Betting on Others' Bets: Unions of Surplus Extraction Mechanisms *

Alia Gizatulina†

This version: March 2016

Abstract

We generalize the classical Crémer-McLean mechanism to allow i's participation fee to depend not only on the valuations reported by $-i$ at the auction stage, but also on the choice of the participation fee by $-i$ at the first stage. Such construction allows to exploit the convex hull property of beliefs whenever it appears in beliefs about beliefs rather than in beliefs about preferences. As the mechanism extracts each agent's entire hierarchy of beliefs, it reveals what is common knowledge among them. Hence for any given finite or countable collection of type spaces $\{\mathcal{T}^k\}_{k\in K}$ each $\mathcal{T}^k$ verifying the convex hull property *within itself*, the designer can propose a union of GCM mechanisms and extract the surplus *across type spaces*, regardless of absence of knowledge by the designer which type space agents share (and without relying on the shoot-the-liar mechanism). We discuss when the technique of using a union of individual mechanisms is extendible to more general cases.

*Preliminary and highly incomplete! For updates of the draft, visit this link. The initial draft of the paper was titled: "Betting on Others' Bets: The Generalized Crémer-McLean Mechanism". This research would not be possible without years of discussion with Martin Hellwig about the subject, in particular about the role of the designer's uncertainty for the question of surplus extraction. I am grateful for questions and comments to the audiences of workshops at the MPI, PSE and University of Cologne.

†University of St. Gallen and Max Planck Institute, Bonn. Email: alia.gizatulina@unisg.ch

1 Introduction

In this paper, we first constructs an extension of the classical Crémer-McLean mechanism that allows to exploit the convex hull property of beliefs whenever this property appears at higher orders rather than at the first order – as it is required for the original construction of Crémer and McLean (1988) to work. Next, we show that because the mechanism induces agents to bet essentially on what is common knowledge among them, it is immune to the uncertainty of the designer about what is actually common knowledge among agents. Particularly, the designer can just use a union of the generalized Crémer-McLean mechanisms where elements of the union are optimal mechanisms each corresponding to a separate type space of agents that the designer perceives as possible. Hence, a union of "local" mechanisms can be assembled in one "global" mechanism to extract surplus at a collection of type spaces considered as possible by the designer.

In a quasi-linear environment, provided agents' beliefs about others' payoff parameters verify a certain condition, it is possible to implement any efficient decision rule via the lottery mechanism constructed in Crémer and McLean (1988). The classical Crémer-McLean mechanism (CCM) could be interpreted as a Vickery-Clarke-Groves (VCG) mechanism augmented by a menu of participation fees that each agent is obliged to choose if wants to access the VCG mechanism (this interpretation is given, e.g., in McAfee and Reny (1992)). Each participation fee in this menu is a pre-commitment to the payment which depends on the reports of $-i$ during the stage of the VCG mechanism.

The condition on agents' beliefs that allows the mechanism to implement the first-best outcome is the so-called convex-hull property. It requires that beliefs of no type t_i can be represented as a convex combination of beliefs of any other types $t_i' \neq t_i$. When agent i's type space verifies this property two things become possible. First, there exists a menu of participation fee schedules such that each type t_i of each agent i has a unique preferred element in this menu that minimizes his expected payment to the designer.[1] Second, the convex hull property implies invertibility of the

[1] When agents' beliefs verify a slightly stronger requirement of linear independence, the

mapping from types to beliefs, each element of the menu can be bundled with an extra constant equal to the expected payoff by type t_i from the VCG mechanism.

Hence, the key element of the classical CM mechanisms is the property of agents' beliefs about others' information revealed during the VCG game. The behaviour during the VCG game reveals, however, only other agents' payoff characteristics and hence when i evaluates which participation fee is the best (the cheapest), only i's first-order beliefs, i.e. beliefs about others' payoffs are involved. This implies, that only if the first order beliefs verify the convex hull property the CCM can bring the surplus extraction result.

In type spaces where agents' information contain more signals than just their own payoff parameters, beliefs have a richer structure. In particular, agents' beliefs may fail to verify the convex hull property at the first order but verify it at higher orders, at the level of beliefs about beliefs. In this case, the classical CM mechanism would not provide a satisfactory answer on the possibility to extract agents' surplus.

That is why, as the first step in this paper we enrich the CM mechanism to allow to extract agents' surplus even if the convex-hull property appears at higher-order beliefs. The main idea of the mechanism is to extend the set of random variables that the participation fee is conditioned upon. In particular, each agent's participation fee is let to depend not only on the information revealed by others at the second stage, but also on the choices of participation fees made by others at the first stage (which at the moment of choice of his own participation fee agent i does not observe). The participation fee payment given this choice is defined as a sum of lotteries conditional on beliefs retrieved from the choices of participation fee schedules.

The first result we obtain is that the generalized Crémer-McLean mechanism can extract agents' full surplus if and only if each agent's belief hierarchy verifies the convex hull property either at some order k or across orders (relative to agents' payoff parameters). The definition of the convex hull property within order k is rather straightforward – belief of level k of

menu of fees reduces to a unique element, i.e. agents do not make any choice among lotteries but are imposed a lottery.

no payoff type θ_i^0 about beliefs of order $k-1$ of other agents can be represented as a convex combination of beliefs of some other payoff types in Θ_i. The convex-hull property across orders is defined in terms of beliefs of some level k of payoff type θ_i^0 about the elements of the product space of beliefs of lower orders.

When the convex hull property holds either at some order of beliefs or across orders, the mechanism of betting on others' bets makes each agent reveal his entire hierarchy of beliefs truthfully in a Bayesian Nash Equilibrium of the surplus extraction game.

A by-product of the GCM mechanism is that it can help to deal with a certain type of designer's uncertainty about agents' beliefs hierarchies. Specifically, the designer may be uncertain which out of finitely or countably many type spaces is the right one. From the perspective of the designer, each type space in the collection has a non-zero probability to be the actual type space, yet he does not know it exactly. We show that provided that each type space in the collection verifies the convex hull property (within or across orders of beliefs, then a union of individual GCM mechanisms extract the full surplus on the entire collection.

Hence, in the second part of this paper we go in the direction of the robust mechanism design. The question of detail-free implementation has been studied extensively in the literature, following the pioneering work of Bergemann and Morris (2005). Most of the subsequent literature has been investigating implementability of social choice functions/correspondences without imposing *any* restriction on the underlying space of beliefs. This program is on another extreme of the classical approach where agents are assumed to share a common prior over the single-dimensional space of payoff parameters. In such settings all differences in beliefs are driven uniquely by the difference in payoff signals. Furthermore, the prior and hence the space of interim beliefs are commonly known not only among agents but also by the designer.

In this paper we follow the program which is in-between of the two extreme cases. Namely, we assume that agents' type space is relatively small, and not equal to the entire universal type space. Yet the designer doesn't know which type space, out of countably many possible type spaces, agents

share. Thus, in principle, there is a scope for agents to fool the designer collectively by pretending, whenever it is individually rational for each agent, to be from a completely different type space.[2] However, as we show, the generalized Crémer-McLean mechanism controls for such incentives – agents can be offered to choose a participation fee from a union of generalized Crémer-McLean mechanisms – for each agent it is optimal to choose the participation fee designer for his type from his actual type space.

2 Setup

2.1 Type Spaces

We assume that there is a finite set of agents indexed by $i = 1, .., I$. Each agent i's private information is encoded in his type $t_i \in T_i$, where T_i is the set of all possible types that agent i can have. Each agent's type t_i includes information about his payoff type, as given by function

$$\hat{\theta}_i : \ T_i \to \Theta_i \tag{1}$$

and so $\hat{\theta}_i(t_i)$ is agent i's payoff type when his type is t_i. A type of agent i also includes a decritpoion of his beliefs about the types of the other agents, as given by function

$$\hat{\pi}_i : \ T_i \to \Delta(T_i) \tag{2}$$

where $\Delta(X)$ stands for a space of probability distributions on X and so $\hat{\pi}_i(t_i)$ corresponds to agent i's belief type when his type is t_i. Thus, $\hat{\pi}_i(t_i)[E]$ is the probability that type t_i of agent i assigns to other agents' types t_{-i} being an element of $E \subseteq T_{-i}$.

Note that such general representation allows for many further specifications, e.g. each belief mapping $\hat{\pi}_i(.)$ can be derived according to the Bayes rule from a common prior over T. Or each agent's type space T_i is just a product space of a set of payoff types and payoff-irrelevant (but strategically relevant) signals, i.e. $T_i = \Theta_i \times S_i$. We will refer to any collection

[2] Note, we do not consider pre-meditated collusive deviations.

$$\mathcal{T}^{int} = \{T_i, \hat{\pi}_i(.), \hat{\theta}_i(.)\}_{i \in I} \tag{3}$$

as agents' *interim type space*. While

$$\mathcal{T}^{ea} = \{T_i, \nu_i\}_{i \in I} \tag{4}$$

denotes the corresponding *ex ante* representation of agents' type space, such for each i $\hat{\pi}$ is the regular conditional distribution derived from prior ν_i. If agents share a common prior, $\nu_i = \nu_j$ for any $i, j \in I$.

Either representation, i.e., $\mathcal{T}^{int}$ or $\mathcal{T}^{ea}$, provides a full description of the smallest set of the states of nature that are commonly known.[3]

As usual, we can unfold any $\mathcal{T}^{int}$ into the Θ-based universal type space of belief hierarchies. That is, given his prior ν_i, priors of other agents ν_{-i} and the information contained in t_i each agent i derives beliefs about other agents' payoffs, about their beliefs about everyone else's payoff, etc. up to infinity. Each order of beliefs is a space of probability measures on its own:

$$H_i^0 := \Theta_i$$

$$\begin{aligned} H_i^1 &:= H_i^0 \times \Delta_i(H_{-i}^0) \\ H_i^2 &:= H_1^1 \times \Delta_i(H_{-i}^1) \\ &= \Theta_1^0 \times \Delta_i(\Theta_{-i}^0) \times \Delta_i \Delta_{-i}(\Theta_j^0) \\ &\dots \\ H_i^k &:= H_i^{k-1} \times \Delta_i(H_{-i}^{k-1}) \end{aligned}$$

A generic element of the belief hierarchy space is $h_i^\infty \in H_i^\infty \equiv H_i^1 \times H_i^2 \times \dots \times H_i^k \times \dots$ with

$$h_i^\infty = (h_i^1, h_i^2, h_i^3, ..., h_i^k, ..) \tag{5}$$

[3] By its very definition, the agent's type space is the smallest set of states of nature (as given by beliefs and payoffs) that is common knowledge among agents.

In addition, instead of dealing with beliefs about the product space of payoffs and beliefs, we will consider the *marginalized* hierarchies of beliefs[4]:

$$\pi_i^0 := \theta_i \text{ with } \theta_i \in \Pi_0$$

$$\pi_i^1 := marg_{\Theta_{-i}} h_i^1 \text{ with } \pi_i^1 \in \Pi_i^1$$

$$\pi_i^2 := marg_{\Pi_{-i}^1} h_i^2 \text{ with } \pi_i^2 \in \Pi_i^2$$

$$...$$

$$\pi_i^k := marg_{\Pi_{-i}^{k-1}} h_i^{k-1} \text{ with } \pi_i^k \in \Pi_i^k$$

Hence, a generic element of the space of hierarchies of marginal beliefs is $\pi_i^\infty \in \Pi_i^1 \times \Pi_i^2 \times ... \times \Pi_i^k \times ...$, with

$$\pi_i^\infty = (\pi^0, \pi_i^1, \pi_i^2, \pi_i^3, ..., \pi_i^k, ..) \tag{6}$$

To any type $t_i \in T_i$, given the type space $\mathcal{T}^{int} = \{T_i, \hat{\theta}_i(.), \hat{\pi}_i(.)\}_{i \in I}$ there corresponds a belief hierarchy $h_i^\infty = h_i^\infty(t_i)$, that is

$$h_i(t_i) = (h_i^0(t_i), h_i^1(t_i), h_i^2(t_i), .., h_i^k(t_i), ...)$$

and its marginalized version $\pi_i^\infty = \pi_i^\infty(t_i)$ with

$$\pi_i(t_i) = (\pi_i^0(t_i), \pi_i^1(t_i), \pi_i^2(t_i), .., \pi_i^k(t_i), ...)$$

.

Vice versa, any explicit representation of belief hierarchies can be implicitly represented as a triple of elements like $\mathcal{T}^{int}$ (cf. Mertens and Zamir (1985) and Brandenburger and Dekel (1993)). [5]

[4]Such restricted version is sufficient for settings where each player behaviour depends only on their expectations about other players average payoff, others' expectation about everyone else's average payoff, etc. See e.g. Morris and Shin (2002). Crémer-McLean is one of such settings.

[5]For a detailed discussion see Bergemann and Morris (2005) and Heifetz and Samet (1998))

2.2 The Classical Crémer-McLean Mechanism

To set the stage for the analysis that follows, in this subsection we briefly overview the main details of the classical, first-order Crémer-McLean mechanism (following its interpretation given in McAfee and Reny (1992)). We also provide an example of a rich type space where the first- and second-order beliefs fail to verify the convex-hull property, while all the remaining higher-order beliefs do verify it. Yet, as the classical CM mechanism is based on the first-order beliefs, it would fail to extract the surplus in this environment.

Following McAfee and Reny (1992) one can consider the following version of the classical Crémer-McLean mechanism, which is strategically equivalent to the direct mechanism of Crémer and McLean (1988). The game is two-stage, at the second stage agents play a VCG mechanism, while at the first stage each agent has to choose from a menu of participation fees $\{\tau_{t_i}\}_{t_i \in T_i}$ where

$$\tau_{t_i} = s_{t_i} + \gamma x_{t_i}(\tilde{\theta}_{-i}) \tag{7}$$

and where $x_{t_i} : \tilde{\Theta}_{-i} \to \Re$ corresponds to a "lottery", i.e. it returns a real number for every announcement by $-i$ at the second stage and this number is the participation fee that the agent is committed to pay to the designer ex post.

The main pitch of the mechanism is that if (and only if) agent i's beliefs verify the convex-hull property, namely for any $t_i \in T_i$ it holds

$$\pi_i(\theta_{-i} \mid t_i) \notin co\{\pi_i(\theta_{-i} \mid t_i')\}_{t_i' \in T_i}$$

the mechanism designer can construct a collection of functions $\{x_{t_i}(.)\}_{t_i \in T_i}$ such that

$$\sum x_{t_i}(\theta_{-i})\pi_i(\theta_{-i} \mid t_i) = 0$$

and

$$\sum x_{t_i'}(\theta_{-i})\pi_i(\theta_{-i} \mid t_i) < 0$$

for $t_i \neq t_i'$

The remaining elements s_{t_i} and γ in (7) are defined as follows. The constant s_{t_i} is equal to the surplus from the participation in the VCG mechanism as expected by type t_i. The multiplier γ is an arbitrary but sufficiently

high number, whose role is to make any incentive to select $\tau(.)$ designed for some other $t_i' \neq t_i$ to be unattractive due to the rise in expected losses from the accompanying misselection of $x_{t_i'}(.)$. As the expected surplus from the participating in the VCG mechanism is finite, while γ can be arbitrary high, Bayesian incentive compatibility of the mechanism follows.

We now provide an example of an abstract type space with 2 agents where the convex hull property fails at the first-order beliefs (i.e. beliefs given by $\pi_i^1(t_i)$ about θ_{-i}) and also at the second-order (beliefs given by $\pi_i^2(t_i)$) because in both cases, the BDP property fails. Yet, at the third-order, agents' beliefs verify the convex-hull, and a fortiori, the BDP property.

EXAMPLE: There are two agents, fully symmetric. Each agent has 4 types in $T_i = \{\theta^1, \theta^2\} \times \{s^1, s^2\}$, where θ_i with $\theta_i^1 < \theta_i^2$ is a payoff parameter, while s_i with $s_1 \neq s_2$ is a payoff-irrelevant but yet informative signal. The belief mapping is given in the following matrix:

	$t_j^1 = (\theta^1, s^1)$	$t_j^2 = (\theta^1, s^2)$	$t_j^3 = (\theta^2, s^1)$	$t_j^4 = (\theta^2, s^2)$
$\pi_i(. \mid \theta_1, s_1)$	1/3	1/3	1/6	1/6
$\pi_i(. \mid \theta_1, s_2)$	1/10	1/10	3/10	5/10
$\pi_i(. \mid \theta_2, s_1)$	$(2/3) - \epsilon$	ϵ	$(1/3) - \epsilon$	ϵ
$\pi_i(. \mid \theta_2, s_2)$	1/6	1/3	1/3	1/6

Table 1: Belief mapping $\hat{\pi}_i : T_i \rightarrow \Delta(T_{-i})$

This abstract type space maps into the following marginalized first- and second-order beliefs, as defined in (6):

In table 2 one can see that the 1st order beliefs of types 1 and 3 are indistinguishable (in particular in any mechanism dependent only on first-order beliefs they cannot be separated from each other). The BDP property fails at the 1st order beliefs in this type space, as $\theta(t_1) \neq \theta(t_3)$. As there

	$\theta^1_j(t^1_j, t^2_j)$	$\theta^2_j(t^3_j, t^4_j)$
$\pi^1_i(t^1_i)$	2/3	1/3
$\pi^1_i(t^2_i)$	1/5	4/5
$\pi^1_i(t^3_i)$	2/3	1/3
$\pi^1_i(t^4_i)$	1/2	1/2

Table 2: 1st-order beliefs

	$\tilde{\pi}^1_j(t^1_j, t^3_j)$	$\pi^1_j(t^2_j)$	$\pi^1_j(t^4_j)$
$\pi^2_i(t^1_i)$	1/2	1/3	1/6
$\pi^2_i(t^2_i)$	4/10	1/10	5/10
$\pi^2_i(t^3_i)$	$1-2\epsilon$	ϵ	ϵ
$\pi^2_i(t^4_i)$	1/2	1/3	1/6

Table 3: 2nd-order beliefs

only 3 distinct belief types of each i at the first order, we can define an equivalence class relative to original types, namely define $\tilde{\pi}^1$, with $\tilde{\pi}^1 := \pi^1_1 = \pi^1_3$.

Table 3 shows that at the 2nd order, e.g. beliefs of t^i_1 and t^i_4 about 1st order beliefs of j are again the same and so BDP fails again (as $\theta(t_1) \neq \theta(t_4)$). Again we can define a corresponding equivalence class $\tilde{\pi}^2 := \pi^2_1 = \pi^2_4$

How about the 3rd order beliefs? As the next table shows, they happen to verify both the convex hull property and the BDP:

	$\tilde{\pi}^2_i(t^1_j, t^4_j)$	$\pi^2_i(t^2_j)$	$\pi^2_i(t^3_j)$
$\pi^3_i(t^1_i)$	1/2	1/3	1/6
$\pi^3_i(t^2_i)$	6/10	1/10	3/10
$\pi^3_i(t^3_i)$	2/3	ϵ	$(1/3)-\epsilon$
$\pi^3_i(t^4_i)$	1/3	1/3	1/3

Table 4: 3rd order beliefs

In this type space, if the designer was to use the classical Crémer-McLean mechanism, he would not be able to distinguish from the choices of lotteries that are designed for types t_1 and t_3 whether the given choice has been made by an agent with θ_1 or with θ_2. Consequently, the designer cannot fine-tune in an incentive compatible and individually rational way the corresponding $s_{t_i(\theta_i)}$ of the participation fee.

We also note that the non-marginalized 2nd order beliefs, as given by (5), i.e. beliefs about the joint realization of payoffs and 1st-orders of beliefs, verify both the CC and the BDP properties:

types	$\theta_1, \hat{\pi}^1$	θ_1, π_2^1	θ_1, π_4^1	$\theta_2, \hat{\pi}^1$	θ_2, π_2^1	θ_2, π_4^1
$h_i^2(t_i^1)$	1/3	1/3	0	1/6	0	1/6
$h_i^2(t_i^2)$	1/10	1/10	0	3/10	0	5/10
$h_i^2(t_i^3)$	(2/3)-ϵ	ϵ	0	(1/3)-ϵ	0	ϵ
$h_i^2(t_i^4)$	1/6	1/3	0	1/3	0	1/6

Table 5: Beliefs of agent i about $\Theta_j \times \Delta_j(\Theta_i)$

3 Results

3.1 The Generalized Crémer-McLean Mechanism

We start with some definitions.

Definition *The set of agent i's belief hierarchies verifies the convex hull property **within** order k if for no $t_i \in T_i$ the following holds*

$$\pi^k(t_i) \in co\{\pi^k(t_i')\}_{t_i' \in T_i}$$

In the following, we will refer to this property as "the within-order CC property".[6] In the example from the previous subsection the CC property is failing for $k = 1$ and $k = 2$, while it holds for $k = 3$.

Definition *The set of agent i's belief hierarchies verifies the convex hull property **across** orders up to k, if for no $t_i \in T_i$ the following holds*

$$h^k(t_i) \in co\{h^k(t_i')\}_{t_i' \in T_i}$$

with $t_i' \neq t_i$.

Similarly, we will refer to this property occasionally as "the across-orders CC property". In the example, this property appears for $k = 2$.

Now everything is ready to present the generalized version of the Crémer-McLean mechanism:

[6] CC stands for "convex combination".

Stage 1: Each agent $i \in I$ is proposed a menu of participation fees $\{\tau_i(t_i)\}_{t_i \in T_i}$ with:

$$\tau_i(t_i) = s_{t_i} + \gamma_{\theta_i} x^1_{t_i}(\theta_{-i}) + \gamma_\tau x^\tau_{t_i}(\tau_{-i})$$

with $s_{t_i} \in \mathbb{R}_-$, $\gamma_{\theta_i}, \gamma_\tau \in \mathbb{R}$, $x^1_{t_i}(\theta_{-i}) : \tilde{\Theta}_{-i} \to \mathbb{R}$, $x^\tau_{t_i}(\tau_{-i}) : L_{-i} \to \mathbb{R}$ with $\tilde{\Theta}_{-i}$ equal to the set of reported valuations by $-i$ at the VCG mechanism of the second stage and $\tilde{L}_i$ corresponds to the set of lotteries chosen by $-i$ at the first stage.

Stage 2: Agents play a (generalized) VCG mechanism, defined by vector of payments

$$p = (p_1(t), .., p_I(t))$$

and allocation

$$q = (q_1(t), .., q_I(t)).$$

The triple $\Lambda(\mathcal{T}) = (\tau(.), p(.), q(.))$ defines the generalized Crémer-McLean mechanism for a given type space $\mathcal{T}$.

There is one observation to be made. The main allocation mechanism does not have to be the VCG or any other dominant strategy mechanism. It can be an arbitrary mechanism which distinguishes players not only according to their payoff parameters (as it happens in the VCG), but also according to their high order beliefs and so the information revealed by each j during the main mechanism is $\bar{\pi}^\ell_j = (\theta_j, \pi^1_j, \pi^2_j, .., \pi^\ell_j)$ for some order l and not just θ_{-i}. In this case the participation fee of i changes to

$$\tau_i(t_i) = s_{t_i} + \gamma_{\pi^\ell_i} x^\ell_{t_i}(\pi^\ell_{-i}) + \gamma_\tau x^\tau_{t_i}(\tau_{-i})$$

i.e the first is the bet on reports in the main mechanism while the second is the bet on the information revealed through the choices of bets by others.

Proposition 3.1 *The full surplus extraction game with a menu of generalized participation fees admits a Bayesian-Nash Equilibrium with the full surplus extraction if and only if for each agent i the set of hierarchies of beliefs verifies either the within-order CC property or the across-orders CC property.*

For the proof we need two, rather straightforward, lemmas outlined just below. As a preparation to meet those lemma, observe that given the regular conditional distribution function $\pi(.) : Z \to \Delta(X)$ we can divide Z into two subsets, Z_π^{ext} and Z_π^{int} defined as follows:

$$Z_\pi^{ext} := \{z \in Z : \pi(. \mid z) \notin co\{\pi(. \mid z')\},\ z' \neq z,\ z' \in Z\}$$

and

$$Z_\pi^{int} := Z \setminus Z^{ext} := \{z \in Z : \pi(. \mid z) \in co\{\pi(. \mid z')\},\ z' \neq z,\ z' \in Z_\pi^{ext}\}.$$

That is, given the belief function $\pi(.)$, we divide Z in two sets. First, a set of "exterior elements under $\pi(.)$", denoted as Z_π^{ext}, where for no $z \in Z_\pi^{ext}$ the conditional measure of X given $z \in Z_\pi^{ext}$ can be represented as a convex combination of conditional measures given some other elements $z' \in Z$. Second, the complementary set $Z_\pi^{int} = Z \setminus Z_\pi^{ext}$ where conditional measure of X given $z \in Z_\pi^{int}$ is a convex combination of conditional measures from the set $\{\pi(. \mid z')\}$ with $z' \in Z_\pi^{ext}$.

Lemma 3.2 *For any $z \in Z_\pi^{ext}$ there exists a profile of functions $\{f_z(.)\}_{z \in Z_\pi^{ext}}$ such that for any $m \geq 0$*

$$\sum_{x \in X} f_z(x)\, \pi(x \mid z) = m$$

and

$$\sum_{x \in X} f_{z'}(x)\, \pi(x \mid z) < 0$$

for $z' \neq z$ with $z' \in Z$.

The proof is the consequence of the Farkas lemma, this lemma has been used in the original Crémer-McLean (1988) construction for $m = 0$.

The second lemma is more subtle and its result is crucial for what follows. It classifies the set of functions that allow to separate elements in Z_π^{int} from any other element of Z, be it Z_π^{int} or Z_π^{ext} .

Lemma 3.3 *For any $\hat{z} \in Z_\pi^{int}$ there exists a profile of functions $\{x_{\hat{z}}(.)\}_{\hat{z} \in Z_\pi^{int}}$ such that*

$$\sum_{x \in X} f_{\hat{z}}(x)\, \pi(x \mid \hat{z}) < 0$$

and

$$\sum_{x \in X} f_{z'}(x)\,\pi(x \mid \hat{z}) < \sum_{x \in X} f_{\hat{z}}(x)\pi(x \mid \hat{z})$$

for any $z' \in Z$ with $z' \neq \hat{z}$.

In words this means that for types having beliefs in a convex hull of beliefs of other types, there exists nonetheless a profile of lottery functions such that the choice of one of those functions by type t_i, for which it was designed, guarantees to this type the best of available payoffs from the menu of lotteries. This result is due to guaranteed existence of *a proper scoring rule* for finite probability measures (and it is omitted, see, e.g. Gneiting and Raftery (2007) and Selten (1998)).

We now ready for the proof of Proposition 3.1.

Proof We start with the sufficiency of the within-order CC property.

"IF" (WITHIN ORDER CC)

We need to show that there exist a profile of fee schedules $\{\tau_i(t_i)\}_{t_i \in T_i}$:

$$\tau_i(t_i) = s_{t_i} + \gamma_{\theta_i} x^1_{t_i}(\theta_{-i}) + \gamma_\lambda x^\lambda_{t_i}(\lambda_{-i})$$

inducing each type to pick up his $\tau_i(t_i)$. The key observation is that a given choice of $\tau_i(.)$ reveals the entire hierarchy of marginalized beliefs of i about Θ. This implies that the designer can define the payment be equal to the following scheme, based on the *inferred higher-order beliefs*

$$\tau_i(t_i) = s_{t_i} + \gamma_1 x^1_{t_i}(\theta_{-i}) + \gamma_2 x^2_{t_i}(\pi^1_{-i}) + ... + \gamma_K x^K_{t_i}(\pi^{K-1}_{-i}) \tag{8}$$

(with K is finite because types are finite).

We start with the sufficiency of the within-order CC property. Suppose w.l.o.g this property appears at order $k' \leq K$ and for all orders k with $k < k'$ either CC fails but BDP holds or both fail.

We proceed by induction. Following the hypothesis suppose the CC property fails at the 1st order (but BDP holds) for some t_i, i.e. consider $\hat{t}_i$ in $\hat{T}_i$. Suppose also for the moment that the lottery payment depends only on retrieved 1st order beliefs.

By Lemma 3.3 there exists a profile of functions $\{x^1_{\hat{t}_i}(\theta_{-i})\}$ such that

$$\sum x^1_{\hat{t}_i}(\theta_{-i})\pi_i(\theta_{-i} \mid \hat{t}_i) < 0 \tag{9}$$

and

$$\sum x_{t'_i}(\theta_{-i})\pi_i(\theta_{-i} \mid \hat{t}_i) < \sum x_{\hat{t}_i}(\theta_{-i})\pi_i(\theta_{-i} \mid \hat{t}_i)$$

for all $t'_i \in T'_i$ with $t'_i \neq \hat{t}_i$.

That is, for each $\hat{t}_i \in \hat{T}_i$ first-order beliefs $\pi^1(\hat{t}_i)$ there exists a unique lottery $x^1_{\hat{t}_i}(\theta_{-i})$ with the highest expected value, even if this value is negative.

Suppose now that also BDP fails at 1st order (hence CC fails as well). For an equivalence class of types $\tilde{t}^1_i$ defined as

$$\tilde{t} = \{t, t' \in T^1 : \pi^1(. \mid t) = \pi^1(. \mid t')\}$$

such that

$$\pi(\theta_{-i} \mid \tilde{t}_i) \notin co\{\pi(\theta_{-i} \mid t''_i\}$$

with $t''_i \neq \tilde{t}$ we can construct a menu of functions with

$$\sum x^1_{\tilde{t}_i}(\theta_{-i})\pi_i(\theta_{-i} \mid \tilde{t}_i) = 0$$

and

$$\sum x^1_{t'_i}(\theta_{-i})\pi_i(\theta_{-i} \mid \tilde{t}_i) < 0$$

for $\tilde{t}_i \neq t'_i$ where $\tilde{t}_i$ is an equivalence class of (first order) types.

Otherwise, if $\pi(\theta_{-i} \mid \tilde{t}_i) \in co\{\pi(\theta_{-i} \mid t'_i\}$, i.e. for a set of types failing BDP and defined by equivalence class $\hat{t}_i$ we have also failure of the CC property, i.e. $\pi(\hat{t}_i)$ can be represented as a convex combination of first-order beliefs of some t'_i we construct the lottery $x_{\hat{t}_i}(.)$ for $\tilde{t}_i$ as in Lemma 3.3.

Finally, for first-order types that are BDP and verify the CC property, we construct the menu of lotteries $\{x_{t_i}(.)\}$ as in Lemma 3.2.

Choosing a lottery among $\{x^1_{t_i}(\theta_{-i})\}_{t^1_i \in T^1_i}$ according to the truthful strategy is a Bayesian Nash equilibrium, so revealed beliefs about θ_{-i} would correspond to the actual distribution of first-order beliefs (recall that for

the moment we do not consider choices at the higher orders). Now, consider the second order beliefs. Anticipating that $-i$ are truthfully revealing their first order beliefs, agent i, when choosing among $\{x_{\pi_i^2}(\pi_{-i})\}$, picks $x_{\pi^2}(.)$ that minimizes the expected payment given his second order beliefs, i.e. beliefs about t_i^k following the same logic as we outlined above for his choice of the lotteries on θ_{-i} given i's first-order beliefs.

By hypothesis, at order k the CC property (and so BDP) holds for a type t_i (note that different t_i can have different orders at which the CC property appears). Following Lemma 3.2 we construct a menu of lotteries $\{x_{t_i}(t_{-i}^{k-1})\}$ such that

$$\sum x_{t_i}(t_{-i}^{k-1})\pi_i^k(t_{-i}^{k-1} \mid t_i) = m \geq 0$$

with m to be specified shortly and

$$\sum x_{t_i'}(t_{-i}^{k-1})\pi_i^k(t_{-i}^{k-1} \mid t_i) < 0$$

for $t_i' \neq t_i$.

So in this way we can construct a sequence of menu of lotteries

$$\{x_{t_i}^1(\theta_{-i})\}_{t_i^1 \in T_i^1}, \{x_{t_i}^2(t_{-i}^1)\}_{t_i^2 \in T_i^2}, .., \{x_{t_i}^K(t_{-i}^{K-1})\}_{t_i^K \in T_i^K}$$

such that each consecutive sequence is a menu of lotteries on choices for the previous order of lotteries. For order k where both BDP and CC holds design $x_{t_i}^k(.)$ with the expected value $m > 0$ enough to compensate negative payments at others orders.

As for each order k', for each type the truthful selection minimizes the expectation over $x_{t_i}^{k'}(.)$, the sum of "minimal" components minimizes the fee as well. Hence, as for each t_i there is a unique vector $(\pi^0, \pi^1, ..., \pi^k, ..)$ and so there is a unique lottery, it means that we can identify the choice of iterative lottery as in (8) with the initial "fixed point" lottery:

$$\tau_i(t_i) = s_{t_i} + \gamma_{\theta_i} x_{t_i}^1(\theta_{-i}) + \gamma_\tau x_{t_i}^\tau(\tau_{-i})$$

With an appropriate choice of all γ, e.g., $\gamma_{\theta_i} = \gamma_1$ and $\gamma_\tau = \gamma_2 = .. = \gamma_K$.

Finally, let $s_{t_i} = EU^{VCG}(\theta_i, \pi_i^1(.))$.

"IF" (ACROSS ORDERS CC) Now suppose the CC property holds across orders, for each type t_i there exists an order $k'(t_i)$ such that, while either CC or BDP fails for all orders $k < k'(t_i)$ separately, considered as a belief over the product space of lower order types $h_i^k(.) \in T_i^k$ verifies the CC property, i.e. it cannot be represented as a convex combination of $h_i^k(t')$ of some other types.

"ONLY IF" PART If CC fails both within each order and also across orders, the full surplus extraction result is impossible due to the argument similar to the only if part of the proof of Theorem 2 in Crémer and McLean (1988). ■

Next we expand the message space of agents to obtain truthful reporting as the unique dominance-solvable Bayesian Nash Equilibrium of the lottery game. Denote as $\Pi_i^k(\mathcal{T}_i)$ the set of k-level marginal beliefs of player i that arise in type space $\mathcal{T}_i$. The result confines itself to the case where the CC property arises within some order k, whether such result is available when only the across-order CC property holds is an open question.

Proposition 3.4 *Let $\mathcal{M}_i = \prod_{k=1}^{K} \Pi_i^k(\mathcal{T}_i)$ and assume that during the VCG stage each agent reports his payoff type truthfully, then provided that for each i the within order CC property holds for some $k_i \leq K$, the surplus extraction game has a unique Bayesian Nash equilibrium where each agent reports his type truthfully.*

Proof The proof of existence of a truthful equilibrium follows exactly the same steps as in Proposition 3.1.

For uniqueness, assume to the contrary that there exists a profile of deception strategies $\alpha = (\alpha_1(.), .., \alpha_I(.))$ with $\alpha_i : T_i \to T_i$ such that choosing $(\tau_1^*(\alpha_1(t_1)), .., \tau_1^*(\alpha_1(t_1)))$ with $\tau_i^*(\alpha_i(t_i)) \neq \tau_i(t_i)$ for at least one agent i is also a BNE of the lottery game. As we will argue, if the message space of each i is rich enough as to allow for an arbitrary combination of beliefs across different orders and if each agent reports truthfully at the VCG stage, there exists agent i and type t_i for whom it is profitable to deviate from deception strategy $\alpha(t_i)$. Furthermore, this deviation triggers deviations by further types of $-i$ towards the truthful reporting.

$$\tau_i(\alpha_i(t_i)) = \alpha_i(s_{t_i}) + \gamma_1 x^1_{\alpha_i(t_i)}(\alpha_{-i}(\theta_{-i})) + \gamma_2 x^2_{\alpha_i(t_i)}(\alpha_{-i}(\pi^1_{-i})) + ... + \gamma_k x^k_{\alpha_i(t_i)}(\alpha_{-i}((\pi^{K-1}_{-i}))$$

as $\alpha_{-i}(\theta_{-i}) = \theta_{-i}$, i.e. $\alpha_{-i}(.)$ is an identity map by the hypothesis, choosing $\tau_i(\alpha_i(t_i))$ with $x^1_{\alpha_i(t_i)}(\alpha_{-i}(\theta_{-i}))$ is dominated by choosing $\tau'_i(\alpha_i(t_i))$ with $x^1_{t_i}(\theta_{-i})$ instead (and keeping the tail part the same).

Now we will argue that

$$\tau'_i(\alpha_i(t_i)) = \alpha_i(s_{t_i}) + \gamma_1 x^1_{t_i}(\theta_{-i}) + \gamma_2 x^2_{\alpha_i(t_i)}(\alpha_{-i}(\pi^1_{-i})) + ... + \gamma_k x^k_{\alpha_i(t_i)}(\alpha_{-i}((\pi^{K-1}_{-i}))$$

cannot be an equilibrium either. Indeed given that for each i and t_i it is individually rational to choose the lottery over θ_i designed for his true type, it is obvious that for any i rather than sticking to $\tau'_i(\alpha_i(t_i))$, it is a best reply to choose $\tau''_i(\alpha_i(t_i))$ with $x^2_{t_i}(\pi^1_{-i})$ instead of $x^2_{\alpha_i(t_i)}(\alpha_{-i}(\pi^1_{-i}))$, as this leads to a smaller participation fee in expectation.

Such reasoning can be continued up to level K of beliefs that differentiate different types in the finite type space in question. For each type t_i, the only lottery that survives iterative elimination of weak best replies is the lottery designed for t_i itself. $\blacksquare$

3.2 Designer Uncertainty

In this subsection we proceed to show that the designer can use a union of the "local" GCM mechanisms despite of his "global" uncertainty about the agents' type space.

The result of Crémer and McLean (1988) has been criticized on the grounds that it relies heavily on the assumption of perfect knowledge by the designer of the environment and in particular, of agents' (first-order) beliefs. In this subsection we are addressing this critique. Instead of assuming that the designer knows perfectly the set of agents mutual beliefs, or, more generally, he knows the set of types that are common knowledge among agents, we allow for possibility that the designer perceives as possible a collection of type spaces. Such form of designer uncertainty has been used in Heifetz and Neeman (2006) to establish generic (measure-theoretic) impossibility of full surplus extraction given a fixed family of priors. The paper of Gizatulina and Hellwig (2013) uses their framework to show instead that among all families of priors, the set of families where the results of Heifetz and Neeman (2006) hold is itself non-generic (from the topological point of view).

We formalize the designer uncertainty as follows. At the interim stage, while each agent knows his type in T_i and furthermore the entire type space $\mathcal{T} = \{T_i, \theta_i(.), \pi_i(.)\}_{i \in I}$, i.e., how each agent's private information maps into this agent's beliefs, the designer only knows that the agents type space is one out of a collection

$$\{(T_i^k, \theta^k(.), \pi^k(.)\}_{k \in K}.$$

By "designer knows" we mean that the designer puts probability one on this set of type spaces (i.e. on this set of what could be common knowledge among agents). We assume also that the beliefs of the designer are correct, i.e. the actual agents' type space $\mathcal{T} = \{T_i, \theta_i(.), \pi_i(.)\}_{i \in I}$ is in $\{(T_i^k, \theta^k(.), \pi^k(.)\}_{k \in K}$. We do not impose any further assumption on the beliefs of the designer except that he puts probability zero on all type spaces outside of the profile $\{\mathcal{T}^k\}_{k \in K}$. In particular we keep an agnostic perspective on subjective probability measures that the designer assigns to individual models in $\{\mathcal{T}^k\}_{k \in K}$. As we show, in fact, actual beliefs of the designer are irrelevant. Furthermore, we assume that for all i and all $k, k' \in K$, $\Theta_i^k = \Theta^{k'}$, i.e. type spaces do not differ in their payoffs.

While designing a menu of lotteries for each agent i, the designer possesses an important piece of knowledge, namely, that agents *commonly know* what the type space is. This follows from the definition of the interim type space as being the smallest set of payoffs and beliefs that are commonly known by agents. The goal of this section is to characterize when the designer can exploit this information of agents about the relevant space of beliefs through the union of GCM mechanisms. We are interested in the setting where the designer can propose *directly a union of GCM mechanisms*, instead of using a two-stage approach of the so-called "shoot-the-liar" mechanisms.

Definition A collection of type spaces $\{\mathcal{T}^k\}_{k \in K}$ is *disjoint* if for all $i \in I$ it holds $T_i^k \cap T_i^{k'}$ for $k \neq k'$.

Lemma 3.5 *A collection of type spaces $\{\mathcal{T}^k\}_{k \in K}$ is disjoint if and only if each $\mathcal{T}^k$ is the minimal common knowledge set.*

Proof For the proof of this result see Gizatulina and Hellwig (2013).

Definition A *union* of generalized Crémer-McLean mechanisms is a collection of tuples $\{(\tau^n(.), p^n(.), q^n(.))\}_{n\in N}$, or, equivalently, $\cup_{n\in N}\Lambda(\mathcal{T}^n)$.

Definition A collection of $\{\mathcal{T}^n\}_{n\in N}$ verifies the convex hull property *within* the models if for any agent i, any model $n \in N$ and any his type t_i^{0n}

$$\pi(. \mid t_i^0) \notin co\{\pi(. \mid t_i^n)\} \tag{10}$$

for $t_i^n \in T^n$ and $t_i^{0n} \neq t_i^k$ with $n \in N$.

Proposition 3.6 *If the convex-hull property is verified for each model in a collection $\{\mathcal{T}^n\}_{n\in N}$ the union of generalized Crémer-McLean mechanisms $\{\Lambda(\mathcal{T}^n)\}_{n\in N}$ extracts the full surplus regardless of the realised type space $\mathcal{T}^n$ and absence of any knowledge by the designer thereof.*

Proof We construct the proof relying on an inductive procedure on the sets of higher-order beliefs. First, for any $n, n' \in N$ it holds that $p^n(.) = p^{n'}(.)$ and $q^n(.) = q^{n'}(.)$. This is because $p(.)$ and $q(.)$, being a part of the second stage VCG mechanism, do not depend on reported beliefs but only on reported elements in Θ and we assumed that $\Theta^n = \Theta^{n'}$ for all $n, n' \in N$. Thus, the union of mechanisms $\cup_{n\in N}\Lambda(\mathcal{T}^n)$ reduces to

$$\cup_{n\in N}\Lambda(\mathcal{T}^n) = (\{\lambda^n(.)\}_{n\in N}, p(.), q(.))$$

Fix a type space $\mathcal{T}^{n*}$ which would be the actual agents' type space. By hypothesis, the mechanism $\Lambda(\mathcal{T}^{n*})$ extracts the full surplus *within* the type space $\mathcal{T}^{n*}$, i.e. no type of no agents have incentives to pick a lottery $\lambda(\hat{t}_i^{n*})$ designed for some $\hat{t}_i^{n*}$.

Hence, to establish the proof, we have to check that when faced with $\cup_{n\in N}\Lambda(\mathcal{T}^n)$ and expecting other agents to report truthfully, each agent i does not have incentives to deviate by choosing a lottery designed for some type t_i^n from a different type space $\mathcal{T}^n$, $n \neq n*$.

(*to be inserted*)

EXAMPLE 2

We construct an example of a family of two type spaces where the union of classical Crémer-McLean mechanisms fails to extract surplus, but the union of extended, generalized CM mechanism achieves this objective.

Consider the following two type spaces, where there are two agents and agents are fully symmetric (so we give directly posterior beliefs given the observed θ_i for each $i \in I$, $I = 1, 2$:

	θ_1^j	θ_2^j	θ_3^j
$\pi_i^1(\theta_1^i)$	1/3	1/3	1/3
$\pi_i^1(\theta_2^i)$	7/45	19/90	19/30
$\pi_i^1(\theta_3^i)$	1/2	1/4	1/4

Table: beliefs of i in type space T^1

	θ_1^j	θ_2^j	θ_3^j
$\pi_i^2(\theta_1^i)$	1/5	1/5	3/5
$\pi_i^2(\theta_2^i)$	5/12	7/24	7/24
$\pi_i^2(\theta_3^i)$	1/9	2/9	2/9

Table: beliefs of i in type space T^2

Here, beliefs $\pi_i^1(\theta_2^i)$ of type θ_i^2 from the type space T^1 are the convex combination of beliefs of types $\pi_i^2(\theta_1^i)$ and $\pi_i^2(\theta_3^i)$. Obviously, offering to i to choose from the menu of 5 functions does not help to extract his surplus.

There are in fact two problems with using the union of classical Crémer-McLean mechanisms: convex hull property may fail to hold *across* type spaces and so constructing instead a "grand" CCM mechanism would not help. Even if the convex hull property holds across type spaces, because for the type space which is a union of type spaces (accounting for the common knowledge degree among agents) there are in general many more rows than columns, it is impossible to construct a profile of lotteries, working for all types (from all type spaces jointly), i.e. verifying incentive constraints across type spaces.

4 Discussion and Extensions

4.1 Robustness to ϵ-misspecification

Suppose the designer knows instead that agents' type space is one from an ϵ-neighbourhood of some benchmark type space $\mathcal{T}$.

4.2 Relationship between the CC property in the abstract type spaces and belief hierarchies

(to be added)

5 Conclusions

The goal of this paper was to give some arguments for the robustness of the Crémer-McLean mechanism to the absence of common knowledge between the designer and agents as it is one of the most common attack on the CM mechanism. We have shown that failure to extract surplus from types in the union of type spaces $\{\mathcal{T}^k\}_{k\in K}$ is due only to failure of the convex hull condition within each T^k, if any. The absence of the common knowledge between the designer and agents alone is irrelevant for failure to extract the surplus. Furthermore, what matters is failure of the CC condition within a model, but not across the models.

References

BERGEMANN, D. AND S. MORRIS (2005): "Robust mechanism design," *Econometrica*, 73, 1521–1534. 4, 7

BRANDENBURGER, A. AND E. DEKEL (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, 189–198. 7

CRÉMER, J. AND R. MCLEAN (1988): "Full extraction of the surplus in Bayesian and dominant strategy auctions," *Econometrica*, 56, 1247–1257. 2, 8, 17, 18

GIZATULINA, A. AND M. HELLWIG (2013): "On the Designer Uncertainty and the Robustness of Full Surplus Extraction in Families of Incomplete-Information Models," Working paper, Max Planck Institute for Research on Collective Goods. 18, 20

HEIFETZ, A. AND Z. NEEMAN (2006): "On the generic (im)possibility of

full surplus extraction in mechanism design," *Econometrica*, 117, 213–233. 18

HEIFETZ, A. AND D. SAMET (1998): "Topology-Free Typology of Beliefs," *Journal of Economic Theory*, 82. 7

MCAFEE, P. AND P. RENY (1992): "Correlated information and mechanism design," *Econometrica*, 60, 395–421. 2, 8

MERTENS, J. AND S. ZAMIR (1985): "Formulation of Bayesian analysis for games with incomplete information," *International Journal of Game Theory*, 10, 619–632. 7

MORRIS, S. AND H. SHIN (2002): "Social Value of Public Information," *American Economic Review*, 92, 1521 – 1534. 7