3106

REPRINT

Claude d'Aspremont and Jacques Crémer

Some remarks on Bayesian mechanism design

In W. Trockel (ed.), *Social Design Studies in Economic Design*. Cham: Springer, 85-98, 2019

CORE
Voie du Roman Pays 34, L1.03.01
B-1348 Louvain-la-Neuve
Tel (32 10) 47 43 04
Email:lidam-library@uclouvain.be
https://uclouvain.be/en/research-institutes/
lidam/core/reprints.html

Some Remarks on Bayesian Mechanism Design

Claude d'Aspremont and Jacques Crémer

1 Introduction

Although the problem of revelation of preferences for public goods had already been brought up in several instances, it was surely the merit of Leo Hurwicz to show that providing incentives was a fundamental problem in the design of any institution for collective decision based on decentralized information and control. In particular, the introduction by Hurwicz (1972) of the concept of incentive compatibility for correct revelation as the Nash equilibria of a family of noncooperative games with complete information, and the difficulty to enforce Pareto-optimality in such a model, suggested to view institutions as rules of a game with incomplete information, as formally defined by Harsanyi (1967, 1968a,b) and to apply his concept of Bayesian equilibrium.

Bayesian mechanism design highlights the importance of individual agents' beliefs in environments where information is incomplete. The prior contribution of Vickrey (1961) had already made this clear for auction design and, more generally for market design, and had, as well, identified the budget balance problem that a marketing agency would encounter in solving the demand revelation problem.

A resulting issue has been to look for the most general class of beliefs allowing to solve the revelation problem without weakening the collective efficiency

C. d'Aspremont
CORE, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
e-mail: claude.daspremont@uclouvain.be

J. Crémer (✉)
Toulouse School of Economics, Toulouse, France
e-mail: jacques.cremer@tse-fr.eu; jacques@cremerline.com

W. Trockel (ed.), *Social Design*, Studies in Economic Design,
https://doi.org/10.1007/978-3-319-93809-7_6

requirement. The objective is to identify the most general conditions on agents' beliefs allowing for the implementation of collective decision rules. Many of these conditions have been presented in their "dual" form, following some variant of the theorem of the alternative. We shall concentrate on the two most general conditions (each in some respect) and show that they can usefully be used in their "primal" form, leading to more constructive arguments. We discuss the issue and analyze some interesting cases.

In Sect. 2, we present the model and formulate both the incentive compatibility problem à la Hurwicz and the Bayesian incentive compatibility problem as the problem of solving a system of linear inequalities and look at the dual consistency conditions. In Sect. 3, we adopt what may be called the "primal approach" showing the prominence of two generic conditions that can be fruitfully imposed on the beliefs. In Sect. 4, we indicate that these two conditions remain central for team adverse selection with participation constraints. This is also true for team moral hazard with stochastic outcome functions, as we mention in the concluding section, where we point out some areas for further research.

2 The Model and the Basic System of Linear Inequalities

The public (or collective) decision-making for which a mechanism has to be designed is very general. It may be a mechanism to provide public goods (or reduce a public bad, such as pollution), or to allocate goods, as in auctions. It is just defined by a set of agents $N = \{1, 2, \cdots, i, \cdots, n\}$, with $n \geq 2$ (for some results we will assume $n \geq 3$) and a compact set X of outcomes. We assume incomplete information about the characteristics of the agents. Adopting the Bayesian point of view (Harsanyi, 1967, 1968a,b), to each agent i is associated a finite set of possible types A_i as well as the beliefs of agent i concerning the types of the other agents. For every $\alpha_i \in A_i$, these beliefs $p_i(\alpha_{-i} \mid \alpha_i)$ are defined on $A_{-i} = \prod_{j \neq i} A_j$. For simplicity, we will assume that these individual beliefs are consistent, i.e., generated from a common prior $p(\alpha)$ defined on the set of states of nature, $A = \prod_j A_j$, but many results do not require this assumption. We call (N, A, p) the *information structure*.

We assume transferable utility: when the public decision is x and he receives a transfer t_i, the utility of agent i of type α_i is $u_i(x; \alpha_i) + t_i$. We will take some liberty with the terminology and call u_i the utility function of the agent. In general, the type α_i determines both the beliefs and the utility of agent i. We call $(N, A, p, X, \{u_i\})$ the *environment*. It is composed of an information structure, a set of outcomes, and the utility functions of the agents.

We study *direct (revelation) mechanisms*, where agents are given incentives to truthfully reveal their types. Formally, a *mechanism* is a pair (s, t), where s is an *outcome function* from $A \equiv \prod_{i=1}^{n} A_i$ to X and where t is a *transfer function* from A to $\Re^N$. The outcome function is *ex post efficient* if, for all $\alpha \in A$,

$$\sum_{i=1}^{n} u_i(s(\alpha); \alpha_i) = \max_{x \in X} \sum_{i=1}^{n} u_i(x; \alpha_i),$$

and *Pareto-optimal* if, in addition, the transfer function is *balanced*, that is if, for all $\alpha \in A$, the following budget balance condition holds:

$$\sum_{i=1}^{n} t_i(\alpha) = 0.$$

A mechanism defines a game of incomplete information, namely a collection of normal form games $\Gamma(\alpha)$, one for each vector of types, where the payoff function of agent i is $u_i(s(\cdot); \alpha_i) + t_i(\cdot)$ and his strategy space is A_i. For each agent i, we can define a set of "normalized" strategies, i.e., functions $a_i(\alpha_i)$ from A_i to itself, and focus on the "truth-telling strategy" : $a_i^*(\alpha_i) = \alpha_i$ for all $\alpha_i \in A_i$.

A first way to introduce incentives, based on the seminal contribution of Hurwicz (1972), is to require truth-telling to be what is now usually called an ex post equilibrium,[1] namely to require the vector $\left(a_1^*(\alpha_1), \ldots, a_n^*(\alpha_n)\right)$ to be a Nash equilibrium of $\Gamma(\alpha)$ for $\alpha \in A$. Since α can be any vector in A, a mechanism (s, t) is *incentive compatible* (abbreviated *IC*) if, for all $i \in N$, all $\alpha_i \in A_i$, all $\tilde{\alpha}_i \in A_i$, and all $\alpha_{-i} \in A_{-i}$)

$$u_i\left(s(\alpha_i, \alpha_{-i}); \alpha_i\right) + t_i(\alpha_i, \alpha_{-i}) \geq u_i\left(s\left(\tilde{\alpha}_i, \alpha_{-i}\right); \alpha_i\right) + t_i(\tilde{\alpha}_i, \alpha_{-i}).$$

In the present framework, where the utility function $u_i(x; \alpha_i)$ of agent i depends only on type α_i (the private value case), IC is equivalent to dominant strategy incentive compatibility (i.e., $\left(a_1^*(\alpha_1), \ldots, a_n^*(\alpha_n)\right)$ are dominant strategies in $\Gamma(\alpha)$ for every $\alpha \in A$) or strong incentive compatibility (SIC) to use the terminology of Green and Laffont (1977). The focal[2] class of efficient SIC mechanisms is the class of Vickrey–Clarke–Groves (VCG) mechanisms where s is an efficient outcome

[1] An ex post equilibrium is called a uniform equilibrium in d'Aspremont and Gérard-Varet (1979a) and incentive compatibility is called uniform incentive compatibility by Holmström and Myerson (1983). See Bergemann and Morris (2005) for more discussion and references.

[2] If the set of types is large enough (i.e., connected), Groves mechanisms are the only efficient SIC mechanisms. This result does not hold in our discrete types of framework. See Green and Laffont (1977), Walker (1978), and Holmström (1979).

function and the transfers are defined by:

$$t_i(\alpha_i, \alpha_{-i}) = \sum_{j \neq i} u_j(s(\alpha_i, \alpha_{-i}); \alpha_j) + h_i(\alpha_{-i}),$$

with h_i any real-valued function defined on A_{-i}. However, as known since Green and Laffont (1979) and Walker (1980), these transfers are in general not balanced, so that Pareto-optimality is not obtained. Finding Pareto-optimal SIC mechanisms is a difficult task even when utility functions are quasi-linear. In this framework, given an efficient outcome function, finding balanced SIC transfers amounts to solving a finite system of linear inequalities and we can apply theorems of the alternative to characterize the consistency of such systems. Using such a method, d'Aspremont et al. (1990) derive the following necessary and sufficient condition imposed on the environment (without involving the beliefs) and the outcome function for an SIC mechanism to exist: for all $\lambda : A_i^2 \times A_{-i} \to \Re_+$ and all $\mu : A \to \Re$ which satisfy

$$\sum_{\tilde{\alpha}_i \neq \alpha_i} \lambda_i(\tilde{\alpha}_i, \alpha_i, \alpha_{-i}) - \sum_{\tilde{\alpha}_i \neq \alpha_i} \lambda_i(\alpha_i, \tilde{\alpha}_i, \alpha_{-i}) = \mu(\alpha) \text{ for all } i \in N \text{ and } \alpha \in A,$$

then

$$\sum_{i=1}^{n} \sum_{(\tilde{\alpha}_i, \alpha_i) \in A_i^2} \sum_{\alpha_{-i}} \lambda_i(\tilde{\alpha}_i, \alpha_i, \alpha_{-i})[u_i(s(\tilde{\alpha}_i, \alpha_{-i}); \alpha_i) - u_i(s(\alpha_i, \alpha_{-i}); \alpha_i)] \leq 0.$$

The family of parameters λ and μ are the dual variables associated, respectively, with the incentive inequalities and the budget balance condition. To illustrate the usefulness of this duality result, that same article shows that a Pareto-optimal SIC mechanism exists whatever the environment and the efficient outcome function, provided that each agent has only two types ($|A_i| = 2$, for any i)[3]—this mechanism may have to be chosen outside the class of VCG mechanisms. The fact that we obtain such a strong result when all agents have only two types should at the minimum make us cautious about the use of the two-type assumption in applied theory.

A second way to introduce incentives for truth revelation in such environments is to reverse the point of view. Instead of looking at incentive compatibility by imposing conditions on the utilities valid whatever the beliefs, one can look at incentive compatibility by imposing conditions on the beliefs valid whatever the utilities. Given some information on structure, the payoff of player i of type α_i is evaluated, for every strategy vector $(a_1(\cdot), \cdots, a_n(\cdot))$, as the conditional

[3] This generalizes a result of Maskin (1986), proved in the case of two agents.

expected utility

$$\sum_{\alpha_{-i}}\Big[u_i(s(a_1(\alpha_1),\cdots,a_n(\alpha_n));\alpha_i)+t_i(a_1(\alpha_1),\cdots,a_n(\alpha_n))\Big]p_i(\alpha_{-i}\mid\alpha_i).$$

A mechanism (s,t) is *Bayesian Incentive Compatible* (BIC) if the truth-telling strategy $a^*(.)$ is a *Bayesian equilibrium* in the sense of Harsanyi; it is characterized by the following inequalities, for all $i\in N$ and all $(\tilde{\alpha}_i,\alpha_i)\in A_i^2$,

$$\sum_{\alpha_{-i}}[u_i(s(\alpha_i,\alpha_{-i});\alpha_i)+t_i(\alpha_i,\alpha_{-i})]\,p_i(\alpha_{-i}\mid\alpha_i)$$
$$\geq\sum_{\alpha_{-i}}\big[u_i(s(\tilde{\alpha}_i,\alpha_{-i});\alpha_i)+t_i(\tilde{\alpha}_i,\alpha_{-i})\big]\,p_i(\alpha_{-i}\mid\alpha_i).$$

Comparing these inequalities with the inequalities for IC, we observe that BIC holds for any information structure such that IC holds.

As above, given an environment and some efficient outcome function, finding balanced BIC transfers amounts to solving a finite system of linear inequalities, but in this case, they depend on the beliefs. Applying a theorem of Fan (1956), d'Aspremont and Gérard-Varet (1979a) obtained a necessary and sufficient condition to be imposed on the environment and the outcome function for a balanced BIC mechanism to exist: for all $\lambda: A_i^2\to\Re_+$ and all $\mu: A\to\Re$, such that for all $i\in N$ and all $\alpha\in A$,

$$p_i(\alpha_{-i}\mid\alpha_i)\sum_{\tilde{\alpha}_i\neq\alpha_i}\lambda_i(\tilde{\alpha}_i,\alpha_i)-\sum_{\tilde{\alpha}_i\neq\alpha_i}\lambda_i(\alpha_i,\tilde{\alpha}_i)p_i(\alpha_{-i}\mid\tilde{\alpha}_i)=\mu(\alpha)$$

we have

$$\sum_{i=1}^{n}\sum_{(\tilde{\alpha}_i,\alpha_i)\in A_i^2}\lambda_i(\tilde{\alpha}_i,\alpha_i)\times$$
$$\left[\sum_{\alpha_{-i}}\big[u_i(s(\tilde{\alpha}_i,\alpha_{-i});\alpha_i)-u_i\left(s(\alpha_i,\alpha_{-i});\alpha_i\right)\big]p_i(\alpha_{-i}\mid\alpha_i)\right]\leq 0.$$

Again, the λs and μs are the dual variables associated, respectively, with the incentive inequalities and the budget balance condition. The main application of this result is to allow for a first formulation of a condition (that we will call condition

C*)[4] imposed only on the information structure and sufficient to guarantee the existence of Pareto-optimal BIC mechanism whatever the utility functions and the *efficient* outcome function. As we will see below, this condition holds generically, is necessary for ensuring the budget balance condition, and, as such, is weaker (or equivalent) to the other conditions that have been proposed in the literature.

An information structure (N, A, p) satisfies Condition C* if whenever $\lambda : A_i^2 \to \Re_+$ and $\mu : A \to \Re$ satisfy

$$p_i(\alpha_{-i} \mid \alpha_i) \sum_{\tilde{\alpha}_i \neq \alpha_i} \lambda_i(\tilde{\alpha}_i, \alpha_i) - \sum_{\tilde{\alpha}_i \neq \alpha_i} \lambda_i(\alpha_i, \tilde{\alpha}_i) p_i(\alpha_{-i} \mid \tilde{\alpha}_i) = \mu(\alpha)$$

$$\text{for all } i \in N \text{ and all } \alpha \in A$$

then $\mu(\alpha) = 0$ for all $\alpha \in A$.

That BIC and budget balance are ensured thanks to condition C is immediate since by efficiency of the outcome function:

$$\sum_{i=1}^{n} \sum_{(\tilde{\alpha}_i, \alpha_i) \in A_i^2} \lambda_i(\tilde{\alpha}_i, \alpha_i) \sum_{\alpha_{-i}} \left[u_i(s(\tilde{\alpha}_i, \alpha_{-i}); \alpha_i) - u_i\,(s(\alpha_i, \alpha_{-i}); \alpha_i)\right] p_i(\alpha_{-i} \mid \alpha_i)$$

$$\leq \sum_{i=1}^{n} \sum_{\alpha} \sum_{j \neq i} u_j(s(\alpha); \alpha_j) \times$$

$$\left[p_i(\alpha_{-i} \mid \alpha_i) \sum_{\tilde{\alpha}_i \neq \alpha_i} \lambda_i(\tilde{\alpha}_i, \alpha_i) - \sum_{\tilde{\alpha}_i \neq \alpha_i} \lambda_i(\alpha_i, \tilde{\alpha}_i) p_i(\alpha_{-i} \mid \tilde{\alpha}_i)\right]$$

$$= 0.$$

Note that, as can be easily verified, Condition C* is satisfied as soon as one agent i has "free beliefs," i.e., as soon as $p(\cdot \mid \tilde{\alpha}_i) \equiv p(\cdot \mid \alpha_i)$, for some i and all pairs $(\alpha_i, \tilde{\alpha}_i)$, and includes, among many others, the so-called independent case where all agents have free beliefs.

3 The "Primal" Approach

As shown by d'Aspremont et al. (2003), Condition C* has an equivalent "primal" version, Condition C, which turns out to be very useful. An information structure (N, A, p) satisfies Condition C if and only if for every function $R : A \to \Re$, there

[4]"C" for "Compatibility condition" the name given by d'Aspremont and Gérard-Varet (1979a) and the star in C* to indicate that it is the "dual" version of the condition. The "primal" version is studied in the next section.

exists a transfer rule t^C such that

$$\sum_{i \in \mathcal{N}} t_i^C(\alpha) = R(\alpha) \text{ for all } \alpha \in A,$$

and

$$\sum_{\alpha_{-i} \in \mathcal{A}_{-i}} t_i^C(\alpha_i, \alpha_{-i})\, p(\alpha_{-i} \mid \alpha_i) \geq \sum_{\alpha_{-i} \in \mathcal{A}_{-i}} t_i^C(\widetilde{\alpha}_i, \alpha_{-i})\, p(\alpha_{-i} \mid \alpha_i)$$
$$\text{for all } i \in N \text{ and all } (\alpha_i, \widetilde{\alpha}_i) \in A_i^2.$$

Using the same theorem of the alternative as above, one can show that this condition is satisfied if and only if, for every $R : A \to \Re$, whenever, for $\lambda : A_i^2 \to \Re_+$ and $\mu : A \to \Re$

$$p_i(\alpha_{-i} \mid \alpha_i) \sum_{\tilde{\alpha}_i \neq \alpha_i} \lambda_i(\tilde{\alpha}_i, \alpha_i) - \sum_{\tilde{\alpha}_i \neq \alpha_i} \lambda_i(\alpha_i, \tilde{\alpha}_i) p_i(\alpha_{-i} \mid \tilde{\alpha}_i) = \mu(\alpha)$$
$$\text{for all } i \in N \text{ and for all } \alpha \in A.$$

then

$$\sum_{\alpha \in \mathcal{A}} \mu(\alpha) R(\alpha) \leq 0.$$

This implies that μ must be identically equal to zero, and therefore C is equivalent to C*.

There is a nice story developed in d'Aspremont et al. (2003) to explain why Condition C guarantees the existence of a Pareto-optimal BIC mechanism whatever the utility functions and the *efficient* outcome function. Imagine a planner consulting two bureaus, the "preference bureau" and the "beliefs bureau." The preference bureau uses a VCG mechanism $t_i^{VCG}(\alpha_i, \alpha_{-i}) = \sum_{j \neq i} u_j(s(\alpha_i, \alpha_{-i}); \alpha_j)$ so that SIC (or IC) is satisfied thanks to efficiency of the outcome function s. But, this leads to a deficit $R(\alpha) = -\sum_{i \in \mathcal{N}} t_i^{VCG}(\alpha)$, for every α. Under condition C, the beliefs bureau can recommend a transfer t^C to cover this deficit while preserving incentives, so that the transfer function $t = t^{VCG} + t^C$ is balanced and ensures BIC and Pareto-optimality. This argument shows more. It shows that, under condition C, it is always possible to supplement transfers ensuring BIC by transfers to balance the budget, while keeping BIC. And, conversely, an information structure guarantees budget balance in that sense only if condition C holds.[5] Using this property of Condition C,

[5] For the proof, see d'Aspremont et al. (2004). Forges et al. (2002) call this property "automatic balance."

d'Aspremont et al. (2004) show that condition C is the weakest among all conditions that have been proposed[6] to implement efficient mechanisms.

Now, in some contexts, one might be interested in implementing mechanisms that are not Pareto-optimal (the outcome function is not efficient). From a normative viewpoint, this would be justified if the decision generates externalities beyond the set of agents who participate in the mechanism. From a positive viewpoint, this would be the case if the principal is at the same time budget constrained (so that he cannot transfer funds to the agents) but had his own, state of nature dependent, preferences over outcomes. In these cases, the mechanism (s, t) should satisfy BIC and the transfer function should be balanced. Assuming $n \geq 3$, Condition B, defined as follows, can be used for that purpose.[7] An information structure (N, A, p) satisfies Condition B if and only if there exists a transfer rule t^B such that for all $\alpha \in A$,

$$\sum_{i \in \mathcal{N}} t_i^B(\alpha) = 0,$$

and, for all $i \in A$ and all $\widetilde{\alpha}_i \in A_i$,

$$\sum_{\alpha_{-i} \in \mathcal{A}_{-i}} t_i^B(\alpha_i, \alpha_{-i})\, p(\alpha_{-i} \mid \alpha_i) > \sum_{\alpha_{-i} \in \mathcal{A}_{-i}} t_i^B(\widetilde{\alpha}_i, \alpha_{-i})\, p(\alpha_{-i} \mid \alpha_i).$$

By applying the same kind of duality argument as above d'Aspremont and Gérard-Varet (1982) (see lemma 3), B is equivalent to the dual condition B* which states that there exist no $\lambda : A_i^2 \to \Re_+$, $\lambda \neq 0$, and no $\mu : A \to \Re$ which satisfy

$$p_i(\alpha_{-i} \mid \alpha_i) \sum_{\tilde{\alpha}_i \neq \alpha_i} \lambda_i(\tilde{\alpha}_i, \alpha_i) - \sum_{\tilde{\alpha}_i \neq \alpha_i} \lambda_i(\alpha_i, \tilde{\alpha}_i) p_i(\alpha_{-i} \mid \tilde{\alpha}_i) = \mu(\alpha)$$

$$\text{for all } i \in N \text{ and for all } \alpha \in A.$$

This immediately shows that B* implies C*, hence that C holds whenever B holds. Moreover, for any environment with an information structure satisfying B, and any outcome function, the BIC constraints can be (strictly) satisfied. It suffices to premultiply the (balanced) transfer function t^B by some large positive number. But, the converse is also true: for any environment, the BIC constraints can be (strictly) satisfied for any outcome function only if the information structure satisfies B (d'Aspremont et al., 2003).

[6]It is strictly weaker than Chung (1999) weak regularity (hence than Matsushima (1991) regularity condition) and Fudenberg et al. (1994) pairwise identifiability. It is equivalent to Johnson et al. (1990) condition called LINK.

[7]Condition B was first defined by d'Aspremont and Gérard-Varet (1982). When $n = 2$, condition C is equivalent to independence of types, and condition B never holds.

The relationship between conditions C and B can be further clarified since condition B implies "No-Freeness": it cannot be the case that two types of an agent generate the same probability distribution over the types of the other agents. Formally, for all $i \in N$ and any two types α_i and $\widetilde{\alpha}_i$, we have $p(\cdot \mid \alpha_i) \neq p(\cdot \mid \widetilde{\alpha}_i)$. Furthermore, it can be shown that condition B is equivalent to C plus No-Freeness: an information structure (N, A, p) satisfies condition B if and only if it satisfies condition C and No-Freeness. Because for $n = 2$, condition C is equivalent to independence of types, that is freeness for any two types of any agent, condition B never holds with $n = 2$. Because, as we will see shortly, B holds generically with $n \geq 3$, this reinforces our words of caution about the generality of results obtained from models with only two types for each agent.

Last, but not least, in "nearly all" environments, we can construct explicitly the transfer function t^B used in the definition of condition B. Consider an information structure such that for all $i \neq j$, all $\alpha_i \in A_i$, and all $\alpha_j \in A_j$, $p(\alpha_j \mid \alpha_i) \equiv \sum_{\left\{\alpha'_{-i} \mid \alpha'_j = \alpha_j\right\}} p\left(\alpha'_{-i} \mid \alpha_i\right) > 0$. Define addition and subtraction on the indices of the agent modulo n, so that $n + 1 \equiv 1$ and $1 - 1 \equiv n$. We let

$$t_i^B(\alpha) = \log p\left(\alpha_{i+1} \mid \alpha_i\right) - \log p\left(\alpha_{i+2} \mid \alpha_{i+1}\right).$$

The negative term is constant in α_i and does not influence the incentives of agent i, but will ensure budget balance. The strict concavity of the function log implies that for all i and all $(\alpha_i, \widetilde{\alpha})$,

$$\begin{aligned}
&\sum_{\alpha_{-i}} \left[\log p\left(\alpha_{i+1} \mid \widetilde{\alpha}_i\right) - \log p\left(\alpha_{i+1} \mid \alpha_i\right)\right] p_i\left(\alpha_{-i} \mid \alpha_i\right) \\
&\qquad = \sum_{\alpha_{i+1}} \left[\log \frac{p\left(\alpha_{i+1} \mid \widetilde{\alpha}_i\right)}{p\left(\alpha_{i+1} \mid \alpha_i\right)}\right] p\left(\alpha_{i+1} \mid \alpha_i\right) \\
&\qquad\qquad < \log \sum_{\alpha_{i+1}} \frac{p\left(\alpha_{i+1} \mid \widetilde{\alpha}_i\right)}{p\left(\alpha_{i+1} \mid \alpha_i\right)} p\left(\alpha_{i+1} \mid \alpha_i\right) = 0,
\end{aligned}$$

whenever $p(\alpha_{i+1} \mid \widetilde{\alpha}_i) \neq p(\alpha_{i+1} \mid \alpha_i)$, for some α_{i+1}. Hence, in these cases, condition B holds, which proves that it holds generically. This also implies that condition C holds generically when $n \geq 3$.

4 Individual Rationality

This short review shows that, for Bayesian incentive compatibility and budget balance to hold whatever the utility functions, two conditions on the information structure seem to emerge. One, condition C (or equivalently C*) requires that the outcome function be efficient and is necessary and sufficient to ensure budget balance. The other, condition B (or equivalently condition B*) is stronger and is necessary and sufficient to implement any outcome function. Both the dual and primal approaches are useful in deriving the results.

We have not, in this note, explicitly introduced individual participation constraints, that is constraints which ensure that the agents are willing to participate in the mechanism. There are two types of individual rationality constraints that correspond to different extensive form games:

- ex ante participation constraints correspond to a game in which: (1) the mechanism is announced; (2) the agents decide whether they are willing to participate; (3) the agents learn their types; and (4) the agents play according to the rules of the mechanism. In this case, the participation constraints are written[8]

$$\sum_{\alpha \in A} \big[u_i(s(\alpha); \alpha_i) + t_i(\alpha)\big] p(\alpha) \geq 0 \text{ for all } i \in N.$$

- *interim* participation constraints correspond to the same game with stages 2 and 3 switched. The participation constraints are written

$$\sum_{\alpha_{-i} \in A_{-i}} \big[u_i(s(\alpha); \alpha_i) + t_i(\alpha)\big] p(\alpha_{-i} \mid \alpha_i) \geq 0 \text{ for all } i \in N \text{ and } \alpha_i \in A_i.$$

In the case of ex ante participation constraints, the agents learn their types only after having accepted to participate in the mechanism, whereas in the case of interim constraints, they learn their types before accepting to participate.

Both types of participation constraint require an ex ante nonnegative aggregate expected surplus condition:

$$\sum_{i \in N} \sum_{\alpha \in A} \big[u_i(s(\alpha); \alpha_i)\big] \, p(\alpha) \geq 0. \tag{1}$$

Actually, this condition is also sufficient, along with either condition C or condition B, for implementation if we impose *ex ante* individual rationality (d'Aspremont et al., 2003).

In the context of a bargaining problem (where the public decision is the final owner of the good) with two agents and independent types, Myerson and Satterthwaite (1983) showed their justly celebrated impossibility result: there exists no Bayesian incentive compatible, efficient, *interim* individually rational mechanism.

Despite the importance of this justly celebrated result, it is important to note that it only holds when there are two agents. In an important paper, Makowski and Mezzetti (1994) assume that there are at least three agents, that each agent has an infinite connected set of types and that types are independent, so that the following

[8] We are assuming that the reservation utilities are equal to 0, both in the case of ex ante and interim participation constraints. It is quite easy to prove that this does not entail any loss of generality.

result[9] holds: all efficient mechanisms are VCG in expectation, i.e., s is an efficient outcome function and the transfers are defined by:[10]

$$t_i(\alpha_i, \alpha_{-i}) = \sum_{\alpha_{-i} \in \mathcal{A}_{-i}} \left[\sum_{j \neq i} u_j(s(\alpha_i, \alpha_{-i}); \alpha_j) \right] p(\alpha_{-i}) + h_i(\alpha),$$

where $\sum_{\alpha_{-i}} h(\alpha_i, \alpha_{-i}) p(\alpha_{-i})$ does not depend on α_i; therefore, it does not affect the incentives of agent i. Then, they show that an efficient *interim* individually rational mechanism holds if, translated in our finite sets of types notation, the following condition holds (with s an efficient outcome function)[11,12]:

$$\begin{aligned}(n-1) &\sum_{\alpha \in A} \left[\sum_{i \in N} u_i(s(\alpha); \alpha_i) \right] p(\alpha) \\ &\leq \sum_{i \in N} \sum_{\alpha_i \in A_i} p_i(\alpha_i) \times \\ &\qquad \min_{\alpha_i' \in A_i} \sum_{\alpha_{-i}} \left[u_i(s(\alpha_i', \alpha_{-i}); \alpha_i') + \sum_{j \neq i} u_j(s(\alpha_i', \alpha_{-i}); \alpha_j) \right] p_i\left(\alpha_{-i}\right).\end{aligned}$$

We refer the reader to Makowski and Mezzetti (1994) and d'Aspremont and Crémer (2018) for more details.

Matsushima (2007) and Kosenok and Severinov (2008) have studied the existence of *interim* incentive compatibility for not necessarily efficient decision functions. Both papers produce conditions on the information structure which guarantee implementation when condition (1) is satisfied, with Kosenok and Severinov's being necessary and sufficient and therefore less restrictive than Matsushima's. Both conditions are strictly more restrictive than condition B.

More discussion of these conditions, showing how conditions B and C are still prominent, as well as exploration of the case where independence of types does not hold, but where there is still some freeness, can be found in d'Aspremont and Crémer (2018).

It may be useful to point out that the auction literature presents other examples of mechanisms where *interim* participation constraints are at the core of the problem. There, the types of the agents are their willingness to pay for the object. At the time at which they decide to participate in the auction, the potential buyers know their

[9] This result was first proved in d'Aspremont and Gérard-Varet (1979b) and generalized in Holmström (1977, 1979).

[10] Because of the independence of types, we can write $p(\alpha_{-i})$ instead of $p\left(\alpha_{-i} \mid \alpha_i\right)$.

[11] Note that with independence, we can write $p(\alpha_{-i})$ without ambiguity as $p(\alpha_{-i} \mid \alpha_i) = p(\alpha_{-i} \mid \widetilde{\alpha}_i)$ for all $\alpha_i, \widetilde{\alpha}_i, \alpha_{-i}$.

[12] To be totally clear, this condition is not necessary in the case of finite sets of types. We are writing it in this way to avoid introducing more notation.

types. One issue that the literature has tackled is the identification of information structures that guarantee that the seller can do as well as if he had full information. He can do so only if he sells the good to the agent with the highest valuation, hence the final allocation is efficient. However, the fact that he desires to extract the whole surplus rather than simply balance the budget is incompatible with free beliefs. See Crémer and McLean (1985, 1988). Interestingly, if Kosenok and Severinov's (2008) condition holds, the condition of Theorem 2 of Crémer and McLean (1988) must hold.

5 Concluding Remarks

The methods used to study "team adverse selection" can also be used to study "team moral hazard," say the sharing of an output (measured in money) depending on individual actions that are not perfectly observable (or controllable). This generates a noncooperative game and the collectively optimal level of output may be unenforceable whatever the proposed sharing rule.[13] A context, where this negative conclusion can be avoided, is when the outcome function is stochastic. We then get a system of linear inequalities in the transfers similar to the ones above, and conditions C (or C^*) and B (or B^*) are transposed as conditions on the stochastic outcome function to allow for a collective optimum to be noncooperatively enforceable (see d'Aspremont and Gérard-Varet, 1998).

Finally, we should point out that in many cases agents are not provided with information exogenously—they need to spend resources in order to acquire information and may choose either to do so or not to do so. There has been a substantial and still increasing body of work on this issue in the principal agent framework (see Crémer and Khalil, 1992; Crémer, Khalil, and Rochet, 1998a,b; Szalay, 2008, and the subsequent literature). It would be of great interest to understand better how the fact that the acquisition of information is endogenous affects the design of multi-agent Bayesian mechanisms.

References

Alchian, A., & Demsetz, H. (1972). Production, information costs and economic organization. *American Economic Review, 62*, 777–805.

Bergemann, D., & Morris, S. (2005). Robust mechanism design. *Econometrica, 73*, 1771–1813.

Chung, K.-S. (1999). A note on Matsushima's regularity condition. *Journal of Economic Theory, 87*, 429–433.

Crémer, J., & Khalil, F. (1992). Gathering information before signing a contract. *American Economic Review, 82*, 566–578.

[13] See Alchian and Demsetz (1972) and Holmström (1982).

Crémer, J., Khalil, F., & Rochet, J.-C. (1998a). Contracts and productive information gathering. *Games and Economic Behavior, 25*, 174–193.

Crémer, J., Khalil, F., & Rochet, J.-C. (1998b). Strategic information gathering before a contract is offered. *Journal of Economic Theory, 81*, 163–200.

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

Crémer, J., & McLean, R. P. (1988). Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica, 56*, 1247–1258.

d'Aspremont, C., & Crémer, J. (2018). Bayesian incentive compatibility with and without free beliefs (in press).

d'Aspremont, C., Crémer, J., & Gérard-Varet, L.-A. (1990). Incentives and the existence of Pareto-optimal revelation mechanisms. *Journal of Economic Theory, 51*, 233–254.

d'Aspremont, C., Crémer, J., & Gérard-Varet, L.-A. (2003). Correlation, independence, and Bayesian incentives. *Social Choice and Welfare, 21*, 281–310.

d'Aspremont, C., Crémer, J., & Gérard-Varet, L.-A. (2004). Balanced Bayesian mechanisms. *Journal of Economic Theory, 115*, 385–396.

d'Aspremont, C., & Gérard-Varet, L.-A. (1979a). Incentives and incomplete information. *Journal of Public Economics, 11*, 25–45.

d'Aspremont, C., & Gérard-Varet, L.-A. (1979b), On Bayesian incentive compatible mechanisms. In J.-J. Laffont (Ed.), *Aggregation and Revelation of Preferences* (pp. 269–288). Amsterdam: North-Holland.

d'Aspremont, C., & Gérard-Varet, L.-A. (1982). Bayesian incentive compatible beliefs. *Journal of Mathematical Economics, 10*, 25–45.

d'Aspremont, C., & Gérard-Varet, L.-A. (1998). Linear inequality methods to enforce partnerships under uncertainty: An overview. *Games and Economic Behavior, 25*, 311–336.

Fan, K. (1956). On systems of linear inequalities. In H. W. Kuhn & A. W. Tucker (Eds.), *Linear inequalities and related systems*. Princeton, NJ: Princeton University Press.

Forges, F., Mertens, J.-F., & Vohra, R. V. (2002). The ex ante incentive compatible core in the absence of wealth effects. *Econometrica, 70*, 1865–1892.

Fudenberg, D., Levine, D. K., & Maskin, E. (1994). The folk theorem with imperfect public information. *Econometrica, 62*, 997–1039.

Green, J., & Laffont, J.-J. (1977). Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica, 45*, 427–438.

Green, J., & Laffont, J.-J. (1979). *Incentives in public decision making*. Amsterdam: North-Holland.

Harsanyi, J. C. (1967). Games with incomplete information played by "Bayesian" players, I-III. Part I. The basic model. *Management Science, 14*, 159–182.

Harsanyi, J. C. (1968a). Games with incomplete information played by "Bayesian" players, Part II. Bayesian equilibrium points. *Management Science, 14*, 320–334.

Harsanyi, J. C. (1968b). Games with incomplete information played by "Bayesian" players, Part III. The basic probability distribution of the game. *Management Science, 14*, 486–502.

Holmström, B. (1977). *On incentive problems in organizations*. Ph.D. thesis, Stanford University.

Holmström, B. (1979). Groves' scheme on restricted domains. *Econometrica, 47*, 1137–1144.

Holmström, B. (1982). Moral hazard in teams. *Bell Journal of Economics, 13*, 324–340.

Holmström, B., & Myerson, R. B. (1983). Efficient and durable decision rules with incomplete information. *Econometrica, 51*, 1799–1819.

Hurwicz, L. (1972). On informationally decentralized systems. In C. B. McGuire, R. Radner, & K. J. Arrow (Eds.) *Decision and organization: A volume in honor of Jacob Marschak* (pp. 297–336). Amsterdam: North-Holland.

Johnson, J. W., Pratt, J. W., & Zeckhauser, R. J. (1990). Efficiency despite mutually payoff-relevant private information: The finite case. *Econometrica, 58*, 873–900.

Kosenok, G., & Severinov, S. (2008). Individually rational, budget-balanced mechanisms and allocation of surplus. *Journal of Economic Theory, 140*, 126–161.

Makowski, L., & Mezzetti, C. (1994). Bayesian and weakly robust first best mechanisms: Characterizations. *Journal of Economic Theory, 64*, 500–519.

Maskin, E. S. (1986). Optimal Bayesian mechanisms. In W. P. Heller, R. M. Starr, & D. A. Starrett (Eds.), *Essays in honor of Kenneth J. Arrow: Volume 3, uncertainty, information, and communication* (pp. 229–238). Cambridge: Cambridge University Press.

Matsushima, H. (1991). Incentive compatible mechanisms with full transferability. *Journal of Economic Theory, 54*, 198–203.

Matsushima, H. (2007). Mechanism design with side payments: Individual rationality and iterative dominance. *Journal of Economic Theory, 133*, 1–30.

Myerson, R., & Satterthwaite, M. (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory, 29*, 265–281.

Szalay, D. (2008). Contracts with endogenous information. *Games and Economic Behaviour, 65*, 586–625.

Vickrey, W. (1961), Counterspeculation, auctions and competitive sealed tenders. *Journal of Finance, 72*, 44–61.

Walker, M. (1978). A note on the characterization of mechanisms for the revelation of preferences. *Econometrica, 46*, 147–152.

Walker, M. (1980). On the nonexistence of a dominant strategy mechanism for making optimal public decisions. *Econometrica, 48*, 1521–1540.