Information acquisition and full surplus extraction

by

Sushil Bikhchandani[†]

Anderson School of Management, UCLA

July 2009

Abstract

It is well-known that when agents' types are correlated, the mechanism designer can extract the entire surplus. This creates an incentive for agents to acquire information about other agents' types. *Robust lotteries* (are payment schemes that) support full extraction and *partially robust lotteries* support efficient implementation in the presence of information acquisition opportunities. Necessary and sufficient conditions for existence of robust and partially robust lotteries are derived. If an agent's information signal spans the set of other agents' types then robust lotteries do not exist. However, by inducing agents to report their signal realizations the mechanism designer may be able to extend the type space so that robust lotteries exist.

[†]I am grateful to John Mamer, Zvika Neeman, Ichiro Obara, David Rahman, Phil Reny, and seminar participants at UCLA and the University of Bonn for helpful comments.

1 Introduction

Consider a mechanism design setting in which agents' types are correlated and each agent is costlessly endowed with information about his own type. Cremer & McLean [9], the seminal paper in this environment, shows that under a generic condition on probability distributions over types the ex post efficient rule is Bayesian incentive compatible and interim individually rational. Moreover, the mechanism designer can extract the entire surplus. Full surplus extraction is implemented through a menu of lotteries (stochastic payment schemes), one lottery for each type of an agent. The ex post payoffs in an agent's lotteries depend on the reports of other agents. These payoffs become unboundedly large as the beliefs of an agent approach independence, creating a strong incentive for an agent to spy on other agents and surreptitiously acquire information about their types in order to thwart full extraction. The private value of such information becomes unboundedly large as beliefs approach independence.[1]

The feasibility of full extraction in the presence of information acquisition opportunities about others' types is the focus of this paper. The timing of moves is as follows. Initially, (i) each agent costlessly and privately learns his own type and (ii) the mechanism designer announces the social choice function and payment scheme to be implemented. Next, agents decide whether to acquire costly and imperfect information about other agents' types. An agent who acquires information observes a signal realization correlated with other agents' types. Each agent decides whether to participate in the mechanism and, if he decides to participate, reports his own type (possibly untruthfully). Finally, the mechanism designer implements the announced social choice function and payments at the reported types. The probability that agents have access to information signals about others is less than one.

The decision to acquire information is covert. A typical lottery that supports full extraction in the absence of information acquisition does not support full extraction if an agent acquires information. There are two deviations that an agent may make after acquiring information: he may decide not to participate or he may misreport his type. Thus, under information acquisition, a full extraction lottery will either not support full extraction or truth-telling.

The existence of lotteries that support full extraction or at least truth-telling in the presence of information acquisition for any mechanism design problem on a given probability structure is examined in this paper. A lottery is *partially robust* (to a given information signal about other agents' types) if it dissuades information

[1] An agent's private value of information about others' types is bounded in mechanisms that do not fully extract surplus.

acquisition no matter how small the cost of information acquisition. A lottery is *robust* if it is partially robust and supports full extraction of surplus by the mechanism designer. Necessary and sufficient conditions for existence of robust lotteries and partially robust lotteries are obtained. If the information signals are sufficiently noisy then partially robust lotteries exist. It is easy to construct examples where robust lotteries do not exist for arbitrarily noisy (but not completely noisy) signals. Robust lotteries do not exist if an agent's information signal satisfies a spanning condition that requires that the number of possible linearly independent beliefs the agent might have is as large as the number of other agents' types. If the number of realizations of an agent's signal is small compared to the number of other agents' types then robust lotteries exist.

As already noted, each agent costlessly knows his own type. If agents do not acquire information about others' types then the ex post efficient rule is implementable. Therefore a mechanism that does not prevent acquisition of costly and socially useless information is inefficient. Partially robust lotteries, if they exist, may be used to pay agents rents not to acquire information and to implement the ex post efficient rule. The rent paid to an agent has to be sufficient to reduce his private value of information below the acquisition cost. If robust lotteries exist, then this rent is zero and full extraction in the presence of information acquisition opportunities is possible.

When two or more symmetric agents have access to information the mechanism designer can induce each agent to report his extended type, which consists of his payoff type and the realization of his costly signal about others' types. This expands the number of other agents' types and robust lotteries exist in this extended type space. Consequently, there exists a mechanism in the extended type space such that agents obtain zero surplus and the mechanism designer receives the entire surplus less information acquisition costs.[2] However, to implement such a mechanism the designer needs to know the joint distributions of agents' signals and types and their costs of information gathering. It is in the interest of each agent not to disclose anything about where or how or whether he acquires information because the mechanism designer may use it against the agent. As a practical matter, the mechanism designer is unlikely to have the knowledge required to implement this zero surplus mechanism.

Full extraction mechanisms are not observed in the real world. Several authors have offered reasons for their impracticality. Robert [26] shows that full extraction is impossible if agents are risk-averse or have limited liability. Neeman [21] notes that it is essential for full extraction that the beliefs of an agent (about others' types) pin

[2]Whether there exists another equilibrium in which agents gather information about others' types but do not report it, thereby restoring spanning and rent extraction, is an open question.

down the payoff type of the agent, and Heifetz and Neeman [14] establish that this assumption is non-generic in the universal type space.

In this paper, the focus is on the robustness of full extraction mechanisms to information acquisition. To this end, it is assumed that agents are risk-neutral and have unlimited liability. In contrast to Heifetz and Neeman, an agent's payoff types can be inferred from his beliefs; however, an agent with one payoff type may have one of several beliefs if he decides to acquire information. This is not the first paper with different beliefs for the same payoff type. Parreiras [25] shows that if an agent has one of two possible probability distributions over other agents, one distribution being Blackwell sufficient for the other, then full extraction is impossible. In Obara [22], agents take actions which jointly determine the distribution of types. If the set of actions is large then full extraction is impossible but approximate full extraction is generically possible. The actions in Obara may be interpreted as information acquisition. The model in my paper is non-generic in Obara's setting and the results I obtain are specialized to a pure information acquisition model.[3] Cremer, Spiegel, and Zheng [10] show that full surplus extraction is possible under (productive) information acquisition. A crucial difference is that ex ante individual rationality is required in [10], with bidders committing to participate before they learn their types, whereas I require the weaker condition of interim individual rationality.

There is a growing literature on information acquisition in a variety of environments. Milgrom [19] and Matthews [18], the first papers in this area, investigate information acquisition by bidders in auctions. Subsequent work by Tan [27], Persico [23], Bergemann and Pesendorfer [2], Jackson [15], Compte and Jehiel [6], and Hagedorn [13] is also in this area. Several of these papers investigate how auction theory is modified when information acquisition is costly. Information acquisition in contracting environments is the subject of Cremer and Khalil [7], Lewis and Sappington [16], and Cremer, Khalil, and Rochet [8]. More recently, information acquisition in committees and in voting models has been considered in Persico [24], Martinelli [17], Gerardi and Yariv [11], Cai [5], and Gershkov and Szentes [12]. Bergemann and Valimaki [3] adopts a general mechanism design approach, but with independent information. My paper differs from this literature in two respects: the acquired information has no social value and information acquisition takes place within full extraction mechanisms.

The model and a description of full extraction mechanisms is presented in Sec-

[3]Unlike in Obara [22], in my paper neither the probability distribution of types nor the utility for outcomes depends on agents' actions. Moreover, an agent chooses his action (whether to acquire information) after learning his own type and not before.

tion 2. The case where only one agent has access to information is analyzed in Section 3. It is shown that in full extraction mechanisms the private value of this information becomes unboundedly large as the agent's beliefs approach independence. Necessary and sufficient conditions for existence of (partially) robust lotteries are obtained in Section 3.1. The use of partially robust lotteries to implement an efficient mechanism is described in Section 3.2. The case where many agents may have access to information is examined in Section 4. Section 5 concludes. The proofs of all lemmas are in an appendix.

2 The model

There are A agents, indexed $a = 1, 2, \dots A$. Agent a's type is $t_a \in T^a$, where T^a is a finite set. Let $T \equiv \Pi_{a=1}^{A} T^a$ and $T^{-a} \equiv \Pi_{\ell \neq a} T^{\ell}$. The vector $t = (t_1, t_2, \dots, t_A) \in T$ denotes the types of all agents and may also be written as $t = (t_a, t_{-a})$. Let $n_a = |T^a| \geq 2$ and $N_a = |T^{-a}| \geq 2$. Agent a of type t_a is referred to as agent (a, t_a).

Each agent knows his own type and is uncertain about other agents' types. The mechanism designer does not know any of the agents' types and has a probability distribution $p(t)$ over T. Agent a's probability beliefs over others' types conditional on his own type being t_a is

$$p(t_{-a}|t_a) \;=\; \frac{p(t_a, t_{-a})}{p(t_a)} \;=\; \frac{p(t_a, t_{-a})}{\sum_{t'_{-a} \in T^{-a}} p(t_a, t'_{-a})}.$$

Note that $p(\cdot|t_a)$ is also the probability distribution that the mechanism designer would have over T^{-a} if he knew that agent a's type was t_a.[4] For every (a, t_a), $p(t_{-a}|t_a) < 1$ for all $t_{-a} \in T^{-a}$. In addition, to simplify the proofs I assume that $p(t_{-a}|t_a) > 0$ for all $t_{-a} \in T^{-a}$.

The mechanism designer asks agents to report their types and then implements an outcome $y \in Y$, where Y is an arbitrary set. Agent a's utility function over outcomes, y, and money, m, is quasilinear and depends on the types vector t:

$$U_a(y, t, m) \;=\; u_a(y, t) + m.$$

An *information structure* is a type space and a probability distribution over the type space: (T, p). A *mechanism design problem* is an information structure together with a set of outcomes and utility functions: $(T, p, Y, u_a(y, t), \forall a \in A, y \in Y, t \in T)$.

[4] If p is a common prior over T between each agent and the mechanism designer, then this condition is satisfied.

A *social choice function* $f : T \to Y$ maps agents' (reported) types to outcomes in Y. A *payment function* $x_a : T \to \Re$ is a function from agents' reported types to an expected payment by agent a to the mechanism designer. Let $x = (x_1, x_2, \ldots, x_A)$. A *mechanism* consists of a social choice function and payment functions for each agent.

A mechanism (f, x) is *Bayesian incentive compatible* if

$$\begin{aligned} & \sum_{t_{-a}} p_a(t_{-a}|t_a)[u_a(f(t_a, t_{-a}), (t_a, t_{-a})) - x_a(t_a, t_{-a})] \\ \geq & \sum_{t_{-a}} p_a(t_{-a}|t_a)[u_a(f(t'_a, t_{-a}), (t_a, t_{-a})) - x_a(t'_a, t_{-a})], \qquad \forall t'_a \in T^a \backslash \{t_a\}. \end{aligned}$$

A mechanism (f, x) is *interim individually rational* if

$$\sum_{t_{-a}} p_a(t_{-a}|t_a)[u_a(f(t_a, t_{-a}), (t_a, t_{-a})) - x_a(t_a, t_{-a})] \ \geq \ 0, \qquad \forall t_a, \forall a.$$

An ex post *efficient* social choice function, denoted f^*, satisfies

$$f^*(t) \ \in \ \arg\max_{y \in Y} \sum_{a=1}^{A} u_a(y, t), \quad \forall t. \tag{1}$$

I assume that ex ante efficiency requires participation by all agents. This is true for auctions and for public good provision.

A mechanism (f, x) is *efficient* if it is Bayesian incentive compatible, interim individually rational, and f is efficient. An efficient mechanism will be denoted as (f^*, x) where f^* satisfies (1).

A *zero-surplus* mechanism (f, x) is Bayesian incentive compatible, interim individually rational, and the expected surplus of each agent is zero. The mechanism designer extracts the entire surplus in a zero-surplus mechanism. Interim individual rationality and zero surplus implies that

$$\sum_{t_{-a}} p_a(t_{-a}|t_a) x_a(t) \ = \ \sum_{t_{-a}} p_a(t_{-a}|t_a) u_a(f(t), t), \qquad \forall (a, t_a).$$

In particular, the following payment function supports a zero-surplus mechanism:

$$x_a(t) = u_a(f(t), t)) + \gamma_a(t_{-a}|t_a), \qquad \forall t = (t_a, t_{-a}) \tag{2}$$

such that $\sum_{t_{-a} \in T^{-a}} p(t_{-a}|t_a)\gamma_a(t_{-a}|t_a) = 0$.

A *full extraction mechanism* is an efficient, zero-surplus mechanism. A full extraction mechanism will be denoted as (f^*, x^*) where x_a^* is defined in (2) with f replaced by f^*.

Cremer and McLean [9] give a necessary and sufficient condition for full extraction on an information structure. This full extraction condition says that the probability beliefs of each type of an agent is not a convex combination of the probability beliefs of other types of the same agent, i.e. there does not exist (a, t_a) and $\lambda(t'_a) \geq 0$ such that[5]

$$p(t_{-a}|t_a) \;=\; \sum_{t'_a \neq t_a} \lambda(t'_a) p(t_{-a}|t'_a), \qquad \forall t_{-a}. \tag{3}$$

Cremer and McLean [9] show that if the full extraction condition is satisfied then every mechanism design problem on information structure (T, p) has a full extraction mechanism; if the full extraction condition is not satisfied then there exists a mechanism design problem on (T, p) which does not have a full extraction mechanism. The full extraction condition is generic in the space of probability distributions over T.[6] McAfee and Reny [20] showed that the full extraction result generalizes to infinite types spaces.

Full extraction mechanisms are built around lotteries whose outcomes depend on reported types of the agents. Let $\gamma_a(\cdot|t_a)$ be a lottery for type t_a of agent a. The stochastic fee that agent a pays if he reports his type as t_a and other agents report their types as t_{-a} is $\gamma_a(t_{-a}|t_a)$. The full extraction condition is necessary and sufficient for the existence of a lottery $\gamma_a(\cdot|t_a)$ for each (a, t_a) such that

$$\begin{aligned}
p(\cdot|t_a) \cdot \gamma_a(\cdot|t_a) \;&\equiv\; \sum_{t_{-a} \in T^{-a}} p(t_{-a}|t_a)\gamma_a(t_{-a}|t_a) \;=\; 0, \qquad \forall t_a \in T^a, \\
p(\cdot|t'_a) \cdot \gamma_a(\cdot|t_a) \;&\equiv\; \sum_{t_{-a} \in T^{-a}} p(t_{-a}|t'_a)\gamma_a(t_{-a}|t_a) \;>\; 0, \qquad \forall t_a \neq t'_a \in T^a.
\end{aligned}$$

Lotteries that satisfy the above two requirements are *full extraction lotteries*. The second requirement implies that $\gamma_a(\cdot|t_a) \not\equiv 0$ and therefore the first implies that there exist $t_{-a}, t'_{-a} \in T^{-a}$ such that $\gamma_a(t_{-a}|t_a) > 0 > \gamma_a(t'_{-a}|t_a)$. Let $\Gamma_a(t_a)$ be the set of full extraction (FE) lotteries for type t_a. $\Gamma_a(t_a)$ is a convex, polyhedral cone.

Figure 1 explains the construction of FE lotteries in an example with two agents, a and b. Each agent can be one of three types with type sets $T^a = T^b = \{t_1, t_2, t_3\}$.

[5] If agents' types are independently distributed then $p(\cdot|t_a) = p(\cdot|t'_a)$ for all t_a, t'_a and the full extraction condition is not satisfied.

[6] A condition on probability distributions holds generically if it is satisfied with probability one by all probability distributions selected by a measure that is absolutely continuous w.r.t. Lebesgue measure.

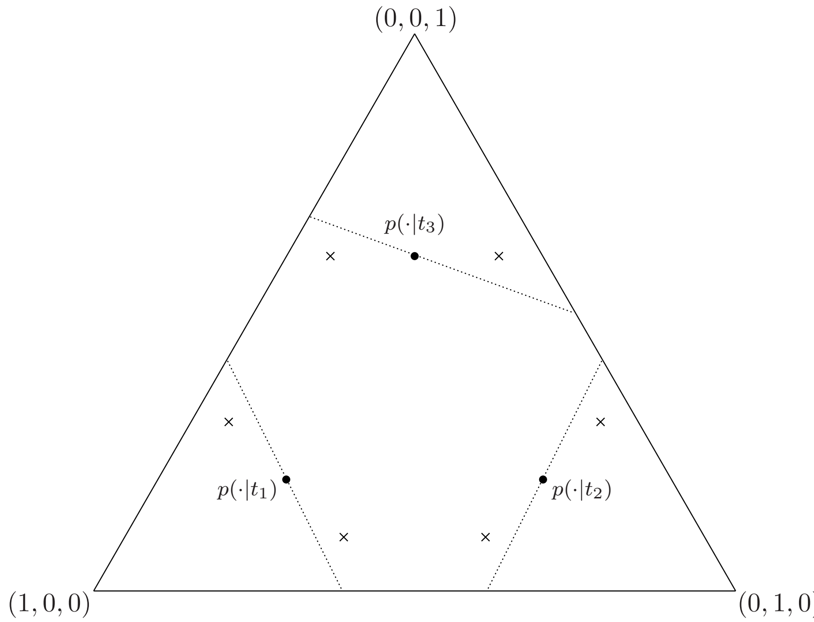

Figure 1: FE lotteries

Agent a's beliefs over T^b are in the probability simplex in $\Re^3$ shown in Figure 1. The lower-left corner of the triangle is the belief that places probability one on the event $\{t_b = t_1\}$, etc. Agent a's beliefs, $p(\cdot|t_a)$, $t_a = t_1$, t_2, and t_3 are indicated. The full extraction condition is satisfied as no belief is a convex combination of the other two, i.e. $p(\cdot|t_1)$ does not lie on the line segment joining $p(\cdot|t_2)$ and $p(\cdot|t_3)$, and so on. Therefore, by a separating hyperplane theorem one can draw a hyperplane through, say $p(\cdot|t_1)$ such that $p(\cdot|t_2)$ and $p(\cdot|t_3)$ lie on the same side of the hyperplane. The coefficients of the hyperplane $\gamma_a(\cdot|t_1)$ may be chosen such that $p(\cdot|t_1) \centerdot \gamma_a(\cdot|t_1) = 0$ and $p(\cdot|t_i) \centerdot \gamma_a(\cdot|t_1) > 0$, $i = 2, 3$. The hyperplane $\gamma_a(\cdot|t_1)$ represent the ex post payments of the FE lottery for type $t_a = t_1$. The dotted line through $p(\cdot|t_1)$ in Figure 1 represents the intersection of this hyperplane with the probability simplex in $\Re^3$. Similar hyperplanes (FE lotteries) can be constructed for $p(\cdot|t_2)$ and $p(\cdot|t_3)$ and are represented by dotted lines through these beliefs. (We shall return to the six points marked $\times$ later.)

The existence of FE lotteries for each (a, t_a) implies that any social choice function f can be implemented and the mechanism designer can extract the entire surplus generated by f. Ask agents to report their types t and implement $f(t)$. To ensure Bayesian incentive compatibility let the payment function for agent a be

$$x_a(t_a, t_{-a}) = u_a(f(t_a, t_{-a}), (t_a, t_{-a})) + \gamma_a(t_{-a}|t_a), \qquad \forall (t_a, t_{-a}).$$

Under this payment function, each agent's interim expected surplus is zero. As $\Gamma_a(t_a)$ is a cone, $\gamma_a(\cdot|t_a)$ may be selected such that $\gamma_a(\cdot|t_a) \cdot p(\cdot|t'_a) > 0$ is arbitrarily large for each $t'_a \neq t_a$. Thus, the expected payment made by agent (a, t'_a) who misreports his type as (a, t_a) can be made large enough to ensure Bayesian incentive compatibility of any social choice rule (including the ex post efficient rule $f^*(t)$).[7] Therefore, the social choice rule to be implemented is unimportant and in the sequel the focus is entirely on special types of FE lotteries that ensure truthful reporting in the presence of possible information acquisition.

Initially, I assume that one agent, after learning his own type, may have the option of acquiring information about others' types at a cost. I restrict attention to the case when this option is available after the agent learns his own type.[8]

3 Information acquisition by a single agent

With probability $\rho \in (0, 1)$, agent a has access to information signal R_a which has realizations r_k, $k = 1, 2, \ldots, K$. The cost of information is $c_a \geq 0$ and its acquisition can be postponed until after agent a learns his type. The timing of moves is as follows:

1. Each agent costlessly learns his own type.

   The mechanism designer announces the social choice function and payment function to be implemented.

2. If agent a has access to information R_a he decides whether to buy it.

3. If agent a buys information he observes a realization of R_a.

4. Each agent decides whether to participate in the mechanism.

5. Each participating agent reports his type to the mechanism designer.

6. The mechanism designer implements (among participating agents) the social outcome and payments at reported types.

[7]However, the existence of other Bayesian equilibria that are Pareto superior for the agents is not ruled out. See Brusco [4].

[8]An agent who does not have an incentive to acquire information no matter what his type will not acquire information if the acquisition has to be made *ex ante*, i.e. if the agent may only acquire information about others before learning his own type. Thus, robust lotteries prevent ex ante information acquisition as well.

The timing of moves is consistent with the assumption that the mechanism designer does not have sufficient knowledge of or control over the agent's information acquisition process. The mechanism designer is aware that agent a may have access to information about other agents' types. However, he does not observe agent a's decision in Step 2 or his signal realization in Step 3. Interim individual rationality implies that the decision to participate in Step 4 follows the decision to acquire information in Steps 2 and 3.[9] The mechanism designer cannot make the agent commit to participation before the agent decides whether to acquire information.

The conditional probability that $R_a = r_k$ when agents' types are (t_a, t_{-a}) is $q(r_k|t_{-a}, t_a)$. The posterior probability distribution over T^{-a} is

$$p(t_{-a}|t_a, r_k) \;=\; \frac{q(r_k|t_{-a}, t_a)p(t_{-a}|t_a)}{\sum_{t'_{-a} \in T^{-a}} q(r_k|t'_{-a}, t_a)p(t'_{-a}|t_a)} \;=\; \frac{p(r_k, t_{-a}|t_a)}{p(r_k|t_a)}. \tag{4}$$

The likelihood function associated with R_a is $Q_a = (q(r_k|t_{-a}, t_a))_{r_k}$, a $K \times (N_a n_a)$ matrix. The kth row of Q_a yields probabilities of observing r_k conditional on each (t_{-a}, t_a). A column of Q_a gives the probability for each r_k conditional on a specific (t_{-a}, t_a); each column sums to one. Let $Q_a(t_a)$ be the $K \times N_a$ submatrix of Q_a obtained by selecting the N_a columns of Q_a that correspond to t_a; $R_a(t_a)$ is the corresponding information signal.

A signal is *uninformative* if the posterior after every signal realization is identical to the prior. Thus, if every entry in the likelihood matrix Q_a is $\frac{1}{K}$ then R_a is uninformative.

Each agent knows his own type and this information, if revealed to the mechanism designer, is sufficient to obtain an efficient outcome. If the full extraction condition is satisfied then a mechanism based on FE lotteries can induce each agent to reveal his type. Therefore, information gathering by agent a about other agents' types has no social value and if $c_a > 0$ then it destroys value. However, this information has private value. Let $\pi_a(t_a)$ be the value of information to agent (a, t_a) in a mechanism (f, x).[10] An agent a has an *incentive to acquire information R_a* under mechanism (f, x) if there exists a type t_a of the agent such that $\pi_a(t_a) > c_a$.

Suppose that in a full extraction mechanism (f^*, x^*), $\gamma_a(\cdot|t_a)$ is the FE lottery for (a, t_a) used in the definition of $x^*_a(t_a, \cdot)$ in (2). If agent (a, t_a) does not acquire

[9] If the participation decision precedes information acquisition then full extraction is possible. A stronger assumption is ex ante individual rationality, which requires the agent to decide on participation before acquiring information about others' types *and* before learning his own type. For information acquisition under ex ante individual rationality see Cremer, Spiegel, and Zhang [10].

[10] The dependence of π_a on (f, x) is suppressed in the notation.

information, then he will truthfully report his type to the mechanism designer. If $\pi_a(t_a) > 0$ and agent (a, t_a) acquires information, then his subsequent actions depend on the signal realization. For at least one signal realization r_k agent (a, t_a) either misreports his type or he does not participate. After observing $R_a(t_a) = r_k$ he misreports his type as t'_a if

$$p(\cdot|t_a, r_k) \cdot \Big[u_a(f^*(t'_a, \cdot), (t_a, \cdot)) - u_a(f^*(t'_a, \cdot), (t'_a, \cdot)) - \gamma_a(\cdot|t'_a)\Big] \;\; > \;\; 0. \qquad (5)$$

If (5) is not true for any t'_a but

$$p(\cdot|t_a, r_k) \cdot \gamma_a(\cdot|t_a) \;\; > \;\; 0 \qquad (6)$$

then (recalling that interim individual rationality is a requirement) agent (a, t_a) does not participate in the mechanism. As $\pi_a(t_a) > 0$ by assumption, there exists r_k such that either (5) or (6) holds.

The next lemma shows that in a full extraction mechanism $\pi_a(t_a)$ is strictly positive. To see why, suppose that in the example of Figure 1, agent a has access to an information signal that takes one of two values, r_1 or r_2. The two points close to $p(\cdot|t_1)$ that are marked $\times$ are $p(\cdot|t_1, r_1)$ and $p(\cdot|t_1, r_2)$. With probability one, $p(\cdot|t_1, r_1)$ and $p(\cdot|t_1, r_2)$ are not on the hyperplane through $p(\cdot|t_a)$ from which $\gamma_a(\cdot|t_1)$ is constructed. Because the zero-surplus condition requires $\gamma_a(\cdot|t_1) \cdot p(\cdot|t_1) = 0$ we have $\gamma_a(\cdot|t_1) \cdot p(\cdot|t_1, r_i) < 0$, $\gamma_a(\cdot|t_1) \cdot p(\cdot|t_1, r_j) > 0$, $i \neq j$. Consequently, information has value to agent (a, t_1).

Lemma 1 *Let (f^*, x^*) be a full extraction mechanism under the assumption that no agent acquires information. For almost every informative signal R_a, the value of information to agent a is strictly positive.*

An appendix contains the proof of Lemma 1 (and of all subsequent lemmas).

Because a full extraction mechanism gives agents zero surplus, any information about other agents' types has positive private value. In contrast, the value of information is less likely to be positive in other types of mechanisms. In Bayesian incentive-compatible, interim individually rational mechanisms in which all agents (except the lowest types) make positive expected surplus, agent a will participate after observing any realization of a sufficiently noisy information. If, in addition, truthful reporting is a unique best response then sufficiently noisy information has value zero. In an ex post incentive compatible and ex post individually rational mechanism (such as an English auction or the public good provision mechanism in

Bergemann and Morris [1]) the value of information, even perfect information, about other agents' types is zero.

Therefore, it is appropriate to focus on the robustness of full extraction mechanisms to information acquisition. Not only is the value of information in such a mechanism positive, it increases without bound as an agent's beliefs approach independence. To establish this trivial cases in which incentive constraints do not bind at full extraction need to be excluded. Let

$$\kappa(t_a, t'_a) \equiv \sum_{t_{-a}} p(t_{-a}|t'_a)[u_a(f^*(t_a, t_{-a}), (t'_a, t_{-a})) - u_a(f^*(t_a, t_{-a}), (t_a, t_{-a}))].$$

$\kappa(t_a, t'_a)$ is the gain (or loss) to (a, t'_a) from misreporting his type as t_a if the mechanism designer attempts full extraction without using FE lotteries.[11]

Non-degeneracy: *There exists a type t_a of agent a such that for some $t'_a \in T^a \backslash \{t_a\}$ $\kappa(t_a, t'_a) > 0$.*

The above assumption rules out degenerate environments in which incentive constraints on implementing an efficient social choice function *and* extracting the entire surplus are not binding. Even if $f^*(t'_a, t_{-a}) = f^*(t_a, t_{-a})$ for all t_{-a}, as long as u_a depends on the agent's type the non-degeneracy assumption is satisfied for all beliefs, except possibly a non-generic set of beliefs. To simplify the proofs, I assume that non-degeneracy is satisfied for all beliefs.

Agent a's beliefs are *ϵ-close* if there exists a probability distribution $p_a(\cdot)$ over T^{-a} such that $||p(\cdot|t_a) - p_a(\cdot)|| \leq \epsilon$ for all t_a.[12] Beliefs are always 1-close to any arbitrary $p_a(\cdot)$. Independent beliefs are 0-close.

Lemmas 2 and 3 establish that ex post payments in FE lotteries and the value of information increase without bound as beliefs approach independence.

Lemma 2 *Consider a (non-degenerate) mechanism design problem on information structure (T, p). Suppose that agent a's beliefs are ϵ-close. There exists $t_a \in T^a$ such that for any FE lottery $\gamma_a(\cdot|t_a)$, $||\gamma_a(\cdot|t_a)|| \to \infty$ as $\epsilon \to 0$.*

Lemma 3 *Consider a mechanism design problem on information structure (T, p) in which the payment scheme is based on FE lotteries. Suppose that agent a's beliefs are ϵ-close. For almost every informative signal R_a there exists t_a such that $\pi_a(t_a) \to \infty$ as $\epsilon \to 0$.*

---

[11] If types are ordered so that higher types obtain greater gross utility at the ex post efficient outcome then higher types have an incentive to imitate lower types at full extraction.

[12] $|| \cdot ||$ stands for the Euclidean norm.

A mechanism has bounded payments if agent a's ex post payments are less than a uniform bound B that does not depend on the agent's beliefs, i.e. $x_a(t_a, t_{-a}) < B$ for all t_a, t_{-a} and beliefs $p(\cdot|t_a)$. It is easy to show that the private value of information in mechanisms with bounded payments is bounded and consequently, such mechanisms are less vulnerable to information acquisition. In addition, any Bayesian incentive compatible mechanism which is *ex post* individually rational (e.g. a first-price auction) has bounded payments. Although it is not ex post individually rational, an all-pay auction has bounded payments.

From Lemma 3 we know that mechanisms based on FE lotteries do not have bounded private value of information. For any cost c_a of acquiring information R_a, if agent a's beliefs are within a small enough neighborhood of each other then $\pi_a(t_a) > c_a$ for any type t_a at which full extraction incentive constraints *bind*.[13] Thus, for any mechanism based on FE lotteries there exist beliefs satisfying the full extraction condition at which agent a will acquire information and, as shown next, undermine full extraction.

Proposition 1 *Let (f^*, x^*) be a full extraction mechanism under the assumption that no agent acquires information. If agent a has an incentive to acquire information then the mechanism designer does not extract the expected entire surplus and (f^*, x^*) is not efficient.*

Proof: Let t_a be agent a's type such that $\pi_a(t_a) > c_a > 0$. Agent a's ex ante expected surplus is at least $[\pi_a(t_a) - c_a]p(t_a) > 0$. Thus, the mechanism designer does not extract the entire expected surplus under (f^*, x^*).

As agent (a, t_a) acquires information that is costly and has no social value, the mechanism is not efficient. There are other sources of inefficiency. There exists at least one signal realization r_k such that either (5) or (6) holds. If (5) holds then, unless $f^*(t_a, t_{-a}) = f^*(t'_a, t_{-a})$ for all t_{-a}, the outcome is not efficient. If, instead, (6) holds then agent (a, t_a) does not participate when $R_a(t_a) = r_k$, which is inefficient. $\square$

The definition of an efficient mechanism needs to be modified to accommodate the possibility of information gathering. A mechanism (f, x) is *efficient* if agent a does not have an incentive to acquire information, the mechanism is Bayesian incentive compatible, interim individually rational, and $f(t)$ is efficient. A *full extraction mechanism* is an efficient (under this new definition), zero-surplus mechanism.

[13] That is, if full extraction is attempted without using FE lotteries then some other type t'_a would profit by imitating t_a.

Lemma 3 and Proposition 1 imply that for any full extraction mechanism (in the absence of information acquisition), almost every R_a and cost c_a, if agent a's beliefs are close enough then he will acquire R_a and undermine full extraction.

In the sequel I investigate whether the mechanism designer can restore full extraction either by extending the type space or (in section 3.1) by carefully choosing full extraction lotteries. It is useful to view options from the stand-point of two mechanism designers who have different objectives. Mechanism designer M_{ER} has lexicographic preferences over efficiency and revenue. M_{ER} prefers an efficient mechanism to any inefficient mechanism; if two mechanism yield the same expected surplus then he prefers the mechanism which yields greater expected revenue. Mechanism designer M_{RE} has lexicographic preferences over revenue and efficiency. Both M_{ER} and M_{RE} prefer a full extraction mechanism, if one exists, to any other mechanism.

Extended Type Space

Can the mechanism designer expand the type space for agent a to accommodate his information acquisition capability and obtain FE lotteries in the expanded type space? That is, agent a of type t_a is replaced by $K+1$ types, $\{(t_a, \emptyset), (t_a, r_1), (t_a, r_2), \ldots, (t_a, r_K)\}$. If agent a does not acquire information his type is $(t_a, \emptyset)$. If agent a acquires information he will be one of K possible types (t_a, r_k), $k = 1, 2, \ldots, r_K$. Therefore,

$$p(\cdot|t_a, \emptyset) \quad = \quad \sum_{k=1}^{K} p(r_k|t_a)p(\cdot|t_a, r_k). \tag{7}$$

Type t_a's beliefs before acquiring information are a convex combination of his K possible beliefs after acquiring information. This violates the full extraction condition. Augmenting the type space does not recover full extraction. The full extraction condition, which is a generic condition in the payoff type space, never holds in the extended type space. Therefore,

Proposition 2 *If the type space of an agent is augmented to include each agent's option of acquiring information signals about other agents then the full extraction condition does not hold.*

If, for all t_a, r_k, beliefs $p(\cdot|t_a, r_k)$ satisfy the full extraction condition and c_a is small enough, mechanism designer M_{RE} may wish to force agent a to acquire information and then extract the entire surplus after compensating agent a with c_a. However, this may not be in M_{RE}'s interest. Suppose that M_{RE} requires agent a to acquire information and report his payoff type and signal realization. If this agent does not have access to information (recall that $\rho < 1$) then participation gives the agent negative surplus. If ρ is not close to 1, the revenue loss from possible non-participation of agent a makes it unprofitable for M_{RE} to require information acquisition.

Mechanism designer M_{ER} would, of course, be interested in efficient mechanisms that prevent information acquisition. If there exist "robust" lotteries (robust in the sense that they permit full extraction in the presence of information acquisition opportunities) then both M_{ER} and M_{RE} prefer a mechanism design based on them. I investigate the existence of such lotteries next.

Before proceeding, the mechanism designer's and other agents' knowledge about agent a needs to be specified. The mechanism designer knows that with probability $\rho \in (0, 1)$, agent a has the capability of acquiring information about others' types. As mentioned earlier, the act of information acquisition by agent a is covert. In the sequel, I shall assume either (I1) or (I2) below:

(I1) The mechanism designer does not know Q_a (the likelihood matrix of R_a).

(I2) The mechanism designer knows Q_a.

Assumption (I2), although consistent with the standard assumption that all participants have a common prior, is strong. Agent a may have a variety of different means of spying on other agents. It is in his interest to prevent the mechanism designer from knowing anything about these sources of information. As shown in section 3.1, the more the mechanism designer knows the greater the possibility agent a obtains zero surplus.

Other agents are also aware of agent a's access to information but do not observe whether he acquires information. I assume that other agents information about Q_a coincides with that of the mechanism designer.

3.1 Robust and Partially Robust Lotteries

Robust lotteries and partially robust lotteries are special types of full extraction lotteries. Robust lotteries make full extraction possible even when agent a has access to information R_a. In mechanisms based on robust lotteries agent a's value of information is zero. In mechanisms based on partially robust lotteries information has value for the participation decision only; if agent a decides to participate after acquiring information then he will not misreport his type. I obtain necessary and sufficient conditions for the existence of robust and partially robust lotteries. Existence of robust lotteries or partially robust lotteries is not enough to ensure that they are used. The mechanism designer must also know the likelihood matrix Q_a to use (partially) robust lotteries.

Equation (7) rules out the existence of a FE lottery $\gamma_a(\cdot|t_a)$ for type $(t_a, \emptyset)$ such that $p(\cdot|t_a) \centerdot \gamma_a(\cdot|t_a) = 0$ and $p(\cdot|t_a, r_k) \centerdot \gamma_a(\cdot|t_a) > 0$ for all k. However, (7) does not rule out the existence of a FE lottery $\gamma_a(\cdot|t_a)$ which in addition satisfies

$$
\begin{aligned}
p(\cdot|t_a, r_k) \centerdot \gamma_a(\cdot|t_a) &= 0, \quad \forall k && (8)\\
p(\cdot|t'_a, r_k) \centerdot \gamma_a(\cdot|t_a) &> 0, \quad \forall k,\ \forall t'_a \neq t_a. && (9)
\end{aligned}
$$

Definition 1 If $\gamma_a(\cdot|t_a)$ satisfies (8) and (9) then $\gamma_a(\cdot|t_a)$ is *robust* to acquisition of information signal $R_a(t_a)$ by agent (a, t_a).

If for each t_a there exists a robust lottery $\gamma_a(\cdot|t_a)$ then agent a's value for information R_a is zero and full extraction is possible. Agents report their type in the extended type space. The mechanism designer disregards the information signal component, if any, of the reported types and implements the efficient rule f^* at the reported payoff types.[14] f^* is supported by a payment rule x^* which is based on robust lotteries for agent a and FE lotteries for other agents. Whether or not agent a acquires information about others' types, his expected surplus (before deducting any information acquisition costs) is zero. Because information acquisition is costly, agent a will not acquire information. As (7) and (8) imply $p(\cdot|t_a) \centerdot \gamma_a(\cdot|t_a) = 0$ and (7) and (9) imply $p(\cdot|t'_a) \centerdot \gamma_a(\cdot|t_a) > 0$, agent a will participate in the mechanism and report his type truthfully. The mechanism (f^*, x^*) is a full extraction mechanism which dissuades agent a from acquiring information.

If robust lotteries do not exist for (a, t_a), then $\pi_a(t_a)$ becomes unboundedly large as beliefs approach independence (see Lemma 3) and for any $c_a > 0$ there exists a positive measure of beliefs for which agent (a, t_a) will acquire information, thus avoiding full extraction.

The next proposition shows that existence of robust lotteries is of no use to a mechanism designer whose state of knowledge is (I1).

Proposition 3 *Suppose that the mechanism designer does not know Q_a. The probability that he selects a robust lottery from $\Gamma_a(t_a)$ is zero.*

[14] $f^*(t)$ depends only on agents' (truthfully reported) payoff types $t = (t_1, t_2, \ldots, t_A)$ and not on the realizations of R_a.

Proof: As $\Gamma_a(t_a)$ is a subset of the null space of the vector $p(\cdot|t_a)$, it has dimension $N_a - 1$. If $\gamma_a(\cdot|t_a) \in \Gamma_a(t_a)$ satisfies (8) then, in addition, $\gamma_a(\cdot|t_a)$ is in the null space of the K vectors $p(\cdot|t_a, r_k)$, $k = 1, 2, \ldots, K$. As $R_a(t_a)$ is informative, at least two of the possible posterior distributions after observing R_a are distinct from the prior distribution. In other words, the subset of $\Gamma_a(t_a)$ that satisfies (8) is either empty or has dimension no more than $N_a - 2$. Therefore, a randomly chosen element of $\Gamma_a(t_a)$ will, with probability one, not satisfy (8). □

If the mechanism designer's state of knowledge is (I1), full extraction is not possible. In the example of Figure 1, there are a continuum of hyperplanes that pass through $p(\cdot|t_1)$ and both $p(\cdot|t_2)$ and $p(\cdot|t_3)$ are on the same side os the hyperplane. Any of these hyperplanes can be used to construct a FE lottery. Only one of these hyperplanes through $p(\cdot|t_1)$ also passes through the two points marked x that represent $p(\cdot|t_1, r_1)$ and $p(\cdot|t_1, r_2)$; the coefficients of this "robust" hyperplane constitute the payments of a robust lottery. However, the mechanism designer needs to know $p(\cdot|t_1, r_1)$ and $p(\cdot|t_1, r_2)$ (or equivalently, the likelihoods of the signal) in order to select the robust hyperplane.

In order to investigate what assumptions are required to restore full extraction, I now make the restrictive assumption that the mechanism designer knows the likelihood matrix Q_a (i.e., his state of knowledge is at least I2). The issue now is whether there exists a robust lottery.

Proposition 4 *Consider an information structure (T, p) in which agent a may have access to information signal R_a. Robust lotteries exist if and only if there does not exist t_a such that some linear combination of beliefs $p(\cdot|t_a, r_k), k = 1, 2, \ldots, K$ is a convex combination of beliefs $p(\cdot|t'_a, r_k)$, $t'_a \in T^a \backslash t_a$, $k = 1, 2, \ldots, K$. That is, there does not exist t_a, $\mu \neq 0$, and $\lambda \geq 0$, such that*

$$\sum_{k=1}^{K} \mu(t_a, r_k) p(t_{-a}|t_a, r_k) = \sum_{t'_a \neq t_a} \sum_{k=1}^{K} \lambda(t_a, t'_a, r_k) p(t_{-a}|t'_a, r_k), \qquad \forall t_{-a} \qquad (10)$$

where $\sum_{t'_a \neq t_a} \sum_{k=1}^{K} \lambda(t_a, t'_a, r_k) = 1$.

Proof: Write (8) and (9) as a linear program:

LP $$\min_{\gamma_a} 0$$

s.t.

$$\sum_{t_{-a}} p(t_{-a}|t_a, r_k)\, \gamma_a(t_{-a}|t_a) = 0, \quad \forall k,\ \forall t_a$$

$$\sum_{t_{-a}} p(t_{-a}|t'_a, r_k)\, \gamma_a(t_{-a}|t_a) \quad \geq \quad 1, \quad \forall k,\ \forall t'_a \neq t_a.$$

Its dual is

DLP
$$\max_{\lambda \geq 0,\, \mu} \sum_{t_a \neq t'_a} \sum_{k=1}^{K} \lambda(t_a, t'_a, r_k)$$

s.t.

$$\sum_{k=1}^{K} \mu(t_a, r_k) p(t_{-a}|t_a, r_k) \;=\; \sum_{t'_a \neq t_a} \sum_{k=1}^{K} \lambda(t_a, t'_a, r_k) p(t_{-a}|t'_a, r_k), \qquad \forall t_{-a}, \quad \forall k, \forall t_a.$$

A robust lottery exists if and only if LP has a feasible solution. Any LP feasible solution is also optimal (with objective function value 0). Therefore, a robust lottery exists if and only if every solution to DLP has $\lambda \equiv 0$.

If DLP has a solution with $0 \neq \lambda \geq 0$ then there exists a solution with $\sum_{t'_a \neq t_a} \sum_{k=1}^{K} \lambda(t_a, t'_a, r_k) = 1$. □

Observe that if agent a does not have access to an informative signal (or, equivalently, R_a is uninformative and therefore $p(\cdot|t_a, r_k) = p(\cdot|t_a)$ for all t_a, r_k) then (10) reduces to the full extraction condition.

The characterization in Proposition 4 can used to obtain several sufficient conditions on prior and posterior beliefs for the existence or non-existence of robust lotteries. For example, if there exist two distinct types t'_a, t_a such that $p(\cdot|t'_a)$ is in the affine space generated by $p(\cdot|t_a, r_k)$, $k = 1, 2, \ldots, K$ then robust lotteries do not exist. The sufficient conditions (for existence and non-existence of robust lotteries, respectively) in Propositions 5 and 6 below do not depend on prior beliefs. The number of realizations of the signal, K, turns out to be important in these propositions.

Proposition 5 *If $N_a \geq n_a K$ then robust lotteries exist for generic prior distributions over the type space $p(t)$ and likelihoods Q_a.*

Proof: Suppose there exists t_a, r_k such that $p(\cdot|t_a, r_k)$ is a linear combination of $p(\cdot|t'_a, r_{k'})$, $\forall (t'_a, r_{k'}) \neq (t_a, r_k)$. That is, there exists $\eta \in \Re^{n_a K - 1}$ such that $\forall t_{-a}$

$$p(t_{-a}|t_a, r_k) \;=\; \sum_{k' \neq k} \eta(t_a, r_{k'})\, p(t_{-a}|t_a, r_{k'}) + \sum_{t'_a \neq t_a} \sum_{k=1}^{K} \eta(t'_a, r_{k'})\, p(t_{-a}|t'_a, r_{k'})$$

$$\Longrightarrow \; \frac{q(r_k|t_a, t_{-a}) p(t_a, t_{-a})}{\sum_{t'_{-a}} q(r_k|t_a, t'_{-a}) p(t_a, t'_{-a})} \;=\; \sum_{k' \neq k} \eta(t_a, r_{k'}) \frac{q(r_{k'}|t_a, t_{-a}) p(t_a, t_{-a})}{\sum_{t'_{-a}} q(r_{k'}|t_a, t'_{-a}) p(t_a, t'_{-a})}$$

$$+ \sum_{t'_a \neq t_a} \sum_{k'=1}^{K} \eta(t'_a, r_{k'}) \frac{q(r_{k'}|t'_a, t_{-a})p(t'_a, t_{-a})}{\sum_{t'_{-a}} q(r_{k'}|t'_a, t'_{-a})p(t'_a, t'_{-a})}$$

Define $\widehat{\eta}(t'_a, r_{k'}) \equiv \eta(t'_a, r_{k'}) \frac{\sum_{t'_{-a}} q(r_k|t_a, t'_{-a})p(t_a, t'_{-a})}{\sum_{t'_{-a}} q(r_{k'}|t'_a, t'_{-a})p(t'_a, t'_{-a})}$, $\forall (t'_a, r_{k'}) \neq (t_a, r_k)$. Noting that $\widehat{\eta}$ does not depend on t_{-a}, we have

$$\begin{aligned} q(r_k|t_a, t_{-a})p(t_a, t_{-a}) &= \sum_{k' \neq k} \widehat{\eta}(t_a, r_{k'}) q(r_{k'}|t_a, t_{-a}) p(t_a, t_{-a}) \\ &\quad + \sum_{t'_a \neq t_a} \sum_{k'=1}^{K} \widehat{\eta}(t'_a, r_{k'}) q(r_{k'}|t'_a, t_{-a}) p(t'_a, t_{-a}), \quad \forall t_{-a} \\ \implies \quad \Pr(r_k, t_a, t_{-a}) &= \sum_{k' \neq k} \widehat{\eta}(t_a, r_{k'}) \Pr(r_{k'}, t_a, t_{-a}) + \sum_{t'_a \neq t_a} \sum_{k'=1}^{K} \widehat{\eta}(t'_a, r_{k'}) \Pr(r_{k'}, t'_a, t_{-a}), \quad \forall t_{-a} \end{aligned}$$

Generic initial beliefs p and likelihoods Q_a give rise to generic joint distributions $[\Pr(r_{k'}, t_a, t_{-a})]$. Therefore, the above equality holds with probability zero. To see this, write the joint distribution $[\Pr(r_{k'}, t'_a, t_{-a})]$ in a $n_a K \times N_a$ matrix with rows corresponding to $\Pr(r_{k'}, t'_a, \cdot)$ and columns corresponding to $\Pr(\cdot, \cdot, t_{-a})$. As $n_a K \leq N_a$, the rank of this matrix is $n_a K$ for generic joint distributions over R_a, T^a, T^{-a}. Therefore, the $n_a K$ possible posterior beliefs over T^{-a}, $p(\cdot|r_{k'}, t_a)$, $k = 1, 2, \ldots, K$, $t_a \in T^a$, are linearly independent for generic beliefs $p(t)$ and likelihoods $q(r_{k'}|t)$. Consequently, Proposition 4 implies that robust lotteries exist. □

If, instead, K is sufficiently large, robust lotteries do not exist for generic prior beliefs. In order to establish this we need the following definition.

Definition 2 The information signal $R_a(t_a)$ *spans* the type space T^{-a} if the rank of the likelihood matrix $Q_a(t_a)$ is N_a.

Recall that $Q_a(t_a)$ is a $K \times N_a$ matrix. If $R_a(t_a)$ spans T^{-a} then $K \geq N_a$ and there exist N_a signal realizations, labeled $k = 1, 2, \ldots, N_a$, such that $q(r_k|\cdot, t_a)$, $k = 1, 2, \ldots, N_a$ (i.e., the first N_a rows of $Q_a(t_a)$) are linearly independent. If $K \geq N_a$ then the spanning condition is generic.

A spanning information signal may be exceedingly noisy in the sense that the posterior distributions $p(\cdot|t_a, r_k)$ for each r_k may be very close to the prior distribution $p(\cdot|t_a)$. That is, the column vectors of $Q_a(t_a)$ can be very close to (but not equal to) $\frac{1}{K}$ and still be linearly independent.

The next lemma, which is proved in an appendix, shows that the spanning condition is equivalent to the existence of N_a linearly independent posteriors.

Lemma 4 *$R_a(t_a)$ spans T^{-a} if and only if for any prior $p(\cdot|t_a)$ there exist N_a linearly independent posteriors after observing $R_a(t_a)$.*

This leads directly to the next result.

Proposition 6 *If $R_a(t_a)$ spans T^{-a} then for any prior $p(\cdot|t_a)$ there does not exist a lottery that is robust to acquisition of $R_a(t_a)$ by agent (a, t_a).*

Proof: By Lemma 4, linear combinations of the posteriors $p(\cdot|t_a, r_k),\ k = 1, 2, \ldots, K$ form a subspace of dimension N_a. This subspace includes $\Delta(T^{-a})$, the set of probability distributions over T^{-a}; therefore it also includes the convex hull of $p(\cdot|t'_a, r_k),\ \forall t'_a \neq t_a,\ k = 1, 2, \ldots, K$. Non-existence of a robust lottery follows from Proposition 4. □

In the above description, agent a has access to one information signal which has $K \geq N_a$ realizations. The same analysis carries over to a setting where agent (a, t_a) has a choice of one of several information signals $R^1(t_a)$, $R^2(t_a)$, $\ldots$, $R^\nu(t_a)$; spanning is satisfied if the possible signal realizations of $R^1(t_a)$, $R^2(t_a)$, $\ldots$, $R^\nu(t_a)$ give rise to N_a possible linearly independent posteriors.

Expectedly, if the cost of a spanning information signal is sufficiently low, full extraction is impossible.

Corollary 1 *Suppose that $R_a(t_a)$ spans T^{-a}. If (f, x) is a mechanism in which x is based on FE lotteries and $c_a \in [0, \pi_a(t_a))$, then (f, x) is not a full extraction mechanism. In particular, if $c_a = 0$ then there does not exist any full extraction mechanism.*

Proof: As $R_a(t_a)$ spans T^{-a} it is informative. Consequently, $p(\cdot|t_a) \centerdot \gamma_a(\cdot|t_a) = 0$ implies there exists r_k such that $p(\cdot|t_a, r_k) \centerdot \gamma_a(\cdot|t_a) > 0$. Thus, (6) is satisfied and $\pi_a(t_a) > 0$. For at least one realization of $R_a(t_a)$ either (5) or (6) is satisfied, undermining Bayesian incentive compatibility or full extraction or both. □

Spanning is not necessary for non-existence of robust lotteries. Even if agent a's information signals lead to less than K possible linearly independent posteriors, there may not exist robust lotteries (see Example 2 below). If robust lotteries do not exist then a mechanism designer who knows Q_a might consider using the following type of lotteries.

Definition 3 If $\gamma_a(\cdot|t_a)$ satisfies $p(\cdot|t_a) \centerdot \gamma_a(\cdot|t_a) = 0$ and (9) then $\gamma_a(\cdot|t_a)$ is *partially robust* to acquisition of information signal $R_a(t_a)$ by agent (a, t_a).

A partially robust lottery is also a full extraction lottery. The next result is a characterization of partially robust lotteries.

Proposition 7 *Consider an information structure (T, p) in which agent a may have access to information signal R_a. Partially robust lotteries exist if and only if there does not exist t_a such that $p(\cdot|t_a)$ is a convex combination of beliefs $p(\cdot|t'_a, r_k)$, $t'_a \in T^a \backslash t_a$, $k = 1, 2, \ldots, K$. That is, there does not exist t_a and $\lambda \geq 0$ such that*

$$p(t_{-a}|t_a) = \sum_{t'_a \neq t_a} \sum_{k=1}^{K} \lambda(t_a, t'_a, r_k) p(t_{-a}|t'_a, r_k), \qquad \forall t_{-a}. \tag{11}$$

where $\sum_{t'_a \neq t_a} \sum_{k=1}^{K} \lambda(t_a, t'_a, r_k) = 1$.

Proof: Write the conditions for a partially robust lottery as a linear program:

$$\min_{\gamma_a} 0$$

s.t.

$$\begin{aligned} \sum_{t_{-a}} p(t_{-a}|t_a)\, \gamma_a(t_{-a}|t_a) &= 0, \quad \forall t_a \\ \sum_{t_{-a}} p(t_{-a}|t'_a, r_k)\, \gamma_a(t_{-a}|t_a) &\geq 1, \quad \forall k,\ \forall t'_a \neq t_a. \end{aligned}$$

Its dual is

$$\max_{\lambda \geq 0,\, \mu} \sum_{t_a \neq t'_a} \sum_{k=1}^{K} \lambda(t_a, t'_a, r_k)$$

s.t.

$$\mu(t_a) p(t_{-a}|t_a) = \sum_{t'_a \neq t_a} \sum_{k=1}^{K} \lambda(t_a, t'_a, r_k) p(t_{-a}|t'_a, r_k), \qquad \forall t_{-a}, \quad \forall k, \forall t_a.$$

The rest of the proof mimics the proof of Proposition 4. □

Proposition 7 implies that if R_a is noisy enough then there exist partially robust lotteries. The proof follows by noting that (11) reduces to the full extraction condition as $p(\cdot|t_a, r_k) \to p(\cdot|t_a)$.

Corollary 2 *Consider an information structure (T, p) which satisfies the full extraction condition. Agent a has access to information signal R_a. There exists $\epsilon > 0$ such that if $||p(\cdot|t_a) - p(\cdot|t_a, r_k)|| < \epsilon$ for all (t_a, r_k) then partially robust lotteries exist.*

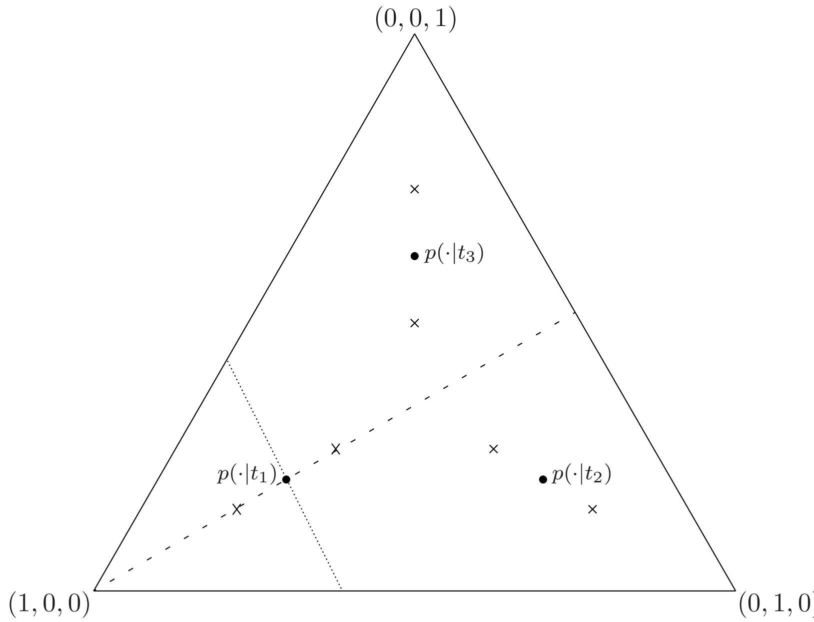

Figure 2: Partially robust lotteries

Figure 2 illustrates an example in which robust lotteries do not exist but partially robust lotteries do. The basic set-up is as in the example of Figure 1 except that the posterior beliefs after observing signals are different. Focusing on $t_a = t_1$, the two points marked $\times$ which are closest to $p(\cdot|t_1)$ are $p(\cdot|t_1, r_1)$ and $p(\cdot|t_1, r_2)$. The (positively-sloped) broken line represents beliefs that are linear combinations of $p(\cdot|t_1)$, $p(\cdot|t_1, r_1)$, and $p(\cdot|t_1, r_2)$. This broken line intersects the convex hull (not shown) of points $p(\cdot|t_2, r_1)$, $p(\cdot|t_2, r_2)$, $p(\cdot|t_3, r_1)$, and $p(\cdot|t_3, r_2)$.[15] Therefore, Proposition 4 implies that robust lotteries do not exist. A geometric proof of non-existence is as follows. Any lottery that satisfies (8) for (t_1, r_1), (t_1, r_2) must be obtained from the coefficients of the hyperplane represented by the broken line. But (t_2, r_1), (t_2, r_2) and (t_3, r_1), (t_3, r_2) are on opposite sides of this hyperplane making it impossible for a lottery that satisfies (8) to also satisfy (9). The points $p(\cdot|t_\ell, r_j)$, $\ell = 2, 3$, $j = 1, 2$ lie on the same side of the (negatively-sloped) dotted line through $p(\cdot|t_1)$. The coefficients of a hyperplane represented by this line can be used to construct a partially robust lottery.

Example 1 below shows that the sufficient condition of Proposition 5 is not necessary for the existence of robust lotteries. Example 2 shows the spanning condition of Proposition 6 is not necessary for non-existence of robust lotteries.

[15]The four points marked $\times$ close to $p(\cdot|t_2)$ and $p(\cdot|t_3)$ respectively.

Example 1: (Robust lotteries exist and $N_a < n_a K$)

There are two symmetric agents and each agent has three possible types t_1, t_2, t_3. Agent a's conditional beliefs about b's type are

$$p(t_b|t_a = t_i) \;=\; \begin{cases} \theta, & \text{if } t_b = t_i \\ \frac{1-\theta}{2}, & \text{otherwise} \end{cases} \tag{12}$$

where $\theta \in (\frac{1}{3}, 1)$. The full extraction condition is satisfied for all $\theta \in (\frac{1}{3}, 1)$. (If $\theta = \frac{1}{3}$ then the agents' types are independent.)

Agent a has access to signal R_a which takes two values r or $\bar{r}$. The likelihood function is

$$\begin{aligned}
q(R_a = r|t_b = t_2, t_a = t_1) &= q(R_a = \bar{r}|t_b = t_3, t_a = t_1) = q \\
q(R_a = r|t_b = t_3, t_a = t_2) &= q(R_a = \bar{r}|t_b = t_1, t_a = t_2) = q \\
q(R_a = r|t_b = t_1, t_a = t_3) &= q(R_a = \bar{r}|t_b = t_2, t_a = t_3) = q \\
q(R_a = r|t_b = t_i, t_a = t_i) &= q(R_a = \bar{r}|t_b = t_i, t_a = t_i) = 0.5, \qquad i = 1, 2, 3
\end{aligned}$$

where $q > 0.5$. Thus, after observing R_a agent (a, t_i)'s belief that $t_b = t_i$ remains unchanged at θ. The posterior distribution is

$$\begin{aligned}
p(\cdot|t_a = t_1, r) &= \Big[\theta,\ q(1-\theta),\ (1-q)(1-\theta)\Big] \\
p(\cdot|t_a = t_1, \bar{r}) &= \Big[\theta,\ (1-q)(1-\theta),\ q(1-\theta)\Big] \\
p(\cdot|t_a = t_2, r) &= \Big[(1-q)(1-\theta),\ \theta,\ q(1-\theta)\Big] \\
p(\cdot|t_a = t_2, \bar{r}) &= \Big[q(1-\theta),\ \theta,\ (1-q)(1-\theta)\Big] \\
p(\cdot|t_a = t_3, r) &= \Big[q(1-\theta),\ (1-q)(1-\theta),\ \theta\Big] \\
p(\cdot|t_a = t_3, \bar{r}) &= \Big[(1-q)(1-\theta),\ q(1-\theta),\ \theta\Big].
\end{aligned}$$

Prior and posterior beliefs in this example are illustrated in Figure 1 of Section 2. As there are only two possible signal realizations and $|T^b| = 3$, $R_a(t_1)$ does not span T^b.

If $\gamma_a = (\gamma_1, \gamma_2, \gamma_3)$ is a FE lottery for (a, t_1) then it may be verified that $\gamma_1 = -(\gamma_2 + \gamma_3)\frac{1-\theta}{2\theta}$, $\gamma_2, \gamma_3 > \gamma_1$, and $\gamma_1 < 0$. If, in addition, $\gamma_a = (\gamma_1, \gamma_2, \gamma_3)$ is a robust lottery then it also satisfies:

$$\begin{aligned}
p(\cdot|t_a = t_1, r) \cdot \gamma_a \;&=\; p(\cdot|t_a = t_1, \bar{r}) \cdot \gamma_a \;=\; 0 \\
p(\cdot|t_a = t_i, r) \cdot \gamma_a \;>\; 0 \;\; &\text{and} \;\; p(\cdot|t_a = t_i, \bar{r}) \cdot \gamma_a \;>\; 0, \quad i = 2, 3.
\end{aligned}$$

It may be verified that $\gamma_a = (-\frac{1-\theta}{\theta}k, k, k,)$, $k > 0$, is a robust lottery provided that either $\theta \geq 1/2$ or $\theta \in (1/3, 1/2)$ and $q < \frac{\theta}{1-\theta}$. Note that a robust lottery has relative measure zero within the set $\{(\gamma_1, \gamma_2, \gamma_3) \mid \gamma_1 = -(\gamma_2 + \gamma_3)\frac{1-\theta}{2\theta} < 0, \gamma_2, \gamma_3 > \gamma_1\}$, the set of all FE lotteries for (a, t_1).[16] $\triangle$

Example 2: (Spanning not satisfied, robust lotteries do not exist, partially robust lotteries may exist)

As in Example 1, there are two symmetric agents, each has three possible types and their beliefs are specified in (12). Agent a has access to signal R_a which takes two values r or $\bar{r}$ with likelihood function

$$q(R = r|t_b, t_a) = \begin{cases} q > 0.5, & \text{if } t_b = t_a \\ 0.5, & \text{otherwise.} \end{cases} \tag{13}$$

R_a does not span T^b.

Agent a's posterior beliefs over t_b when $t_a = t_1$ and either $R_a = r$ or $R_a = \bar{r}$ are

$$\begin{aligned} p(\cdot|t_a = t_1, r) &= \frac{1}{2q\theta + (1-\theta)}\Big[2q\theta, \frac{1-\theta}{2}, \frac{1-\theta}{2}\Big] \\ p(\cdot|t_a = t_1, \bar{r}) &= \frac{1}{2(1-q)\theta + (1-\theta)}\Big[2(1-q)\theta, \frac{1-\theta}{2}, \frac{1-\theta}{2}\Big]. \end{aligned}$$

Prior and posterior beliefs in this example are illustrated in Figure 2.

Any robust lottery $\gamma_a = (\gamma_1, \gamma_2, \gamma_3)$ satisfies equation (8):

$$\begin{aligned} 4q\theta\gamma_1 + (1-\theta)\gamma_2 + (1-\theta)\gamma_3 &= 0 \\ 4(1-q)\theta\gamma_1 + (1-\theta)\gamma_2 + (1-\theta)\gamma_3 &= 0. \end{aligned}$$

Substituting $\gamma_1 = -(\gamma_2 + \gamma_3)\frac{1-\theta}{2\theta}$ into the first equation and noting that $q \neq 0.5$ yields

$$(\gamma_2 + \gamma_3)(1-\theta) = 0.$$

$\theta < 1$ implies $\gamma_2 + \gamma_3 = 0$ which in turn implies that $\gamma_1 = 0$. But that contradicts $\gamma_1 < \gamma_2, \gamma_3$. Hence, for any $q > 0.5$ and $\theta \in (\frac{1}{3}, 1)$ there do not exist robust lotteries.

However, partially robust lotteries exist for some parameter values. Note that

$$\begin{aligned} p(\cdot|t_a = t_2, r) &= \frac{1}{2q\theta + (1-\theta)}\Big[\frac{1-\theta}{2}, 2q\theta, \frac{1-\theta}{2}\Big] \\ p(\cdot|t_a = t_2, \bar{r}) &= \frac{1}{2(1-q)\theta + (1-\theta)}\Big[\frac{1-\theta}{2}, 2(1-q)\theta, \frac{1-\theta}{2}\Big]. \end{aligned}$$

[16] Robust lotteries for (a, t_2) and (a, t_3) are obtained by symmetry.

If there exists a partially robust lottery, then there exists one of the form $\gamma_a = (-\frac{1-\theta}{\theta}k, k, k, k)$, $k > 0$. Equation (9) for t_2 implies that

$$\begin{aligned} -\frac{1-\theta}{\theta}\frac{1-\theta}{2} + 2q\theta + \frac{1-\theta}{2} &> 0 \\ -\frac{1-\theta}{\theta}\frac{1-\theta}{2} + 2(1-q)\theta + \frac{1-\theta}{2} &> 0. \end{aligned}$$

These inequalities are satisfied if

$$\begin{aligned} \text{either} \quad & \theta \geq 1/2 \\ \text{or} \quad & \theta \in (1/3,\, 1/2) \ \text{ and } \ q < \bar{q} \equiv 0.5 + \frac{3\theta - 1}{4\theta^2}. \end{aligned} \tag{14}$$

Note that $\bar{q} \downarrow 0.5$ as $\theta \downarrow \frac{1}{3}$. As beliefs approach independence partially robust lotteries exist only for very noisy signals. △

3.2 Informational rents

If robust lotteries do not exist but partially robust lotteries do then a mechanism designer can pay agent a informational rents to dissuade him from acquiring information. Let $\gamma_a^\Delta(\cdot|t_a)$ be defined by subtracting Δ from each outcome of a lottery $\gamma_a(\cdot|t_a)$:

$$\gamma_a^\Delta(t_{-a}|t_a) \ \equiv \ \gamma_a(t_{-a}|t_a) - \Delta, \qquad \forall t_{-a} \in T^{-a}.$$

If $\gamma_a(\cdot|t_a)$ is a partially robust lottery then for sufficiently large Δ, $\gamma_a^\Delta(\cdot|t_a)$ ensures Bayesian incentive compatibility (even if R_a spans T^{-a}). This is shown in the next lemma. In the statement of this lemma $\pi_a(t_a)$ and $\pi_a^\Delta(t_a)$ are the private values of information when the payments are $x_a(t) = u_a(f(t), t) + \gamma_a(t_{-a}|t_a)$ and $x_a^\Delta(t) = u_a(f(t), t) + \gamma_a^\Delta(t_{-a}|t_a)$ respectively.

Lemma 5 *Let (T, p) be an information structure and let $\gamma_a(\cdot|t_a)$ be a partially robust lottery w.r.t. $R_a(t_a)$ on this information structure. Consider a social choice rule f which may be implemented by either $x_a(t)$ or $x_a^\Delta(t)$ defined above. Then $\pi_a^\Delta(t_a) < \pi_a(t_a)$. If Δ is large enough, then $\pi_a^\Delta(t_a) = 0$ for all t_a.*

From the proof of Lemma 5 it is clear that if a partially robust lottery exists then by choosing large enough Δ any social choice function may be implemented, including the efficient rule. Therefore,

Corollary 3 *Consider an information structure (T, p). Suppose that lotteries that are partially robust to R_a exist. Then for any mechanism design problem on (T, p) an efficient mechanism exists.*

The next lemma, which is proved in an appendix, raises the possibility that the information rent Δ required to implement a social choice rule may be large.

Lemma 6 *Let $\gamma_a(\cdot|t_a)$ be a partially robust lottery for a mechanism design problem in which full extraction constraints bind towards t_a. Suppose that agent a's beliefs are ϵ-close and that $\gamma_a^\Delta(\cdot|t_a) = \gamma_a(\cdot|t_a) - \Delta$ implements the mechanism. Then $||\gamma_a(\cdot|t_a)|| \to \infty$ as $\epsilon \to 0$.*

Example 3: (Information rents for very noisy signals)

There are two agents and the distribution of types is as in Example 2. Agent a has access to signal R_a with the likelihoods specified in (13). We know from Example 2 that when θ and q satisfy (14) there exists a partially robust lottery of the form $\gamma_a(\cdot|t_a = t_1) = (-\frac{1-\theta}{\theta}k, k, k)$, $k > 0$. Let $\gamma_a^\Delta(\cdot|t_a = t_1) = (-\frac{1-\theta}{\theta}k - \Delta, k - \Delta, k - \Delta)$; rent Δ is paid to agent (a, t_1) to not acquire information and truthfully report his type.

The value of information $\pi_a^\Delta(t_a)$ at the smallest Δ at which agent (a, t_1) will not gather information satisfies $\pi_a^\Delta(t_a) = c_a$. Therefore,

$$\begin{aligned} \pi_a^\Delta(t_1) &= -\min\{[p(\cdot|t_a = t_1, r) \cdot \gamma_a^\Delta(\cdot|t_a = t_1)], 0\}p(r|t_a = t_1) \\ &\quad -\min\{[p(\cdot|t_a = t_1, \bar{r}) \cdot \gamma_a^\Delta(\cdot|t_a = t_1)], 0\}p(\bar{r}|t_a = t_1) - \Delta \;=\; c_a. \end{aligned}$$

At this smallest Δ if (a, t_1) were to gather information he would participate if and only if $R_a = r$. Hence, noting that $p(t_b = \cdot, R_a = r|t_a = t_1) = (q\theta, \frac{1-\theta}{4}, \frac{1-\theta}{4})$, we have

$$\begin{aligned} \pi_a^\Delta(t_1) &= -(q\theta, \frac{1-\theta}{4}, \frac{1-\theta}{4}) \cdot \gamma_a^\Delta(\cdot|t_a = t_1) - \Delta \\ &= [q(1-\theta) - \frac{1-\theta}{2}]k - [(1-q)\theta + \frac{1-\theta}{2}]\Delta \;=\; c_a \\ \implies \quad \Delta &= \frac{[q(1-\theta) - \frac{1-\theta}{2}]k - c_a}{(1-q)\theta + \frac{1-\theta}{2}}. \end{aligned} \tag{15}$$

For any fixed $q > 0.5$, as beliefs approach independence, i.e. $\theta \to \frac{1}{3}$, (14) is violated and partially robust lotteries fail to exist. In order to ensure existence of partially robust lotteries and to show that information rents can be bounded away from zero for signals that are almost uninformative suppose that $\theta = \frac{1+\epsilon}{3}$ and $q = \frac{1+\epsilon}{2}$. These values satisfy (14). Substituting in (15), we have

$$\Delta = \frac{\epsilon(2-\epsilon)k - 6c_a}{3-\epsilon-\epsilon^2}.$$

Suppose that if full extraction is attempted without an FE lottery then type t_3 (or t_2) gains $\bar{\kappa} > 0$ from misreporting his type as t_1. It may be verified that $k > \frac{\bar{\kappa}}{\epsilon}$ is necessary for incentive compatibility. Therefore,

$$\Delta > \frac{(2-\epsilon)\bar{\kappa} - 6c_a}{3-\epsilon-\epsilon^2} \geq \frac{(2-\epsilon)\bar{\kappa} - 6c_a}{3}.$$

Suppose that $c_a < \frac{1}{3}\bar{\kappa}$. Then as $\epsilon \to 0$, R_a approaches uninformativeness but the informational rent Δ is bounded away from zero. △

To summarize, as beliefs approach independence agent a has an incentive to acquire spanning information at any cost. This leads to non-existence of robust lotteries. If the spanning information is noisy enough then partially robust lotteries exist and any social choice function may be implemented using these lotteries. However, the rent extracted by the agent can be substantial.

3.3 The mechanism designer's options

If robust lotteries exist then both types of mechanism designers, M_{ER} and M_{RE}, will use them. However, as already noted, when beliefs approach independence the private value of information that spans N_a becomes unboundedly large. Therefore, it is not unreasonable that an agent will seek out such information.

If robust lotteries do not exist but partially robust lotteries do, M_{ER} will use them to implement an efficient mechanism, unless another efficient mechanism yields greater revenues. Mechanism designer M_{RE} may or may not want to use partially robust lotteries. If the information rent earned by agent a in the implementation of the efficient mechanism using partially robust lotteries exceeds agent a's marginal product, M_{RE} would rather exclude agent a and, instead, implement a full extraction mechanism among agents $A \backslash a$.[17]

Providing incentives for information acquisition

I close this section with a mechanism that is important when there are many agents with information access. It has been noted that because $\rho < 1$ the mechanism

[17] However, suppose that the mechanism designer knows that one agent has access to information about other agents but he does not know the identity of this agent. Then a mechanism that excludes the unknown agent with access to an information signal is not an option.

designer cannot force agent a to acquire information. However, it may be possible to design a mechanism such that if agent a has access to information he acquires it, reports his signal realization and is reimbursed c_a. If agent a does not have access to an information signal he reports a null signal $\emptyset$. Thus agent (a, t_a) makes one of $K+1$ possible reports: $(t_a, \emptyset)$ or (t_a, r_k), $k = 1, 2, \ldots, K$. If he reports $(t_a, \emptyset)$ then his transfer is based on robust lottery $\gamma_a(\cdot|t_a)$. If he reports (t_a, r_k) then his transfer is based on a FE lottery $\gamma_a(\cdot|t_a, r_k)$ such that

$$p(\cdot|t_a, r_k) \centerdot \gamma_a(\cdot|t_a, r_k) = -c_a, \qquad \forall (t_a, r_k), \tag{16}$$

$$p(\cdot|t'_a, r'_k) \centerdot \gamma_a(\cdot|t_a, r_k) > 0, \qquad \forall (t'_a, r'_k) \neq (t_a, r_k). \tag{17}$$

The lotteries $[\gamma_a(\cdot|t_a),\ \gamma_a(\cdot|t_a, r_k),\ k = 1, 2, \ldots, K]$, where $\gamma_a(\cdot|t_a)$ is a robust lottery and each $\gamma_a(\cdot|t_a, r_k)$ satisfies (16) and (17), make it incentive compatible for agent a to acquire information if he is able to and report it.[18] As shown in the next lemma, existence of robust lotteries is necessary and sufficient for the existence of such a mechanism.

Lemma 7 *There exists a zero-surplus mechanism which provides agent a incentives to acquire and reveal information to the mechanism designer if and only if robust lotteries exist.*

If robust lotteries exist then full extraction is possible. The mechanism of Lemma 7 is less efficient (because information gathering cost is incurred) and yields less revenue (because agent a is reimbursed c_a) than the full extraction mechanism. Consequently, neither M_{ER} nor M_{RE} would want to use such a mechanism. However, if many agents have access to information there may be a reason to.

4 Information acquisition by many agents

We now consider the case where each agent $a \in A$ has, with probability $\rho < 1$, access to information signal R_a that is informative about T^{-a}. If robust lotteries exist for each agent then full extraction is possible. If only partially robust lotteries exist, then M_{ER} designs a mechanism with these lotteries unless another efficient mechanism yields greater revenues. The necessary and sufficient conditions for existence of robust lotteries and partially robust lotteries for an agent are exactly as in Section 3.1.

[18]If $p(\cdot|t_a, r_k) \centerdot \gamma_a(\cdot|t_a, r_k) = -c_a - \epsilon$ for each (t_a, r_k) then agent a with access to information strictly prefers to acquire and report it truthfully rather than not acquire information or pretend not to have access to information.

The main difference when more than one agent has access to information is that a sufficiently well-informed mechanism designer may be able to design a mechanism that induces agents with information access to buy information and report their signal realization. The construction, outlined in the proof of the next proposition, exploits the fact that when several agents have information access the extended type space $\tilde{T}$ (which consists player payoff types and signal realizations) is large.

Proposition 8 *Suppose that agents are symmetric and that the mechanism designer knows the likelihood matrix and cost of each agent's signal. Then there exists a zero-surplus mechanism in which the mechanism designer extracts the efficient surplus less information acquisition costs.*

Proof: Let $n = |T^a|$ for each a and let K be the number of realizations of each agent's signal. As agents are symmetric, n and K do not depend on the agent. Therefore, the cardinality of the extended type space $\tilde{T}^{-a}$ of agents other than agent a exceeds nK: $|\tilde{T}^{-a}| = (|A| - 1)n(K + 1) > nK$. Proposition 5 implies that robust lotteries exist in the extended type space. Repeated application of Lemma 7 for each agent completes the proof. □

If information acquisition costs are small then the mechanism designer does almost as well as in a full extraction mechanism. By acquiring and reporting their information, agents are collectively responsible for restoring (almost) full extraction. However, in the mechanism of Proposition 8 there may exist another equilibrium in which each agent earns positive surplus. Suppose that robust lotteries do not exist in the payoff type spaces T^{-a}. In this purported alternative equilibrium (the existence of which is an open question), agents always report the null signal and collect information if they are able to. Therefore, the type space is not extended and robust lotteries do not exist. Agents acquire information if they can and participate only if they earn positive surplus. Such an equilibrium is more likely to exist in a private values model so that participating agents do not draw payoff-relevant inferences from non-participation decisions by others.

5 Concluding remarks

In this model, it is socially wasteful for an agent to acquire costly information about other agents' types. However, because agents earn zero surplus, information about other agents has positive private value and this value increases without bound as beliefs approach independence. Therefore, it is important to investigate the robustness of full extraction mechanisms to information acquisition by agents.

The robustness of full extraction mechanisms depends on the mechanism designer's knowledge of agents' information signals. If the mechanism designer does not know the probability distribution of signals then full extraction is impossible. If agents are asymmetric and, in particular, one agent has possible access a rich variety of information signals (i.e., a large number of possible posteriors that span other agents' extended type space) then this agent earns information rents even when the mechanism designer knows signal distributions and costs. If agents are not very asymmetric then a mechanism designer who knows the probability distributions and costs of agents' information signals can, by inducing agents to gather and report their information, extract the entire surplus less information acquisition costs. However, in practice it is unlikely that a mechanism designer has sufficient knowledge about agents' information sources to implement such a mechanism.

In an information acquisition model with independent information, Bergemann and Valimaki [3] show that agents acquire an efficient amount of information only in a private values setting. One's initial intuition might be that in an information acquisition model with correlated information, an adaptation of the Cremer-McLean mechanism achieves efficient information acquisition and full extraction of surplus. That is, the mechanism designer asks each agent to acquire an efficient amount of information, compensates each agent for his information acquisition cost, and implements a full extraction mechanism. However, because private and social values of information diverge, things are not so straightforward. In the model of this paper, it is inefficient to acquire any information. If only one agent has access to a spanning information signal then full extraction is impossible (Proposition 6); efficient information acquisition is possible if partially robust lotteries exist (Corollary 3). When multiple symmetric agents have access to information and the mechanism designer is well-informed about agents' information signals, then zero-surplus mechanisms can be designed. However, to the extent that information is costly these mechanisms are neither efficient nor do they fully extract surplus (Proposition 8). It would be interesting to investigate the possibility of efficient information acquisition in a model in which information is both correlated and socially useful.

6 Appendix: Proofs of lemmas

Proof of Lemma 1: It is enough to show that for at least one signal realization, agent (a, t_a) will not participate. Let $\gamma_a^*(\cdot|t_a) \in \Gamma_a(t_a)$ be the full extraction lottery which is part of x_a^* [i.e., $x_a^*(t_a, t_{-a}) = u_a(f^*(t_a, t_{-a}), (t_a, t_{-a})) + \gamma_a^*(t_{-a}|t_a)$]. Let r_k be a realization of R_a. Equation (4) implies that $p(\cdot|t_a, r_k) \cdot \gamma_a^*(\cdot|t_a) = 0$ if and only if the vector $q(r_k|\cdot, t_a)p(\cdot|t_a)$ is in the null space of $\gamma_a^*(\cdot|t_a)$. The set of likelihoods $q(r_k|\cdot, t_a)$ for which this is true is of measure zero. Therefore, as $p(\cdot|t_a) \cdot \gamma_a^*(\cdot|t_a) = 0$, for almost every R_a there exists a realization r_ℓ such that $p(\cdot|t_a, r_\ell) \cdot \gamma_a(\cdot|t_a) > 0$ (and another realization $r_{\ell'}$ such that $p(\cdot|t_a, r_{\ell'}) \cdot \gamma_a(\cdot|t_a) < 0$). Agent a will not participate if he observes $R_a = r_\ell$. $\square$

Proof of Lemma 2: Suppose that agent a's beliefs are in a ϵ neighborhood of $p_a(\cdot)$. Let $N(p_a, \epsilon)$ denote a closed set of probability distributions on T^{-a} that are in a ϵ neighborhood of $p_a(\cdot)$. Let t_a, t_a' satisfy the non-degeneracy assumption. Define

$$\bar{\kappa}_a(t_a, t_a') \equiv \min_{p(\cdot|t_a') \in N(p_a, \epsilon)} \kappa_a(t_a, t_a').$$

For small enough ϵ, the set $N(p_a, \epsilon)$ contains only strictly positive beliefs and therefore $\bar{\kappa}_a(t_a, t_a') > 0$. Consider $p(\cdot|t_a')$, $p(\cdot|t_a)$ that are ϵ-close to $p_a(\cdot)$. To dissuade agent (a, t_a') from misreporting his type as t_a it is necessary that $p(\cdot|t_a') \cdot \gamma_a(\cdot|t_a) \geq \bar{\kappa}_a(t_a, t_a')$. As $p(\cdot|t_a) \cdot \gamma_a(\cdot|t_a) = 0$, this implies $[p(\cdot|t_a') - p(\cdot|t_a)] \cdot \gamma_a(\cdot|t_a) \geq \kappa_a(t_a, t_a')$. The fact that $\bar{\kappa}_a(t_a, t_a')$ does not depend on $p(\cdot|t_a), p(\cdot|t_a')$, implies $||\gamma_a(\cdot|t_a)|| \geq \frac{\bar{\kappa}_a(t_a, t_a')}{2\epsilon} \to \infty$ as $\epsilon \to 0$. $\square$

Proof of Lemma 3: A lower bound on the value of information that applies to all mechanism design problems on (T, p) is obtained by considering only the decision to participate.

Let $\gamma_a(\cdot|t_a)$ be a FE lottery under no information acquisition. We know that

$$\begin{aligned} \sum_{t_{-a}} p(t_{-a}|t_a)\gamma_a(t_{-a}|t_a) &= 0 \\ \Longrightarrow \quad E[p(\cdot|t_a, R_a) \cdot \gamma_a(\cdot|t_a)] &= \sum_{k=1}^{K} p(r_k|t_a) \sum_{t_{-a}} p(t_{-a}|t_a, r_k)\gamma_a(t_{-a}|t_a) = 0. \end{aligned}$$

The proof of Lemma 1 implies that almost every information signal has a realization r_k such that $\sum_{t_{-a}} p(t_{-a}|t_a, r_k)\gamma_a(t_{-a}|t_a) \neq 0$. Thus, the above equations imply that there exist r_ℓ, $r_{\ell'}$ such that

$$\sum_{t_{-a}} p(t_{-a}|t_a, r_\ell)\gamma_a(t_{-a}|t_a) < 0 < \sum_{t_{-a}} p(t_{-a}|t_a, r_{\ell'})\gamma_a(t_{-a}|t_a).$$

Let $L^+ \subset \{r_1, r_2, \ldots, r_K\}$ be such that $\sum_{t_{-a}} p(t_{-a}|t_a, r_\ell)\gamma_a(t_{-a}|t_a) < 0$ for all $r_\ell \in L^+$. A lower bound on $\pi_a(t_a)$ is the expected surplus for (a, t_a) by participating in the mechanism if and only if he observes a signal $r_\ell \in L^+$. Thus,

$$\pi_a(t_a) \quad \geq \quad -\sum_{r_\ell \in L^+} p(r_\ell|t_a) \sum_{t_{-a}} p(t_{-a}|t_a, r_\ell)\gamma_a(t_{-a}|t_a).$$

As $\epsilon \to 0$, $||\gamma_a(\cdot|t_a|| \to \infty$ by Lemma 2. Consequently, the right hand side of the above inequality becomes unboundedly large and $\pi_a(t_a) \to \infty$. $\square$

Proof of Lemma 4: Suppose there exists r_k such that posterior beliefs after observing r_k are a linear combination of posterior beliefs after observing r_ℓ, $\ell \neq k$. That is, there exists $\eta \in \Re^{K-1}$ such that

$$\begin{aligned}
p(t_{-a}|t_a, r_k) &= \sum_{\ell \neq k} \eta_\ell \, p(t_{-a}|t_a, r_\ell), \qquad \forall t_{-a} \\
\Longrightarrow \quad \frac{q(r_k|t_{-a}, t_a)p(t_{-a}|t_a)}{p(r_k|t_a)} &= \sum_{\ell \neq k} \eta_\ell \frac{q(r_\ell|t_{-a}, t_a)p(t_{-a}|t_a)}{p(r_\ell|t_a)}, \qquad \forall t_{-a} \\
\Longrightarrow \quad q(r_k|t_{-a}, t_a) &= \sum_{\ell \neq k} \eta_\ell \frac{p(r_k|t_a)}{p(r_\ell|t_a)} \, q(r_\ell|t_{-a}, t_a), \qquad \forall t_{-a} \\
&= \sum_{\ell \neq k} \widehat{\eta}_\ell \, q(r_\ell|t_{-a}, t_a), \qquad \forall t_{-a}
\end{aligned}$$

where $\widehat{\eta}_\ell = \eta_\ell \frac{p(r_k|t_a)}{p(r_\ell|t_a)}$. Therefore, $q(r_k|\cdot, t_a)$ is a linear combination of $q(r_\ell|\cdot, t_a)$, $\ell \neq k$. Hence, if there are fewer than N_a linearly independent posteriors then $Q_a(t_a)$ does not span T^{-a}.

Suppose that $q(r_k|\cdot, t_a)$ is a linear combination of the likelihoods of the other signals. That is, there exists a vector $\zeta \neq 0$ such that $q(r_k|t_{-a}, t_a) = \sum_{\ell \neq k} \zeta_\ell q(r_\ell|t_{-a}, t_a)$, for all $t_{-a} \in T^{-a}$. Then

$$\begin{aligned}
p(t_{-a}|t_a, r_k) &= \frac{q(r_k|t_{-a}, t_a)p(t_{-a}|t_a)}{\sum_{t'_{-a} \in T^{-a}} q(r_k|t'_{-a}, t_a)p(t'_{-a}|t_a)} \\
&= \frac{\sum_{\ell \neq k} \zeta_\ell q(r_\ell|t_{-a}, t_a)p(t_{-a}|t_a)}{\sum_{t'_{-a} \in T^{-a}} \sum_{\ell \neq k} \zeta_\ell q(r_\ell|t'_{-a}, t_a)p(t'_{-a}|t_a)} \\
&= \sum_{\ell \neq k} \zeta_\ell \frac{\sum_{t'_{-a} \in T^{-a}} q(r_\ell|t'_{-a}, t_a)p(t'_{-a}|t_a)}{\sum_{t'_{-a} \in T^{-a}} \sum_{\ell \neq k} \zeta_\ell q(r_\ell|t'_{-a}, t_a)p(t'_{-a}|t_a)} \; \frac{q(r_\ell|t_{-a}, t_a)p(t_{-a}|t_a)}{\sum_{t'_{-a} \in T^{-a}} q(r_\ell|t'_{-a}, t_a)p(t'_{-a}|t_a)} \\
&= \sum_{\ell \neq k} \overline{\zeta}_\ell p(t_{-a}|t_a, r_\ell)
\end{aligned}$$

where $\overline{\zeta}_\ell \equiv \zeta_\ell \frac{\sum_{t'_{-a} \in T^{-a}} q(r_\ell|t'_{-a}, t_a) p(t'_{-a}|t_a)}{\sum_{t'_{-a} \in T^{-a}} \sum_{\ell \neq k} \zeta_\ell q(r_\ell|t'_{-a}, t_a) p(t'_{-a}|t_a)}$. Therefore, if $Q_a(t_a)$ does not span T^{-a} then fewer than N_a posteriors are linearly independent. $\square$

Proof of Lemma 5: The partially robust lottery $\gamma_a(\cdot|t_a)$ may be chosen such that for all $t_a \neq t'_a$ and for all r_k,

$$p(\cdot|t_a, r_k) \cdot \left[u_a(f(t'_a, \cdot), (t_a, \cdot)) - u_a(f(t'_a, \cdot), (t'_a, \cdot)) - \gamma_a(\cdot|t'_a)\right] \ < \ 0.$$

Because $p(\cdot|t_a, r_k) \cdot \gamma_a^\Delta(\cdot|t'_a) = p(\cdot|t_a, r_k) \cdot \gamma_a(\cdot|t'_a) - \Delta$

$$p(\cdot|t_a, r_k) \cdot \left[u_a(f(t'_a, \cdot), (t_a, \cdot)) - u_a(f(t'_a, \cdot), (t'_a, \cdot)) - \gamma_a^\Delta(\cdot|t'_a)\right] \ < \ \Delta \ = \ -p(\cdot|t_a) \cdot \gamma_a^\Delta(\cdot|t_a).$$

Thus, agent (a, t_a) does not have an incentive to acquire information in order to misreport his type (whether $x_a(t_a, \cdot)$ is used or $x_a^\Delta(t_a, \cdot)$).

Next, consider the decision to participate. Noting that

$$-p(\cdot|t_a, r_k) \cdot \gamma_a^\Delta(\cdot|t_a) \ = \ -p(\cdot|t_a, r_k) \cdot \gamma_a(\cdot|t_a) + \Delta,$$

whenever agent (a, t_a) decides not to participate under $\gamma_a^\Delta(\cdot|t_a)$, he would not participate and avoid greater payments under $\gamma_a(\cdot|t_a)$. Thus, $\pi_a^\Delta(t_a) < \pi_a(t_a)$.

Suppose that $\Delta \geq -p(\cdot|t_a, r_k) \cdot \gamma_a(\cdot|t_a)$ for all r_k, t_a. Hence $p(\cdot|t_a, r_k) \cdot \gamma_a^\Delta(\cdot|t_a) \leq 0$. Even if he acquires information agent a will participate and truthfully report his type regardless of the realization of $R_a(t_a)$. Therefore $\pi_a^\Delta(t_a) = 0$. $\square$

Proof of Lemma 6: Suppose that t'_a has an incentive to report t_a at full extraction. Let $\bar{\kappa}_a(t_a, t'_a)$ be as in Lemma 2. To dissuade misreporting by agent (a, t'_a) it is necessary that

$$\bar{\kappa}_a(t_a, t'_a) - p(\cdot|t'_a) \cdot \gamma_a^\Delta(\cdot|t_a) \leq \Delta = -p(\cdot|t_a) \cdot \gamma_a^\Delta(\cdot|t_a).$$

This implies $[p(\cdot|t'_a) - p(\cdot|t_a)] \cdot \gamma_a^\Delta(\cdot|t_a) \geq \bar{\kappa}_a(t_a, t'_a)$. The rest of the proof is identical to that of Lemma 2. $\square$

Proof of Lemma 7: A zero surplus mechanism that provides incentives for information acquisition by agent a is supported by lotteries $[\gamma_a(\cdot|t_a), \gamma_a(\cdot|t_a, r_k)]$, for all t_a, r_k, where $\gamma_a(\cdot|t_a)$ is robust and $\gamma_a(\cdot|t_a, r_k)$ satisfy (16) and (17). Clearly existence of robust lotteries is necessary for the existence of such a mechanism.

To prove sufficiency, suppose that $\gamma_a(\cdot|t_a, r_k)$ that satisfying (16) and (17) do not exist. Hence, there exists $(t'_a, r'_k) \neq (t'_a, \emptyset)$ such that $p(\cdot|t'_a, r'_k)$ is a convex combination of $p(\cdot|t_a, r_k)$, $\forall (t_a, r_k) \neq (t'_a, r'_k)$, $r_k \neq \emptyset$. Therefore, a linear combination of beliefs $p(\cdot|t'_a, r_k)$, $k = 1, 2, \ldots, K$ is a convex combination of beliefs $p(\cdot|t_a, r_k)$, $t_a \neq t'_a$, $k = 1, 2, \ldots, K$. Proposition 4 implies that robust lotteries do not exist. $\square$

References

[1] Bergemann, D. and S. Morris (2008). "Ex post implementation," *Games and Economic Behavior*, 63:527-566.

[2] Bergemann, D. and M. Pesendorfer (2003). "Information structures in optimal auctions," *Journal of Economic Theory*, 137:580-609.

[3] Bergemann, D. and J. Valimaki (2002). "Information acquisition and efficient mechanism design," *Econometrica*, 70:1007-1033.

[4] Brusco, S. (1998). "Unique implementation of the full surplus extraction outcome in auctions with correlated types," *Journal of Economic Theory*, 80:185-200.

[5] Cai, H., (2009). "Costly participation and heterogenous preferences in informational committeees," *Rand Journal of Economics*, 40:173-189.

[6] Compte, O. and P. Jehiel (2007). "Auctions and information acquisition: Sealed-bid or dynamic formats?" *Rand Journal of Economics*, 38:355-372.

[7] Cremer, J. and F. Khalil (1992). "Gathering information before signing a contract," *American Economic Review*, 82:566-578.

[8] Cremer, J., F. Khalil, and J.-C. Rochet (1998). "Strategic information gathering before a contract is offered," *Journal of Economic Theory*, 81:163-200.

[9] Cremer, J. and R. McLean (1988). "Full extraction of surplus in Bayesian, dominant strategy auctions," *Econometrica*, 56:1247-1258.

[10] Cremer, J., Y. Spiegel, and C. Zheng (2009). "Auctions with costly information acquisition," *Economic Theory*, 38:41-72.

[11] Gerardi, D. and L. Yariv (2008). "Information acquisition in committees," *Games and Economic Behavior*, 62:436-459.

[12] Gershkov, A. and B. Szentes (2009). "Optimal voting schemes with costly information acquisition," *Journal of Economic Theory*, 144:36-68.

[13] Hagedorn, M. (2009). "The value of information for auctioneers," forthcoming *Journal of Economic Theory*.

[14] Heifetz, A. and Z. Neeman (2006). "On the generic (im)possibility of full surplus extraction in mechanism design," *Econometrica*, 74:213-233.

[15] Jackson, M. (2003). "Efficiency and information aggregation in auctions with costly information," *Review of Economic Design*, 8:121-141.

[16] Lewis, T. and D. Sappington (1997). "Information management in incentive problems," *Journal of Political Economy*, 105:796-821.

[17] Martinelli, C. (2005). "Would rational voters acquire costly information?" *Journal of Economic Theory*, 129:225-251.

[18] Matthews, S. (1984). "Information acquisition in competitve bidding process," in M. Boyer and R. Kihlstrom (eds.) *Bayesian Models in Economic Theory*, pp. 181-207. Amsterdam:North-Holland.

[19] Milgrom, P. (1981). "Rational expectations, information acquisition, and competitive bidding," *Econometrica*, 49:921-943.

[20] McAfee, R.P. and P. Reny (1982). "Correlated information mechanism design," *Econometrica*, 60:395-421.

[21] Neeman, Z., (2004). "The relevance of private information in mechanism design," *Journal of Economic Theory*, 117:55-77.

[22] Obara, I., (2008). "The full extraction theorem with hidden actions," *Advances in Theoretical Economics*, 8: No. 1, Article 8, http://www.bepress.com/bejte/vol8/iss1/art8/.

[23] Persico, N. (2000). "Information acquisiton in auctions," *Econometrica*, 68:135-148.

[24] Persico, N. (2004). "Committee design with endogenous information," *Review of Economic Studies*, 71:165-191.

[25] Parreiras, S.O., (2005). "Correlated information, mechanism design and informational rents," *Journal of Economic Theory*, 123:210–217.

[26] Robert, J., (1991). "Continuity in auction design," *Journal of Economic Theory*, 55:169-179.

[27] Tan, G. (1992). "Entry and R&D in procurement contracting," *Journal of Economic Theory*, 58:41-60.