cannot be extracted. Their model is the only other example we are familiar with where beliefs do not determine preferences.

public good is $C>0$. If the public good is provided with probability $p$, then the agents must pay together an expected sum of $pC$, or else the good cannot be provided.[6] We also require the decision about providing the public good to be *voluntary*. That is, each agent retains the right to withhold its contribution (with the understanding that doing so may decrease the probability of provision). In the context of this example, full extraction of the agents' information rents implies that the public good should be provided if and only if it is efficient to do so, and the agents should be made to pay the full value they obtain from the provision of the public good. In particular, each agent should be prevented from free-riding on the other agent's contribution.

Suppose that each of the two agents is equally likely to be either one of the following two types: low ($L$) with willingness to pay $v(L)=0$ for the public good and high ($H$) with willingness to pay $v(H)=1$. The payoff to agent $i$, $i \in \{1,2\}$, of type $t^i \in \{L,H\}$ when it pays $x^i$ and the public good is provided with probability $p$, is given by $p \cdot v(t^i) - x^i$. Agents know their own types and seek to maximize their expected payoff given their beliefs, and everything above is commonly known among the agents.

It can be shown that if agents' types are independent, which in this simple example implies that they must both believe that the other agent is equally likely to be of either type, and the cost $C$ of providing the public good is between $\frac{2}{3}$ and 1, then it is impossible to provide the public good with probability one whenever it is efficient to do so.[7] Intuitively, this result, which holds whenever agents' types are independent and costs are neither too high nor too low, is a consequence of the fact that if the public good is provided whenever it is efficient to do so, or whenever at least one agent indicates it has a high willingness to pay for it, then agents cannot be prevented from free-riding on each other's contribution.

However, if agents' types are correlated, then ex-post efficiency may be attained in a dominant-strategy equilibrium. Suppose that $C<1$ and that each agent believes that the other agent is of the same type as itself with probability $\frac{2}{3}$ and of the other

[6] This (ex-ante) budget balance constraint is weaker than the requirement that agents' pay the sum $C$ whenever the public good is provided, or ex-post budget balance. It is the appropriate constraint for situations in which the two agents have access to a well-functioning credit market where they can insure themselves against not having sufficient funds to pay for the public good in some cases in return for the surplus they collect in other cases. In any case, because we obtain an impossibility result, requiring ex-ante budget balance only strengthens it.

[7] For $\frac{2}{3}<C\leqslant 1$ the optimal mechanism assumes the following form: both agents report their types; the probability of providing the public good as a function of the agents' reports is given by $p(L,L)=0$, $p(H,H)=1$, and $p(L,H)=p(H,L)=\rho<1$; an agent who reports $L$ pays nothing, and an agent who reports $H$ pays $\frac{C}{2}$ if the other agent reports $H$, and $\rho C$ if the other agent reports $L$. The parameter $\rho$ is set such that the incentive compatibility constraint for type $H$ is binding.

For $C\leqslant\frac{2}{3}$ the cost is so low that free-riding on the other agent's contribution is not worth the reduction in the probability of provision from 1 to $\frac{1}{2}$; the optimal mechanism is ex-post efficient in this case. For $C>1$, the public good should be provided if and only if both agents have a high valuation for it. This makes both agents pivotal, which implies that efficient provision is possible.

type with probability $\frac{1}{3}$.[8] We call this environment *environment* 1. Consider the following Clarke–Groves mechanism. Agents are required to report their types. Their reports are denoted $\hat{t}^1$ and $\hat{t}^2$, respectively. The probability of provision and agents' payments are given by

$$p(\hat{t}^1, \hat{t}^2) = \begin{cases} 0, & \hat{t}^1 = \hat{t}^2 = L, \\ 1, & \hat{t}^1 = L, \hat{t}^2 = H, \\ 1, & \hat{t}^1 = H, \hat{t}^2 = L, \\ 1, & \hat{t}^1 = \hat{t}^2 = H, \end{cases} \qquad x^i(\hat{t}^i, \hat{t}^j) = \begin{cases} \frac{1}{2} - \frac{C}{3}, & \hat{t}^i = \hat{t}^j = L, \\ \frac{2C}{3} - \frac{1}{2}, & \hat{t}^i = L, \hat{t}^j = H, \\ \frac{1}{2} + \frac{2C}{3}, & \hat{t}^i = H, \hat{t}^j = L, \\ \frac{2C}{3} - \frac{1}{2}, & \hat{t}^i = \hat{t}^j = H. \end{cases}$$

It can be easily verified that this mechanism is ex-post efficient, ex-ante budget balanced, and induces truthful reporting as a dominant strategy. The only problem with this mechanism is that it violates the voluntary participation constraint of type $L$. While it gives type $H$ a positive expected payoff of $\frac{7}{6} - \frac{2C}{3}$, the expected payoff it gives to type $L$ is $-\frac{1}{6}$. However, the fact that the two types have different beliefs implies that this problem can be easily fixed by simply adding to each agent's payment the following lottery: if the other agent reports $L$, then the agent receives an amount of $\frac{3}{2} - \frac{2C}{3}$, but if the other agent reports $H$, then the agent has to pay an additional amount of $\frac{5}{2} - \frac{4C}{3}$. Because the outcome of this lottery depends only on the other agent's report, adding this lottery to agents' payments does not change their incentives – truthfully reporting their type is still a dominant strategy. But because

$$\begin{array}{cc} & \begin{array}{cc} L & H \end{array} \\ \begin{array}{c} L \\ H \end{array} & \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \end{array} \begin{pmatrix} -\frac{3}{2} + \frac{2C}{3} \\ \frac{5}{2} - \frac{4C}{3} \end{pmatrix} = \begin{pmatrix} -\frac{1}{6} \\ \frac{7}{6} - \frac{2C}{3} \end{pmatrix}$$

the expected payoff from participating in the mechanism becomes zero for both agents' types.

The intuition for what makes full rent extraction possible in this example is that the fact that different types have different beliefs facilitates the making of targeted monetary transfers to specific agents' types. A perhaps more straightforward way in which this could be seen is the following. It is always possible to construct participation fees that induce risk-neutral agents to reveal their beliefs at zero cost to the agents themselves. In the example above, this can be simply done by asking agents to choose one of the two lotteries $\tilde{L}_L$ and $\tilde{L}_H$, and then report their types, paying one agent according to the other agent's report. The lotteries should be

[8] The common prior that describes the joint distribution of agents' types that induces these beliefs is $\Pr(L, L) = \Pr(H, H) = \frac{1}{3}$ and $\Pr(L, H) = \Pr(H, L) = \frac{1}{6}$. Note that the two agents' two types are equally likely.

designed so that they solve the following equation:

$$\begin{array}{cc} & \begin{array}{cc} L & H \end{array} \\ \begin{array}{c} L \\ H \end{array} & \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \end{array} \begin{pmatrix} \tilde{L}_L(L) & \tilde{L}_H(L) \\ \tilde{L}_L(H) & \tilde{L}_H(H) \end{pmatrix} = \begin{pmatrix} 0 & \lambda \\ \lambda & 0 \end{pmatrix}$$

for some large $\lambda > 0$. It can be readily verified that if an agent of either type chooses the lottery designed for its type, its expected payment is zero, but if it chooses the other lottery, then its expected payment is $\lambda > 0$. What makes full rent extraction possible is that the agents' beliefs determine their willingness to pay. So, after the agents have been costlessly induced to reveal their beliefs/willingness to pay, it is possible to implement the ex-post efficient decision and simply ask the agents to pay their expected gains when this decision is made.[9]

Obviously, neither this mechanism nor the one above can work if agents' beliefs do not uniquely determine their willingness to pay. To see that consider *environment* 2 below. *Environment* 2 is constructed so that it is similar to *environment* 1 except that instead of two possible agents' types, there are three: $L$, $H$ and $LH$. An agent of type $LH$ has the same low willingness to pay as an agent of type $L$ but the same belief as an agent of type $H$. Specifically, $v(L) = v(LH) = 0$ and $v(H) = 1$; the belief of an agent of type $L$ is that the other agent is of type $L$, $LH$, and $H$ with probabilities $b(L) = (\frac{2}{3+a}, \frac{a}{3+a}, \frac{1}{3+a})$, respectively; and the belief of agents of types $H$ and $LH$ is that the other agent is of type $L$, $LH$, and $H$ with probabilities $b(H) = b(LH) = (\frac{1}{3+2a}, \frac{2a}{3+2a}, \frac{2}{3+a})$, respectively. The parameter $a$, which is assumed to be nonnegative, affects both the agents' beliefs and the relative likelihood of type $LH$ with respect to type $H$ ($LH$ is $a$ as likely as $H$). The smaller $a$ is, the closer environment 2 is to environment 1; when $a = 0$, environment 2 coincides with environment 1.[10] Environment 2 is an environment where agents' beliefs do not determine their willingness to pay for the public good. An agent who holds the belief $b(L)$ must have willingness to pay zero, but an agent who holds the belief $b(H)$ could have willingness to pay one (with probability $\frac{1}{1+a}$) or willingness to pay zero (with probability $\frac{a}{1+a}$).

There is an optimal mechanism for environment 2 that is similar to the second optimal mechanism described for environment 1.[11] It also requires agents to choose one of two lotteries whose payments depend on the other agent's choice. As before,

[9] This mechanism is simpler than the one described above, but because choosing the "right" lottery is not a dominant strategy, it is also more susceptible to collusion.

[10] The common prior that describes the joint distribution of agents' types is $\Pr(L,L) = \Pr(H,H) = \frac{2}{6+6a+2a^2}$, $\Pr(L,H) = \Pr(H,L) = \frac{1}{6+6a+2a^2}$, $\Pr(LH,H) = \Pr(H,LH) = \frac{2a}{6+6a+2a^2}$, $\Pr(LH,L) = \Pr(L,LH) = \frac{a}{6+6a+2a^2}$, and $\Pr(LH,LH) = \frac{2a^2}{6+6a+2a^2}$.

[11] It is possible to establish the optimality of this mechanism by formulating the problem as a linear programming problem of maximizing the expected sum of agents' payments subject to incentive compatibility and individual rationality constraints.

the lotteries are constructed so that they costlessly induce the agents to reveal their beliefs. However, in environment 2 knowing an agent's belief is not sufficient to extract the agent's full information rent. If both agents reveal they hold the belief $b(L)$, then the public good is not provided and no payments are made. No rents can be extracted in this case anyway, and ex-post efficiency is easily achieved. If the agents reveal that they hold the beliefs $b(L)$ and $b(H)$, respectively, then the agent who reported the belief $b(H)$ is asked to supply the public good on its own. The fact that there is exactly one such agent ensures that it would make the ex-post efficient decision. However, when both agents reveal that they hold the belief $b(H)$, then because conditional on this revelation, each agent's willingness to pay is independent of the other's (each is equal to 1 with probability $\frac{1}{1+a}$ and to 0 with probability $\frac{a}{1+a}$), free-riding cannot be prevented, and ex-post efficiency cannot be achieved.[12] It is possible to show that the implied inefficiency decreases to zero with $a$, that is, as environment 2 approaches environment 1 where beliefs do "determine" willingness to pay. But only when $a = 0$, when *environment* 2 coincides with *environment* 1, is it possible to extract the entire surplus from the agents and implement the ex-post efficient decision.

## 3. Voluntary public good provision

Consider an economy with a private and a public good and $n$ agents. Each agent $i$ is characterized by its (nonnegative) willingness to pay for the public good, and by its beliefs about other agents' willingness to pay for the public good, other agents' beliefs about other agents' willingness to pay for the public good, other agents' beliefs about other agents' beliefs, and so on. The agents' infinite hierarchy of beliefs about beliefs about beliefs … about the relevant variables in the economy may be encoded in an interactive belief system (Aumann and Brandenburger, [2]) which we adapt for our purposes.[13] Such an interactive belief system consists of:

- for each agent $i$, a set $T^i$ of agent $i$'s types;

and for each type $t^i$ of $i$:

- a probability distribution $b^i(t^i) \in B^i$ on the set $T^{-i}$ of $(n-1)$-tuples of types of the other agents ($t^i$'s belief),[14] and
- a willingness to pay $v^i(t^i) \in V^i \subseteq \mathbb{R}_+$ ($t^i$'s willingness to pay for the public good).

[12] The optimal mechanism in this case is similar to the one described in footnote 7 for the case of independent agents' types.

[13] Any such hierarchy that satisfies minimal coherency requirements may be encoded in this way [15] (see also [5]).

[14] Aumann and Brandenburger [2] call $b^i(t^i)$, $t^i$'s theory.

An interactive belief system contains an explicit description of the agents' beliefs. It encompasses within it the "traditional model" employed in mechanism design literature (see, e.g., [6,7,19]) in which each agent is characterized by its willingness to pay $v^i \in V^i$, and where a common prior describes the joint distribution of the agents' willingness to pay. In this "traditional model," each agent is assumed to know its own willingness to pay and to obtain its beliefs about the other agents' types by conditioning the common prior on its own willingness to pay, using Bayes' rule. Because in the traditional model, the agents' beliefs can be obtained in a straightforward way from the common prior, their explicit description is redundant and consequently not usually specified in the description of the information structure.

For simplicity, we assume that the type spaces $T^i$ and hence also the belief and willingness to pay spaces $B^i$ and $V^i$, are finite. Each set $V^i$ is assumed to include the number zero, and to be uniformly bounded. That is, we assume the existence of some $\bar{v} < \infty$ such that every $v^i \leqslant \bar{v}$. We also assume that different agent's types are indeed different. That is, if $t^i, t^{i\prime} \in T^i$ are different, then either $v^i(t^i) \neq v^i(t^{i\prime})$ or $b^i(t^i) \neq b^i(t^{i\prime})$. In other words, each type determines a belief and a willingness to pay, and each belief and willingness to pay determine a (possibly null) type.

Let $T \equiv T^1 \times \cdots \times T^n$. The members $t = (t^1, \dots, t^n)$ of $T$ are called states of the world. An event is a subset $E$ of $T$. Agent $i$ is said to know an event $E$ at $t = (t^1, \dots, t^n)$ if $b^i(t^i)$ assigns probability one to the set $\{t^{-i} \in T^{-i} : (t^i, t^{-i}) \in E\}$.[15] Thus in every state of the world $t$, each agent knows its own type $t^i$, its beliefs about other agents' types $b^i(t^i)$, and its willingness to pay for the public good $v^i(t^i)$.

Each type $t^i$'s belief, $b^i(t^i)$, has domain $T^{-i}$; define an extension $p(\cdot; t^i)$ of $t^i$'s belief to $T$ as follows. If $E$ is an event, then $p(E; t^i)$ is defined as the probability that $t^i$'s belief assigns to the set $\{t^{-i} \in T^{-i} : (t^i, t^{-i}) \in E\}$.

A probability distribution Pr on $T$ is called a common prior if for every agent $i$ and agent $i$'s type $t^i$, the conditional distribution of Pr given $t^i$ is $p(\cdot; t^i)$. We assume that a common prior exists for this economy.[16]

The interactive belief system for this economy may be embedded in a probability space $(T, 2^T, \Pr)$. Consequently, functions defined on $T$ may be viewed as random variables in probability theory. In particular, agent $i$'s type may be viewed as the random variable $\tilde{t}^i : T \to T^i$ that projects the set $T$ into the set $T^i$, and agent $i$'s willingness to pay and belief may be viewed as random variables $\tilde{v}^i = v^i(t^i) : T \to V^i$ and $\tilde{b}^i = b^i(t^i) : T \to B^i$, respectively.

Note that agents' private information (types) may be informative about or correlated with other agents' private information in the following two senses: (1) every agent's willingness to pay $\tilde{v}^i$ may be correlated with the agent's own belief $\tilde{b}^i$.

[15] More generally, individual $i$ can be said to $p$-believe an event $E$ at $t$ if $b^i(t^i)$ assigns probability at least $p$ to the set $\{t^{-i} \in T^{-i} : (t^i, t^{-i}) \in E\}$ [17].

[16] Without a common prior, agents may assign zero probability to the true state of the world and other pathological phenomena may occur [15]. For this reason, the existence of a common prior is assumed in much of applied game theory.

In other words, it may be likely that an agent with a certain willingness to pay $v^i$ will also hold a certain belief $b^i$ and vice versa; and (2) the prior Pr that describes the joint distribution of agents' types may exhibit correlation among different agents' types. However, as the next lemma demonstrates, certain independence among the agents' willingness to pay and beliefs must remain.[17]

**Lemma 1.** *The agents' willingness to pay for the public good are independent conditional on the profile of agents' beliefs. That is, for every two profiles of willingness to pay* $v \in V^1 \times \cdots \times V^n$ *and beliefs* $b \in B^1 \times \cdots \times B^n$,

$$\Pr(\tilde{v}^1 = v^1, \ldots, \tilde{v}^n = v^n | \tilde{b}^1 = b^1, \ldots, \tilde{b}^n = b^n)$$
$$= \prod_{i=1}^{n} \Pr(\tilde{v}^i = v^i | \tilde{b}^1 = b^1, \ldots, \tilde{b}^n = b^n)$$

*whenever these two expressions are well defined.*

Note that because for every two events $A$ and $B$, $\Pr(A, B|A) = \Pr(B|A)$ whenever the two are well defined, the lemma implies that the agents' types are independent conditional on the agents' beliefs.

The next assumption simplifies the proof of the main theorem below.

**Assumption A.** There exists some $\varepsilon > 0$ such that

$$\Pr(\tilde{b}^i = b^i) \geqslant \varepsilon$$

and

$$\Pr(\tilde{v}^i = v^i | \tilde{b}^i = b^i) \geqslant \varepsilon$$

for every agent $i$ and agent $i$'s belief $b^i \in B^i$ and willingness to pay $v^i \in V^i$.

Assumption A ensures that agents' beliefs do not uniquely determine their willingness to pay which we formally define as follows.[18]

**Definition.** For every agent $i$ and agent $i$'s belief $b^i \in B^i$, we say that *agent $i$'s belief $b^i$ uniquely determines its willingness to pay* if

$$\Pr(\tilde{v}^i = v^i | \tilde{b}^i = b^i) > 0 \text{ for some } v^i \in V^i \Rightarrow \Pr(\tilde{v}^i = v^{i\prime} | \tilde{b}^i = b^i) = 0$$
$$\text{for every } v^{i\prime} \neq v^i.$$

[17] This lemma is trivially satisfied in [6,7,14] because in these models, conditional on the profile of agents' beliefs, each agent's willingness to pay is a constant random variable.

[18] Assumption A could be replaced with the weaker assumption that agents' beliefs do not determine their willingness to pay provided it is assumed, in addition, that the cost of providing the public good is uniformly larger than the expected sum of the agents' lowest possible willingness to pay that is consistent with any profile of beliefs.

If every agent $i$'s belief $b^i$ uniquely determines its willingness to pay, then we say that *agent $i$'s beliefs determine its willingness to pay.*

**Remark 1.** If agent $i$'s own beliefs do not determine its willingness to pay, then the entire profile of agents' beliefs cannot determine it either. The fact that agent $i$'s beliefs do not determine its willingness to pay implies that there exist two different types of agent $i$, $t^i$ and $t^{i\prime}$, that hold the same belief but two different willingness to pay. If it was commonly known that the profile of agents' beliefs determined agent $i$'s willingnesses to pay, then the two types $t^i$ and $t^{i\prime}$ would have had to hold different beliefs about the beliefs of other agents, contradicting the assumption that they hold the same belief.

**Remark 2.** Assumption A, and more generally, the property that beliefs do not determine willingness to pay, fails to be satisfied in the entire literature that describes the possibility of full rent extraction (see footnote 1). This failure is generic in finite versions of the "traditional model" mentioned above [6,7], but whether or not beliefs generically determine preferences in infinite interactive belief systems remains an open question. For additional discussion of this issue, see the concluding section.

Agents are assumed to have von Neumann–Morgenstern utility functions that are given by $p \cdot v^i(t^i) - x^i$, respectively, where $x^i$ denotes agent $i$'s payment and $p$ denotes the probability that the public good is provided.[19] The cost of providing the public good is denoted by $C(n)$. We assume that $C(n) \geqslant c \cdot n$ for some $c > 0$.

We study the probability with which the public good can be provided in large economies such as those described above by voluntary and budget-balanced mechanisms. According to the *revelation principle* (e.g., [20]), no loss of generality is entailed by assuming that agents rely on voluntary and budget-balanced incentive compatible direct revelation mechanisms to provide the public good. A direct revelation mechanism, $\langle p, x \rangle$, is composed of a decision rule $p : T \to [0,1]$ that describes the probability that the public good is provided as a function of agents' reports of their types, and of payments $x = (x^1, \ldots, x^n)$ where for every $i$, $x^i : T \to \mathbb{R}$, describes agent $i$'s payment as a function of agents' reports of their types.

**Definition.** A direct revelation mechanism $\langle p, x \rangle$ is incentive compatible if it induces an equilibrium in which all the agents truthfully report their types. Namely, it is a best response for each agent to truthfully report its type provided everyone else does, or,

$$\sum_{t^{-i} \in T^{-i}} [v^i(t^i)p(t) - x^i(t)] b^i(t^i)(\{\tilde{t}^{-i} = t^{-i}\})$$
$$\geqslant \sum_{t^{-i} \in T^{-i}} [v^i(t^i)p(t^{i\prime}, t^{-i}) - x^i(t^{i\prime}, t^{-i})] b^i(t^i)(\{\tilde{t}^{-i} = t^{-i}\})$$

for every agent $i$ and agent $i$'s types $t^i, t^{i\prime} \in T^i$.

[19] The probability $p$ can also be interpreted as the quantity of the public good provided.

A mechanism for providing the public good is voluntary or individually rational if agents cannot be coerced to contribute to the public good. We interpret this requirement as implying that the mechanism must make agents weakly better off relative to the situation in which they do not contribute anything and the public good is not provided.

**Definition.** A direct revelation mechanism $\langle p, x \rangle$ is voluntary or individually rational if

$$\sum_{t^{-i} \in T^{-i}} [v^i(t^i)p(t) - x^i(t)]b^i(t^i)(\{\tilde{t}^{-i} = t^{-i}\}) \geqslant 0$$

for every agent $i$ and agent $i$'s type $t^i \in T^i$.

We also require that the mechanism be budget-balanced. The sum of agents' payments for the public good must be sufficient to cover the costs of providing it. We impose a weak version of this constraint that only requires the expected sum of agents' payments to be larger than or equal to the cost of providing the public good.

**Definition.** A direct revelation mechanism $\langle p, x \rangle$ is (ex-ante) budget-balanced if

$$E[p(\tilde{t})]C(n) \leqslant \sum_{i=1}^{n} E[x^i(\tilde{t}^i)].$$

We obtain the following main result.

**Theorem.** *The probability that the public good is provided by any individually rational and budget-balanced mechanism tends to zero as the number of agents, n, increases.*

In fact, we prove a somewhat stronger result. As we show in the appendix, the rate at which the probability of providing the public good converges to zero is asymptotically proportional to $n^{-\frac{1}{12}}$. The proof of the theorem builds on the proof that appears in the appendix of Mailath and Postlewaite [13].[20] It is based on the observation that a random variable such as the decision about whether or not to provide the public good, viewed as a point in a linear space, cannot have a high correlation with many members of an orthogonal basis for this space. Because by Lemma 1, the agents' types are independent conditional on their beliefs, it follows that for every possible profile of agents' beliefs $b = (b^1, \ldots, b^n)$, the expected probability that the public good is provided cannot be highly correlated with the expectations of many agents. This implies that when the number of agents is large,

[20] Mailath and Postlewaite [13] considered a similar model but assumed that agents' types are independent. In their model, the probability with which the public good is provided decreases to zero at the asymptotic rate $n^{-\frac{1}{4}}$. Recently, Al-Najjar and Smorodinsky [1] showed that the rate of convergence for the case of independent types is in fact proportional to $n^{-\frac{1}{2}}$.

then for most agents, telling the truth about their willingness to pay for the public good has a negligibly small effect on the probability the public good is provided. Incentive compatibility, individual rationality, and the fact that agents' beliefs do not determine their willingness to pay, then imply that most agents can ensure they will not pay anything for the public good by pretending that their willingness to pay is zero, without causing any significant change in the probability the public good is provided. In other words, when $n$ is large, most agents are not pivotal and therefore have an incentive to free-ride on other agents' contributions. Consequently, the group of agents is not capable of raising enough contributions to finance the public good, and the probability that the public good is provided decreases to zero.

The impossibility of constructing an effective mechanism for curbing agents' opportunistic behavior is obviously harmful for the group of agents. It implies that when the number of agents $n$ is large, the public good cannot be provided even when it is extremely inefficient not to do so. If, for example, all agents are ex-ante identical, $C(n) = cn$, and the expected willingness to pay of each agent is much larger than the cost per agent of providing the public good, $c$, then the ex-ante (expected) inefficiency per agent from not providing the public good tends to $E[v^i(\tilde{t}^i)] - c \gg 0$ as the number of agents in the economy increases.

## 4. Conclusion

By demonstrating the importance of the relationship between beliefs and willingness to pay, or preferences, our observations raise the question of how "likely" are beliefs to determine preferences in a "general" model of incomplete information. For the case of the universal type space with consistent beliefs [15], this question remains open. Recently, Bergemann and Morris [4] have showed that models in which beliefs do not determine preferences are dense in the universal type space. Their argument is quite simple. Fix a model in which beliefs do determine preferences. From each type $t^i$ of each agent $i$ create $|V^i|$ new types in the following way: with probability $1 - \varepsilon$ (for some small $\varepsilon > 0$), suppose that $t^i$ retains its original preferences or willingness to pay; and with probability $\varepsilon > 0$, suppose that $t^i$ has a different willingness to pay that is chosen randomly (and independently) from the set $V^i$. Update all the agents' types' beliefs to take account of this change.[21] Obviously, after the change, all the new types that are constructed from any type $t^i$ hold the same beliefs, but have different preferences, and hence, the agents' beliefs do not determine their preferences in the new model. And, by choosing $\varepsilon$ sufficiently small, the new model can be made arbitrarily close to the original model (for additional discussion of these issues, see also [18]).

Issues of generality notwithstanding, preferences and beliefs have traditionally been considered to be independent of one another in both economic and decision theory. This tradition presumably reflects the idea that the processes that generate

[21] Note that in the example presented in Section 2, types $LH$ and $H$ in environment 2 were created from type $H$ in environment 1 in this way.

utilities and beliefs are cognitively distinct and causally “independent,” or at least should be treated as such.

## Acknowledgments

I am grateful for the suggestions made by Itzhak Gilboa, Hsueh-Ling Huynh, Stephen Morris, Bob Rosenthal, Steve Tadelis, Shmuel Zamir, and seminar audiences at the Center for Rationality at the Hebrew University, Econometric Society Meeting in Montreal, Harvard/MIT, Northwestern Summer Microtheory Workshop, Penn, and Rice. An associate editor and two anonymous referees provided particularly useful comments. Financial support from the NSF under Grant SBR-9806832 is gratefully acknowledged.

## Appendix. Proofs

**Proof of Lemma 1.** The definition of $p(\cdot;t^i)$ implies that for every event $E \subseteq T^{-i}$, $p(F;t^i) = b^i(t^i)(E)$ where $F = \{t \in T : t^{-i} \in E\}$, for every type $t^i$ of agent $i$. The existence of a common prior implies that for every two events $E \subseteq T^{-i}$ and $F = \{t \in T : t^{-i} \in E\}$, if two different non-null types $t^i$ and $t^{i\prime}$ of agent $i$ hold the same beliefs $b^i(t^i) = b^i(t^{i\prime})$, then

$$\Pr(F|\tilde{t}^i = t^i) = p(F;t^i) = b^i(t^i)(E) = b^i(t^{i\prime})(E) = p(F;t^{i\prime}) = \Pr(F|\tilde{t}^i = t^{i\prime}).$$

Let $T^i(b^i(t^i)) = \{t^{i\prime} \in T^i : b^i(t^{i\prime}) = b^i(t^i)\}$ denote the set of types of agent $i$ who hold the same belief as $t^i$. Because $\{t' \in T : t^{i\prime} = t^i\} \subseteq \{t' \in T : b^i(t^{i\prime}) = b^i(t^i)\}$, $\Pr(F|\tilde{t}^i = t^{i\prime}, \tilde{b}^i = b^i(t^i)) = \Pr(F|\tilde{t}^i = t^{i\prime})$ for every event $F \subseteq T$ of the form $F = T^i \times E$, where $E \subseteq T^{-i}$, and $i$'s type $t^{i\prime} \in T^i(b^i(t^i))$. Thus, for every non-null type $t^i$ and event $F = T^i \times E$ such that $E \subseteq T^{-i}$,

$$\begin{aligned}
\Pr(F|\tilde{b}^i = b^i(t^i)) &= \sum_{t^{i\prime} \in T^i(b^i(t^i))} \Pr(F|\tilde{t}^i = t^{i\prime}, \tilde{b}^i = b^i(t^i)) \Pr(\tilde{t}^i = t^{i\prime}|\tilde{b}^i = b^i(t^i)) \\
&= \sum_{t^{i\prime} \in T^i(b^i(t^i))} \Pr(F|\tilde{t}^i = t^{i\prime}) \Pr(\tilde{t}^i = t^{i\prime}|\tilde{b}^i = b^i(t^i)) \\
&= \Pr(F|\tilde{t}^i = t^i) \sum_{t^{i\prime} \in T^i(b^i(t^i))} \Pr(\tilde{t}^i = t^{i\prime}|\tilde{b}^i = b^i(t^i)) \\
&= \Pr(F|\tilde{t}^i = t^i).
\end{aligned}$$

In particular, for every non-null type $t^i$ and vectors of beliefs and willingness to pay of agents different from $i$, $b^{-i} \in B^{-i}$ and $v^{-i} \in V^{-i}$, respectively,

$$\Pr(\tilde{b}^{-i} = b^{-i}, \tilde{v}^{-i} = v^{-i}|\tilde{t}^i = t^i) = \Pr(\tilde{b}^{-i} = b^{-i}, \tilde{v}^{-i} = v^{-i}|\tilde{b}^i = b^i(t^i))$$

or

$$\frac{\Pr(\tilde{t}^i = t^i, \tilde{b}^{-i} = b^{-i}, \tilde{v}^{-i} = v^{-i})}{\Pr(\tilde{t}^i = t^i)} = \frac{\Pr(\tilde{b}^i = b^i(t^i), \tilde{b}^{-i} = b^{-i}, \tilde{v}^{-i} = v^{-i})}{\Pr(\tilde{b}^i = b^i(t^i))}.$$

Similarly, for every non-null type $t^i$ and vector of beliefs $b^{-i} \in B^{-i}$,

$$\frac{\Pr(\tilde{t}^i = t^i, \tilde{b}^{-i} = b^{-i})}{\Pr(\tilde{t}^i = t^i)} = \frac{\Pr(\tilde{b}^i = b^i(t^i), \tilde{b}^{-i} = b^{-i})}{\Pr(\tilde{b}^i = b^i(t^i))}.$$

It follows that for every non-null type $t^i \in T^i$ and vectors of beliefs and willingness to pay $b^{-i} \in B^{-i}$, and $v^{-i} \in V^{-i}$, respectively,

$$\begin{aligned}
&\Pr(\tilde{v}^{-i} = v^{-i} | \tilde{t}^i = t^i, \tilde{b}^{-i} = b^{-i}) \\
&\quad = \frac{\Pr(\tilde{t}^i = t^i, \tilde{b}^{-i} = b^{-i}, \tilde{v}^{-i} = v^{-i})}{\Pr(\tilde{t}^i = t^i)} \cdot \frac{\Pr(\tilde{t}^i = t^i)}{\Pr(\tilde{t}^i = t^i, \tilde{b}^{-i} = b^{-i})} \\
&\quad = \frac{\Pr(\tilde{b}^i = b^i(t^i), \tilde{b}^{-i} = b^{-i}, \tilde{v}^{-i} = v^{-i})}{\Pr(\tilde{b}^i = b^i(t^i))} \cdot \frac{\Pr(\tilde{b}^i = b^i(t^i))}{\Pr(\tilde{b}^i = b^i(t^i), \tilde{b}^{-i} = b^{-i})} \\
&\quad = \Pr(\tilde{v}^{-i} = v^{-i} | \tilde{b}^i = b^i(t^i), \tilde{b}^{-i} = b^{-i}),
\end{aligned}$$

whenever these expressions are well defined, or, recalling that an agent's belief and willingness to pay determine its type (namely, $\tilde{t}^i = t^i$ if and only if $\tilde{v}^i = v^i(t^i)$ and $\tilde{b}^i = b^i(t^i)$),

$$\begin{aligned}
&\Pr(\tilde{v}^{-i} = v^{-i} | \tilde{v}^i = v^i(t^i), \tilde{b}^i = b^i(t^i), \tilde{b}^{-i} = b^{-i}) \\
&\quad = \Pr(\tilde{v}^{-i} = v^{-i} | \tilde{b}^i = b^i(t^i), \tilde{b}^{-i} = b^{-i}).
\end{aligned}$$

Now, for every profile $v \in V$ and $b \in B$,

$$\begin{aligned}
\Pr(\tilde{v} = v | \tilde{b} = b) &= \Pr(\tilde{v}^{-i} = v^{-i} | \tilde{v}^i = v^i, \tilde{b} = b) \Pr(\tilde{v}^i = v^i | \tilde{b} = b) \\
&= \Pr(\tilde{v}^{-i} = v^{-i} | \tilde{b} = b) \Pr(\tilde{v}^i = v^i | \tilde{b} = b),
\end{aligned}$$

whenever these expressions are well defined and the argument can be repeated for every $k \neq i$ to obtain

$$\begin{aligned}
\Pr(\tilde{v}^{-i} = v^{-i} | \tilde{b} = b) &= \Pr(\tilde{v}^{-i,j} = v^{-i,j} | \tilde{b} = b) \Pr(\tilde{v}^j = v^j | \tilde{b} = b) \\
&\;\;\vdots \\
&= \prod_{k \neq i} \Pr(\tilde{v}^k = v^k | \tilde{b} = b). \qquad \square
\end{aligned}$$

**Proof of the Theorem.** Recall that for every $n \in \mathbb{N}$, the economy consisting of $n$ agents may be embedded in a probability space $(T^1 \times \cdots \times T^n, 2^{T^1 \times \cdots \times T^n}, \Pr)$. The sets of agents' types, the common prior, and everything else in the economy may generally depend on the number of agents in the economy, $n$, but in order to simplify the notation, this dependence is not reflected in the notation below. Direct revelation mechanisms for providing the public good in the economy that consists of $n$ agents may thus be viewed as random variables $\tilde{p}: T^1 \times \cdots \times T^n \to [0,1]$ and $\tilde{x} = (\tilde{x}^1, \dots, \tilde{x}^n) : T^1 \times \cdots \times T^n \to \mathbb{R}^n$.

Let $\tilde{p}_b : \{t \in T^1 \times \cdots \times T^n : \tilde{b}(t) = b\} \to [0,1]$ denote the restriction of $\tilde{p}$ to the event where agents' beliefs are given by $b$, and let $E[\tilde{p}_b | \tilde{v}^i] : \{t \in T^1 \times \cdots \times T^n : \tilde{b}(t) = b\} \to [0,1]$ denote the random variable that describes the expected probability the public good is provided as perceived by agent $i$ with willingness to pay $\tilde{v}^i$ under the assumption that the profile of agents' beliefs is given by $b$. Similarly, let $\tilde{p}_{b^i}$ : $\{t \in T^1 \times \cdots \times T^n : \tilde{b}^i(t) = b^i\} \to [0,1]$ denote the restriction of $\tilde{p}$ to the event where agent $i$'s belief is given by $b^i$, and let $E[\tilde{p}_{b^i} | \tilde{v}^i] : \{t \in T^1 \times \cdots \times T^n : \tilde{b}^i(t) = b^i\} \to [0,1]$ denote the random variable that describes the expected probability the public good is provided as perceived by agent $i$ with willingness to pay $\tilde{v}^i$ and belief $b^i \in B^i$. Notice that because agents' types are characterized by their willingness to pay and beliefs, $E[\tilde{p}_{b^i(t^i)} | \tilde{v}^i = v^i(t^i)] = E[\tilde{p} | \tilde{t}^i = t^i]$ for every $t^i \in V^i$,

$$E[\tilde{p}_{b^i} | \tilde{v}^i] = \sum_{b \in B} \Pr(\tilde{b} = b | \tilde{b}^i = b^i) E[\tilde{p}_b | \tilde{v}^i] \tag{A.1}$$

and

$$\begin{aligned} E[E[\tilde{p}_{b^i} | \tilde{v}^i]] &= E[\tilde{p}_{b^i}] \\ &= \sum_{b \in B} \Pr(\tilde{b} = b | \tilde{b}^i = b^i) E[\tilde{p}_b]. \end{aligned} \tag{A.2}$$

The random variables $\tilde{x}^i_{b^i}$ and $E[\tilde{x}^i_{b^i} | \tilde{v}^i]$ are similarly defined.

The random variable $E[\tilde{p}_b | \tilde{v}^i]$ is a projection of $\tilde{p}_b$ into the space of random variables that are defined on $\{t \in T^1 \times \cdots \times T^n : \tilde{b}(t) = b\}$ and measurable with respect to $\tilde{v}^i$. We therefore have the following lemma.

**Lemma A.1.** *For every $n \in \mathbb{N}$, $i \in \{1, \dots, n\}$, and $b \in B$,*

$$\operatorname{cov}(\tilde{p}_b, E[\tilde{p}_b | \tilde{v}^i]) = \operatorname{cov}(E[\tilde{p}_b | \tilde{v}^i], E[\tilde{p}_b | \tilde{v}^i]) = \operatorname{Var}(E[\tilde{p}_b | \tilde{v}^i]).$$

**Proof.** For every two random variables $X$ and $Y$, $E[E[X|Y]] = E[X]$. Thus, for every two bounded random variables $X$ and $Y$,

$$\begin{aligned} \operatorname{cov}(X, E[X|Y]) &= E[(X - E[X])(E[X|Y] - E[E[X|Y]])] \\ &= E[(X - E[X])(E[X|Y] - E[X])] \\ &= E[XE[X|Y] + E[X]^2 - E[X]E[X|Y] - XE[X]] \\ &= E[XE[X|Y]] - E[X]^2. \end{aligned}$$

The fact that $E[X|Y]$ is measurable with respect to $Y$ implies that

$$E[XE[X|Y]] - E[X]^2 = E[E[XE[X|Y]|Y]] - E[X]^2$$

$$
\begin{aligned}
&= E[E[X|Y]E[X|Y]] - E[X]^2 \\
&= E[E[X|Y]^2] - E[E[X|Y]]^2 \\
&= \mathrm{Var}(E[X|Y]). \qquad \square
\end{aligned}
$$

The next lemma relies on Lemma 1 and on the fact, mentioned in [13] appendix, that a random variable $\tilde{p}_b$, viewed as a point in a linear space, cannot have a high correlation with many members of an orthogonal basis $\{E[\tilde{p}_b|\tilde{v}^i]\}_{i=1}^n$ for this space.

**Lemma A.2.** *For every $n \in \mathbb{N}$ and $b \in B$,*

$$\sum_{i=1}^{n} \mathrm{Var}(E[\tilde{p}_b|\tilde{v}^i]) \leqslant 1.$$

**Proof.** Fix a profile of beliefs $b \in B$. By Lemma 1, conditional on $\tilde{b} = b$, $\{\tilde{v}^i\}_{i=1}^n$ are independent random variables. Since functions of independent random variables are independent random variables, $\{E[\tilde{p}_b|\tilde{v}^i]\}_{i=1}^n$ are independent random variables and $\mathrm{cov}(E[\tilde{p}_b|\tilde{v}^i], E[\tilde{p}_b|\tilde{v}^j]) = 0$ for every $i \neq j$. It follows that

$$
\begin{aligned}
0 &\leqslant \mathrm{Var}\left(\tilde{p}_b - \sum_{i=1}^{n} E[\tilde{p}_b|\tilde{v}^i]\right) \\
&= \mathrm{Var}(\tilde{p}_b) + \sum_{i=1}^{n} \mathrm{Var}(E[\tilde{p}_b|\tilde{v}^i]) - 2\sum_{i=1}^{n} \mathrm{cov}(\tilde{p}_b, E[\tilde{p}_b|\tilde{v}^i]).
\end{aligned}
$$

Upon rearranging,

$$2\sum_{i=1}^{n} \mathrm{cov}(\tilde{p}_b, E[\tilde{p}_b|\tilde{v}^i]) - \sum_{i=1}^{n} \mathrm{Var}(E[\tilde{p}_b|\tilde{v}^i]) \leqslant \mathrm{Var}(\tilde{p}_b) \leqslant 1$$

and by Lemma A.1,

$$\sum_{i=1}^{n} \mathrm{Var}(E[\tilde{p}_b|\tilde{v}^i]) \leqslant 1. \qquad \square$$

The next lemma relies on the previous one to show that for most agents and most belief profiles $b \in B$, $\mathrm{Var}(E[\tilde{p}_b|\tilde{v}^i])$ is small. Define the set

$$\mathscr{B}^i = \left\{ b \in B : \mathrm{Var}(E[\tilde{p}_b|\tilde{v}^i]) < \frac{1}{n^{\frac{5}{8}}} \right\}.$$

**Lemma A.3.** *For every $n \in \mathbb{N}$, there exists a set $\hat{N} \subseteq \{1, \ldots, n\}$ that includes at least $n - n^{\frac{3}{4}}$ agents, such that for every agent $i \in \hat{N}$, $\Pr(\mathscr{B}^i) > 1 - \frac{1}{n^{\frac{1}{8}}}$.*

**Proof.** We prove a more general result. For every $x \geqslant 0$, for at least $n-x$ agents, $\Pr(\mathscr{B}^i) > 1 - \frac{n^{\frac{5}{8}}}{x}$. The lemma follows upon plugging $x = n^{\frac{3}{4}}$.

For $i \in \{1, \dots, n\}$ and $b \in B$, let

$$Q^i(b) = \begin{cases} \Pr(b) & \text{if } \operatorname{Var}(E[\tilde{p}_b | \tilde{v}^i]) \geqslant \frac{1}{n^{\frac{5}{8}}}, \\ 0 & \text{otherwise.} \end{cases}$$

By the previous lemma, for every $b \in B$, there can be at most $n^{\frac{5}{8}}$ agents for which $\operatorname{Var}(E[\tilde{p}_b|\tilde{v}^i]) \geqslant \frac{1}{n^{\frac{5}{8}}}$, therefore, $\sum_{b \in B, i \in \{1,\dots,n\}} Q^i(b) \leqslant n^{\frac{5}{8}} \sum_{b \in B} \Pr(b) \leqslant n^{\frac{5}{8}}$. It follows that the set $B \backslash \mathscr{B}^i$, or the set of $b$'s for which $\operatorname{Var}(E[\tilde{p}_b|\tilde{v}^i]) \geqslant \frac{1}{n^{\frac{5}{8}}}$, can have probability $\Pr(B \backslash \mathscr{B}^i) \geqslant \frac{n^{\frac{5}{8}}}{x}$ for at most $x$ agents. Otherwise, $\sum_{b \in B, i \in \{1,\dots,n\}} Q^i(b) > x \sum_{b \in B \backslash \mathscr{B}^i} \Pr(b) \geqslant x \cdot \frac{n^{\frac{5}{8}}}{x} = n^{\frac{5}{8}}$, contradiction. Therefore, for at least $n-x$ agents, $\operatorname{Var}(E[\tilde{p}_b|\tilde{v}^i]) < \frac{1}{n^{\frac{5}{8}}}$ for every $b \in \mathscr{B}^i$ where $\Pr(\mathscr{B}^i) > 1 - \frac{n^{\frac{5}{8}}}{x}$. □

The next lemma demonstrates that for most agents, the effect that an agent can have on the level of public good through its willingness to pay is negligible when $n$ is large. That is, when $n$ is large, for most agents $\operatorname{Var}(E[\tilde{p}_{b^i}|\tilde{v}^i])$ is small. Intuitively, the fact that Lemma A.3 shows that for most agents and most belief profiles $b$, conditional on $b$, the variance of the decision to provide the public good as a function of agent $i$'s willingness to pay $v^i$ is small, implies that this should also be the case when, given a specific $b^i$, we average agent $i$'s influence on the probability the public good is provided over $b^{-i}$.

**Lemma A.4.** *For every $n \in \mathbb{N}$, there exists a set $N^* \subseteq \{1, \dots, n\}$ that includes at least $n - n^{\frac{3}{4}}$ agents, such that for every agent $i \in N^*$,*

$$\operatorname{Var}(E[\tilde{p}_{b^i}|\tilde{v}^i]) \leqslant \frac{9}{\varepsilon^2 n^{\frac{1}{4}}}$$

*for every $b^i \in B^i$.*

**Proof.** For every $i \in \{1, \dots, n\}$ and $b^i \in B^i$, we may write,

$$\begin{aligned} \operatorname{Var}(E[\tilde{p}_{b^i}|\tilde{v}^i]) &= \sum_{v^i \in V^i} (E[\tilde{p}_{b^i}|\tilde{v}^i] - E[E[\tilde{p}_{b^i}|\tilde{v}^i]])^2 \Pr(\tilde{v}^i = v^i | \tilde{b}^i = b^i) \\ &= \sum_{v^i \in V^i} \left( \sum_{b \in B} (E[\tilde{p}_b|\tilde{v}^i] - E[\tilde{p}_b]) \Pr(\tilde{b} = b | \tilde{b}^i = b^i) \right)^2 \\ &\quad \times \Pr(\tilde{v}^i = v^i | \tilde{b}^i = b^i), \end{aligned}$$

where the first equality follows from the definition of variance and the second from (A.1) and (A.2). By Chebyshev's inequality, for every $b \in B$,

$$\Pr\left( |E[\tilde{p}_b|\tilde{v}^i] - E[\tilde{p}_b]| > \frac{1}{n^{\frac{1}{4}}} \right) \leqslant n^{\frac{1}{2}} \operatorname{Var}(E[\tilde{p}_b|\tilde{v}^i]).$$

Therefore, for $i \in N^* \equiv \hat{N}$ and $b \in \mathscr{B}^i$,

$$\Pr\left( |E[\tilde{p}_b|\tilde{v}^i] - E[\tilde{p}_b]| > \frac{1}{n^{\frac{1}{4}}} \right) \leqslant \frac{1}{n^{\frac{1}{8}}}$$

and

$$\begin{aligned} |E[\tilde{p}_b|\tilde{v}^i] - E[\tilde{p}_b]| \leqslant & \Pr\left( |E[\tilde{p}_b|\tilde{v}^i] - E[\tilde{p}_b]| > \frac{1}{n^{\frac{1}{4}}} \right) \cdot 1 \\ & + \Pr\left( |E[\tilde{p}_b|\tilde{v}^i] - E[\tilde{p}_b]| \leqslant \frac{1}{n^{\frac{1}{4}}} \right) \cdot \frac{1}{n^{\frac{1}{4}}} \\ \leqslant & \frac{1}{n^{\frac{1}{8}}} + \left(1 - \frac{1}{n^{\frac{1}{8}}}\right) \frac{1}{n^{\frac{1}{4}}} \\ \leqslant & \frac{2}{n^{\frac{1}{8}}}. \end{aligned}$$

And for $b \notin \mathscr{B}^i$,

$$|E[\tilde{p}_b|\tilde{v}^i] - E[\tilde{p}_b]| \leqslant 1.$$

So, for $i \in N^*$,

$$\begin{aligned} & \operatorname{Var}(E[\tilde{p}_{b^i}|\tilde{v}^i]) \\ & = \sum_{v^i \in V^i} \left( \sum_{\{b \in \mathscr{B}^i\}} (E[\tilde{p}_b|\tilde{v}^i] - E[\tilde{p}_b]) \Pr(\tilde{b} = b|\tilde{b}^i = b^i) \right. \\ & \quad \left. + \sum_{\{b \notin \mathscr{B}^i\}} (E[\tilde{p}_b|\tilde{v}^i] - E[\tilde{p}_b]) \Pr(\tilde{b} = b|\tilde{b}^i = b^i) \right)^2 \Pr(\tilde{v}^i = v^i|\tilde{b}^i = b^i) \\ & \leqslant \sum_{v^i \in V^i} \left( \frac{2}{n^{\frac{1}{4}}} \sum_{\{b \in \mathscr{B}^i\}} \frac{\Pr(\tilde{b}^{-i} = b^{-i}, \tilde{b}^i = b^i)}{\Pr(\tilde{b}^i = b^i)} \right. \\ & \quad \left. + \sum_{\{b \notin \mathscr{B}^i\}} \frac{\Pr(\tilde{b}^{-i} = b^{-i}, \tilde{b}^i = b^i)}{\Pr(\tilde{b}^i = b^i)} \right)^2 \Pr(\tilde{v}^i = v^i|\tilde{b}^i = b^i) \end{aligned}$$

$$\begin{aligned}
&\leqslant \sum_{v^i\in V^i}\left(\frac{2}{n^{\frac{1}{4}}}\cdot\frac{1}{\varepsilon}\sum_{\{b\in\mathscr{B}^i\}}\Pr(\tilde{b}^{-i}=b^{-i},\tilde{b}^{\,i}=b^i)\right.\\
&\qquad\left.+\frac{1}{\varepsilon}\sum_{\{b\notin\mathscr{B}^i\}}\Pr(\tilde{b}^{-i}=b^{-i},\tilde{b}^{\,i}=b^i)\right)^2\Pr(\tilde{v}^{\,i}=v^i|\tilde{b}^{\,i}=b^i)\\
&\leqslant \sum_{v^i\in V^i}\left(\frac{2}{\varepsilon n^{\frac{1}{4}}}+\frac{1}{\varepsilon n^{\frac{1}{8}}}\right)^2\Pr(\tilde{v}^{\,i}=v^i|\tilde{b}^{\,i}=b^i)\\
&\leqslant \frac{9}{\varepsilon^2 n^{\frac{1}{4}}}. \qquad \square
\end{aligned}$$

That is, no more than $n^{\frac{3}{4}}$ agents may have a significant effect on the probability of providing the public good. Most agents, or more precisely, $n-n^{\frac{3}{4}}$ agents, are likely to be of a type whose influence on the level of provision is small, or for which $\mathrm{Var}(E[\tilde{p}_{b^i}|\tilde{v}^{\,i}])\leqslant\frac{9}{\varepsilon^2 n^{\frac{1}{4}}}$. We denote this set of agents by

$$I=\left\{i\in\{1,\ldots,n\}:\mathrm{Var}(E[\tilde{p}_{b^i}|\tilde{v}^{\,i}])\leqslant\frac{9}{\varepsilon^2 n^{\frac{1}{4}}}\text{ for every } b^i\in B^i\right\}.$$

For every belief $b^i\in B^i$ of agent $i$, define the set of states of the world where agent $i$ with belief $b^i$ has a significant impact on the probability the public good is provided by

$$C^i_{b^i}=\left\{t\in T: b^i(t^i)=b^i \text{ and } v^i(t^i) \text{ is such that } |E[\tilde{p}_{b^i}|\tilde{v}^{\,i}=v^i(t^i)]-E[\tilde{p}_{b^i}]|>\frac{3}{n^{\frac{1}{12}}}\right\}.$$

For every $i\in\{1,\ldots,n\}$ and $b^i\in B^i$,

$$\begin{aligned}
&\mathrm{Var}(E[\tilde{p}_{b^i}|\tilde{v}^{\,i}])\\
&\quad=E[(E[\tilde{p}_{b^i}|\tilde{v}^{\,i}]-E[\tilde{p}_{b^i}])^2]\\
&\quad\geqslant \sum_{\{t\in T:b^i(t^i)=b^i \text{ and } (E[\tilde{p}_{b^i}|\tilde{v}^{\,i}=v^i(t^i)]-E[\tilde{p}_{b^i}])^2>\frac{9}{n^{\frac{1}{6}}}\}}(E[\tilde{p}_{b^i}|\tilde{v}^{\,i}=v^i(t^i)]-E[\tilde{p}_{b^i}])^2\\
&\quad\times\Pr(\tilde{t}=t|\tilde{b}^{\,i}=b^i)
\end{aligned}$$

$$\geqslant \frac{9}{n^{\frac{1}{6}}} \sum_{\{t \in T : b^i(t^i)=b^i \text{ and } (E[\tilde{p}_{b^i}|\tilde{v}^i=v^i(t^i)]-E[\tilde{p}_{b^i}])^2 > \frac{9}{n^{\frac{1}{6}}}\}}$$

$$\times \frac{\Pr(\tilde{t}^{-i} = t^{-i} \text{ and } t^i \text{ is such that } \tilde{b}^i = b^i)}{\Pr(\tilde{b}^i = b^i)}$$

$$\geqslant \frac{9\Pr(C^i_{b^i})}{\varepsilon n^{\frac{1}{6}}}$$

and therefore $\Pr(C^i_{b^i}) \leqslant \frac{\varepsilon n^{\frac{1}{6}}}{9} \mathrm{Var}(E[\tilde{p}_{b^i}|\tilde{v}^i])$. For an agent $i \in I$ with belief $b^i \in B^i$, the probability that agent $i$'s willingness to pay has a significant impact on the level of production is $\Pr(C^i_{b^i}) \leqslant \frac{\varepsilon n^{\frac{1}{6}}}{9} \mathrm{Var}(E[\tilde{p}_{b^i}|\tilde{v}^i]) \leqslant \frac{1}{\varepsilon n^{\frac{1}{12}}}$.

By Assumption A, for any $b^i \in B^i$ the event $D^i_{b^i} = \{\tilde{b}^i = b^i \text{ and } \tilde{v}^i = 0\}$ in which agent $i$ holds the belief $b^i$ and willingness to pay zero has probability at least $\varepsilon^2 > 0$. Therefore, for all $n$ large enough and $i \in I$, Pr $(D^i_{b^i} \backslash C^i_{b^i}) > 0$. That is, for $i \in I$, $|E[\tilde{p}_{b^i}|\tilde{v}^i = 0] - E[\tilde{p}_{b^i}]| \leqslant \frac{3}{n^{\frac{1}{12}}}$ for all $n$ sufficiently large. In other words, for an agent $i \in I$, having willingness to pay zero for the public good does not affect the mechanism's decision by much. The fact that such an agent $i$ can always report that its willingness to pay is zero, pay zero, and not decrease by much the probability that the public good is provided, implies that $i$'s payment cannot be very large.[22] We make this argument precise by using the incentive compatibility and individual rationality constraints to bound $E[\tilde{x}^i_{b^i}]$ from above.

Individual rationality implies that the expected payment of an agent with willingness to pay zero should be nonpositive, or

$$E[\tilde{x}^i_{b^i}|\tilde{v}^i = 0] \leqslant 0$$

for every $i \in \{1, \ldots, n\}$ and $b^i \in B^i$. Together with incentive compatibility this implies

$$E[\tilde{x}^i_{b^i}|\tilde{v}^i = v^i] \leqslant v_i(E[\tilde{p}_{b^i}|\tilde{v}^i = v_i] - E[\tilde{p}_{b^i}|\tilde{v}^i = 0])$$

for every $i \in \{1, \ldots, n\}$, $b^i \in B^i$, and $v^i \in V^i$. If in addition $i \in I$ and $v^i \notin C^i_{b^i}$, the triangle inequality implies that $|E[\tilde{p}_{b^i}|\tilde{v}^i = v_i] - E[\tilde{p}_{b^i}|\tilde{v}^i = 0]| \leqslant \frac{6}{n^{\frac{1}{12}}}$, and

$$E[\tilde{x}^i_{b^i}|\tilde{v}^i = v^i] \leqslant \frac{6v^i}{n^{\frac{1}{12}}}$$

[22] This is where we rely on the fact that agents' beliefs do not determine their willingness to pay for the public good. If they did, then an agent with belief $b^i$ would not have been able to claim that its willingness to pay is zero when it is in fact positive.

for every $i \in I$, $b^i \in B^i$, and $v^i \notin C^i_{b^i}$. In case $v^i \in C^i_{b^i}$, because $v^i \leqslant \bar{v}$, $i$ can never be made to pay more than $\bar{v}$ for the public good. Thus, we have that for every $i \in I$,

$$\begin{aligned}
E[\tilde{x}^i] &= \sum_{t^i=(v^i,b^i)\in T^i} E[\tilde{x}^i_{b^i}|\tilde{v}^i = v^i]\Pr(t^i) \\
&\leqslant \sum_{b^i\in B^i}\left[\sum_{v^i\notin C^i_{b^i}} E[\tilde{x}^i_{b^i}|\tilde{v}^i = v^i]\Pr(\tilde{v}^i = v^i|\tilde{b}^i = b^i) + \Pr(C^i_{b^i})\bar{v}\right]\Pr(\tilde{b}^i = b^i) \\
&\leqslant \sum_{b^i\in B^i}\sum_{v^i\notin C^i_{b^i}} E[\tilde{x}^i_{b^i}|\tilde{v}^i = v^i]\Pr(\tilde{v}^i = v^i|\tilde{b}^i = b^i)\Pr(\tilde{b}^i = b^i) + \frac{\bar{v}}{\varepsilon n^{\frac{1}{12}}} \\
&\leqslant \sum_{b^i\in B^i}\left(\frac{6\bar{v}}{n^{\frac{1}{12}}}\right)\Pr(\tilde{b}^i = b^i) + \frac{\bar{v}}{\varepsilon n^{\frac{1}{12}}} \\
&\leqslant \frac{7\bar{v}}{\varepsilon n^{\frac{1}{12}}}.
\end{aligned}$$

If $i \notin I$, $E[\tilde{x}^i] \leqslant \bar{v}$. Thus,

$$\begin{aligned}
\sum_{i=1}^{n} E[\tilde{x}^i] &\leqslant \sum_{i\in I}\frac{7\bar{v}}{\varepsilon n^{\frac{1}{12}}} + \sum_{i\notin I}\bar{v} \\
&\leqslant \sum_{i=1}^{n}\frac{7\bar{v}}{\varepsilon n^{\frac{1}{12}}} + \bar{v}n^{\frac{3}{4}} \\
&\leqslant \frac{8\bar{v}n^{\frac{11}{12}}}{\varepsilon}.
\end{aligned}$$

Finally, budget balance implies

$$\begin{aligned}
E[\tilde{p}] &\leqslant \frac{\sum_{i=1}^{n} E[\tilde{x}^i]}{C(n)} \\
&\leqslant \frac{8\bar{v}n^{\frac{11}{12}}}{\varepsilon cn} \\
&= \frac{8\bar{v}}{\varepsilon cn^{\frac{1}{12}}} \searrow_{n\to\infty} 0. \qquad \square
\end{aligned}$$

## References

[1] N. Al-Najjar, R. Smorodinsky, Pivotal players and characterization of influence, J.Econ. Theory 92 (2000) 318–342.

[2] R. Aumann, A. Brandenburger, Epistemic conditions for nash equilibrium, Econometrica 63 (1995) 1161–1180.

[3] E. Auriol, J.-J. Laffont, Regulation by duopoly, J. Econ. Manage. Strategy 1 (1992) 507–533.

[4] D. Bergemann, S. Morris, Robust mechanism design, Cowles Foundation Discussion Paper No. 1421, Yale University, 2003.

[5] A. Brandenburger, E. Dekel, Hierarchies of beliefs and common knowledge, J. Econ. Theory 59 (1993) 189–198.

[6] J. Crémer, R. McLean, Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent, Econometrica 53 (1985) 345–361.

[7] J. Crémer, R. McLean, Full extraction of the surplus in Bayesian and dominant strategy auctions, Econometrica 56 (1988) 1247–1257.

[8] D. Fudenberg, J. Tirole, Game Theory, MIT Press, Cambridge, MA, 1991.

[9] P. Jehiel, B. Moldovanu, Efficient design with interdependent valuations, Econometrica 69 (2001) 1237–1260.

[10] S. Johnson, J. Pratt, R. Zeckhauser, Efficiency despite mutually payoff-relevant private information: the finite case, Econometrica 58 (1992) 873–900.

[11] S. Johnson, N. Miller, J. Pratt, R. Zeckhauser, Efficient design with interdependent valuations and an informed center, Mimeo, Kennedy School of Government, Harvard University, 2002.

[12] J.-J. Laffont, D. Martimort, Mechanism design with collusion and correlation, Econometrica 68 (2000) 309–342.

[13] G. Mailath, A. Postlewaite, Asymmetric information bargaining problems with many agents, Rev. Econ. Stud. 57 (1990) 351–368.

[14] P. McAfee, P. Reny, Correlated information and mechanism design, Econometrica 60 (1992) 395–421.

[15] J.-F. Mertens, S. Zamir, Formulation of Bayesian analysis for games with incomplete information, Int. J. Game Theory 10 (1985) 619–632.

[16] C. Mezzetti, Mechanism design with interdependent valuations: efficiency and full surplus extraction, working paper, University of North Carolina, Chapel Hill, 2003.

[17] D. Monderer, D. Samet, Approximating common knowledge with common beliefs, Games Econ. Behav. 1 (1989) 170–190.

[18] S. Morris, Typical types, Mimeo, Yale University, 2002.

[19] R. Myerson, Optimal auction design, Math. Oper. Res. 6 (1981) 58–73.

[20] R. Myerson, Bayesian equilibrium, incentive compatibility: an introduction, in: L. Hurwicz, et al., (Eds.), ‘Social Goals and Social Organization’, Cambridge University Press, Cambridge, 1985.

[21] J. Robert, Continuity in auction design, J. Econ. Theory 55 (1991) 169–179.