# Full Surplus Extraction and Consideration Sets[*]

Nicolas Pastrian[†]

November 28, 2024

**Abstract**

We analyze the surplus extraction problem in a mechanism design setting with consideration sets. We study a bounded rationality version of a general mechanism design environment with correlation in which the agent evaluates only a subset of types as possible deviations. We call these subsets the agent's consideration sets. We identify the inverse consideration sets as the key elements that determine whether full extraction is feasible in this setting and characterize the conditions beliefs need to satisfy to guarantee full surplus extraction. These conditions require the beliefs of each type to be separated from the beliefs of types in his inverse consideration set only. This relaxes the independence condition in Crémer and McLean (1988), which remains sufficient in our setting. Finally, we discuss some applications and limitations of our model.

JEL codes: D82, D83, D40

Keywords: *mechanism design, surplus extraction, consideration sets, bounded rationality*

[*] I thank Luca Rigotti and Richard Van Weelden for providing guidance in this project. I also thank Nicolas Figueroa, Svetlana Kosterina, Alexey Kushnir, Benjamin Matta, Tymofiy Mylovanov, and participants of the Microeconomic Theory Brownbag at University of Pittsburgh, the 2021 and 2023 Pennsylvania Economic Theory Conference, the 32nd and 33rd Stony Brook International Conferences on Game Theory, the 3rd Economics Alumni Workshop at Pontificia Universidad Católica de Chile, and the 2023 Annual Meeting of the Economic Society of Chile (SECHI 2023) for useful comments on the current and previous versions of the paper. This paper supersedes *Surplus Extraction with Behavioral Types*.

[†] Department of Economics, Social Studies, Applied Mathematics and Statistics (ESOMAS), University of Turin. Email: nicolasraul.pastriancontador@unito.it

# 1 Introduction

We study the surplus extraction problem in a setting with correlation and bounded rationality in the form of partial consideration: the agent only considers a subset of types as a potential deviation. We identify these types as this type's *consideration set* and introduce them in a general environment. We show that *inverse consideration sets* are the key elements that determine whether full extraction is feasible in this setting and characterize the conditions beliefs need to satisfy to achieve full extraction for any payoff structure. These conditions require the beliefs of each type to be separated from the beliefs of types in his inverse consideration set only. This relaxes the independence condition in Crémer and McLean (1988), which remains sufficient in our setting. This allows us to obtain a positive result in more general settings.

We define the inverse consideration set for a type $t$ as the set of types which consider $t$ as a possible deviation. Inverse consideration sets are key in our characterization because building an incentive compatible mechanism that extracts all the surplus from type $t$ requires that any other type $t'$ prefers his own contract over the contract for type $t$, and this must hold regardless of what other contracts type $t$ would consider. This property is not identified in models with full-consideration due to the symmetry imposed in such models: under full-consideration consideration sets and inverse consideration sets coincide. Hence, the distinction between them is inconsequential. Our framework shows that this distinction is indeed crucial once we break down this symmetry. We consider a discrete version of the reduced model used by McAfee and Reny (1992) and more recently by Lopomo et al. (2022), where in an unmodeled stage an initial allocation determines the surplus an agent generates based on his private information or type. There is also an exogenous source of uncertainty correlated with the private information which contracts implemented in a following stage could condition on. The analysis focuses on the design of the contracts in this second stage. This modeling approach allows to apply our results to a variety of applications without limiting the characteristics of the environment considered.

As in McAfee and Reny (1992), we study whether full surplus extraction can be guaranteed and characterize the conditions that make this possible. Myerson (1981) was the first to notice that in a mechanism design problem it is possible to extract all the surplus from the agents if their private information is correlated. Later, Crémer and McLean (1988) identified the key condition that guarantees that the full surplus could be extracted from agents in expectation in standard mechanism design settings. More recently, Farinha

Luz (2013), Krähmer (2020), Fu et al. (2021), and Lopomo et al. (2022) have shown that this convex independence condition remains key to guarantee full surplus extraction in more general environments. The same condition remains sufficient in our model but could be relaxed in a way made precise below.

Different notions of bounded-rationality in mechanism design settings have been considered before in the context of implementation by Eliaz (2002), de Clippel (2014), and Clippel et al. (2018). There are also studies looking at more specific settings and behavioral assumptions. For example, Severinov and Deneckere (2006) and Saran (2011) study problems in which a fraction of agents are honest and always reveal their information truthfully. In both, there is no correlated information and no general behavioral assumptions are explored. We consider the surplus extraction problem in a correlated information setting and allow for more general behavioral assumptions on boundedly rational agents, considering the case of honest types as a special case.

Our work is not the first to study consideration sets in a theoretical economic setting: Eliaz and Spiegler (2011), Manzini and Mariotti (2014), Caplin et al. (2018), Fershtman and Pavan (2022), among others, have studied environments in which full-consideration fails and agents only consider a subset of alternatives.[1] Most of them focus on the analysis of the agent's decision problem without discussing the design problem embedded in such environments. An exception is Eliaz and Spiegler (2011) which studies a competition model between two firms choosing both which product to offer and what marketing strategy to use. In their model, buyers have an exogenous consideration function which determines whether they would observe the product offered by each firm or only one of them. They found that in equilibrium firms cannot do better than if buyers always observe the offers of both firms. Our paper also studies a model in which agents have consideration sets from the perspective of the designer but differs from Eliaz and Spiegler (2011) by considering a general mechanism design setting with correlation and heterogeneity.

In our model consideration sets are exogenous and have mainly a behavioral or bounded rationality motivation: they arise from limitations in the capacity of agents to evaluate all available alternatives. However, our model is also compatible with the interpretation that there are restrictions on the reports that the agent could communicate: there is a technological limitation (e.g., evidence or certification that need to be presented by the agent) that impose restrictions on the types an agent could convincingly "misreport". Search

[1] For an econometric perspective on search and consideration sets, see Honka et al. (2019).

frictions and inattention could also give rise to consideration sets as pointed out by Fershtman and Pavan (2022) and Caplin et al. (2018) respectively. The frictions studied in those models give rise to random consideration sets in which the actual "choices" available to each type are not deterministic. Instead, in our approach consideration sets are deterministic and each type has a clearly defined set of types he could imitate. Nevertheless, random consideration sets could be partially accommodated in our framework by enlarging the type space to include different types with the same surplus and beliefs but different consideration sets. As long as the inverse consideration sets in this extended environment respect the structure required in our main theorem, surplus extraction could still be feasible. Hence, while our principal motivation for consideration sets comes from a bounded rationality perspective, alternative non-behavioral interpretations are also compatible with our model.

Due to the modeling approach used for consideration sets, our work is also related to the literature on mechanism design with partially verifiable information initiated by Green and Laffont (1986). Green and Laffont (1986) studies an implementation problem in which the agent's message space depends on their true characteristics. They focus on direct mechanisms and characterize necessary and sufficient conditions to obtain a revelation principle in this environment. These conditions require different types' messages spaces to be nested. In a related setting, Bull and Watson (2007) introduces hard evidence to mechanism design problems. They show that under an equivalent nested condition the model of hard evidence and partially verifiable information are equivalent. While their general environment is similar to our implementation of consideration sets, in our model consideration sets of different types are not required to be nested. Therefore, the revelation principle in Green and Laffont (1986) fails in our environment, and there are settings in which an allocation could be implemented with non-truthful mechanisms but fails to be implementable with truthful mechanisms. This also implies that the equivalence between hard evidence and partially verifiable information also breaks down in our setting. Strausz (2017) proposes an alternative revelation principle by considering an extended environment in which the agent is required to send a message to verify his type as part of the allocation. Reuter (2022) uses this approach to study revenue maximization in auctions in which bidders have some partially verifiable private information. However, using this revelation principle is problematic in our model as asking the agent to verify his type with an untruthful type requires him to send an inconsistent report in our reduced form setting. We argue that the restriction to truthful direct mechanisms still provides interesting insights, even though the revelation principle fails and there is potential loss on restricting to this class of mechanisms. In particular, since we focus on a reduced form environment, the restriction to truthful direct mechanisms allows intervention to be minimal in the sense that the interactions that determine this reduced form is not altered. Using non-truthful mechanisms will involve the agent to report inconsistent types in both interactions, making our approach invalid as the initial types contained in each type consideration sets became meaningless. Moreover, since our main result focus on sufficient conditions for full surplus extraction, focusing on this restricted class makes our result stronger.[2]

The remainder of this paper is organized as follows. We describe the model and our main theorem in Section 2. In Section 3 we discuss some applications of our theorem, including two simpler environments in which the conditions for full surplus extraction are easier to interpret. Finally, Section 4 concludes.

## 2 Model

We consider the problem of a principal or designer interacting with a single agent.[3] The agent has a type $t$ in a finite set of types $T$[4]. Each type $t$ is associated with three elements: (1) the surplus type $t$ generates in the interaction with the principal $v_t \in \mathbb{R}_+$, (2) his belief $p_t$ over a finite set of exogenous states $\Omega$[5], and (3) a subset of types $C_t \subseteq T$ which identifies the types he could pretend to be or convincingly report. We will refer to $C_t$ as the consideration set of type $t$, and assume $t \in C_t$ for all $t \in T$.

The surplus $v_t$, beliefs $p_t$, and consideration set $C_t$ for each type $t$ are all assumed to be exogenous and fixed. As in McAfee and Reny (1992), we could think about $v_t$ and $p_t$ as coming from the interaction between the agent and the principal in a previous unmodeled mechanism. This allows us to consider a general mechanism design problem without explicitly defining the environment details. Similarly, our treatment of consideration sets $C_t$ follows the partially verifiable information in Green and Laffont (1986) and allows us

[2] Other reasons for considering only direct mechanisms is that they are closer to mechanisms implemented in practice as argued by Crawford (2021), and that the use of more general mechanisms could involve very complex implementations as argued by Ollár and Penta (2017, 2023).

[3] Alternatively, we could interpret this as the interaction between a designer and continuum of mass one of agents. We will use this interpretation in some of the applications below.

[4] Results could be extended to infinite types and virtual surplus extraction as in McAfee and Reny (1992).

[5] States could contain any information potentially correlated with the agent's information but which the agent has no control over. For example, it could contain the valuations of other bidders in a multi-bidder auction environment with correlated valuations.

to treat them also as a consequence of the previous interaction between the agent and the principal.[6]

We now proceed to formally define contracts and mechanisms in this setting.

A contract $x$ is a mapping from states to transfers such that $x(\omega) \in \mathbb{R}$ is the transfer required from the agent in state $\omega \in \Omega$. We do not impose restrictions on the sign of these transfers, allowing them to be negative, i.e., to flow from the designer to the agent. A (direct) mechanism or menu is a collection of contracts $\{x_t : t \in T\}$ with $x_t : \Omega \to \mathbb{R}$ the contract for type $t$.[7]

The payoff of an agent with type $t$ under a contract $x$ is given by

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

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

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

Note that the first part of the agent's payoff does not depend on the contract he faces, this is because the surplus generated by the interaction between the agent and the designer is independent of the contracts offered in this stage.

Our main goal is to characterize under which conditions the designer is able to collect all the surplus $v_t$ from the agents (in expectation over $\Omega$) simultaneously using a direct mechanism.[8]

We introduce the definition of full surplus extraction formally.

**Definition 1.** *A mechanism $\{x_t : t \in T\}$ achieves full surplus extraction if for all $t \in T$*

$$\langle p_t, x_t \rangle = v_t.$$

In a traditional mechanism design problem, it is required that all types prefer their own contract to the contracts designed for others. Hence, incentive compatibility con-

[6] Under a more general model, the characteristics of the mechanism implemented would influence the alternatives an agent would take into account before deciding what alternative to pick or what message to report to the designer. Hence, the consideration sets would depend on the actual characteristic of the mechanism analyzed.

[7] Note that translations to/from non-direct mechanisms is not trivial in this setting since consideration sets are defined directly in terms of types instead of characteristics of the mechanism itself (e.g., allocation and prices).

[8] As Börgers (2015) notes, the focus on surplus extraction is arbitrary and the same results could be applied to implement any particular profile of payoffs or even allocations in a more general context.

straints must be imposed over all combinations of types. Since in our setting agents will be able to deviate over a subset of types, only some of those incentive compatibility constraints need to be satisfied. The definition of incentive compatibility must be adjusted accordingly.

**Definition 2.** *A mechanism $\{x_t : t \in T\}$ is incentive compatible if each type $t$ has no incentive to imitate any other type $t'$ in his consideration set $C_t$, i.e., if for all types $t \in T$,*

$$v_t - \langle p_t, x_t \rangle \geq v_t - \langle p_t, x_{t'} \rangle, \quad \forall t' \in C_t$$

Note that in the definition of incentive compatibility above $v_t$ plays no role since the "allocation", which determines $v_t$, is fixed regardless of the behavior of the agent in this stage. Hence, the incentive compatibility conditions above could alternatively be written as requiring each type $t \in T$ to choose his cost minimizing contract given his belief over $\Omega$, i.e.,

$$\langle p_t, x_t \rangle \leq \langle p_t, x_{t'} \rangle, \quad \forall t' \in C_t$$

The goal of the designer is to find an incentive compatible mechanism that extracts the full surplus from all types simultaneously.

**Definition 3.** *Full surplus extraction is feasible if there exists an incentive compatible mechanism $\{x_t : t \in T\}$ which achieves full surplus extraction.*

Crémer and McLean (1988) have shown that in a setting with full-consideration in which there are no restrictions on the types a particular type could imitate (i.e., $C_t = T$ for all types $t \in T$), full surplus extraction is guaranteed to be feasible only 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 condition, known as *convex independence*, is a linear independence condition over the set of beliefs. For finite settings, it also coincides with the more general condition of probabilistic independence used by McAfee and Reny (1992) and Lopomo et al. (2020).

We assume that different types hold different beliefs, that is, $p_t \neq p_{t'}$ if $t \neq t'$, and denote by $P^X$ the set of beliefs associated to types in $X \subseteq T$. Note that having $p_t \neq p_{t'}$ introduces correlation in our environment, i.e., types and states are not independent.[9]

[9] Having $p_t \neq p_{t'}$ is a necessary condition for full surplus extraction in environments with full-consideration, but it is not necessary in our environment with partial consideration and results could be easily extended to settings in which some types hold the same beliefs at the cost of increasing notation complexity.

Our characterization shows that the key element to identify is not the types a particular type $t$ could imitate (i.e., his consideration set) but the types that could pretend to be $t$ (i.e., the "inverse" of the consideration sets). More formally, we define the set of potential imitators or inverse consideration set for type $t$ as

$$D_t = \{t' \in T : t \in C_{t'} \text{ and } t \neq t'\}.$$

Note that the inverse consideration set $D_t$ for type $t$ is not determined by his own consideration set $C_t$ but the consideration sets of the types different from $t$.

Using the inverse consideration sets, we can rewrite the incentive compatibility constraints as requiring that for each type $t \in T$,

$$\langle p_{t'}, x_{t'} \rangle \leq \langle p_{t'}, x_t \rangle, \quad \forall t' \in D_t$$

Note that the conditions above are just a rearrangement of the incentive compatibility conditions described before, no new conditions are imposed and only the vacuous conditions are discarded (i.e., comparing the contract for type $t$ to itself).

We now proceed to present our main result.

**Theorem 1.** *Suppose $p_t \notin co\left(P^{D_t}\right)$ for all $t \in T$. Then, full surplus extraction is feasible.*

*Proof.* Consider any type $t \in T$. We will be looking for a function $z_t : \Omega \to \mathbb{R}$ which allows us to separate $t$ from the types that could pretend to be $t$. That is,

$$\langle p_t, z_t \rangle = 0$$

$$\langle p_{t'}, z_t \rangle > 0, \quad \forall t' \in D_t$$

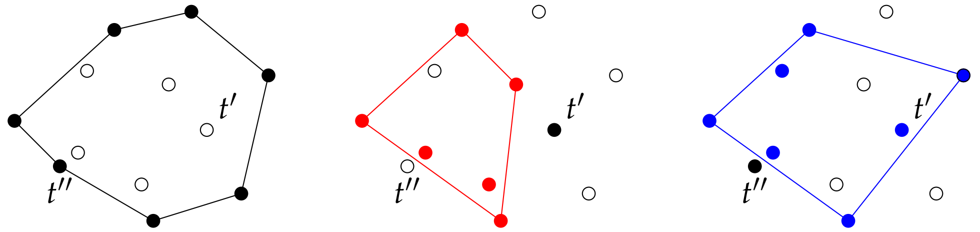
If $p_t \notin co\left(P^{D_t}\right)$ then existence of such a $z_t$ is guaranteed by Farkas' lemma. Then, we could build the contract for $t$ as follows

$$x_t = v_t + \alpha_t z_t$$

where $\alpha_t = \max_{t' \in D_t} \frac{v_{t'} - v_t}{\langle p_{t'}, z_t \rangle}$. Note this contract $x_t$ satisfies $\langle p_t, x_t \rangle = v_t$ for all and $\langle p_{t'}, x_t \rangle > v_{t'}$ for $t' \in D_t$. We can repeat this process for all other types and obtain a contract $x_{t'}$ for each type $t'$ in $T$. Note that $\langle p_{t'}, x_{t'} \rangle = v_{t'}$ and $\langle p_{t'}, x_t \rangle > v_{t'}$ implies that incentive compatibility for type $t'$ with respect to $t$ is satisfied.

Finally, the collection of contracts identified above satisfies the incentive compatibility constraints with respect to the relevant consideration sets, and achieves full surplus extraction. □

The proof of Theorem 1 follows the same structure as the proofs in Crémer and McLean (1985, 1988), however it highlights the importance of the inverse consideration sets in the characterization of the solution: only the set of beliefs associated with the inverse consideration sets are used in each step of the construction, being the key objects that determine whether a positive result could be obtained. Consideration sets are only indirectly relevant as they determine the structure of the inverse consideration sets. Note that the set of beliefs associated with each inverse consideration set is not required to satisfy any (convex or linear) independence condition, and there are no other direct restrictions on the relationship between these sets for different types beyond the conditions imposed in the main theorem.



(a) Failure of convex independence over $P^T$

(b) Type $t'$ and its inverse consideration set $D_{t'}$

(c) Type $t''$ and its inverse consideration set $D_{t''}$

Figure 1: Failure of convex independence and inverse consideration sets

Figure 1 represents an environment where the standard convex independence condition in Crémer and McLean (1988) fails. In particular, in panel $(a)$, there are types residing inside the convex hull of the beliefs of all types. While it is possible to build a contract which extracts all the rents in expectation for a type like $t''$, a type whose belief is an extreme point of the convex hull of $P^T$, the presence of types like $t'$, a type whose belief is in the interior of the convex hull of $P^T$, makes overall full surplus extraction unfeasible under full-consideration (i.e., if $C_t = T$ for all $t$) since any hyperplane going through $t'$ will leave types to both sides of such hyperplane. Hence, by using this hyperplane to punish types in one side we will be rewarding types in the other side, leading to potential violations of incentive compatibility, and making full surplus extraction unfeasible. However, if the set of types each type could imitate is constrained then full surplus extraction could still be feasible as our main theorem states. Panel $(b)$ and panel $(c)$ depict inverse consideration sets for types $t'$ and $t''$ respectively such that Theorem 1 holds. Note that types in the inverse consideration set of type $t$ do not need to satisfy convex independence. Moreover, the relation between different inverse consideration sets is not clearly determined by the conditions in Theorem 1. Indeed, the inverse consideration sets for both types share several common types without blocking the possibility of achieving full surplus extraction.

The failure of incentive compatibility in Figure 1 under full-consideration also identifies why inverse consideration sets are key instead of consideration sets directly: the reward-punishment scheme using an hyperplane going through the belief of type $t'$ makes some other types prefer the contract of type $t'$ over the contract designed for their true type in the case of full-consideration. Note this is a problem even if $t'$ is allowed to report or consider his true type $t'$ only, i.e., if $C_{t'} = \{t'\}$. Moreover, incentive compatibility would fail as long as there is another type for which the contract for type $t'$ is preferred to his designed contract, but discarding $t'$ from the consideration sets of those types eliminates such concern. Then, it follows that it is not the direct structure of the consideration sets that determine whether an incentive compatible full extracting mechanism could be constructed or not, but the structure of the inverse consideration sets.

## Necessity

In Theorem 1 we established sufficient conditions for obtaining full surplus extraction. We now turn to the question on whether such conditions could be relaxed. We show in result below that when these conditions fail, full surplus extraction and incentive compatibility are indeed incompatible for some surplus structures.

**Theorem 2.** *Suppose $p_t \in co\left(P^{D_t}\right)$ for some type $t$. Then, there exists a surplus vector $(v_{t'})_{t' \in T}$ such that there is no incentive compatible mechanism that achieves full surplus extraction.*

*Proof.* Suppose $p_t \in co\left(P^{D_t}\right)$ for type $t$ and that there is an incentive compatible mechanism $\{x_{t'}\}_{t' \in T}$ which achieves full surplus extraction. From incentive compatibility, we have

$$\langle p_{t'}, x_{t'} \rangle \leq \langle p_{t'}, x_t \rangle, \forall t' \in D_t.$$

Now, since $p_t \in co(P^{D_t})$, there exist a vector $\lambda \in \mathbb{R}^{D_t}$ such that $p_t = \sum_{t' \in D_t} \lambda_{t'} p_{t'}$, $\sum_{t' \in D_t} \lambda_{t'} = 1$, and $\lambda_{t'} \geq 0$ for all $t' \in D_t$. Then, we can combine these inequalities to obtain

$$\sum_{t'\in D_t} \lambda_{t'} \left\langle p_{t'}, x_{t'} \right\rangle \leq \sum_{t'\in D_t} \lambda_{t'} \left\langle p_{t'}, x_t \right\rangle .$$

As this mechanism achieves full surplus extraction, we have $\langle p_{t'}, x_{t'} \rangle = v_{t'}$, for all $t' \in T$. Thus, we can rewrite the inequality above as

$$\sum_{t'\in D_t} \lambda_{t'} v_{t'} \leq \sum_{t'\in D_t} \lambda_{t'} \left\langle p_{t'}, x_t \right\rangle .$$

For the right-hand side of this inequality, note that

$$\begin{aligned} \sum_{t'\in D_t} \lambda_{t'} \left\langle p_{t'}, x_t \right\rangle &= \sum_{t'\in D_t} \lambda_{t'} \sum_{\omega\in\Omega} p_{t'}(\omega) x_t(\omega) \\ &= \sum_{\omega\in\Omega} \sum_{t'\in D_t} \lambda_{t'} p_{t'}(\omega) x_t(\omega) \\ &= \left\langle \sum_{t'\in D_t} \lambda_{t'} p_{t'}, x_t \right\rangle = \left\langle p_t, x_t \right\rangle \\ &= v_t \end{aligned}$$

Hence, for this mechanism to be incentive compatibility and achieve full surplus extraction, we need

$$\sum_{t'\in D_t} \lambda_{t'} v_{t'} \leq v_t$$

to be satisfied. But for $v_t < \min_{t'\in D_t} v_{t'}$ this condition is violated. Hence, no incentive compatible mechanism could achieve full surplus extraction in this case. □

This results then shows that the conditions in Theorem 1 are not only sufficient but also necessary if we want to establish the feasibility of full surplus extraction for all surplus structures.

## Failure of the Revelation Principle

The previous discussion focused on the use of truthful direct mechanisms without addressing whether such restriction is without loss, i.e., whether a revelation principle holds in our environment. Below, we show that this is not the case and the restriction to truthful direct mechanisms, in the form of a menu and consideration sets as defined in our model, is with potential loss of generality since there are instances in which a non-truthful mechanism could implement a fully extracting scheme but all truthful mechanisms fail to do so.

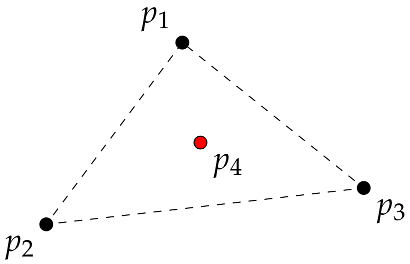


Figure 2: Failure of full surplus extraction with a truthful mechanism

Consider the following environment illustrated in Figure 2: there are four types $t_1, t_2, t_3$, and $t_4$, with associated beliefs $p_1, p_2, p_3$, and $p_4$ as depicted in the figure. Suppose types $t_1$ and $t_2$ share the same consideration set $C_1 = C_2 = \{t_1, t_2, t_4\}$: they could imitate each other and also type $t_4$. Types $t_3$ and $t_4$ also share a common consideration set $C_3 = C_4 = \{t_3, t_4\}$, so they can imitate each other but do not consider the contract of any other type. This induces inverse consideration sets $D_1 = \{t_2\}$, $D_2 = \{t_1\}$, $D_3 = \{t_4\}$, and $D_4 = \{t_1, t_2, t_3\}$. Since $p_4 \in co(P^{D_4})$, Theorem 1 fails, and full surplus extraction cannot be guaranteed for an arbitrary surplus structure $\{v_1, v_2, v_3, v_4\}$.

However, it is possible to create a mechanisms which exchanges the targeted type for contracts $x_3$ and $x_4$ so that $t_3$ chooses $x_4$ while $t_4$ chooses $x_3$, and satisfies adjusted notions of incentive compatible and full surplus extraction. This is possible given that the contract of type $t_4$ ($x_4$), is evaluated by all types but the belief of type $t_3$ is a extreme point of the set of beliefs, and hence could be separated from the beliefs of the other types. Given this, it is possible to tailor contract $x_4$ to obtain all the surplus from type $t_3$ and make any other type unwilling to choose $x_4$. Instead, for contract $x_3$ we have that only types $t_3$ and $t_4$ are able to evaluate it as a potential deviation. Hence, as long as $p_3 \neq p_4$ it is possible to configure $x_3$ in a way that makes type $t_3$ unwilling to accept it while extracts all the surplus from type $t_4$. Pairing these contracts with the construction in the proof of Theorem 1 for contracts $x_1$ and $x_2$, we obtain a mechanism that achieves full surplus extraction and satisfy an appropriate notion of incentive compatibility. Moreover, this could be achieved for any fixed profile of values $\{v_1, v_2, v_3, v_4\}$. However, no truthful mechanism is able to achieve both conditions simultaneously unless very specific surplus structures are assumed.

While in the example above relabeling types allows us to achieve full surplus extraction, the use of non-truthful mechanisms could be problematic in our setting. Note that this relabeling requires types and consideration sets to be "purely cheap talk" in the sense that labels have no associated features or value. But, this is inconsistent with having consideration sets determined by the interaction in the first (unmodeled) stage since the inclusion of a type in another types' consideration sets could be linked to the allocation associated to them (Section 3.3 offers an example of this dependence in a screening problem). Moreover, implementing such relabeling requires types to report inconsistently among stages: they initially report their true type but some of them fail to do so in the second stage.

### Genericity and beliefs-determine-preferences settings

We finish this section discussing how flexible or generic is our model compared to a standard setting with full-consideration.

In the standard full-consideration framework as in Crémer and McLean (1988) or McAfee and Reny (1992), having two types $t$ and $t'$ with the same beliefs, $p_t = p_{t'}$, but different surplus or valuations, $v_t \neq v_{t'}$, makes full surplus extraction impossible. Indeed, having such types would immediately violate the CM condition since $p_{t'} \in P^{T \setminus t}$. In contrast, this is not necessarily a problem in our general model: having $v_t \neq v_{t'}$ and $p_t = p_{t'}$ is compatible with full surplus extraction as long the associated inverse consideration sets $D_t$ and $D_{t'}$ (and the rest of the model) satisfy the conditions in Theorem 1. Moreover, this is possible even if their consideration sets also coincide everywhere else, i.e., if $C_t \setminus \{t\} = C_{t'} \setminus \{t'\}$, as this does not necessarily has an impact on the inverse consideration sets of types $t$ and $t'$. However, having $t' \in C_t$ does make full surplus extraction impossible in our model since it implies $t \in D_{t'}$ which in turn implies $p_{t'} \in co(P^{D'_t})$, violating the required condition in this case.

Hence, our model allows some flexibility in terms of the structure of beliefs, surplus and consideration sets that the standard framework does not, extending surplus extraction results beyond beliefs-determine-preferences environments (Chen and Xiong, 2011) as long as some violations of rationality are allowed. However, our model still involves limitations in terms of the general feasibility of full surplus extraction since the structure of the inverse consideration sets is not allowed to vary freely.

## 3 Applications

In this section we present some examples in which our main result could be applied. We start with a procurement problem in which firms can only partially misreport their risk level. Then, we define separable environments and characterize sufficient conditions that guarantee full surplus extraction in those settings. We then study a screening problem in which the consideration sets depend on the quality provided under different contracts instead of depending of types directly, offering a form of partial endogeneization of the consideration sets. Finally, we apply our main theorem to characterize the conditions required to guarantee full surplus extraction in a setting in which some types always report truthfully, and use this result to characterize the revenue maximizing mechanism for an auction with correlation and behavioral bidders.

## 3.1 Procurement under partial risk misrepresentation

We consider the problem of a buyer hiring a firm for completing a project. The project could either be completed on time or be delayed in which case its cost of completion increases. Delays depend on the type of firm working on the project. If the project is completed on time then its cost is $\omega_0$, while if its delayed then its cost increases by $\omega_1$. There are a $N$ types of firms, and the type of the firm is private information. Then, a firm with type $t$ will have a probability of facing delays of $p_t$, which we refer as its risk level. Types are ordered according to their risk level so $p_t > p_{t'}$ if and only if $t > t'$. Firms could also differ on an initial or fixed cost $K_t \geq 0$ of starting the project. A contract $x = (x_0, x_1)$ specifies the compensation without and with delays. Note there is no restriction on the sign of these compensations, so they could either flow from the buyer to the seller (if positive) or from the seller to the buyer (if negative). The buyer is able to screen the firm offering a menu of contracts $\{x^t : t \in T\}$. The firm exhibits partial consideration and only is able to evaluate or fulfill contracts designed for a subset of types. In particular, a firm of type $t$ will only consider contracts associated with risk levels $p \in [p_t, p_{t+2}]$, i.e., she cannot claim having a lower level of risk than the level she really has, nor reporting risk levels way beyond her true risk level. No under-reporting could be due to the requirement of presenting some type of certification (which could not be falsified), while no over-reporting of risk could be due to reputation concerns.

The goal of the buyer is to hire the firm at the least expected cost possible, while the firm will choose the best contract considered which has an expected benefit above her expected cost.

The following proposition establishes that in this setting, the buyer can always design a menu of contracts in which the expected payment to each type of firm is exactly her expected cost of completing the project.

**Proposition 1.** *Consider the procurement problem described above. Then, there exists a menu of contracts such that each type of firm chooses a contract with expected payment exactly equal to her expected cost of completing the project.*

*Proof.* We start by defining the consideration set for each type in the space of types as $C_t = \{t, t+1, t+2\} \bigcap T$ for any type $t \in T$. Then, the inverse consideration set for type $t$ is $D_t = \{t-1, t-2\} \bigcap T$. Let's denote the expected cost of type $t$ as $v_t = K_t + \omega_0 + p_t \omega_1$.

Since $p_t > p_{t'}$ for any $t > t'$, $p_t \neq \alpha p_{t-1} + (1-\alpha) p_{t-2}$ for any $\alpha \in [0,1]$. This implies $p_t \notin co\left(P^{D_t}\right)$, so the condition in Theorem 1 holds.

Then, from Theorem 1, there exists an incentive compatible mechanism $\{x_t : t \in T\}$ such that $\langle p_t, x_t \rangle = v_t$ for all $t \in T$. Under this mechanism a firm of type $t$ receives an expected compensation exactly equal to her expected cost of completing the project as required. □

Note that the restricted structure of the consideration sets is required to obtain a menu in which each type of firm compensation is just her expected cost of completing the project. For example, if any type could report types above and below her true type then Theorem 1 would fail. However, existence of a "fully extracting" mechanism could be recovered in this context if there are more states available to condition contracts on and alternative orderings of types are imposed.[10] Hence, the requirement of single-sided misrepresentation is a consequence of having only two states (i.e., cost levels) and not a general restriction on the model.

## 3.2 Separable environments

In this section we consider a special structure of consideration sets which allows for a simpler characterization of the conditions for surplus extraction. This environment classify types into different groups for which each type only considers some types in his same group as a potential deviation.

We consider a permissive definition of a separable environment: consideration sets are not required to be identical among types in the same group, and they could be strict subsets of the block they belong to. Such flexibility comes at the cost of identifying only sufficient conditions for full surplus extraction instead of a full characterization.

**Definition 5.** *We say that an environment is separable if there exists a partition of $T$, $\{T_1, T_2, ...\}$ such that $C_t \subseteq T_i$ for all $t \in T_i$.*

[10] An example of this is the screening model in Section 3.3.

In separable environments, for each group $T_i$ an isolated design problem is faced since there is no interaction with types in other groups. Hence, if the standard independence condition is satisfied for each group then full surplus extraction could be obtained.

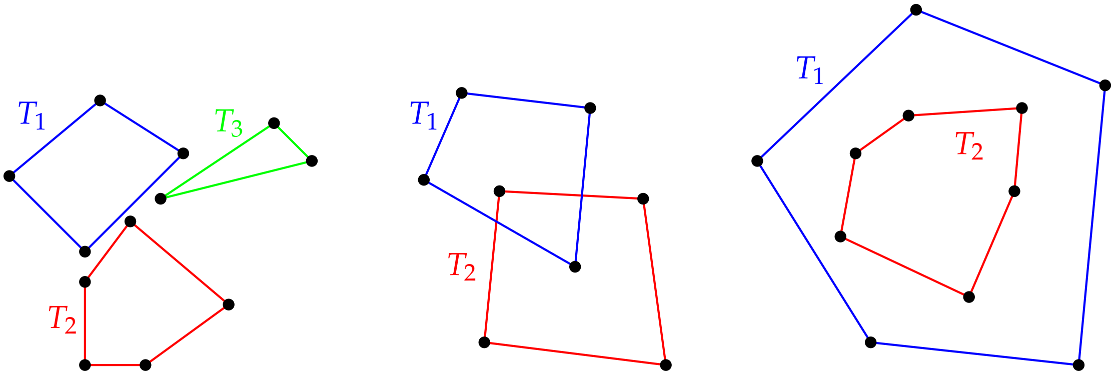
**Corollary 1.** *Consider a separable environment indexed by $\mathcal{I}$. Suppose $P^{T_i}$ satisfies the CM condition for each $i \in \mathcal{I}$, then full surplus extraction is feasible.*

*Proof.* Pick an element $T_i$ from the partition of $T$, and consider the restricted problem in which the set of types is restricted to $T_i$ only. Since $P^{T_i}$ satisfies the CM condition, there exists a mechanism $\{x_t : t \in T_i\}$ such that $\langle p_t, x_t \rangle = v_t$ for all $t \in T_i$ which satisfies incentive compatibility for all the types in $T_i$. Since no incentive compatibility constraint involving types in $T_i$ relates to types outside $T_i$, such restricted mechanism satisfies all relevant incentive compatibility constraints for types in $T_i$.

Repeating this process for all elements of the partition generates a sequence of restricted mechanisms $(\{x_t : t \in T_i\})_{i \in \mathcal{I}}$. Since $(T_i)_{i \in \mathcal{I}}$ is a partition, no type $t$ appears in two elements of such sequence. Hence, we can aggregate the contracts in each mechanism in $\{x_t : t \in T\}$. This resulting mechanism is incentive compatible and achieves full surplus extraction as required.

□

The interpretation of this conditions is straightforward since they translate to a collection of separated surplus extraction problem which could be handled in complete isolation to each other.



(a) Fully separated groups

(b) Overlapping convex hull for different groups

(c) Convex hull of a group containing another group

Figure 3: Three examples of separable environments in which full surplus extraction is feasible

Figure 3 illustrate different structures of beliefs which are separable and compatible with full surplus extraction in the sense of Corollary 1. Panel (a) describes the most intuitive separable environment: each group convex hull of beliefs is completely separated from the convex hull of beliefs of other groups. Panels (b) and (c) instead illustrate that in a separable environment the convex hull of different groups can intersect (b), or even be completely contained in the convex hull of other groups (c), as long as different groups do not share types. Hence, in a separable environment having $co\left(P^{D_t}\right) \cap co\left(P^{D_{t'}}\right) \neq \emptyset$ for $t \neq t'$ is possible.

A natural application of a separable environment is a setting in which consumers could be classified into different groups according to their characteristics as in standard third degree price discrimination strategies. Our model allows heterogeneity to remain among each group, having different contracts offered to different members of the same group.

As we mentioned before, Corollary 1 identifies only sufficient conditions for having full surplus extraction in this environments. We illustrate this in the following example. Consider the environment depicted in Figure 4. Suppose that initially $C_t = \{t_1, t_2, t_3\}$ for $t \in \{t_1, t_2, t_3\}$, $C_t = \{t_4, t_5, t_6\}$ for $t \in \{t_4, t_5, t_6\}$, and $C_t = \{t_7, t_8\}$ for $t \in \{t_7, t_8\}$. Clearly, this environment is separable: let $T_1 = \{t_1, t_2, t_3\}$, $T_2 = \{t_4, t_5, t_6\}$, and $T_3 = \{t_7, t_8\}$, then $C_t \subseteq T_i$ for all $t \in T_i$ and $i \in \{1, 2, 3\}$. Since, $P^{T_1}$, $P^{T_2}$, and $P^{T_3}$ satisfy the CM condition, full surplus extraction is feasible in this case.

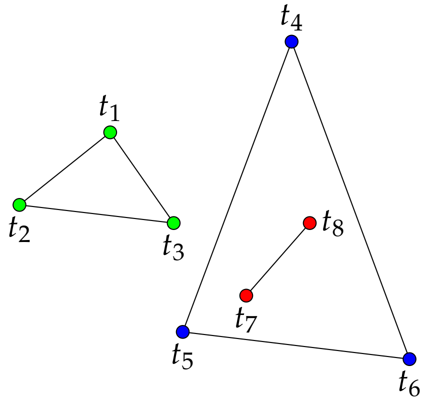


Figure 4: A separable environment consistent with Corollary 1

Now suppose that types $t_2$ and $t_4$ also consider type $t_7$ as a possible deviation. This new environment is still separable: let $T_1 = \{t_1, t_2, t_3\}$ and $T_2 = \{t_4, t_5, t_6, t_7, t_8\}$, then $C_t \subseteq T_i$ for all $t \in T_i$ and $i \in \{1, 2\}$. Note that now the condition in Corollary 1 is violated since $p_{t_7} \in co\left(P^{T_1 \setminus \{t_7\}}\right)$ as could be seen in Figure 5. However, full surplus extraction is still feasible since $D_{t_7} = \{t_4, t_6, t_8\}$, and $p_{t_7} \notin co\left(P^{D_{t_7}}\right)$ as Theorem 1 requires.

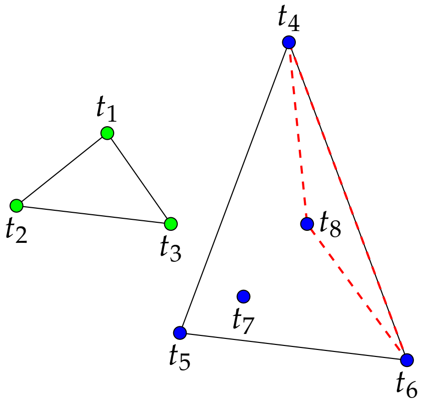


Figure 5: A separable environment violating Corollary 1

## 3.3 Screening with partially endogenous consideration sets

Below we analyze a setting in which the consideration sets (in terms of types) are partially endogenized by allowing them to be formed as a response to evaluating only products close to the ideal product for a particular type. However, even in this model the primitive motivator of consideration sets formation is exogenous: each type will only look at products/contracts in the neighborhood of his ideal quality level, and the size of this neighborhood is exogenously given.

In particular, we consider a monopolistic screening problem in which the allocation in the first stage explicitly determines all the conditions for the second stage: surplus, beliefs, and consideration sets. This way, the model presented in this section offers consideration sets which are partially endogenous by allowing the types belonging to a consideration set to be determined by the allocation they are associated with.

Consider initially a setting without the exogenous source of uncertainty and full-consideration, i.e., a setting in which transfers have no state to condition on (beyond types) and all types could evaluate all the alternatives offered by the mechanism. Consider the following quadratic utility function for the agents

$$u(\theta, q, \overline{x}) = \theta \cdot q - \frac{q^2}{2} - \overline{x}$$

where $\theta$ is the valuation, $q$ is the quality of the product, and $\overline{x}$ is the transfer paid by the agent. We consider, as before, a finite set of types $T$, and assume that each type $t$ has a different valuation $\theta_t > 0$. We further assume that $T$ is ordered according to the valuations of different types, so $\theta_t > \theta_{t-1}$ for all $t > 1$.

We are interested in a mechanism which implements the allocation that maximizes the surplus generated for each type (i.e., the first best allocation): $q_t^* = \theta_t$ for all $t$. Setting $\overline{x}_t = \overline{x}_{t-1} + \frac{(\theta_t - \theta_{t-1})^2}{2}$ and $\overline{x}_1 = \frac{\theta_1^2}{2}$, allows this allocation to be implemented.

Then, the surplus generated for each type $t$ by this mechanism is[11]

$$v_t = \theta_t q_t^* - \frac{q_t^{*2}}{2}$$

Since, $q_t^* = \theta_t$, we have $v_t = \frac{\theta_t^2}{2}$.

For the second stage, we define consideration sets as follows: each type $t$ only looks at the neighborhood of his ideal quality level. Due to the quadratic utility specification the ideal quality for type $t$ matches his valuation $\theta_t$. In particular, lets assume type $t$ considers contracts which offer quality in the interval $[q_{t-1}^*, q_{t+1}^*]$ for $t > 1$, and $[0, q_2^*]$ for $t = 1$.[12] Given the allocation above, this is equivalent to allowing each type to consider the contracts for the types immediately above and below his true type in addition to the contract for his true type, i.e., $C_t = \{t-1, t, t+1\}$.[13]

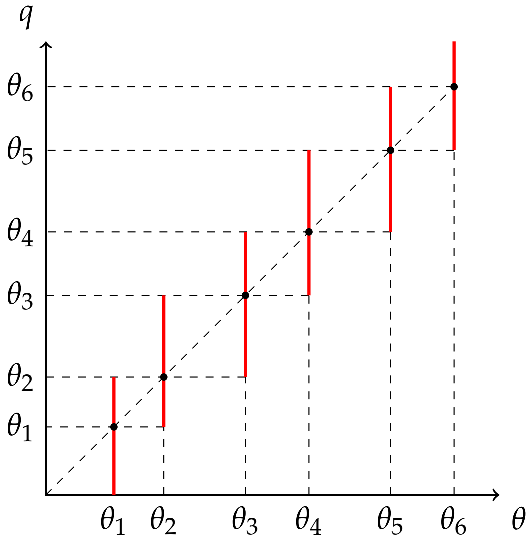


Figure 6: First stage allocation and associated consideration sets

Figure 6 depicts the quality provided in the first stage to each type as a function of their valuation (black), and the consideration set for each type in terms of quality (red).

Finally, we introduce correlated states in the model: each type $t$ will be associated

[11] Note that we can consider the net surplus instead without changes in the results of this section.

[12] More generally, the quality interval for type $t$ could take the form $[\underline{q}_t, \overline{q}_t]$ with $\underline{q}_t \in (\theta_{t-1}, \theta_t]$ and $\overline{q}_t = [\theta_t, \theta_{t+1})$ without changing the consideration sets in terms of types.

[13] The same consideration sets structure could be induced if instead of quality each type looks only at an interval of prices close to $\overline{x}_t$ under some conditions.

with a belief or distribution over a finite set of exogenous states $\Omega$. Let $p_t \in \Delta(\Omega)$ be such distribution for type $t$. As before, we assume different types hold different beliefs over $\Omega$, so $p_t \neq p_{t'}$ if $t \neq t'$.

Hence, the model described above matches an instance of the general model described in Section 2.

The inverse consideration sets in this case are

$$D_t = \{t-1, t+1\}$$

for $t > 1$, and $D_t = \{2\}$ for $t = 1$. As Theorem 1 states, if for each type $t$ his belief $p_t$ does not belong to the convex hull of the beliefs of the types in his inverse consideration set $D_t$, then all the surplus could be extracted.

**Proposition 2.** *Consider the monopolistic screening problem described above. Suppose that for each type $t > 1$, there is no $\alpha \in [0, 1]$ such that $p_t = \alpha p_{t+1} + (1-\alpha)p_{t-1}$. Then, full surplus extraction is feasible.*

*Proof.* Suppose that for all $t > 1$, there is no $\alpha \in [0, 1]$ such that $p_t = \alpha p_{t+1} + (1-\alpha)p_{t-1}$. Note that for any type $t > 1$ this condition is equivalent to $p_t \notin P^{D_t}$ since $D_t = \{t-1, t+1\}$. For type $t$, we have $D_t = \{2\}$ and $p_2 \neq p_1$ which imply $p_t \notin P^{D_t}$ as well.

Hence, the environment in the second stage satisfies the conditions in Theorem 1. Then, from Theorem 1 we have that there exists an incentive compatible mechanism $\{x_t : t \in T\}$ such that $\langle p_t, x_t \rangle = v_t$ for all $t \in T$. □

This model allows us to partially endogenize the consideration sets in terms of types. However, the underlying cause generating the structure of consideration remains exogenous: the agent will only look at contracts with quality around his ideal point.

In Figure 7 we illustrate a belief structure which is consistent with full surplus extraction in this setting. Note that there is no obvious ordering of these beliefs with respect to their associated type. However, for any triple of consecutive types, the belief of the "middle" type is never a convex combination of the beliefs of the "extreme" types. Hence, it is possible to find a mechanism that is both incentive compatible and collects all the surplus.

### 3.4 Honest and sophisticated types

In this section we study a simple environment in which types could be classified into two groups: honest and sophisticated. Honest types will be unable to imitate any other

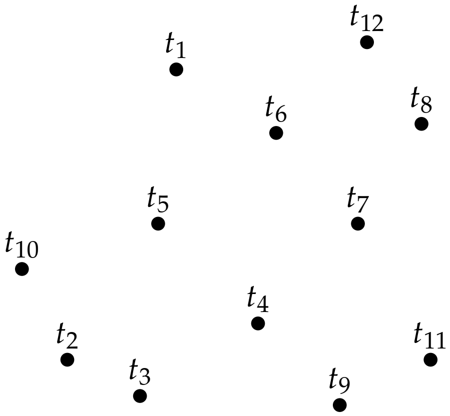


Figure 7: A beliefs structure consistent with full surplus extraction in this setting

type and always report their type truthfully. Sophisticated types instead will be fully rational and will be able to imitate any other type in $T$, including honest types. So, for an honest type $t$ his consideration set is $C_t = \{t\}$, while $C_{t'} = T$ is the consideration set for a sophisticated type with type $t'$.

Then, for a set of honest types $H$, the inverse consideration set for an honest type will include only the sophisticated types, i.e., $D_t = T \backslash H$ for all honest types $t \in H$. For a sophisticated type $t' \in T \backslash H$, his inverse consideration set would include all other sophisticated types, also excluding all honest types, i.e., $D_{t'} = T \backslash \left( H \cup \{t'\} \right)$.

Thus, the conditions that guarantee full surplus extraction in Theorem 1 reduce to determine whether for all types $t \in T$, $p_t$ is inside the convex hull of $P^{T \backslash H}$ or not. Moreover, these condition could be separated into two sets of conditions.

**Corollary 2.** *Consider an environment with a set of honest types $H \subseteq T$. Suppose $P^{T \backslash H}$ satisfies the CM condition, and $p_t \notin co\left(P^{T \backslash H}\right)$ for each $t \in H$, then full surplus extraction is feasible.*

*Proof.* Consider first types in $H$. Since, $D_t = P^{T \backslash H}$ for any $t \in H$, the condition above for these types is equivalent to $p_t \notin co\left(P^{D_t}\right)$ for each $t \in H$.

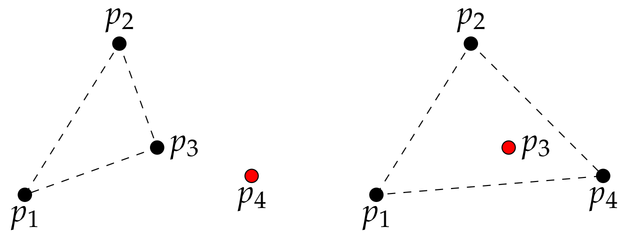
Now for any type $t \in T \backslash H$ note that the $P^{T \backslash H}$ satisfying the CM condition implies that $p_t \notin co\left(P^{T \backslash (H \cup \{t\})}\right)$, and $D_t = T \backslash (H \cup \{t\})$. Together these imply $p_t \notin co\left(P^{D_t}\right)$.

Hence, the conditions in Theorem 1 holds and the existence of an incentive compatible mechanism which achieves full extraction follows from Theorem 1.

□

Here checking the conditions is less cumbersome than in the general environment. It reduces to check whether the subset of sophisticated types satisfies the CM condition, and then check whether for each behavioral type his belief is inside the convex hull of beliefs of the sophisticated types. Note the evaluation for each behavioral type could be carried over in complete isolation of other behavioral types.

In this environment, the convex independence condition among sophisticated types is necessary but not sufficient to guarantee full surplus extraction to be feasible. Indeed, beliefs of honest types are required to be outside the convex hull of the beliefs of the sophisticated types as otherwise a contract offered to an honest type could still be preferred by a sophisticated type (at least for some payoff structures).



(a) Only $t = 4$ being an honest type

(b) Only $t = 3$ being an honest type

Figure 8: Representation of Corollary 2.

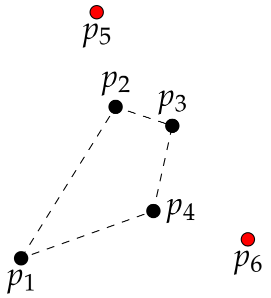


Figure 9: More sophisticated and honest types

Figures 8 and 9 illustrate the result in Corollary 2 considering two different cases. Beliefs of sophisticated types are depicted in black while beliefs of honest types are depicted in red. Note that in all cases $P^T$ violates the CM condition, hence Crémer and McLean (1988) theorem fails. In Figure 8, panel (a) shows that if only type $t = 4$ is honest, then full surplus extraction is feasible, while panel (b) shows that if only $t = 3$ is honest full extraction is not feasible since any contract aimed to fully extract type $t = 3$ could induce deviations from some of the sophisticated types. Figure 9 shows that having more sophisticated and honest types does not change the result as long as the conditions in Corollary 2 still hold.

While we presented this case as an environment where some types are honest, any environment in which there is a class of types which can be easily identified or cannot falsify their report are behaviorally equivalent to this class of honest types, and the result in this section carries over to such environments as well.

Note that this environment is not separable since sophisticated types consider not only the contracts offered to other sophisticated types but also to honest types. This shows that there are interesting environments beyond the separable case which could also allow simpler characterizations.

A particularly salient application of the model in this section are markets in which some groups of agents exhibit inertia. For example, Abaluck and Adams-Prassl (2021) and Abaluck and Gruber (2022) discuss evidence of inertia in health insurance markets. Under this interpretation, honest types decide to keep their current policy ignoring potentially better available policies, while sophisticated types evaluate all available policies before making their choice.

## 3.5 Auction with behavioral bidders (and correlation)

In this section we introduce an auction environment to illustrate the main result in a model with honest types. We start by formally describing the auction model, and then apply our main theorem to characterize the fully extracting mechanism.

Consider a standard private values auction environment with correlation: there is a single item which could be allocated to one of $n \geq 2$ bidders. Each bidder has a valuation $\theta_i$ for the item. This valuation is the bidder's private information, observed only by himself. There is a finite set of potential valuations for each bidder $i$, which we denote by $\Theta_i$. We also define $\boldsymbol{\Theta} = \times_{i \in N} \Theta_i$ and $\Theta_{-i} = \times_{j \neq i} \Theta_j$, with general elements $\boldsymbol{\theta}$ and $\theta_{-i}$ respectively. There is a common prior $F$ over the vector of valuations $\boldsymbol{\theta}$, i.e., $F \in \Delta(\boldsymbol{\Theta})$. A bidder $i$ with valuation $\theta_i$ holds beliefs $F(\cdot|\theta_i) \in \Delta(\Theta_{-i})$ over the valuations of the other bidders.

We introduce *behavioral types* among the bidders. Here a behavioral type will be determined by his valuation-belief pair. We denote by $B_i \subseteq \Theta_i$ the set of behavioral types for player $i$. Behavioral types will be honest and always report truthfully, while non-behavioral types will report what is best for them under the mechanism implemented.

A complete mechanism here is an allocation rule $\{q_i : \Theta \Rightarrow [0,1]\}_{i \in \{1,\ldots,n\}}$ and a transfer rule $\{x_i : \Theta \Rightarrow \mathbb{R}\}_{i \in \{1,\ldots,n\}}$ where $\sum_{i=1}^{n} q_i(\theta) = 1$ for all $\theta \in \Theta$.

For the discussion of this section, we focus on the symmetric case in which all bidders share the same space of valuations, that is for any $i$ and $j$, $\Theta_i = \Theta_j = \Theta$, and their beliefs are also symmetric, that is for any $t \in \Theta$ and $\omega \in \Theta^{n-1}$, $F(\theta_{-i} = \omega | \theta_i = t) = F(\theta_{-j} = \omega | \theta_j = t)$ for all $i$ and $j$. Moreover, we assume $B_i = B_j = B$ for all bidders $i$ and $j$ as well.[14]

We further assume that each valuation generates a different distribution over the valuations of the other bidders, so $F(\cdot|\theta_i = t) \neq F(\cdot|\theta_i = t')$ for all $t, t' \in \Theta$.

Note that by our symmetry assumption, each valuation will not only determine the beliefs but also the degree of sophistication of a particular bidder of type $\theta$. Hence, a bidder with valuation $\theta \in B$ holds beliefs $F(\cdot|\theta)$ and always report truthfully, while a bidder with valuation $\theta' \notin B$ has beliefs $F(\cdot|\theta')$ and is fully strategic.

In order to apply our main result to characterize the optimal mechanism in this setting, we first need to follow three steps: (1) fix the allocation rule, (2) transform the multi-bidder problem into the problem of a single bidder, and (3) impose restrictions on the distribution of valuations.

For the first step, we need to fix the allocation rule since in our framework only transfers are allowed. Since we are interested on discussing the conditions that allows the seller to extract all the surplus from the bidders, we will fix the allocation rule to be the allocation that maximizes the total surplus: the bidder with the highest valuation gets the item. We assume any tie is resolved in favor of a particular bidder $i$ when he is part of the set of winners, and in favor of the bidder with the lowest index otherwise.[15]

For the next step we need to reduce the multi-bidder problem to the problem of a single bidder. This a necessary step since our main theorem applies to settings with a single agent. We transform the multi-bidder auction into a single bidder auction by focusing in the problem of bidder $i$ and taking expectations over the valuations of the other types.

In particular, we denote valuation of bidder $i$ by $t$ and the vector of valuations of the other bidders different from $i$ by $\omega$. We let the probability of the profile $\omega$ conditional on valuation $t$ for bidder $i$ to be $p_t(\omega) = F(\theta_{-i} = \omega | \theta_i)$.

We will denote the gross expected utility of bidder $i$ with valuation $t$ by $v_t$. Hence, under the efficient allocation rule $v_t$ is equal to the valuation $t$ multiplied by the probability he has the highest valuation.[16]

[14] This is just for simplifying the exposition, and all the discussion extends directly to the asymmetric case as the main proposition in this section shows.

[15] Results extend directly to alternative tie-breaking rules as usual.

[16] Given the notation of this section, we have

$$v_t = t \cdot \left( \sum_{\{\theta_{-i} : \max_{j \neq i} \theta_j \leq t\}} F(\theta_{-i} | \theta_i = t) \right).$$

Finally, $x_t(\omega)$ represents the transfer made by the bidder if his reported valuation is $t$ and the vector of valuations reported by other bidders is $\omega$.

For the final step, we impose restrictions over the conditional distribution of valuations. In particular, we impose that for all $t \in \Theta$

$$p_t \notin co(p_{t'} : t' \notin B \text{ and } t' \neq t).$$

These conditions are equivalent to the conditions in Theorem 1 (and Corollary 2 for $B = H$). Hence, we can apply directly Theorem 1 to guarantee full surplus extraction in this setting. Moreover, we can use the construction in the proof of Theorem 1 to compute the transfers required to extract all the rents. Since this transfer rule is incentive compatible and extracts all the surplus under the surplus maximizing allocation, the optimal mechanism will indeed extract all the informational rents in expectation as long as the conditions above hold.

Note that the above arguments do not rely on the symmetry assumption imposed before and could be easily extended to the asymmetric case. We state this general result formally below, using the original notation for the auction environment without the symmetry restriction.

**Proposition 3.** *Consider the auction environment. Let $B_i$ the set of behavioral types for bidder $i$. If for all bidders $i \in \{1, ..., n\}$, and valuations $\theta_i \in \Theta_i$,*

$$F(\cdot|\theta_i) \notin co(\{F(\cdot|\theta_i') : \theta_i' \notin B_i \text{ and } \theta_i' \neq \theta_i\}),$$

*then the optimal mechanism achieves full surplus extraction.*

*Proof.* Assume the conditions over the conditional distributions hold.

Let $q_i(\theta)$ denote the allocation for bidder $i$ for a vector of types $\theta$ when the good is allocated to the bidder with the highest valuation and ties are broken arbitrarily.

Then, the expected payoff of bidder $i$ if his valuation is $\theta_i$ is

$$v_i(\theta_i) = \sum_{\theta_-i \in \Theta_{-i}} F(\theta_{-i}|\theta_i) q_i(\theta_i, \theta_{-i})$$

Let $C_i(\theta_i) \subseteq \Theta_i$ denote the consideration set of bidder $i$ if his valuation is $\theta_i$. Then, $C_i(\theta_i) = \{\theta_i\}$ for $\theta_i \in B_i$ and $C_i(\theta_i) = \Theta_i$ for $\theta_i \notin B_i$. The corresponding inverse consideration sets are given by $D_i(\theta_i) = \Theta_i \backslash B_i$ for $\theta_i \in B_i$ and $D_i(\theta_i) = \Theta_i \backslash (B_i \bigcup \{\theta_i\})$ for $\theta_i \notin B_i$. Therefore, we have that the environment $\left(\Theta_i, (v_i(\theta_i), F(\cdot|\theta_i), C_i(\theta_i))_{\theta_i \in \Theta_i}\right)$ matches the structure of an environment with a set of honest types $B_i$ as in Corollary 2.

It remains to show that the conditions in Corollary 2 are satisfied under the assumptions over the conditional distributions for bidder $i$.

Note that for any valuation $\theta_i \in B_i$ from a behavioral type, the set of conditional distributions

$$\{F(\cdot|\theta_i') : \theta_i' \notin B_i \text{ and } \theta_i' \neq \theta_i\} = \{F(\cdot|\theta_i') : \theta_i' \notin B_i\} = P^{\Theta_i \backslash B_i},$$

since the second condition is redundant.

Hence, the condition over the conditional distribution for this valuation reduces to $F(\cdot|\theta_i) \notin co\left(P^{\Theta_i \backslash B_i}\right)$.

Now, for a valuation $\theta_i \notin B_i$ from a non-behavioral type we have,

$$\{F(\cdot|\theta_i') : \theta_i' \notin B_i \text{ and } \theta_i' \neq \theta_i\} = \{F(\cdot|\theta_i') : \theta_i' \notin B_i\} \backslash \{F(\cdot|\theta_i)\} = P^{\Theta_i \backslash B_i} \backslash \{F(\cdot|\theta_i)\},$$

since this holds for any $\theta_i \notin B_i$, $P^{\Theta_i \backslash B_i}$ satisfies the CM condition.

Hence, the set of conditional distributions for bidder $i$ satisfy the two conditions in Corollary 2. This implies there exists an incentive compatible mechanism $\{x_i(\theta_i) \in \mathbb{R}^\Omega : \theta_i \in \Theta_i\}$ such that $\langle F(\cdot|\theta_i), x_i(\theta) \rangle = v_i(\theta_i)$ for each $\theta_i \in \Theta_i$.

Since this holds for an arbitrary bidder $i$, it holds for any such bidder. Then, the mechanism $(q_i, x_i)_{i \in \{1,\ldots,n\}}$ is incentive compatible and achieves full surplus extraction. Since under this mechanism the seller obtains the maximum revenue among the mechanisms that satisfies participation, it is indeed the optimal mechanism for the seller.

□

Notice that if in addition to the private information of other bidders, we include other variables correlated with the valuation of bidder $i$ on which the auction payments (and allocation) could condition on then after adjusting the notation to accommodate such variables essentially the same result holds.

Under the symmetry assumption we have discussed above and a symmetric tie-breaking rule, the optimal mechanism is also anonymous since each bidder will face the same allocation and transfer rules regardless of his index $i$. Without symmetry the optimal mechanism is no longer anonymous and would require to offer personalized allocation and transfer rules based on each particular bidder index $i$. In such a case, the optimal auction will not only require personalized allocation and transfer rules accordingly to the bidders valuations distribution but also with respect to their behavioral status since it allows behavioral types to represent different characteristic based on the index of the bidder. That is, bidder $i$ could be behavioral if his type is $t_i = t$ but for index $j \neq i$ a bidder with type $t_j = t$ could be fully strategic. Thus, the induced personalized mechanism could create new challenges or advantages depending on the particular application we are dealing with.

## 4 Concluding remarks

We examined the full surplus extraction problem with boundedly-rational agents which can only imitate a subset of types. We characterize the conditions that guarantee full surplus extraction regardless of the valuations of the agents. While the key condition identified by Crémer and McLean (1988) is still sufficient in this setting, we show that such condition could be relaxed and a more general characterization is feasible. Our characterization highlights the importance of focusing on the set of potential imitators or inverse consideration sets instead of the considerations sets directly.

Our result suggests an alternative way to look at incentive compatibility problems: instead of looking at the potential deviations for a type, look at the types that could deviate to a particular type. While this interpretation could also be applied under a full-consideration model, having restricted consideration sets highlight the importance of the inverse consideration sets over the direct consideration sets. This importance remains hidden in the traditional model due to the "symmetry" between consideration and inverse consideration sets in the case of full-consideration.

While treating consideration sets as exogenous is a strong assumption, it allows us to provide a full characterization of the conditions to guarantee full surplus extraction. Fully endogenizing the consideration sets is a natural and appealing future path of research that could provide a broader view of the problem of surplus extraction in mechanism design settings with correlation.

## References

ABALUCK, JASON AND ABI ADAMS-PRASSL (2021) "What do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses*," *The Quar-*

*terly Journal of Economics*, 136 (3), 1611–1663, 10.1093/qje/qjab008. [22]

Abaluck, Jason and Jonathan Gruber (2022) "When Less is More: Improving Choices in Health Insurance Markets," *The Review of Economic Studies*, 10.1093/restud/rdac050, rdac050. [22]

Börgers, Tilman (2015) *An introduction to the theory of mechanism design*, New York, NY: Oxford University Press. [5]

Bull, Jesse and Joel Watson (2007) "Hard evidence and mechanism design," *Games and Economic Behavior*, 58 (1), 75–93, https://doi.org/10.1016/j.geb.2006.03.003. [3]

Caplin, Andrew, Mark Dean, and John Leahy (2018) "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *The Review of Economic Studies*, 86 (3), 1061–1094, 10.1093/restud/rdy037. [2], [3]

Chen, Yi-Chun and Siyang Xiong (2011) "The genericity of beliefs-determine-preferences models revisited," *Journal of Economic Theory*, 146 (2), 751–761, https://doi.org/10.1016/j.jet.2010.12.005. [12]

de Clippel, Geoffroy (2014) "Behavioral Implementation," *American Economic Review*, 104 (10), 2975–3002, 10.1257/aer.104.10.2975. [2]

de Clippel, Geoffroy, Rene Saran, and Roberto Serrano (2018) "Level-$k$ Mechanism Design," *The Review of Economic Studies*, 86 (3), 1207–1227, 10.1093/restud/rdy031. [2]

Crawford, Vincent P. (2021) "Efficient mechanisms for level-k bilateral trading," *Games and Economic Behavior*, 127, 80–101, https://doi.org/10.1016/j.geb.2021.02.005. [4]

Crémer, Jacques and Richard P. McLean (1985) "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent," *Econometrica*, 53 (2), 345–361, `http://www.jstor.org/stable/1911240`. [8]

——— (1988) "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56 (6), 1247–1257, `http://www.jstor.org/stable/1913096`. [1], [6], [8], [12], [21], [26] Eliaz, Kfir (2002) "Fault Tolerant Implementation," *The Review of Economic Studies*, 69 (3), 589–610, 10.1111/1467-937X.t01-1-00023. [2]

Eliaz, Kfir and Ran Spiegler (2011) "Consideration Sets and Competitive Marketing," *The Review of Economic Studies*, 78 (1), 235–262, 10.1093/restud/rdq016. [2]

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. [1]

Fershtman, Daniel and Alessandro Pavan (2022) "Searching for "Arms": Experimentation with Endogenous Consideration Sets." [2], [3]

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. [2]

Green, Jerry R. and Jean-Jacques Laffont (1986) "Partially Verifiable Information and Mechanism Design," *The Review of Economic Studies*, 53 (3), 447–456, 10.2307/2297639. [3], [4]

Honka, Elisabeth, Ali Hortaçsu, and Matthijs Wildenbeest (2019) "Chapter 4 - Empirical search and consideration sets," in Dubé, Jean-Pierre and Peter E. Rossi eds. *Handbook of the Economics of Marketing, Volume 1*, 1 of Handbook of the Economics of Marketing, 193–257: North-Holland, https://doi.org/10.1016/bs.hem.2019.05.002. [2]

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. [2]

Lopomo, Giuseppe, Luca Rigotti, and Chris Shannon (2020) "Detectability, Duality, and Surplus Extraction." [6]

——— (2022) "Uncertainty and robustness of surplus extraction," *Journal of Economic Theory*, 199, 105088, https://doi.org/10.1016/j.jet.2020.105088, Symposium Issue on Ambiguity, Robustness, and Model Uncertainty. [1], [2]

Manzini, Paola and Marco Mariotti (2014) "Stochastic Choice and Consideration Sets," *Econometrica*, 82 (3), 1153–1176, https://doi.org/10.3982/ECTA10575. [2]

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], [4], [6], [12]

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

Ollár, Mariann and Antonio Penta (2017) "Full Implementation and Belief Restrictions," *American Economic Review*, 107 (8), 2243–77, 10.1257/aer.20151462. [4]

——— (2023) "A Network Solution to Robust Implementation: The Case of Identical but Unknown Distributions," *The Review of Economic Studies*, 90 (5), 2517–2554, 10.1093/restud/rdac084. [4]

Reuter, Marco (2022) "Revenue Maximization with Partially Verifiable Information." [3]

Saran, Rene (2011) "Bilateral trading with naive traders," *Games and Economic Behavior*, 72 (2), 544 – 557, https://doi.org/10.1016/j.geb.2010.09.009. [2]

Severinov, Sergei and Raymond Deneckere (2006) "Screening when some agents are nonstrategic: does a monopoly need to exclude?" *The RAND Journal of Economics*, 37 (4), 816–840, 10.1111/j.1756-2171.2006.tb00059.x. [2]

Strausz, Roland (2017) "Mechanism Design with Partially Verifiable Information." [3]