Surplus Extraction with Behavioral Types*
PRELIMINARY AND INCOMPLETE

Nicolas Pastrian†

This version: June 21, 2021
For the most recent version click here

Abstract

We reexamine the surplus extraction problem in a mechanism design setting with behavioral types. We focus on an extreme class of behavioral types which always perfectly reveal their private information. However, incentive compatibility constraints remain necessary for strategic types with respect to the contract of all types. We characterize the sufficient conditions that guarantee full extraction in the reduced form environment of McAfee and Reny (1992). The standard convex independence condition identified in Crémer and McLean (1988) remains necessary only among the beliefs of strategic types, while a weaker condition is required for the beliefs of behavioral types.

1 Introduction

In mechanism design settings, private information leads agents to retain informational rents if information is independently distributed. However, we know since Myerson (1981) that under correlation it is possible to extract all the informational rents from agents. Hence, in presence of correlation, private information not necessarily lead to agents obtaining information rents. This result is what is usually called *full (surplus) extraction*. Crémer and McLean

*I thank Luca Rigotti and Richard Van Weelden for providing guidance in this project. I also thank Svetlana Kosterina, Benjamin Matta, participants of the Microeconomic Theory Brownbag at University of Pittsburgh, the 2021 Pennsylvania Economic Theory Conference and the 32nd Stony Brook International Conference on Game Theory for useful comments.

†Department of Economics, University of Pittsburgh. Email: nip59@pitt.edu

(1988) have identified the key independence condition that guarantees full extraction (convex independence) in mechanism design settings.

In this project we examine the full surplus extraction problem considering the presence of *behavioral types*. That is, types that not react optimally to incentives. We focus on a particular extreme case of such behavioral types: types which doesn't react to the mechanisms implemented and always reveal their private information perfectly. While extreme, this assumption allows for a simple characterization and also works as a benchmark for the designer problems since it gives him the most advantage setting to operate. The key features of this assumption is that it allows for a perfect identification of each behavioral type, and it allows us to characterize their behavior regardless of the mechanism implemented. While extracting rents from this types should be easier, incentive compatibility still requires their contract to be non-attractive to strategic types, imposing constraints to the designer.

We consider a reduced form environment similar to the one used by McAfee and Reny (1992) and Lopomo et al. (2020), where in an unmodeled stage an agent is left with informational rents which depends on his private information. There is an exogenous source of uncertainty in which the current stage contracts could condition on. We will refer to the current stage contracts simply as contracts and usually ignore any reference to the mechanism that generates the informational rents. The main differences with respect to previous works are that we restrict to finite types and that we introduce a class of behavioral types which doesn't react optimally to incentives.

The next section describe the model and the main result. A short review of related literature is provided in the final section.

2 Model

There is a finite set of types T and a finite set of states Ω. Each type t is associated with a valuation (i.e., informational rents) $v_t \in \mathbb{R}_+$ and beliefs $p_t \in \Delta(\Omega)$. We define a contract c as a mapping from states into transfers, such that $c(\omega) \in \mathbb{R}$ is the transfer required in state $\omega \in \Omega$. A contract menu $\mathbf{c}$ is a collection $\{c_t : t \in T\}$ such that $c_t : \Omega \to \mathbb{R}$, i.e., a collection of contracts.

There is a single agent with quasilinear preferences. Hence, from a contract c, type t's payoff is given by

$$v_t - \langle p_t, c \rangle$$

where $\langle p_t, c \rangle$ denotes the expected value of c under p_t, that is

$$\langle p_t, c \rangle = \sum_{\omega \in \Omega} p_t(\omega)\, c(\omega).$$

We are interested on whether the principal or designer is able to extract all the remaining rents from the agent using a contract menu $\mathbf{c}$.

We introduce the definition of full extraction formally below.

Definition 1. *A contract menu* $\mathbf{c}$ *achieves full extraction if for all* $t \in T$

$$\langle p_t, c_t \rangle = v_t$$

In the traditional setting, it is required that all types prefer their own contract to the contract offered to others.

We depart from the standard model by introducing a particular class of *behavioral* types. In particular, a behavioral type here is a type for which no incentive compatibility constraint is considered. That is, a behavioral type always report his type truthfully. While this assumption could seem extreme, we claim that this is without loss due to a revelation principle argument as long as the strategy of the behavioral types is independent of the mechanism implemented.

Let $B \subseteq T$ be the set of behavioral or unsophisticated types. Similarly, let $S = T \backslash B$ be the set of *standard, sophisticated or strategic* types.

Since in our model behavioral types don't respond to incentives, we will require incentive compatible constraints only for strategic types. Moreover, we will use a restrictive incentive compatibility notion, requiring that potential deviations have no impact in the informational rents of the agent. That is, we will require that each type chooses his cost minimizing contract.

Definition 2. *A contract menu* $\mathbf{c}$ *is incentive compatible if for each strategic type* $s \in S$,

$$c_s \in \arg\min_{t \in T} \langle p_s, c_t \rangle$$

We will be looking for a incentive compatible contract menu that fully extract the informational rents from the agent.

Definition 3. *Full extraction with behavioral types is feasible if there exists a incentive compatible contract menu* $\mathbf{c}$ *which achieves full extraction*

Crémer and McLean (1988) have shown that in a setting without behavioral types full extraction is feasible if the set of beliefs satisfies the independence condition below.

Definition 4. *A set of beliefs* P *satisfies the CM condition if for any* $p \in P$, $p \notin co(P \backslash \{p\})$

This is condition is known as the convex independence condition, and it is a linear independence condition over the set of beliefs. It also coincides with the more general condition of probabilistic independence used by McAfee and Reny (1992) and Lopomo et al. (2020) applied to a setting with finite types as ours.

For any subset of types $X \subseteq T$, P^X denote the set of beliefs associated to types on X.

We state our main result below.

Theorem 1. *Full extraction with behavioral types is feasible if*

(i) P^S *satisfies the CM condition, and*

(ii) For all types $b \in B$, $p_b \notin co(P^S)$

Proof. Step 1: from (i), we now from Crémer and McLean (1988) that we can find a contract that fully extract if we restrict types to S. Notice that such contract remains incentive compatible and reaches full extraction among types in S as long as the contracts offered to types in B doesn't generate incentives to any type in S to deviate.

Step 2: now, we construct the contract for types in B. Consider a single behavioral type b. For this type, it suffices to find z_b such that

$$\langle p_b, z_b \rangle = 0$$

$$\langle p_s, z_b \rangle > 0, \quad \forall s \in S$$

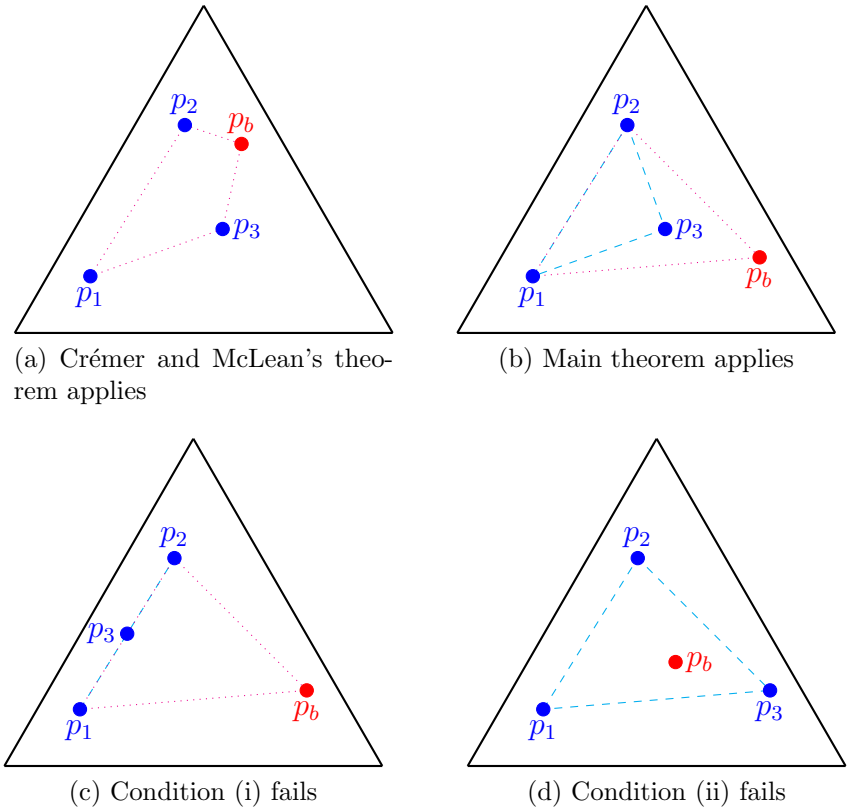
Due to condition (ii), the solution to this system of inequalities is guaranteed by a separation argument (see for example Farkas' lemma). Hence, we can use z_b to construct the extracting contract for the behavioral type in a similar way we construct the contract for types in S. In particular,

$$c_b = v_b + \alpha_b z_b$$

with $\alpha_b = \max_{s \in S} \frac{v_s - v_b}{\langle p_s, z_b \rangle}$ satisfies all the required conditions.

Note that the contract designed for type b has no impact on the contract of any other type in B or S. Hence, we can repeat the process for all others types in B to construct the contract for each remaining type.

□



(a) Crémer and McLean's theorem applies

(b) Main theorem applies

(c) Condition (i) fails

(d) Condition (ii) fails

Figure 1: Representation of an environment with three states, three strategic types (1,2, and 3), and a single behavioral type (b). In (a), P^T satisfies the CM condition so full extraction is feasible. In (b), only P^S satisfies the CM condition; since $p_b \notin co(P^S)$, full extraction remains feasible. In (c), P^S no longer satisfies the CM condition (condition (i) in the main theorem fails), so full extraction cannot be guaranteed. Finally, in (d) P^S satisfies the CM condition but $p_b \in co(P^S)$ (condition (ii) fails), hence full extraction cannot be be guaranteed.

Notice that from the proof above, it is without loss to look at environments where $|B| = 1$ since the problem for each behavioral type could be looked at in complete isolation from other behavioral types. This is possible since there is no "cross" incentive compatibility conditions.

So, in the presence of the behavioral types studied here, a slightly relaxed convex independence conditions is required to guarantee full extraction. This highlight in particular that incentive compatibility and full extraction are not entirely isolated features of a mechanism. Note that the second condition on the theorem above allows us to separate the behavioral type from strategic types. This translates into a particular direction of incentive compatibility: that no strategic type wants to deviate to the contract of the behavioral type

considered. This is in contrast to the first condition which requires, as in the standard extraction problem, to consider incentive compatibility in both directions. This is why the augmented condition required for behavioral types is weaker than the one required over strategic types (i.e., the standard CM condition).

As was argued before, having P^S satisfying the CM condition is a necessary condition for full extraction even with behavioral types. Hence, considering environments where this condition is violated will automatically rule out the possibility of fully extracting.

Clearly, if the CM condition is satisfied not only over P^S but over the whole set of beliefs P^T, then fully extraction holds trivially since the same proof from the original work from Crémer and McLean (1988) works in this context. If such condition only holds over P^S and in addition $p_b \notin co(P^S)$, then full extraction is still feasible. Hence, the conditions identified in Theorem 1 relaxes the condition required for full extraction without behavioral types. However, if $p_b \in co(P^S)$ full extraction fails again even if the only types violating the independence condition are behavioral and do not need to satisfy any incentive compatibility constraints themselves. The reason why full extraction fails here is that even if these types don't need to satisfy incentive compatibility constraint, other types could be attracted by their contracts, limiting the power of the designer to extract rents from them.

3 A numerical example

Note that Theorem 1 only establishes sufficient conditions for full extraction with behavioral types. A follow up question is whether such conditions are also necessary for full extraction. The example below shows that it is not the case and even if such conditions fail to be satisfied a fully extracting mechanism could indeed exist. However, it also shows that in order to obtain such contract we need to impose additional restrictions on the informational rents structure, which are absent in Theorem 1.

Consider the following numeric example based on Myerson (1981). There are two (strategic) types with valuations $v_1 = 10$, and $v_2 = 100$, and two states ω_1 and ω_2. Beliefs and valuations are correlated,

$$p_1 = (2/3, 1/3)$$

$$p_2 = (1/3, 2/3)$$

Notice that the CM condition holds here since $p_1 \neq p_2$, so full extraction is possible (as Myerson (1981) have shown).

For example, the contract menu $c_1 = (-80, 190)$, $c_2 = (100, 100)$ is incentive compatible and achieves full extraction.

In general, any contract such that

$$c_1(\omega_1) \leq -80 \leq c_2(\omega_2),$$

$$c_1(\omega_2) = \frac{10 - (2/3)c_1(\omega_1)}{1/3}$$

$$c_2(\omega_2) = \frac{100 - (1/3)c_2(\omega_1)}{2/3}$$

would work in this example.

Let's introduce our behavioral type in this example. That is a type b with valuation v_b and beliefs p_b. Notice that by Theorem 1, if $p_b \notin co\{p_1, p_2\}$ then fully extraction is possible. In particular, this requires either $p_b(\omega_1) < 1/3$ or $p_b(\omega_1) > 2/3$.

Suppose this is not the case, so $p_b(\omega_1) \in [1/3, 2/3]$. Hence, there exists $\lambda \in [0, 1]$ such that $p_b = \lambda\frac{1}{3} + (1 - \lambda)\frac{2}{3}$. Clearly, the second condition on Theorem 1 fails so its result doesn't apply. Would full extraction be still possible here?

First, let's consider two cases where is easy to see that full extraction remain feasible.

1. $v_b = \lambda 10 + (1 - \lambda)100$, that is b's valuation follow exactly the same convex combination over the valuations of the strategic types as his beliefs over their beliefs. Due to the linearity of the payoffs, then the convex combination of the contracts c_1 and c_2 will work for b and guarantee incentive compatibility as well (for types 1 and 2 this new contract is a randomization between a contract with net payoff zero and a contract with negative net payoff, while for b it offers expected net payoff equal to 0). Notice that this would hold even if b is not behavioral.

2. $v_b \geq 100$, then a deterministic contract $c_b = v_b$ doesn't mess up incentives since no strategic type would prefer such contract since their payoff will always be non-positive under such contract. This contract would be feasible only if b is behavioral, otherwise b would prefer one of the contracts offered to the other types.

Ruling out the cases above, a fully extracting contract exists only if

$$(1/3)\left(\frac{\lambda 10 + (1 - \lambda)100 - v_b}{1 - \lambda}\right) \leq (2/3)\left(\frac{v_b - (\lambda 10 + (1 - \lambda)100)}{\lambda}\right)$$

From which is clear that

$$v_b \geq \lambda 10 + (1-\lambda)100$$

is a necessary and sufficient condition for such contract to exists[1].

3.1 A numerical example where Theorem 1 works

If we change the beliefs of the behavioral type above, and consider $p_b = 1/4$ instead, then an optimal contract could take the form

$$\begin{aligned} c_1 &= (-260, 550) \\ c_2 &= (100, 100) \\ c_b &= (900 - 8v_b, 4v_b - 300) \end{aligned}$$

and c_b will be necessarily stochastic as long as $v_b < 100$.

For example, if $v_b = 50$ then

$$c_b = (500, -100)$$

Note that this contract is not unique and we can construct several alternative contracts which also implements full extraction. All of them generate the same expected revenue but promise different expected values for deviators.

4 Further results

In this section we present two more results in this setting.

The first one is a direct consequence of Theorem 1, and shows that the same contract works for behavioral types even if full extraction could not be achieved for strategic types.

Corollary 1. *If $p_b \in co(P^S)$, then a contract which extracts all rents of type b exists.*

Proof. Follows directly from the proof of Theorem 1. □

[1]The general condition for a fully extracting contract to exist in a binary value, binary state, single behavioral type environment for any v_1, v_2, p_1, p_2 with $p_b \in co(p_1, p_2)$ could be written as

$$v_b \geq \left(\frac{p_b - p_1}{p_1 - p_2}\right) v_1 + \left(\frac{p_2 - p_b}{p_1 - p_2}\right) v_2$$

The second result extends full extraction when the second condition of Theorem 1 fails beyond the example on the previous section.

Proposition 1. *Suppose P^S satisfies the CM condition. Let $\hat{B} = \{b \in B : p_b \in co(P^S)\}$. Then, full extraction with behavioral types is feasible if for each $b \in \hat{B}$*

$$v_b \geq \sum_{s \in S} \lambda^b(s) v_s,$$

where $\lambda^b \in \Delta(S) : p_b = \sum_{s \in S} \lambda^b(s) p_s$.

Proof. The contracts for types $s \in S$ and $b \in B \backslash \hat{B}$ remain the same as in Theorem 1.

For types $b \in \hat{B}$, there are two cases to consider:

1. If $v_b \geq \max_{s \in S} v_s$ then a flat constant contract $c_b = v_b$ doesn't violate any incentive compatibility condition and extract all rents from type b.

2. For $v_b < \max_{s \in S} v_s$ we construct the contract for b as follows.

   First, $p_b \in co(P^S)$ implies there exist weights $(\lambda_s)_{s \in S}$ such that $\sum_{s \in S} \lambda_s = 1$, $\lambda_s \geq 0$ for all $s \in S$ and $p_b = \sum_{s \in S} \lambda_s p_s$.

   Let $v_{max} = \max_{s \in S} v_s$ and $\overline{v}_b = \sum_{s \in S} \lambda_s v_s$. Now, consider the contract

   $$c_b = \alpha_b \sum_{s \in S} \lambda_s c_s + (1 - \alpha_b) v_m ax$$

   with $\alpha_b = \frac{v_b - \overline{v}_b}{v_m ax - \overline{v}_b}$. Note $\langle p_b, c_b \rangle = v_b$ and $\langle p_s, c_b \rangle \geq v_s$ for all $s \in S$. Hence c_b fully extract the rents from b and doesn't violate any incentive compatibility constraint.

Repeating this process for all other $b' \in \hat{B}$ we obtain a feasible contract menu which achieves full extraction. □

5 Related literature

Myerson (1981) characterize the optimal auction assuming bidder's valuation are continuous and independently distributed, but also construct an example with discrete valuations and correlation where full extraction is feasible.

Crémer and McLean (1988) characterize the full extraction problem in an auction environment with correlation and discrete types. They originally

identify the convex independence condition, and shows that such condition is key to obtain full extraction.

The studies above restrict the space of types to the set of payoff types or small type spaces. More general type spaces are considered by Farinha Luz (2013) which studies the surplus extraction problem in rich type spaces. The author characterizes an upper bound to the revenues studying a relaxed problem and shows that under a linear independence condition this upper bound is achieved by the optimal mechanism in the complete problem.

A different approach is taken by McAfee and Reny (1992) which studies the surplus extraction problem with both discrete and continuous types using a reduced form approach. Recently, Lopomo et al. (2020) revisits both the continuous and discrete problems studied in McAfee and Reny (1992) providing an alternative proof of the original results and considering infinite menus instead of finite menus as in the original work. In this paper we also use a reduced form approach but restrict to finite type spaces.

Fu et al. (2021) shows that the full extraction results holds even if the correlated distribution is unknown but the designer has access to samples from the true distribution. They show that the solution involve using the samples as a correlating device instead of a learning process.

Krähmer (2020) studies the joint problem of designing the information structure and the mechanism, and shows that extracting the full surplus is possible (under partial information control) if beliefs satisfy a convex independence condition.

References

Crémer, Jacques and Richard P. McLean (1988) "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56 (6), 1247–1257, http://www.jstor.org/stable/1913096. [1], [3], [4], [6], [9]

Farinha Luz, Vitor (2013) "Surplus extraction with rich type spaces," *Journal of Economic Theory*, 148 (6), 2749–2762, https://doi.org/10.1016/j.jet.2013.07.016. [10]

Fu, Hu, Nima Haghpanah, Jason Hartline, and Robert Kleinberg (2021) "Full surplus extraction from samples," *Journal of Economic Theory*, 193, 105230, https://doi.org/10.1016/j.jet.2021.105230. [10]

Krähmer, Daniel (2020) "Information disclosure and full surplus extraction in mechanism design," *Journal of Economic Theory*, 187, 105020, https://doi.org/10.1016/j.jet.2020.105020. [10]

Lopomo, Giuseppe, Luca Rigotti, and Chris Shannon (2020) "Detectability, Duality, and Surplus Extraction." [2], [4], [10]

McAfee, R. Preston and Philip J. Reny (1992) "Correlated Information and Mechanism Design," *Econometrica*, 60 (2), 395–421, http://www.jstor.org/stable/2951601. [1], [2], [4], [10]

Myerson, Roger B. (1981) "Optimal Auction Design," *Mathematics of Operations Research*, 6 (1), 58–73, http://www.jstor.org/stable/3689266. [1], [6], [9]