Professor Lars Stole
University of Chicago
June 6, 2019,   (v.1.06)

Part VI: Designing efficient mechanisms

Please do not distribute to anyone outside of class.

1 General framework

We assume throughout that there are n agents, $i = 1, \ldots, n$, each with payoffs that are quasi-linear in money. Player i's payoff is given by

$$u_i(x, \theta_i) - t_i,$$

where $x \in \mathcal{X} \equiv \{x_,, \ldots, x_k\}$ represents a "social state" or "outcome" and $\theta_i \in \Theta_i$ represents player i's preferences over $\mathcal{X}$. We assume that θ_i is private information to player i, but its probability distribution function $f_i(\cdot)$ is commonly known by all. In general we will take expectations without specifying whether or not θ_i is continuously distributed or a discrete distribution, unless it is needed for the proof. We also assume that the types are independently distributed across agents. Thus, $\theta \equiv (\theta_1, \ldots, \theta_n) \in \Theta \equiv \Theta_1 \cdots \Theta_n$ is distributed on Θ according to the probability function $f(\theta) \equiv f_1(\theta_1) \cdots f_n(\cdot)$. (Otherwise, we would typically achieve the full-information outcome by designing Cremer-McLean style mechanisms with side bets.) For now, we do not make any assumptions about single-crossing in (x, θ_i).

Examples:

- Auctions: There are n bidders and the social states $k = 1, \ldots, n$ represent who gets the good;
- Bilateral trade: $x = 0$ corresponds to no trade and $x = 1$ corresponds to trade;
- Public goods: $\{x_0, \ldots, x_m\}$ represent m different mutually exclusive public works projects, where $x = 0$ corresponds to no project.

Our main focus in these notes is in implementing efficient allocations when agents have private information that is relevant for efficiency. [If you are interested more generally in what inefficient (but perhaps profit-maximizing) allocations are implementable using dominant strategies, take a look at Börgers (2015, ch 4). If we have time at the end of these lectures, I will say some more about this.]

1.1 Pareto-efficient allocations

Definition 1. *An allocation $\hat{x} : \Theta \to \mathcal{X}$ is* **ex post efficient** *iff*

$$\hat{x}(\theta) \in \arg\max_{x \in \mathcal{X}} \sum_{i=1}^{n} u_i(x, \theta_i).$$

We will assume in these notes, for simplicity, that there is a unique $\hat{x}$ fo each profile of types, θ.

Note that this allocation is Pareto efficient in a full-information world in the following sense: For any allocation $\tilde{x}$ that is not ex post efficient, there exists a set of transfers $(t_1, \ldots, t_n)$ such that $\sum_i t_i = 0$ and every player is strictly better of with the allocation $(\hat{x}(\theta), t_1, \ldots, t_n)$ than the original social state $\tilde{x}$. This can be proven by construction. Let

$$t_i = u_i(\hat{x}(\theta), \theta_i) - u_i(\tilde{x}, \theta_i) - \frac{1}{n} \sum_{j=1}^{n} \left(u_j(\hat{x}(\theta), \theta_j) - u_i(\tilde{x}, \theta_j) \right).$$

By construction, $\sum_i t_i = 0$. Moreover, for any player i, the payoff under the new allocation is

$$\begin{aligned} u_i(\hat{x}(\theta), \theta_i) - t_i &= u_i(\hat{x}(\theta), \theta_i) - u_i(\hat{x}(\theta), \theta_i) + u_i(\tilde{x}, \theta_i) + \frac{1}{n} \sum_{j=1}^{n} \left(u_j(\hat{x}(\theta), \theta_j) - u_i(\tilde{x}, \theta_j) \right) \\ &= u_i(\tilde{x}, \theta_i) + \frac{1}{n} \sum_{j=1}^{n} \left(u_j(\hat{x}(\theta), \theta_j) - u_i(\tilde{x}, \theta_j) \right). \end{aligned}$$

Because $\hat{x}(\theta)$ maximizes the sum of the agents' utilities, the righthand side is strictly greater than $u_i(\tilde{x}, \theta_i)$, and thus every player i has a strict preference for the new allocation $\hat{x}$ for the given transfers. One can also show that the reverse is true. Given the initial allocation is ex post efficient, there does not exist a new allocation and a set of transfers that can make everyone at least as well off.

2 Two notions of incentive compatibility

Throughout we will consider direct-revelation mechanism of the form $\{\phi(\cdot|\cdot), t_1, \ldots, t_n\}$ where $\phi(x|\hat{\theta})$ gives the probability that social state $x \in \mathcal{X}$ is chosen when the agents report the type profile $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_n)$. Hence, for all $\hat{\theta} \in \Theta$, $\sum_{x \in \mathcal{X}} \phi(x|\hat{\theta}) = 1$ and $\phi(x|\hat{\theta}) \in [0, 1]$ for all $x \in \mathcal{X}$. $t_i : \Theta \to \mathbb{R}$ is the transfer that player i must pay (think of t_i as player i's tax).

As usual, a few definitions will be helpful. For all $i = 1, \ldots, n$, and for all $\theta_i, \hat{\theta}_i \in \Theta_i$, define the following interim functions:

$$\bar{t}_i(\hat{\theta}_i) \equiv E_{\theta_i}[t_i(\hat{\theta}_i, \theta_{-i})],$$

$$\overline{\phi}_i(x|\hat{\theta}_i) \equiv E_{\theta_i}[\phi(x|\hat{\theta}_i, \theta_{-i})],$$
$$U_i(\hat{\theta}_i|\theta_i) \equiv -\overline{t}_i(\hat{\theta}_i) + \sum_{x \in \mathcal{X}} \overline{\phi}_i(x|\hat{\theta}_i) u_i(x, \theta_i),$$
$$U_i(\theta_i) \equiv U_i(\theta_i|\theta_i).$$

2.1 (Bayesian) incentive compatibility (BIC)

Our previous study of auctions and bilateral trade mechanisms focused on a notion of incentive compatibility that is referred to as **Bayesian incentive compatibility (BIC)**. BIC requires that a player is willing to tell the truth given her *expected* payoffs when the *other agents are also telling the truth*. In the present context we have the following definition of BIC:

> **Definition 2.** *A direct mechanism $\{\phi, t_1, \ldots, t_n\}$ is* **Bayesian incentive compatible** *iff for all i*
> $$U_i(\theta) \geq U_i(\hat{\theta}_i|\theta_i), \text{ for all } \theta_i, \hat{\theta}_i \in \Theta_i.$$

The associated revelation principle for this notion of incentive compatibility is the one we have used so far in class. In the present context, the BIC revelation principle follows:

> **Proposition 1. (BIC Revelation Principle)** *Let Γ be an n-player Bayesian-Nash game in which each player chooses a strategy $s_i : \Theta_i \to S_i$, and the strategies determine a distribution, $\phi(\cdot|s_1, \ldots, s_n)$ over $\mathcal{X}$ and a set of payments $\{t_i(s_1, \ldots, s_n)\}_i$. Let $\{s_1^*, \ldots, s_n^*\}$ be an equilibrium of Γ. Denote the equilibrium allocation as*
> $$\phi^*(x|s_1^*(\theta_1), \ldots, s_n^*(\theta_n)), \text{ for all } x \in \mathcal{X} \text{ and } \theta \in \Theta$$
> *and the equilibrium transfers, for each i, as*
> $$t_i^*(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)), \text{ for all } \theta \in \Theta.$$
> *Then there exists a direct-revelation game, $\tilde{\Gamma}$, in which player's strategies are reported types, $\tilde{s}_i : \Theta_i \to \Theta_i$, such that there is a truthful equilibrium (i.e., for all i and $\theta_i \in \Theta_i$, , $\tilde{s}_i^*(\theta_i) = \theta_i$), the equilibrium allocation is*
> $$\tilde{\phi}(x|\theta_1, \ldots, \theta_n) = \phi^*(x|s_1^*(\theta_1), \ldots, s_n^*(\theta_n)), \text{ for all } x \in \mathcal{X} \text{ and } \theta \in \Theta,$$
> *and the equilibrium transfer for each i is*
> $$\tilde{t}_i(\theta_1, \ldots, \theta_n) = t_i^*(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)), \text{ for all } \theta \in \Theta.$$

The proof of the revelation principle is by construction. The direct mechanism is constructed so as to embed the equilibrium strategies of the original game:

$$\tilde{\phi}(x|\theta_1, \ldots, \theta_n) \equiv \phi^*(x|s_1^*(\theta_1), \ldots, s_n^*(\theta_n)), \quad \text{for all } x \in \mathcal{X} \text{ and } \theta \in \Theta,$$

$$\tilde{t}_i(\theta_1, \ldots, \theta_n) \equiv t_i^*(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)), \quad \text{for all } \theta \in \Theta.$$

If all players other than player i report their types truthfully, then player i can achieve the same equilibrium payoffs as in the original by also reporting truthfully. Moreover, any non-truthful report by player i will correspond to choosing the equilibrium strategy of another type of player i in the original game. Because the original allocation is a BNE, it must be that choosing a different strategy is not preferred to reporting truthfully.

Because of the revelation principle, if we are interested in allocations that are achievable as Bayesian-Nash equilibria in a large class of games, we may restrict attention to truthful equilibria in direct-revelation mechanism games of the form $\{\phi, t_1, \ldots, t_n\}$.

2.2 Dominant-strategy incentive compatibility (DSIC)

There is a stronger notion of incentive compatibility than BIC which is *dominant-strategy incentive compatibility* (also known as *strategy-proofness*). The motivation for using a stronger concept is that BNE's rely on common knowledge of the distributions of types, and they require that the players are reasonably sophisticated (recall the difficulty of finding an equilibrium bidding function in a first-price auction, even when distributions are symmetric and uniform). Compare the first-price auction to the second-price auction or the ascending bid auction. In the latter auction formats, it is a dominant strategy to bid your type in the second-price auction, and it is a dominant strategy to keep bidding in the ascending-bid auction as long as the active bid is below your (independent-private) value. In particular, you do not need to know the distribution of types – in fact the players could disagree about the distributions – and the equilibrium strategy does not depend upon whether or not you were bidding against rational bidders. In this sense, the second-price and ascending bid auctions are robust to the details of the environment.

Definition 3. *We say that a direct mechanism* $\{\phi, t_1, \ldots, t_n)$ *is* **dominant-strategy incentive compatible** *iff for all* i*,*

$$\begin{aligned} & -t_i(\theta_i, \hat{\theta}_{-i}) + \sum_{x \in \mathcal{X}} \phi(x|\theta_i, \hat{\theta}_{-i}) u_i(x, \theta_i) \\ & \qquad \geq -t_i(\hat{\theta}_i, \hat{\theta}_{-i}) + \sum_{x \in \mathcal{X}} \phi(x|\hat{\theta}_i, \hat{\theta}_{-i}) u_i(x, \theta_i) \ \ \textit{for all } \theta_i, \hat{\theta}_i \in \Theta_i \textit{ and } \hat{\theta}_{-i} \in \Theta_{-i}. \end{aligned}$$

Notice that DSIC requires IC for player i to hold for any reports of the other players, and not simply in expectation. If DSIC holds for every player, then every player has an incentive to tell the truth. In the truth-telling equilibrium, it will be the case that player i prefers to tell the truth regardless of the profile of other types.

Remarks:

1. DSIC mechanisms are also called **strategy-proof** mechanisms or **straightforward**

mechanisms.

2. BIC requires that i prefers to tell the truth after taking expectations (using a commonly known probability distribution, $f(\cdot)$). DSIC is weaker. *Every DSIC mechanism is a BIC mechanism, but the converse is not true when there are multiple agents.*

3. In the monopoly-screening environment with a single agent of unknown type, BIC and DSIC are equivalent.

4. There is a third notion of IC that is slightly weaker than DSIC and generally much stronger the BIC: *ex post incentive compatibility*. Ex post incentive compatibility requires that it is a dominant-strategy *in equilibrium* for i to report truthfully for any θ_{-i} (i.e., assuming that $\hat{\theta}_{-i} = \theta_{-i}$). That is, if after all of the reports are revealed, if $\hat{\theta}_{-i} = \theta_{-i}$, then player i has no regret about reporting $\hat{\theta}_i = \theta_i$. Research in robust mechanism design has shown that *ex post incentive compatibility* has key properties one would like in an environment where agents have arbitrary beliefs about each others' type distributions. Of course, DSIC implies ex-post IC.

If we wish to restrict attention to dominant-strategy equilibria in a class of games, the DSIC Revelation Principle tells us it is without loss of generality to restrict attention to direct-mechanism games in which truth-telling is a dominant strategy.

Proposition 2. (DSIC Revelation Principle) *Let Γ be an n-player Bayesian-Nash game in which each player chooses a strategy $s_i : \Theta_i \to S_i$, and the strategies determine a distribution, $\phi(\cdot|s_1, \ldots, s_n)$ over $\mathcal{X}$ and a set of payments $\{t_i(s_1, \ldots, s_n)\}_i$. Let $\{s_1^*, \ldots, s_n^*\}$ be a dominant-strategy equilibrium of Γ. Denote the equilibrium allocation as*

$$\phi^*(x|s_1^*(\theta_1), \ldots, s_n^*(\theta_n)), \text{ for all } x \in \mathcal{X} \text{ and } \theta \in \Theta$$

and the equilibrium transfers, for each i, as

$$t_i^*(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)), \text{ for all } \theta \in \Theta.$$

Then there exists a direct-revelation game, $\tilde{\Gamma}$, in which player's strategies are reported types, $\tilde{s}_i : \Theta_i \to \Theta_i$, such that there is a dominant-strategy truthful equilibrium (i.e., for all i and $\theta_i \in \Theta_i$, , $\tilde{s}_i^*(\theta_i) = \theta_i$), the equilibrium allocation is

$$\tilde{\phi}(x|\theta_1, \ldots, \theta_n) = \phi^*(x|s_1^*(\theta_1), \ldots, s_n^*(\theta_n)), \text{ for all } x \in \mathcal{X} \text{ and } \theta \in \Theta,$$

and the equilibrium transfer for each i is

$$\tilde{t}_i(\theta_1, \ldots, \theta_n) = t_i^*(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)), \text{ for all } \theta \in \Theta.$$

The proof here is very similar to the BIC revelation principle. (See MWG, Proposition 23.C.1). Construct the direct mechanism exactly as in the statement so that it matches the equilibrium outcome when agents report truthfully. Let $\hat{\theta}_{-i}$ be any report (possibly untruthful), which corresponds to some $s^*_{-i}(\hat{\theta}_{-i})$ in the original equilibrium. Because $s^*_i(\theta_i)$ weakly optimal by θ_i for any s_{-i}, it is also weakly optimal against the subset of strategies, $s^*_{-i}(\Theta_{-i})$. Hence, player i does best by reporting $\hat{\theta}_i = \theta_i$ and inducing $s^*_i(\theta_i)$.

3 DSIC implementation (VCG mechanisms)

Before we get started, note that if we restrict attention to DSIC mechanisms, the Gibbard-Satterthwaite theorem tells us that there is not much we can implement in general environments (i.e., only dictatorial allocations). Recall that in the proof of GS, however, the assumption that any preferences are possible was required, and so he negative result of GS applies only to settings in which the preference space is incredibly rich. In our present context, we have assumed the agents' preferences are quasi-linear in money. Because of this restriction, the negative result of GS does not apply.

We are most interested in the following question:

> Can we implement the ex post efficient allocation, $\hat{x}(\theta)$, for any $\theta \in \Theta$ using a DISC mechanism?

The answer is "yes". The idea is to construct transfers in such a way that each agent pays an amount equal to the impact of the agent's report on social welfare (agent i pays her social externality), where we evaluate the externality using the reports of the other agents. This is easier understood by showing the construction and verifying that it works. The general construction is due to Groves (1973), so this is sometimes called a Groves mechanism:

$$t_i^g(\hat{\theta}) = -\sum_{j \neq i} u_j(\hat{x}(\hat{\theta}_i, \hat{\theta}_{-i}), \hat{\theta}_i) + h_i(\hat{\theta}_{-i}),$$

where h_i is some arbitrary function that is independent of i's report, $\hat{\theta}_i$.[1] Given this construction, it is straightforward to see that $\{\hat{x}, t_1^g, \ldots, t_n^g\}$ is a DSIC mechanism.

> **Theorem 1.** *$\{\hat{x}, t_1^g, \ldots, t_n^g\}$ is a DSIC mechanism for any $h_i(\cdot)$ that is independent of $\hat{\theta}_i$.*

[1]Because $\hat{x}(\theta)$ is deterministic, the corresponding random allocation requires $\phi(x|\theta) = 1$ iff $x = \hat{x}(\theta)$. Because we want to implement $\hat{x}(\theta)$, we don't have to consider more general, random allocations and can economize on notation by dropping our use of $\phi(x|\theta)$ for now.

Proof: Using t_i^g, we can write i's payoff for any reports $\hat{\theta}_{-i}$ as

$$\begin{aligned} U_i(\hat{\theta}_i|\theta_i, \hat{\theta}_{-i}) &\equiv u_i(\hat{x}(\hat{\theta}_i, \hat{\theta}_{-i}), \theta_i) - t_i^g(\hat{\theta}_i, \hat{\theta}_{-i}) \\ &= -h_i(\hat{\theta}_{-i}) + u_i(\hat{x}(\hat{\theta}_i, \hat{\theta}_{-i}), \theta_i) + \sum_{j \neq i} u_j(\hat{x}(\hat{\theta}_i, \hat{\theta}_{-i}), \hat{\theta}_j). \end{aligned}$$

Notice that the second and third terms on the righthand side reflect the social surplus of the type profile $(\theta_i, \hat{\theta}_{-i})$ when the social state is $\hat{x}(\hat{\theta}_i, \hat{\theta}_{-i})$. By the definition of $\hat{x}$, however, we know that the social surplus for type profile $(\theta_i, \hat{\theta}_{-i})$ is maximized by $\hat{x}(\theta_i, \hat{\theta}_{-i})$ which is obtainable when player i reports truthfully, $\hat{\theta}_i = \theta_i$. Thus,

$$U_i(\theta_i|\theta_i, \hat{\theta}_{-i}) \geq U_i(\hat{\theta}_i|\theta_i, \hat{\theta}_{-i})$$

for any θ_i, $\hat{\theta}_i$ and $\hat{\theta}_{-i}$. We conclude that truth telling is a dominant strategy. □

Remarks:

1. Notice that $\{\hat{x}, t_1^g, \ldots, t_n^g\}$ is DSIC for any $h_i(\cdot)$, providing that h_i is independent of player i's reported type.
2. Green and Laffont (1979) show that if payoffs are linear in money and if the space of preferences over $\mathcal{X}$, $\mathcal{U} \equiv \{u_i(\cdot, \theta_i)\}_{\theta_i \in \Theta_i}$, is sufficiently rich such that every possible valuation function from $\mathcal{X}$ to $\mathbb{R}$ is possible (i.e., θ_i reflects an arbitrary K-tuple of values from $\mathbb{R}^K$ for the K social states), then any ex-post efficient DSIC mechanism must have the form of a Groves mechanism. See MWG, Proposition 23.C.5.
3. Green and Laffont (1979) show under the same hypotheses that any ex-post efficient DSIC mechanism violates budget balance: $\sum_i t_i^g(\theta) \neq 0$ for all $\theta \in \Theta$. See MWG, Proposition 23.C.6.

Clarke (1971) independently discovered the idea of Groves, but with a specific h_i function,

$$h_i(\hat{\theta}_{-i}) = \max_{x \in \mathcal{X}} \sum_{j \neq i} u_j(x, \hat{\theta}_j).$$

Define

$$\hat{x}_{-i}(\hat{\theta}_{-i}) \equiv \arg\max_{x \in \mathcal{X}} \sum_{j \neq i} u_j(x, \hat{\theta}_j).$$

We can thus write h_i as

$$h_i(\hat{\theta}_{-i}) = \sum_{j \neq i} u_j(\hat{x}_{-i}(\hat{\theta}_{-i}), \hat{\theta}_j).$$

Substituting this into the Groves formula, we obtain a mechanism that is referred to as the Vickrey-Clarke-Groves (VCG) mechanism:

$$t_i^{vcg}(\hat{\theta}) \equiv \sum_{j \neq i} u_j(\hat{x}_{-i}(\hat{\theta}_{-i}), \hat{\theta}_j) - \sum_{j \neq i} u_j(\hat{x}(\hat{\theta}_i, \hat{\theta}_{-i}), \hat{\theta}_i).$$

Remarks:

1. Given our previous result for arbitrary h_i, it follows immediately that the VCG mechanism, $\{\hat{x}, t_1^{vcg}, \ldots, t_n^{vcg}\}$, is also DSIC.
2. Notice that the definition of the VCG mechanism is based on implementing the ex post efficient social state, $\hat{x}(\theta)$. We have not said anything yet about DSIC mechanisms for social states that are not ex post efficient.
3. Clarke's choice of h_i yields a payment by player i that exactly reflects the externality that i's report has on the other players. That is, the payment amounts to the change in aggregate welfare of the $j \neq i$ agents moving from $\hat{x}_{-i}(\hat{\theta}_{-i})$, which maximizes their welfare ignoring i, to the allocation $\hat{x}(\hat{\theta})$, which maximizes everyone's utility, including i's. In this sense, t_i^{vcg} charges i for the lost utility to the other players given i's presence in the mechanism.
4. Note that the idea in Clarke's mechanism appears in Vickrey's (1961) second-price auctions. In a second price auction, the winner i is required to pay the externality of her winning, which is the loss to the second-highest type bidder who would have consumed the good had player i not participated in the auction. For this reason, the mechanism above is referred to as the Vickrey-Clarke-Groves (or VCG) mechanism.
5. The VCG mechanism is also referred to as the **pivot mechanism**, because i makes a payment if and only if i's presence is pivotal: i.e., $\hat{x}(\hat{\theta}) \neq \hat{x}_{-i}(\hat{\theta}_{-i})$ given the reports. If i's report does not change the social state, then $t_i = 0$. If i's report changes the social state, i's payment is exactly the cost it imposes on the other players.

   You should *carefully* work through the running example in JR (chapter 9.5) that involves the swimming pool and bridge. Using this idea of constructing payments from pivotal reports allows you to fully characterize t_i^{vcg}.

3.1 Properties of the VCG mechanism

We have established that VCG is ex post efficient and DSIC. There are two other properties worth noting.

Property 1: VCG transfers are nonnegative. Whenever agents are pivotal, $\sum_i t_i^{vcg}(\theta) > 0$.

This can easily be seen by noting that

$$t_i^{vcg}(\hat{\theta}) \equiv \sum_{j \neq i} u_j(\hat{x}_{-i}(\hat{\theta}_{-i}), \hat{\theta}_j) - \sum_{j \neq i} u_j(\hat{x}(\hat{\theta}_i, \hat{\theta}_{-i}), \theta_i)$$

must be nonnegative because $\hat{x}_{-i}(\hat{\theta}_{-i})$ maximizes $\sum_{j\neq i} u_j$ and $\hat{x}(\hat{\theta})$ does not (necessarily). Indeed, if $\hat{x}_{-i}(\hat{\theta}_{-i}) \neq \hat{x}(\hat{\theta})$ (i.e., player i is pivotal), then the transfer must be strictly positive.

Remark: This property implies that a VCG mechanism will generally run a budget surplus. From the point of view of the n agents, *this is bad*. It is inefficient because the mechanism has taken away $\sum_j t_j^{vcg}$ units of utility. Unless that is somehow efficiently used elsewhere (e.g., given to some player $n+1$), it is welfare reducing. Pareto efficiency requires not only that $x = \hat{x}(\theta)$, but that $\sum_i t_i = 0$ (budget balance); the VCG mechanism is inefficient because it wastes money. Recall from Green and Laffont (1979) that any Groves mechanism (and this includes the specific VCG mechanism) fails to achieve budget balance (BB). Below, we will see that this is a consequence of DSIC; if we only require that the mechanism satisfy BIC, then it is possible to achieve BB.

> **Property 2:** Suppose that if player i does not participate in the VCG mechanism, then the VCG mechanism is played for the remaining $n-1$ players, $\hat{x}_{-i}(\theta_{-i})$ is implemented, and player i pays nothing. In this case, it is an equilibrium for all agents to voluntarily participate in a VCG mechanism.

To see this, consider player i. If player i does not participate, the VCG mechanism implements $\hat{x}_{-i}(\theta)$ and player i obtains the payoff $u_i(\hat{x}_{-i}(\theta), \theta_i)$. If player i participates, in the truth-telling equilibrium, player i obtains

$$u_i(\hat{x}(\theta), \theta_i) + \sum_{j\neq i} u_j(\hat{x}(\theta), \theta_j) - \sum_{j\neq i} u_j(\hat{x}_{-i}(\theta_{-i}), \theta_j).$$

The difference between participating and not participating is therefore

$$\begin{aligned}\left(u_i(\hat{x}(\theta), \theta_i) + \sum_{j\neq i} u_j(\hat{x}(\theta), \theta_j) - \sum_{j\neq i} u_j(\hat{x}_{-i}(\theta_{-i}), \theta_j)\right) - u_i(\hat{x}_{-i}(\theta), \theta_i)& \\ = \sum_{j=1}^{n} u_j(\hat{x}(\theta), \theta_j) - \sum_{j=1}^{n} u_j(\hat{x}_{-i}(\theta_{-i}), \theta_j),&\end{aligned}$$

which is nonnegative (and strictly positive if i's participation is pivotal).

The intuition is that the government (or mechanism designer) will go ahead and choose the social state with or without i's participation. If i participates, i has the opportunity to be pivotal which will improve i's payoff. If i's participation is not pivotal, then i does not care about participating because $t_i = 0$ in that case.

Remark: The preceding result is entirely built on the assumption that the designer has full control rights over $x \in \mathcal{X}$. If an agent controls some aspect of $\mathcal{X}$ and by withdrawing participation can prevent the designer from choosing some $x \in \mathcal{X}$, then it will no longer be the case that an agent is always willing to participate. For example, consider the Myerson-Satterthwaite bilateral trade game. The social states are $x = 0$ (no trade) and $x = 1$ (trade). If the seller chooses not to participate in that game, the designer cannot force $x = 1$. This is very different from the current assumption where we allow the designer full control.

We will introduce agent control (i.e., agent property rights) over $\mathcal{X}$ below. This will allow us to place the bilateral trade model of MS into a broader framework with VCG mechanisms.

3.2 Example: Bilateral trade

Consider the bilateral trade model (one buyer and one seller) in which the designer can force trade ($x = 1$) or not ($x = 0$) as a function of reports. Types are uniformly distributed on $[0, 1]$. We take the utility functions to be $u_b(x, \theta_b) = x\theta_b$ and $u_s(x, \theta_s) = -x\theta_s$.[2] What does the VCG mechanism look like? The ex post efficient trading allocation is $\hat{x}(\theta_b, \theta_s) = 1$ iff $\theta_b \geq \theta_s$. Using our formula for transfer t_i^{vcg}, we have

$$t_b^{vcg}(\hat{\theta}_b, \hat{\theta}_s) = u_s(0, \hat{\theta}_s) - u_s(\hat{x}(\hat{\theta}_b, \hat{\theta}_s), \hat{\theta}_s) = \hat{\theta}_s \hat{x}(\hat{\theta}_b, \hat{\theta}_s) \geq 0,$$

$$t_s^{vcg}(\hat{\theta}_b, \hat{\theta}_s) = u_b(1, \hat{\theta}_b) - u_b(\hat{x}(\hat{\theta}_b, \hat{\theta}_s), \hat{\theta}_b) = \hat{\theta}_b(1 - \hat{x}(\hat{\theta}_b, \hat{\theta}_s)) \geq 0.$$

Notice in the above expressions that if the buyer is absent, $\hat{x}_{-b}(\hat{\theta}_s) = 0$, while if the seller is absent, $\hat{x}_{-s}(\hat{\theta}_b) = 1$. Also note that we are using the convention that each t_i^{vcg} is a payment to the designer from agent i. In particular, t_b is not a payment to the seller, and t_s is a nonnegative *payment* to the designer (not a nonnegative transfer that is received by the seller).

As a verification of our earlier result, let's check that this mechanism is DSIC for the buyer. The buyer's payoff under the mechanism will be

$$\hat{x}(\hat{\theta}_b, \hat{\theta}_s)\theta_b - t_b^{vcg}(\hat{\theta}_b, \hat{\theta}_s) = \hat{x}(\hat{\theta}_b, \hat{\theta}_s)(\theta_b - \hat{\theta}_s),$$

which is maximized when $\hat{\theta}_b = \theta_b$. A similar argument verifies DSIC for the seller; the seller's payoff under the mechanism will be

$$-\hat{x}(\hat{\theta}_b, \hat{\theta}_s)\theta_s - t_s^{vcg}(\hat{\theta}_b, \hat{\theta}_s) = -\hat{\theta}_b + \hat{x}(\hat{\theta}_b, \hat{\theta}_s)(\hat{\theta}_b - \theta_s).$$

As with the buyer, it is a dominant strategy for the seller to report $\hat{\theta}_s = \theta_s$. Finally, note that if the seller does not participate, then the buyer gets the good and the seller suffers $-\theta_s$. If the seller participates, however, the seller earns

$$\hat{x}(\theta_b, \theta_s)(-\theta_s) + \theta_b(1 - \hat{x}(\theta_b, \theta_s)).$$

[2] Note that our choice of payoff for the seller implicitly treats the seller's type as the cost of producing the good, as in Myerson and Satterthwaite (1983). We could also take the seller's utility to be the payoff from consuming the good herself, θ_s, rather than giving it to the buyer. In this case, it would be natural to write the seller's payoff as $u_s(x, \theta_b) = (1 - x)\theta_s = \theta_s - x\theta_s$. The is how the bilateral trade payoff to the seller is modeled in JR, Chapter 9. Regardless of how we write the seller's preferences, however, the VCG transfer will be the same because VCG transfers are based on a comparison of utilities across social states, so the extra θ_s in the second expression will net out. Verify this for yourself.

Computing the net utility from participating, we have

$$\hat{x}(\theta_b, \theta_s)(\theta_b - \theta_s) \geq 0.$$

4 BIC implementation (EE/AGV mechanisms)

We now return to our typical setting of Bayesian incentive constraints and begin by asking whether or not we can achieve budget balance (BB) by replacing the DSIC requirement with the weaker notion of BIC. The answer is "yes" and is an implication of a more general result which we state and prove now.

Theorem 2. *Let $\{\phi, t_1, \ldots, t_n\}$ be any BIC mechanism that runs an expected budget surplus*

$$E_\theta \left[\sum_i t_i(\theta) \right] \geq 0.$$

Then there exists another BIC mechanism, $\{\tilde{\phi}, \tilde{t}_1, \ldots, \tilde{t}_n\}$, such that $\sum_i \tilde{t}_i(\theta) = 0$ for all $\theta \in \Theta$ (BB) and is weakly preferred by every agent. Moreover, if the expected surplus in the original mechanism is strictly positive, then $\{\tilde{\phi}, \tilde{t}_1, \ldots, \tilde{t}_n\}$ is strictly preferred by every agent i and type θ_i.

Proof: We prove this by construction. Define

$$\bar{t}_i(\theta_i) \equiv E_{\theta_{-i}}[t_i(\theta_i, \theta_{-i})]$$

and

$$\bar{t}_i \equiv E_\theta[t_i(\theta)].$$

Construct for $i < n$,

$$\tilde{t}_i(\theta) \equiv \bar{t}_i(\theta_i) + \left(\bar{t}_{i+1} - \bar{t}_{i+1}(\theta_{i+1})\right) - \frac{1}{n} \sum_{j=1}^{n} \bar{t}_j,$$

and for $i = n$ construct

$$\tilde{t}_n(\theta) \equiv \bar{t}_n(\theta_n) + \left(\bar{t}_1 - \bar{t}_1(\theta_1)\right) - \frac{1}{n} \sum_{j=1}^{n} \bar{t}_j,$$

By construction, $\sum_i \tilde{t}_i(\theta) = 0$:

$$\sum_i \tilde{t}_i(\theta) = \left(\sum_{i=1}^{n} \bar{t}_i(\theta_i) \right) + \sum_{i=1}^{n} \bar{t}_i - \left(\sum_{i=0}^{n-1} \bar{t}_{i+1}(\theta_{i+1}) \right) - \sum_{i=1}^{n} \bar{t}_i = 0.$$

Now consider the payoff to agent i with type θ_i when reporting $\hat{\theta}_i$ in the original mechanism:

$$U_i(\hat{\theta}_i | \theta_i) = \sum_{x \in \mathcal{X}} \overline{\phi}_i(x | \hat{\theta}_i) u_i(x, \theta_i) - \bar{t}_i(\hat{\theta}_i).$$

Contrast this with the payoff agent i with type θ_i receives when reporting $\hat{\theta}_i$ in the new mechanism:

$$\begin{aligned}
\tilde{U}_i(\hat{\theta}_i|\theta_i) &= \sum_{x\in\mathcal{X}} \overline{\phi}_i(x|\hat{\theta}_i)u_i(x,\theta_i) - E_{\theta_{-i}}[\tilde{t}_i(\hat{\theta}_i,\theta_{-i})] \\
&= U_i(\hat{\theta}_i|\theta_i) + \overline{t}(\hat{\theta}_i) - E_{\theta_{-i}}[\tilde{t}_i(\hat{\theta}_i,\theta_{-i})] \\
&= U_i(\hat{\theta}_i|\theta_i) + E[\overline{t}_{i+1}(\theta_{i+1}) - \overline{t}_{i+1}] + \frac{1}{n}\sum_{j=1}^{n}\overline{t}_j \quad \text{(where } i+1=1 \text{ if } i=n\text{)} \\
&= U_i(\hat{\theta}_i|\theta_i) + \frac{1}{n}\sum_{j=1}^{n}\overline{t}_j.
\end{aligned}$$

Thus, we conclude

$$\tilde{U}_i(\hat{\theta}_i|\theta_i) = U_i(\hat{\theta}_i|\theta_i) + \frac{1}{n}\sum_{j=1}^{n}\overline{t}_j.$$

Because $\tilde{U}_i$ and U_i differ by a constant, if $\{\phi, t_1, \ldots, t_n\}$ is BIC, then $\{\phi, \tilde{t}_1, \ldots, \tilde{t}_n\}$ is also BIC.

Moreover, comparing the truth-telling equilibrium payoffs in each mechanism, we have

$$\tilde{U}_i(\theta_i) - U_i(\theta_i) = \frac{1}{n}\sum_{j=1}^{n}\overline{t}_j.$$

If the original mechanism runs a positive budget surplus in expectation, $\frac{1}{n}\sum_{j=1}^{n}\overline{t}_j > 0$, then all agent types (i.e., all i and all $\theta_i \in \Theta_i$) strictly prefer the new mechanism $\{\phi, \tilde{t}_1, \ldots, \tilde{t}_n\}$ over the original mechanism $\{\phi, t_1, \ldots, t_n\}$. □

Remarks:

1. This result is not limited to ex post efficient mechanisms and applies to any ϕ allocation for which $\{\phi, t_1, \ldots, t_n\}$ is BIC and $E\left[\sum_i t_i(\theta)\right] \geq 0$ for all $\theta \in \Theta$.
2. Because the VCG mechanism is DSIC, it is also trivially BIC. Hence, we can construct a budget-balanced, BIC mechanism that is weakly preferred by all agents to the original VCG mechanism. This deserves special attention.

Corollary 1. *There exists a budget-balanced, Bayesian incentive compatible mechanism which implements the ex post efficient allocation.*

Let's construct the ex post efficient, BIC mechanism that balances the VCG mechanism using the construction in Theorem 2. We'll call this the **expected-externality (EE) mechanism**; the moniker will become clear shortly:

Definition 4. *The* **Expected Externality (EE) mechanism** *is an ex post efficient, budget-balanced, BIC mechanism with payments for* $i = 1, \ldots, n$

$$t_i^{ee}(\theta) \equiv \bar{t}_i^{vcg}(\theta_i) + \left(\bar{t}_{i+1}^{vcg} - \bar{t}_{i+1}^{vcg}(\theta_{i+1})\right) - \frac{1}{n}\sum_{j=1}^{n} \bar{t}_j^{vcg},$$

where

$$\bar{t}_i^{vcg}(\theta_i) \equiv E_{\theta_{-i}}[t_i^{vcg}(\theta_i, \theta_{-i})], \quad \textit{and} \quad \bar{t}_i^{vcg} \equiv E_\theta[t_i^{vcg}(\theta)].$$

Remarks:

1. Why is it the *expected externality mechanism*? Observe that once we substitute for t_i^{vcg} and take expectations over θ_{i+1}, we can write i's payment as

$$\boxed{\bar{t}_i^{ee}(\theta_i) = E_{\theta_{-i}}\left[\sum_{j \neq i} u_j(\hat{x}_{-i}(\theta_{-i}), \theta_j) - \sum_{j \neq i} u_j(\hat{x}(\theta), \theta_j)\right] - \left(\frac{1}{n}\sum_{j=1}^{n} \bar{t}_j^{vcg}\right), \qquad (1)}$$

   which is simply the expected externality that i imposes on the other $n-1$ participants, minus a social dividend equal to the sum of the player's expected payments.

2. The EE mechanism is of fundamental importance in allocation problems. It was first studied by d'Aspremont and Gerard-Varet, and is often called an **AGV mechanism**. It was also independently noted by Arrow, so arguably AAVG is a better acronym. Because the structure of the payments require that player i now pays her *expected* externality to the remaining $(n-1)$ agents, the mechanism is frequently called the **Expected Externality (EE) mechanism**, which is the phrase we will use.

3. There are alternative derivations of t_i^{ee} which differ only in how the budget surplus is divided among the agents. These alternative derivations have different transfers than $t_i^{ee}(\theta)$, but have the identical interim payment functions, $\bar{t}_i^{ee}(\theta_i)$. In Definition 4 and in JR, chapter 9, the division is done by having agent i take the transfer of the next agent in line, $i+1$, while paying back the expected value. In the original paper and in MWG, the division is implemented by having agent i take an equal share of all the other $n-1$ agent's transfers, while paying back the expected value. In both derivations, after taking expectations, the interim payments are identical and as in (1), which is what we really care about in any economic analysis. Hence, the approaches are economically equivalent. JR, Exercise 9.29 asks you to prove this equivalence.

4. EE mechanisms can solve a lot of interesting economic problems. For example, suppose that we are back in the Myerson-Satterthwaite bilateral trade example, but the agents can commit to the mechanism *before* learning their types. In this case, they can implement the first best by constructing an expected-externality mechanism. At

the ex ante stage, one of the parties may have a negative expected payoff from the mechanism, but in this case the other party can make an ex ante payment to satisfy everyone's ex ante IR constraints. (We know such a payment is possible because the sum of the players ex ante expected payoffs is positive as long as it is efficient to sometimes trade.)

Example: Returning to our bilateral trade example, recall that the VCG payments are

$$t_b^{vcg}(\theta_b, \theta_s) = \theta_s \hat{x}(\theta_b, \theta_s),$$
$$t_s^{vcg}(\theta_b, \theta_s) = \theta_b(1 - \hat{x}(\theta_b, \theta_s)).$$

Taking expectations, we have

$$\bar{t}_b^{vcg}(\theta_b) = \int_0^{\theta_b} \theta_s d\theta_s = \frac{1}{2}\theta_b^2,$$
$$\bar{t}_s^{vcg}(\theta_s) = \int_0^{\theta_s} \theta_b d\theta_b = \frac{1}{2}\theta_s^2.$$

Taking expectations again, we have

$$\bar{t}_b^{vcg} = \int_0^1 \frac{1}{2}\theta_b^2 d\theta_b = \frac{1}{6},$$
$$\bar{t}_s^{vcg} = \int_0^1 \frac{1}{2}\theta_s^2 d\theta_s = \frac{1}{6}.$$

Using our formula for $t_i^{ee}(\theta)$, we have

$$\begin{aligned} t_b^{ee}(\theta_b, \theta_s) &= \bar{t}_b^{vcg}(\theta_b) + \bar{t}_s^{vcg} - \bar{t}_s^{vcg}(\theta_s) - \frac{1}{2}(\bar{t}_b^{vcg} + \bar{t}_s^{vcg}) \\ &= \frac{1}{2}\theta_b^2 + \frac{1}{6} - \frac{1}{2}\theta_s^2 - \frac{1}{2}\left(\frac{1}{6} + \frac{1}{6}\right) \\ &= \frac{1}{2}\theta_b^2 - \frac{1}{2}\theta_s^2. \end{aligned}$$

Similarly,

$$t_s^{ee}(\theta_b, \theta_s) = \frac{1}{2}\theta_s^2 - \frac{1}{2}\theta_b^2.$$

Let's verify BIC for the buyer. The buyer's payoff is

$$U_b(\hat{\theta}_b|\theta_b) = \hat{\theta}_b\theta_b - \frac{1}{2}\hat{\theta}_b^2 + \frac{1}{2}\theta_s^2,$$

which is maximized at $\hat{\theta}_b = \theta_b$. What about ex ante individual rationality? Let's compute the expected payoffs to the EE mechanism, $\{\hat{x}, t_b^{ee}, t_s^{ee}\}$. The buyer obtains (in expectation)

$$E\left[\frac{1}{2}\theta_b^2 + \frac{1}{2}\theta_s^2\right] = \frac{1}{3}.$$

The seller's payoff is

$$U_s(\hat{\theta}_s|\theta_s) = -(1-\hat{\theta}_s)\theta_s - \frac{1}{2}\hat{\theta}_s^2 + \frac{1}{2}\hat{\theta}_b^2,$$

which is maximized at $\hat{\theta}_s = \theta_s$ regardless of $\hat{\theta}_b$. Hence, it is DISC. The seller with type θ_s obtains

$$(1-\theta_s)(-\theta_s) - \frac{1}{2}\theta_s^2 + \frac{1}{2}E[\theta_b^2] = \frac{1}{6} - \frac{1}{2}\theta_s^2 - (1-\theta_s)\theta_s.$$

Taking expectations over θ_s, we have an interim expected payoff to the seller of

$$E_{\theta_s}\left[\frac{1}{6} - \frac{1}{2}\theta_s^2 - (1-\theta_s)\theta_s\right] = -\frac{1}{6} < 0.$$

These two expected surpluses add up to the surplus generated by the efficient trading rule, $E[\max\{\theta_b - \theta_s, 0\}] = \frac{1}{6}$. Because the buyer's ex ante surplus exceeds the seller's ex ante loss, the buyer can make an ex ante payment to the seller (e.g., $\frac{1}{4}$) which leaves both sides with positive expected gains to playing the EE mechanism once they learn their types. Thus, if the parties can agree to contract before learning types, they can achieve efficient trade even if the players have control over $\mathcal{X}$, as in the original MS bilateral trade setting.

5 Individually-rational, budget-balanced, BIC mechanisms

In this section, we will continue to explore BIC mechanisms, but we now suppose that each agent i has a type-dependent interim IR constraint given by $\underline{U}_i(\theta_i)$. This IR constraint may capture an exogenous requirement for agent payoffs, or it may capture the value the agent could obtain by exercising control over some components of $\mathcal{X}$.

As a motivating example, in the Myerson-Satterthwaite (1983) bilateral trade model, the seller has the option to not participate and keep the good from being transferred to the buyer. This control over the social state $x = 1$ can be modeled by requiring that the designer guarantee the seller a type-dependent payoff in the mechanism that reflects the seller's outside option of non-participation.

To avoid confusion, note that there are two equally valid ways to incorporate individual rationality constraints in the bilateral trade example. The variation depends on whether we view θ_s as the opportunity cost of producing the good or view θ_s as the value to the seller of keeping the good. (The opportunity cost, properly defined, of course captures exactly the lost value to the seller, so these approaches are equivalent.) In the cost-based approach (analogous to Myerson and Satterthwaite), the seller's payoff in the mechanism is

$$-\hat{x}(\theta_b, \theta_s)\theta_s - t_s,$$

and this must exceed the outside option of zero, $\underline{U}_s(\theta_s) = 0$ (not producing the good). Effectively, we have embedded the outside option into the payoff function by using the correct notion of opportunity cost. This is how we have been treating the MS bilateral trade model in these notes. Alternatively, in the value-based approach (analogous to the approach in JR, chapter 9), the seller obtains

$$(1 - \hat{x}(\theta_b, \theta_s))\theta_s - t_s,$$

but this needs to be weighed against not participating and keeping the good with probability 1, yielding a value of $\underline{U}_s(\theta_s) = \theta_s$. In both cases, the *net utility* to participating is the same. You can choose either one, but you will need to be consistent.

Definition 5. *A mechanism* $\{\phi, t_i, \ldots, t_n\}$ *is* **(interim) IR** *with respect to the outside options,* $\{\underline{U}_i, \ldots, \underline{U}_n\}$*, iff*

$$\sum_{x \in \mathcal{X}} \overline{\phi}_i(x|\theta_i) u_i(x, \theta_i) - \overline{t}_i(\theta_i) \geq \underline{U}_i(\theta_i), \ \ \textit{for all } \theta_i \in \Theta_i.$$

We are going to modify the VCG mechanism in order to guarantee the IR constraints are satisfied. Define the interim payoff of agent i with type θ_i who plays the original VCG mechanism.

$$U_i^{vcg}(\theta_i) \equiv E_{\theta_{-i}}\left[u_i(\hat{x}(\theta_i, \theta_{-i}), \theta_i) - t_i^{vcg}(\theta_i, \theta_{-i})\right].$$

Next, define the minimum payment one must give i at the ex ante stage (before learning θ_i) to guarantee that i's interim IR constraint is satisfied once she learns her type. That is, we want to find the smallest ψ_i such that

$$U_i^{vcg}(\theta_i) + \psi_i \geq \underline{U}_i(\theta_i), \ \ \text{for all } \theta_i \in \Theta_i.$$

This minimum value is defined by

$$\psi_i^* \equiv \max_{\theta_i \in \Theta_i} \underline{U}_i(\theta_i) - U_i^{vcg}(\theta_i).$$

Note that $\{\psi_1^*, \ldots, \psi_n^*\}$ are constants and are type independent. They are computed using the type distributions, however. Paying ψ_i^* guarantees that the interim IR constraint for player i is satisfied when playing the modified VCG mechanism.

We are now ready to define the IR-VCG mechanism for a given set of outside options, $\{\underline{U}_i, \ldots, \underline{U}_n\}$:

Definition 6. *The* **individually-rational VCG mechanism**, $\{\hat{x}, t_1^{ir}, \ldots, t_n^{ir}\}$, *implements the ex post efficient allocation* $\hat{x}$ *with transfers*

$$t_i^{ir}(\theta) \equiv t_i^{vcg}(\theta) - \psi_i^*,$$

where

$$\psi_i^* \equiv \max_{\theta_i \in \Theta_i} \underline{U}_i(\theta_i) - E_{\theta_{-i}}\left[u_i(\hat{x}(\theta_i, \theta_{-i}), \theta_i) - t_i^{vcg}(\theta_i, \theta_{-i})\right].$$

Remarks:

1. The IR-VCG mechanism is ex post efficient and (by construction) it is interim-IR relative to the outside options $\{\underline{U}_i, \ldots, \underline{U}_n\}$.
2. Because the IR-VCG mechanism differs from the VCG mechanism only in the constants $\{\psi_1^*, \ldots, \psi_n^*\}$, it is also DSIC.
3. Generally speaking, the IR-VCG mechanism is not going to satisfy budget balance.
4. We can rewrite the definition of ψ_i^* as

$$\psi_i^* = \max_{\theta_i \in \Theta_i} \underline{U}_i(\theta_i) - U_i^{vcg}(\theta_i),$$

where $U_i^{vcg}(\theta_i) \equiv E_{\theta_{-i}}\left[u_i(\hat{x}(\theta_i, \theta_{-i}), \theta_i) - t_i^{vcg}(\theta_i, \theta_{-i})\right]$. As we discussed in footnote 2, in some settings like bilateral trade, there may be two equally valid ways to write an agent's payoff function (e.g., $u_s = -x\theta_s$ versus $u_s = (1-x)\theta_s$), but for each of the approaches, there will correspond different outside option functions (e.g., $\underline{U}_s(\theta_s) = 0$ and $\underline{U}_s(\theta_s) = \theta_s$, respectively) so that the difference $\underline{U}_i(\theta_i) - U_i^{vcg}(\theta_i)$ is identical in the two approaches.

5.1 Sufficiency of IR-VCG expected surplus for BB-IR-BIC mechanisms

The IR-VCG mechanism is BIC (because it is DSIC). If it runs an expected budget surplus, we can apply Theorem 2. This allows us to construct an IR-expected-externality mechanism that is ex post efficient, BIC, and weakly preferred by all agents and all types. Because of the last property, we know the new mechanism will also satisfy IR. Thus, we have an important and useful corollary to Theorem 2.

Corollary 2. (Sufficiency of IR-VCG expected surplus). *Suppose that the IR-VCG mechanism runs an expected surplus,*

$$E_\theta\left[\sum_i t_i^{ir}(\theta)\right] = E_\theta\left[\sum_i t_i^{vcg}(\theta) - \psi_i^*\right] \geq 0,$$

then there exists a BIC mechanism that is budget balanced, individually rational, and that is weakly preferred by all types of all agents.

Remark:

1. This corollary is very useful! The corollary says that a sufficient condition for the existence of an ex post efficient, BIC, IR mechanism is that the IR-adjusted VCG mechanism runs an expected budget surplus.
2. Of course in the bilateral trading game in MS, using the seller's IR constraint of $\underline{U}_s = \theta_s$, the required ψ_s^* is so high that the IR-VCG mechanism runs an expected deficit.

Example: Dissolving a partnership efficiently. As an example of the usefulness of the corollary, take a look at JR, Exercise 9.36. This is based on a paper by Cramton, Gibbons and Klemperer (*Ema*, 1987) entitled "Dissolving a partnership efficiently." One can show that if n agents all own equal shares of a venture ($s_i = \frac{1}{n}$, or even close to equal shares), and each agent's outside option is $s_i\theta_i$, then there exists a BB, BIC and IR mechanism that dissolves the partnership efficiently (ie., sells all of the ownership shares to the agent with the highest θ_i). You can prove this by showing that the ex post efficient IR-VCG mechanism runs an expected surplus if each agent's share is sufficiently close to $\frac{1}{n}$.

Recall the partnership-dissolution setting that we considered briefly in class. There, we showed that efficient dissolution is possible for $n = 2$ when θ_i distributed uniformly on $[0, 1]$ for each type and when each agent has originally owns half the shares. We did this by checking that the following is ex post efficient, BIC and IR.

- the partnership is sold entirely to player 1 if $\theta_1 \geq \theta_2$, and is sold to player 2 otherwise; if player i obtains control
- player i pays player j the amount $t_i(\hat{\theta}_1, \hat{\theta}_2) = \frac{1}{3}\hat{\theta}_i > 0$ and if player j obtains control, then player i pays player j the amount $t_i(\hat{\theta}_1, \hat{\theta}_2) = -\frac{1}{3}\hat{\theta}_2 < 0$ (negative, so this is actually a payment to i from j).

We checked in class that this is BIC and IR (and it is clearly BB). Where did the mechanism come from? Why does it work?

To answer these questions, let's apply our results so far. Define the ex post efficient allocation as $\hat{x}(\theta_1, \theta_2) = 1$ if $\theta_1 \geq \theta_2$ and let $\hat{x} = 0$ otherwise. Using our definition, the VCG payments to implement the efficient partnership dissolution are given by

$$t_1^{vcg}(\hat{\theta}_1, \hat{\theta}_2) = u_2(0, \hat{\theta}_2) - u_2(\hat{x}(\hat{\theta}_1, \hat{\theta}_2), \hat{\theta}_2) = \hat{x}(\hat{\theta}_1, \hat{\theta}_2)\hat{\theta}_2,$$

$$t_2^{vcg}(\hat{\theta}_1, \hat{\theta}_2) = u_1(1, \hat{\theta}_1) - u_1(\hat{x}(\hat{\theta}_1, \hat{\theta}_2), \hat{\theta}_1) = (1 - \hat{x}(\hat{\theta}_1, \hat{\theta}_2))\hat{\theta}_1.$$

We want to compute each agent's interim payoff from participating in the VCG mechanism. Taking interim expectations, we have

$$\bar{t}_i^{vcg}(\theta_1, \theta_2) = \frac{1}{2}\theta_i^2.$$

From here, we can determine interim VCG payoffs for each player as

$$U_i^{vcg}(\theta_i) = \text{Prob}[\theta_j < \theta_i]\theta_i - \bar{t}_i^{vcg}(\theta_i) = \theta_i^2 - \frac{1}{2}\theta_i^2 = \frac{1}{2}\theta_i^2.$$

We next compute the minimal payment to satisfy agent i's IR constraint which is

$$\psi_i^* = \max_{\theta_i \in [0,1]} \underline{U}_i(\theta_i) - U_i^{vcg}(\theta_i) = \max_{\theta_i \in [0,1]} \frac{1}{2}\theta_i - \frac{1}{2}\theta_i^2 = \frac{1}{8}.$$

Thus, the required (interim) IR-VCG payments are

$$\bar{t}_i^{ir}(\theta_i) = \bar{t}_i^{vcg}(\theta_i) - \psi_i^* = \frac{1}{2}\theta_i^2 - \frac{1}{8}.$$

Taking expectations over θ_i, we see that

$$\bar{t}_i^{ir} = E[\bar{t}_i^{ir}(\theta_i)] = \frac{1}{6} - \frac{1}{8} > 0,$$

hence, $\sum_i \bar{t}_i^{ir} > 0$, and we have a strictly positive expected surplus. We can therefore conclude, using Corollary 2, that there exists an ex post efficient, BIC, IR mechanism. Indeed, the corresponding budget-balanced, expected-externality (EE) interim transfers for the IR-VCG mechanism are

$$t_i^{ee}(\theta) = \bar{t}_i^{ir}(\theta_i) - (\bar{t}_{-i}^{ir}(\theta_{-i}) - \bar{t}_{-i}^{ir}) - \frac{1}{2}\sum_j \bar{t}_j^{ir}.$$

Therefore, the interim value of these transfers are

$$\bar{t}_i^{ee}(\theta_i) = \frac{1}{2}\theta_i^2 - \frac{1}{6}.$$

Now reconsider the mechanism from class. In that mechanism, player 1 pays $\frac{1}{3}\theta_1$ whenever he receives the partnership and receives $\frac{1}{3}\theta_2$ whenever he loses the partnership. The expected transfer for player 1 is

$$E_{\theta_2}[t_1(\theta_1, \theta_2)] = \theta^1\left(\frac{1}{3}\theta_1\right) + (1 - \theta_1)\left(-\frac{1}{3}E[\theta_2 | \theta_2 \geq \theta_1]\right) = \frac{1}{2}\theta_i^2 - \frac{1}{6}.$$

But this is exactly the same as $\bar{t}_i^{ee}(\theta_i)$, and thus the example in class is a form of the EE mechanism that improves on the IR-VCG mechanism. (Note that the transfers in class were modified with the property that payments go in the opposite direction of the shares, but the same expected EE payments are made.)

5.2 Necessity of IR-VCG expected surplus for BB-IR-BIC mechanisms

There is a converse of the corollary for some environments. The converse holds that if the IR-VCG mechanism runs an expected budget deficit, then there does not exist an ex post efficient, BIC, IR mechanism. In other words, a **necessary condition** for the existence of an ex post efficient, BIC and IR mechanism is that the IR-VCG mechanism runs an expected budget surplus.

Note that if the converse is true, and if we establish that the IR-VCG bilateral trade mechanism runs an expected deficit, then we will have established an alternative proof to the MS impossibility theorem.

The proof of the necessary statement uses the standard integral condition implied by incentive compatibility. In discrete-type settings, unfortunately, BIC does not imply an analogous summation condition because we generally do not know which adjacent IC constraints (upper or lower) are binding. Indeed, the converse does not hold for discrete-type settings, but it does hold for continuous-type settings with some additional structure (e.g., one-dimensional type, multiple dimensional type if utility is linear in types).

For simplicity, let's consider the one-dimensional type case. That is, assume for each i that θ_i is distributed according to F_i on the support $\Theta_i = [\underline{\theta}_i, \overline{\theta}_i]$. We want to prove

Theorem 3. *Suppose that for each i, θ_i is distributed continuously on support $[\underline{\theta}_i, \overline{\theta}_i] \subset \mathbb{R}$ and $\frac{\partial}{\partial \theta_i} u_i(x, \theta_i)$ is bounded on Θ_i. An ex post efficient, budget-balanced, individually rational, BIC mechanism exists if and only if the IR-VCG mechanism runs an expected budget surplus.*

Proof: Corollary 2 establishes the sufficiency of the theorem. To prove necessity, note that because $\frac{\partial}{\partial \theta_i} u_i(x, \theta_i)$ is bounded on Θ_i, the envelope theorem implies that for any mechanism, $\{\hat{x}, t_1, \ldots, t_n\}$, the associated indirect utility function $U_i(\theta_i)$ is absolutely continuous and

$$U_i(\theta_i) = U_i(\underline{\theta}) + \int_{\underline{\theta}_i}^{\theta_i} E_{\theta_{-i}} \left[\frac{\partial}{\partial \theta_i} u_i(\hat{x}(s, \theta_{-i}), s) \right] ds.$$

Hence, any two BIC mechanisms that implement $\hat{x}(\cdot)$, say $\{\hat{x}, t_1, \ldots, t_n\}$ and $\{\hat{x}, \tilde{t}_1, \ldots, \tilde{t}_n\}$, differ in terms of the utility that they give player i by a fixed constant, δ: i.e., $U_i(\theta_i) = \tilde{U}_i(\theta_i) + \delta$ for all θ_i. But the IR-VCG mechanism was designed to have to smallest payment ψ_i^* so that $U_i^{ir}(\theta_i) \geq \underline{U}_i(\theta_i)$ with equality for some type(s). Because of this, any other BIC mechanism must give the same or higher level of utility to player i. Because for any BIC mechanism transfers are negatively related to $U_i(\underline{\theta}_i)$, i.e.

$$\begin{aligned}\bar{t}_i(\theta_i) &\equiv E_{\theta_{-i}}[u_i(\hat{x}(\theta_i,\theta_{-i}),\theta_i)] - U_i(\theta_i)\\ &= E_{\theta_{-i}}[u_i(\hat{x}(\theta_i,\theta_{-i}),\theta_i)] - \int_{\underline{\theta}_i}^{\theta_i} E_{\theta_{-i}}\left[\frac{\partial}{\partial\theta_i}u_i(\hat{x}(s,\theta_{-i}),s)\right]ds - U_i(\underline{\theta}),\end{aligned}$$

the fact that $U_i^{ir}(\underline{\theta}_i) \leq \tilde{U}_i(\underline{\theta}_i)$ for any BIC-IR mechanism $\{\hat{x},\tilde{t}_1,\ldots,\tilde{t}_n\}$ further implies $\bar{t}_i^{ir}(\theta_i) \geq \tilde{t}_i(\theta_i)$ as well. Hence, no BIC-IR mechanism can have greater expected budget surplus than the IR-VCG mechanism. It follows that if the IR-VCG mechanism runs a strictly negative expected deficit, then all BIC mechanisms must run a strictly negative expected deficit. Hence there cannot be an ex post efficient, IR, BIC mechanism which does not require external funds (i.e., it must run a strict deficit). □

Remarks:

1. We can use this result to more easily prove the impossibility result of MS, as well as other impossibility results (e.g., public goods games with more than two players). See for example the numerous examples in Börgers (2015, ch. 3).
2. Krishna and Perry (1998) provide a general framework with multidimensional types (and a mutli-dimensional version of the BIC integral condition) to arrive at the same conclusion as in Theorem 3.

Bilateral trade example: Return to our bilateral trade setting in which types are uniformly distributed on $[0,1]$. Previously we calculated

$$\bar{t}_b^{vcg}(\theta_b) = \int_0^{\theta_b} \theta_s d\theta_s = \frac{1}{2}\theta_b^2,$$

$$\bar{t}_s^{vcg}(\theta_s) = \int_0^{\theta_s} \theta_b d\theta_b = \frac{1}{2}\theta_s^2.$$

We want to compute ψ_i^* for each player in the IR-VCG mechanism. For the buyer, the interim payoff from the VCG mechanism is

$$U_b^{vcg}(\theta_b) = \theta_b^2 - \bar{t}_b^{vcg}(\theta_b),$$

where the first term is the probability that the buyer gets the good multiplied by the value of the good. Hence,

$$U_b^{vcg}(\theta_b) = \frac{1}{2}\theta_b^2.$$

Because the buyer's outside option is $\underline{U}_b(\theta_b) = 0$, we find ψ_b^* as

$$\psi_b^* = \max_{\theta_b\in[0,1]} \underline{U}_b(\theta_b) - U_b^{vcg}(\theta_b) = \max_{\theta_b\in[0,1]} -\frac{1}{2}\theta_b^2 = 0.$$

We now want to compute ψ_s^*. As we mentioned at the outset, there are two approaches (equally valid) for characterizing $U_s^{vcg}(\theta_s)$ and $\underline{U}_s(\theta_s)$ in this context. In the cost-based approach, the seller's outside option is $\underline{U}_s(\theta_s) = 0$ and the seller earns the interim VCG payoff of

$$U_s^{vcg}(\theta_s) = -(1-\theta_s)\theta_s - \bar{t}_s^{vcg}(\theta_s) = -\theta_s + \frac{1}{2}\theta_s^2,$$

where the first term captures the probability the buyer's type exceeds θ_s multiplied by the cost of supplying the good to the buyer. We find ψ_s^* by solving

$$\psi_s^* = \max_{\theta_s \in [0,1]} 0 - U_s^{vcg}(\theta_s) = \max_{\theta_s \in [0,1]} \theta_s - \frac{1}{2}\theta_s^2 = \frac{1}{2}.$$

Alternatively, in the value-based approach, the seller's outside option is $\underline{U}_s(\theta_s) = \theta_s$ and the seller's VCG interim payoff is

$$U_s^{vcg}(\theta_s) = \theta_s\theta_s - \bar{t}_s^{vcg}(\theta_s) = \frac{1}{2}\theta_s^2,$$

where the first term represents the probability the buyer's type is below θ_s multiplied by the value of the good to the seller. We again find ψ_s^* by solving

$$\psi_s^* = \max_{\theta_s \in [0,1]} \theta_s - U_s^{vcg}(\theta_s) = \max_{\theta_s \in [0,1]} \theta_s - \frac{1}{2}\theta_s^2 = \frac{1}{2}.$$

Thus, with either approach, $\psi_s^* = \frac{1}{2}$.

We can now compute the expected surplus of the IR-VCG mechanism. We know from before that $\bar{t}_b^{vcg} = \bar{t}_s^{vcg} = \frac{1}{6}$, and thus

$$E_\theta[t_b^{ir}(\theta_b, \theta_s) + t_s^{ir}(\theta_b, \theta_s)] = (\bar{t}_b^{vcg} - \psi_b^*) + (\bar{t}_s^{vcg} - \psi_s^*) = \frac{1}{6} - 0 + \frac{1}{6} - \frac{1}{2} = -\frac{1}{3} < 0.$$

We conclude from Theorem 3 that there is no budget-balanced, Bayesian incentive compatible, individually rational ex post efficient mechanism.

6 DSIC implementation, more generally

Throughout these notes we have considered mechanisms to implement the efficient allocation and we have used the VCG mechanism as a building block for finding BIC, IR mechanisms that are ex post efficient. An alternative line of inquiry is to ask "What can be implemented with DSIC mechanisms that are also ex post IR?" An excellent source for the answer to this question is Börgers (2015, chapter 4). Here, I will simply give you a flavor of the kinds of results that have been proven.

1. If the standard single-crossing property is satisfied, there is a characterization theorem for DSIC mechanisms that shares similarities with our well-used BIC theorem. First, there is a stronger monotonicity condition that requires allocations are appropriately monotone in θ_i for any θ_{-i} and not just in expectation over θ_{-i}. Second, the

integral condition for utility (and transfers) is stated for each θ_{-i}. For example, in the case of auctions, DSIC requires that $\phi_i(\theta_i, \theta_{-i})$ is nondecreasing in θ_i for any θ_{-i},

$$U_i(\theta_i, \theta_{-i}) = U_i(\underline{\theta}_i, \theta_{-i}) + \int_{\underline{\theta}_i}^{\theta_i} \phi_i(x, \theta_{-i}) dx,$$

and

$$t_i(\theta_i, \theta_{-i}) = \phi(\theta_i, \theta_{-i})\theta_i - U_i(\underline{\theta}_i, \theta_{-i}) - \int_{\underline{\theta}_i}^{\theta_i} \phi_i(x, \theta_{-i}) dx.$$

2. For auctions, any allocation that awards the good to the agent with the highest $V_i(\theta_i)$, where V_i is an arbitrary increasing function of type, can be implemented with a DISC/ex-post-IR mechanism. This is not very demanding and, as we know, the second-price auction is a DSIC auction.

3. For bilateral trade, the requirements are DSIC are much more demanding. Specifically, assuming we impose ex post budget balance and ex post IR, the only mechanisms that are feasible are those which trade at an exogenous (independent of type) fixed price. Thus, the discontinuous trading equilibria in Chatterjee-Samuelson (1983) are the only DSIC/ex-post-BB/ex-post-IR allocations.