Mechanism Design

Bruno Salcedo*

Winter 2020

So far in the course we have analyzed the behavior of agents taking the institutional arrangements as given. For example, we studied the price setting behavior of firms when they set prices independently, when they set the prices sequentially, and when they set prices repeatedly. For the remainder of the course we will no longer take institutional arrangements as given. Instead, we will address the issue of how to design institutions that deliver desired outcomes. The art of doing so is called *mechanism design*.

1. A Roommates' Dilemma

Frank and Gary are roommates and coffee snobs. They are considering buying a top of the line espresso machine, which costs 1000 dollars. The espresso machine would be a public good in the apartment.[1] They have to decide whether they *buy* the machine or *not*. If they decide to buy it, they also have to decide how much each will pay. That is, they need to come up with transfers t_F and t_G such that $t_F + t_G = 1000$.

How should the roommates make this decision? In class, you proposed some mechanisms including: fixed-split mechanisms, mechanisms with price splits based

*Department of Economics, Western University · brunosalcedo.com · bsalcedo@uwo.ca

[1] It order to keep matters tractable, lets make some assumptions. The machine never has to be washed or repaired, it never breaks, and it has no resell value. There is no way to prevent one another from using the machine. There is no way of monitoring how much each roommate uses the machine.

on past behavior, alternating bargaining protocols, and a simultaneous contributions game. Which of these mechanisms are better for the roommates?

Let us start by figuring out which outcomes are desirable for the roommates. For that purpose, we need a more precise specification of the roommates' preferences. Let v_i be the *value* that roommate i would derive from using the machine. Suppose that the roommates preferences over money are quasilinear. In other words, the utility for roommate i is given by

$$u_i = \begin{cases} v_i - t_i & \text{if they buy the machine} \\ 0 & \text{otherwise} \end{cases} . \tag{1}$$

Recall that in environments with quasilinear preferences, an outcome is Pareto efficient if and only if it is utilitarian. That is, if and only if it maximizes the sum of the individual utilities. The sum of individual utilities is given by

$$u_F + u_G = \begin{cases} v_F + v_G - 1000 & \text{if they buy the machine} \\ 0 & \text{otherwise} \end{cases} . \tag{2}$$

If $v_F + v_G > 1000$ then the total utility from buying is greater than than of not buying, and vice versa. Hence, the optimal outcome is to buy the machine if the sum of values is greater than the price, and to not buy otherwise. That is, it is efficient to buy if and only if the values of the roommates lie within the blue region in Figure 1.

Here is a mechanism that would deliver this outcome. The roommates buy the machine if and only if it is efficient to do so. In case they buy, the roommate with the lowest value pays either 500 or their value, whichever is less. The other roommate pays the rest. For example, if $v_G = 400$ and $v_F = 800$, then the roommates would buy the machine. Gary would pay 400 and Frank 600.

There is a problem with this mechanism. In order to implement it, each of the roommates must know *exactly* what is the value of the other roommate. In reality, it is reasonable to assume that the roommates values are only known privately. If the roommates had to *report* their values, they might have incentives to lie. Suppose that Frank truthfully reports his value to be 800. If Gary reported his value to be only 200, they would still buy the machine, but he would pay 200 instead of 400. Gary has incentives to lie.

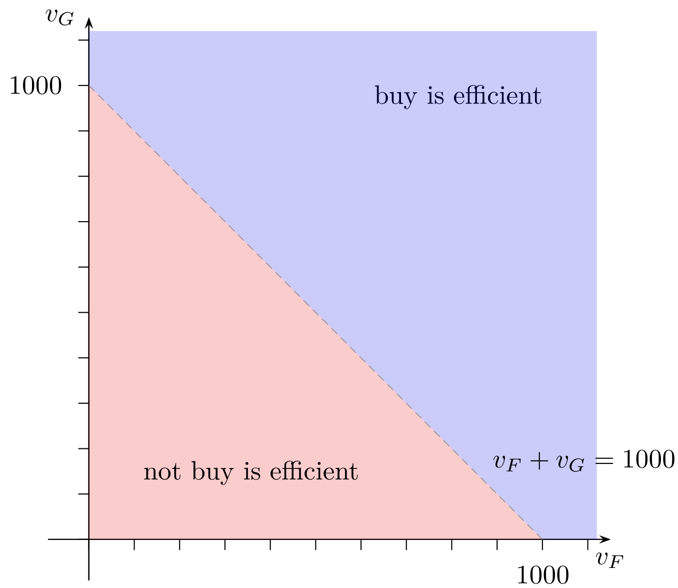

Figure 1 – Efficient outcomes for the roommate problem.

Some of the mechanisms proposed in class do not have this problem. Take for instance the 50-50 split mechanism. According to this mechanism, the machine is bought only if each roommate is willing to pay for half the price. In case the machine is bought, each roommate pays 500. Even if the values were private information, the roommates would have no incentives to lie if they were asked what their value is. If a roommate with a value less than 500 said that his value is greater than 500, he would risk having to pay too much for the machine. Likewise, a roommate with a value greater than 500 would not want to say that his value is less than 500. The roommates have incentives to truthfully report their values in all fixed-split mechanisms. Mechanisms with this property are called *incentive compatible*.

Using the 50-50 fixed-split mechanism, the machine is bought only when *both* roommates have a value greater than 500. This leaves many combinations of values for which the machine is not bought, but it would be efficient to do so. See Figure 2. For example, suppose that their respective values are $v_F = 300$ and $v_G = 900$. Since Frank is not willing to pay 500 for the machine, they would not but it. This outcome is inefficient, because if they bought the machine with Frank paying 200, and Gary paying 800, they would both be strictly better off.

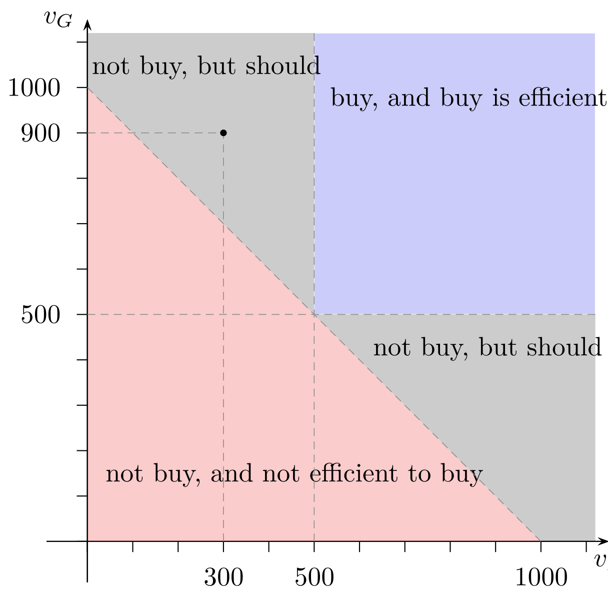

Figure 2 – 50-50 mechanism for roommate problem.

2. The Revelation Principle

Recall the notion of *social choice problems* that we studied at the beginning of the course. Each social choice problem models a situation in which a choice is to be made and the outcome of the choice can potentially affect many individuals. We will focus on transferable-utility environments. In such environments, it is possible to transfer money across individuals, and the individuals have quasilinear utilities.

Formally, a social choice problem consists of three components: (*i*) the set $I = \{1, \dots, n\}$ of individuals involved; (*ii*) the set A of alternatives available to chose from; and (*iii*) for each individual i, a function $w_i(\,\cdot\,)$ that specifies the gross utility that the individual realizes from each possible alternative without taking monetary transfers into account. The quasilinearity assumption means that i's utility can be expressed as

$$u_i = w_i(a) - t_i, \tag{3}$$

where t_i denotes i's transfer, that is, how much i has to pay. In the roommates example we have $w_i(\text{buy}) = v_i$ and $w_i(\text{not buy}) = 0$ for each roommate i.

A mechanism is a institutional arrangement that maps individual decisions into social choices. For example, in the 50-50 split mechanism, each of the roommates reports whether he is willing to pay 500 or not for the machine, and an outcome is determined as a result. Formally, a mechanism consists of three components: (i) for each individual i, a set of actions or messages M_i for i to choose from; (ii) a function $\alpha(\,\cdot\,)$ mapping profiles of action into alternatives; and (iii) for each player i, a function $t_i(\,\cdot\,)$ that specifies how much i has to pay as a function of the actions of all players. A mechanism induces a game in which each player chooses an action, and an alternative and a list of transfers are realized as a result. Assumptions 1–4 in the Game Theory notes allow us to make predictions for these games. Hence, we can map each mechanism into a social outcome.

If we have a way to rank outcomes, we can use it to rank mechanisms based on which outcomes they deliver. One possibility is to look for *optimal mechanisms* from the perspective of the mechanism designer. For example, suppose a monopolist gets to choose between using a fixed-price mechanism or perfect price discrimination. We have learned that perfect price discrimination results in higher profits. Hence, the monopolist would prefer the price-discrimination mechanisms. Another possibility is to rank mechanisms from a social perspective. In that case, we might be interested in mechanisms that deliver Pareto efficient outcomes. For simplicity, we assume throughout these notes that there is always a *unique* Pareto efficient alternative.

Definition 1 A mechanism is *efficient* if it always delivers the Pareto efficient outcome.

The problem of mechanism design can appear to be overwhelming at first sight. There are too many different mechanisms to consider. How are we supposed to keep track of all of them? As we shall see, it suffices to consider a simple and tractable class of mechanisms called *direct mechanisms.* In direct mechanisms, each agent simply reports all of their private information. This information is sometimes called the agent's *type.* Reports are simultaneous and independent. The mechanism determines an outcome based solely on the reports. One can ask whether players have the incentive to report truthfully in a direct mechanism.

Definition 2 A direct mechanism is *incentive compatible* if lying is weakly dominated by telling the truth for all players.

We can think of the 50-50 split mechanism for the roommate problem as a direct mechanism. Each roommate is asked to report their value. The machine is bought if and only if both reports are greater than 500, in which case each roommate pays 500. Moreover, as we have argued before, this mechanism is incentive compatible. A formal justification of this claim is left as a homework exercise.

As it turns out, there is no loss of generality focusing on direct mechanisms in the following sense. If there exists a mechanism that delivers certain outcome, then there also exists a direct incentive-compatible mechanism that delivers the same outcome. This is a powerful result known as the revelation principle. A formal statement is beyond the scope of this class. The following informal claim is sufficient for our purposes.

Claim 1 (Revelation Principle) *Restricting attention to direct incentive-compatible mechanisms is often without loss of generality.*

Different authors contributed to establishing the revelation principle, including Gibbard (1973), Holmström (1977), Dasgupta et al. (1979), and Myerson (1979).

3. The Vickrey Mechanism

We have already analyzed in class an efficient incentive-compatible direct mechanism: the sealed-bid second-price auction. Sometimes, this mechanism is also called the *Vickrey* mechanism or Vickrey auction in honor of Vickrey (1961). Suppose Anna inherited a piece of art from a distant relative. She has no use for this artwork ($v_A = 0$), and would like to make sure that it goes to the person that would enjoy it the most. Bob, Charlie, and David are interested in the artwork. Suppose that their values for receiving the object are $v_B = 7$, $v_C = 10$, and $v_D = 4$, respectively. Of course if Anna simply asked them who values it the most, they would have incentives to lie. Instead, she could use a Vicrey auction.

The Vickrey auction is a direct mechanism for the allocation of a single object. Each player as asked to report their value for the object. The reports are simultaneous and are called *bids*. The object is allocated to the person with the *highest* *bid*. This person pays the auctioneer an amount equal to the *second-highest bid*. The rest of the participants do not pay anything.

In the context of our example. The person with the highest value is Charlie, with a value of 10. The person with the second highest value is Bob, with a value of 7. Hence, if everyone reports truthfully, then Charlie would get the artwork and would pay 7 to Anna.

Claim 2 *Under some conditions, the Vickery auction is efficient and incentive-compatible.*

Justification. We have argued in the game theory notes, that bidding truthfully in sealed-bid second-price auctions is weakly dominant. Hence, the mechanism is incentive compatible. It only remains to show that the outcome is efficient.

The sum of utilities is maximized when the person with the highest value receives the object. The Vickery auction allocates the object to the person with the highest bid. Because of incentive compatibility, the person with the highest bid is also the person with the highest value. Therefore the mechanism is efficient. ■

3.1. Winner's Curse

Claim 2 mentions some conditions. Which conditions are these? One thing that is important is that each person has a different value and knows their own value. This is a plausible assumption when talking about a piece of art that will be used for personal consumption. Some people might like it, while others do not. Each person knows how much they like it.

In contrast, suppose that instead of a piece of art, the auctioneer is auctioning the right to drill for oil in a specific location. The value of drilling right is not determined by subjective tastes, but rather by how much oil there is and how costly it is to extract. It is thus reasonable to assume that the value of the drilling rights is the same (or at least similar) for all the bidders. This kind of settings are called *common-value* environments.

Let us also assume that each bidder observes a *noisy private* signal regarding how much oil there is, but does not know the exact amount. I claim that, in this case, bidding truthfully is no longer weakly dominant. To see this, suppose that everybody bids truthfully. Then, upon winning, the winner of the auction would realize his bid was the highest. This would reveal to him that his private signal was more optimistic than those observed by other firms. Since his signal is as noisy as everyone else's, this realization would lead him to revise his beliefs about how much oil there is. Their new beliefs would be more pessimistic, which means that his initial assessment—and thus bid—was too high. This phenomenon is called the *winner's curse*.

The winner's curse arises in very general settings, but it is easier to understand with a specific example. Suppose that the value oil field either has oil ($v > 0$) or not ($v = 0$), with each of these possibilities being equally likely. Moreover, suppose that each bidder privately runs a test before bidding. When there is oil, the test always comes back positive. However, the test is noisy. Even if there is no oil, the test can return a false positive with some small (bust strictly positive) probability.

Suppose that everyone bids truthfully. That is, everyone bids exactly the expected profitability of the field conditional on the result of their own test. Suppose that at least two bidders observed a positive test result, and at least one bidder observed a negative test result. Negative test results are only possible in fields without oil. Hence, those bidders who observed a negative result would bid 0. On the other hand, those who observed a positive result would bid a positive amount. Because there are at least two such bidders, the winner of the auction would pay a positive price. However, they would also realize that at least one bidder bid zero, which would reveal the fact that there is no oil and the drilling rights are worthless. Anticipating this possibility, bidders with optimistic signals should bid less than what they expect the oil field to be worth.

3.2. Consumption Externalities

Another important condition for Claim 2 to hold is that there should *not* be any consumption externalities. Suppose for instance that Bob would have a dissutility of -5 if Charlie gets the painting. This could arise perhaps because Bob and Charlie are neighbors and Bob plans to locate the artwork in a way that would block the sunlight to Bob's garden. This information is summarized in Figure 3. Each column represents one of the possible alternatives, and each row corresponds to the utility that each of the bidders would receive if such alternative was chosen.

	b	c	d
w_B	7	-5	0
w_C	0	10	0
w_D	0	0	4

Figure 3 – Negative consumption externality

With this externality present, it is no longer weakly dominant for Bob to bid truthfully in a Vickrey auction. If everyone were to bid truthfully, then Charlie would win the auction and Bob's utility would be -5. Bob would rather outbid Charlie. He would have to pay more than his value for the object, but her total utility would be $7 - 10 = -3$, which is greater than -5.

4. The Vickrey–Clarke–Groves Mechanism

The Vickrey auction works great under some conditions. It is weakly dominant for bidders to bid truthfully, and the object is allocated efficiently. But this no longer true in general environments, e.g., with common values or consumption externalities. Moreover, the Vickrey auction is only defined for environments in which a single object is to be allocated to a single person. It cannot be applied, for instance, to the roommate problem. This section analyzes the *Vickrey–Clarke–Groves* (VCG) mechanism, named after Vickrey (1961), Clarke (1971), and Groves (1973). It is a generalization of the Vickrey mechanism that works well in very general settings.

4.1. Individual Contribution to Society

In the VCG mechanism, transfers are determined based on each individual's contribution to the rest of society. Fix an arbitrary individual j. The contribution of j to society is defined as the difference between (i) the total utility that people other than j would receive in a Pareto efficient outcome with j as part of society, and (*ii*) the maximum total utility they could receive if j was not a member of society. The contribution of individual j can be computed in five steps:

Step 1— Find the utilitarian alternative, i.e., the alternative that maximizes the sum of utilities $\sum_i w_i(a)$. Call this alternative a^*.

Step 2— Compute $\sum_{i \neq j} w_i(a^*)$, i.e., the corresponding sum of the utilities of everyone *except* j. This is the total utility that people other than j would receive if j was a member of society and a Pareto efficient alternative was chosen.

Step 3— Find the alternative that would be the utilitarian alternative if j was *not* a member of society, i.e., the alternative that maximizes $\sum_{i \neq j} w_i(a)$. Call this alternative b^*.

Step 4— Compute $\sum_{i \neq j} w_i(b^*)$, i.e., the corresponding sum of the values of everyone *except* j. This is the maximum total utility that people other than j could receive if j was *not* a member of society.

Step 5— j's contribution to society is defined to be the difference between the number you found in step 2 and the number you found in step 4.

Consider for example the art-allocation problem without externalities from Section 3. The utilitarian alternative is to give the object to Charlie, $a^* = c$. The utility of everyone except for Charlie is $v_B(c) + v_D(c) = 0$. If Charlie was not a member of society, the utilitarian alternative would be to give the object to Bob, $b^* = b$. The corresponding sum of utilities is $v_B(b) + v_D(b) = 7$. Hence, Charlie's contribution to society is -7.

What about Bob's contribution? Again, the utilitarian alternative is to give the object to Charlie, $a^* = c$. The utility of everyone except for Bob is $v_C(c) + v_D(c) = 10$. If Bob was not a member of society, the utilitarian alternative would still be to give the object to Charlie, $b^* = c$. The corresponding sum of utilities is $v_C(c) + v_D(c) = 10$. Hence, Bob's contribution to society is 0. Note that both Charlie's and Bob's contribution correspond to how much they would have to pay in a Vickrey auction.

4.2. The VCG Mechanism

The VCG mechanism is the direct mechanism that always chooses the efficient outcome and compensates individuals in accordance with their contribution to society, given the reported types. Let $\hat{w}$ be the utility functions reported by the individuals. The alternative chosen by the VCG mechanism is the one that maximizes $\sum_i \hat{w}_i(a)$. Call this alternative $a^*(\hat{w})$ to emphasize the fact that it is determined by the reports, and not on the true utilities. The transfer to player j when the VCG mechanism is being used is

$$t_j^{\text{VCG}}(\hat{w}) = \sum_{i \neq j} \hat{w}_i \left[a^*(\hat{w})\right] - \sum_{i \neq j} \hat{w}_i \left[b^*(\hat{v})\right]. \tag{4}$$

In the case of a single-object private-value auction without externalities, the VCG mechanism coincides with the Vickery auction. But the VCG mechanism can be defined and works well in general settings.

Claim 3 *The VCG mechanism is always incentive compatible and efficient.*

Justification. By construction, the VCG mechanism chooses the utilitarian alternative given the *reported* values. Hence, as long as the individuals report truthfully, the VCG mechanism will be efficient. It remains to show that each individual has incentives to report truthfully even when she believes that others are lying,

Individual j's utility is determined by the chosen alternative $a^*(\hat{w})$, and by her transfer $t_j^{\text{VCG}}(\hat{w})$, in accordance with the following equation

$$u_j = v_j \left[a^*(\hat{w})\right] - t_j^{\text{VCG}}(\hat{w}). \tag{5}$$

Substituting with the formula for the VCG transfers (4) in (5), j's utility equals

$$u_j = \underbrace{v_j \left[a^*(\hat{w})\right] + \sum_{i \neq j} \hat{w}_i \left[a^*(\hat{w})\right]}_{\text{sum of everyone's values at } a^*(\hat{v})} - \underbrace{\sum_{i \neq j} \hat{w}_i \left[a^*_{-j}(\hat{w})\right]}_{\text{does } \textit{not} \text{ depend on } \hat{v}_j}. \tag{6}$$

Notice that the last term does not depend on j's report (why?). Hence, it will not affect j's incentives to report truthfully. The first two terms, correspond to the sum of j's value and everybody else's reported values, evaluated at $a^*(\hat{w}_j, \hat{w}_{-j})$. The utilitarian action $a^*(\hat{w}_j, \hat{w}_{-j})$ is precisely the action which maximizes the sum of those values. Hence, the sum of these first two terms is maximized when $\hat{w}_j = w_j$. Therefore, j's utility is maximized when j reports truthfully, regardless of the report of other individuals. ■

Hence, unlike the Vickrey auction, the VCG mechanism is efficient both for common value auctions, and for auctions with consumption externalities. Let us now examine how does the VCG mechanism look like for the auction with consumption externalities as in Figure 3. The first step is to figure out what the utilitarian alternative is. Note that the sum of values from alternative a for this example can be written as

$$\sum_i v_i(a) = v_B(a) + v_C(a) + v_D(a) = q_B(v_B - 5) + q_C v_C + q_D v_D. \tag{7}$$

We can adjust the bidders values to take into account consumption externalities as follows $\tilde{v}_B = v_B - 5$, $\tilde{v}_C = v_C$, and $\tilde{v}_D = v_D$. The sum of values can be expressed in terms of adjusted values as follows

$$\sum_i v_i(a) = q_B \tilde{v}_B + q_C \tilde{v}_C + q_D \tilde{v}_D. \tag{8}$$

Hence, the efficient outcome is to give the object to the bidder with the highest *adjusted* value.

As for transfers, we need to consider different cases. First, suppose that $\tilde{v}_B > \tilde{v}_C > \tilde{v}_D$. In that case, using formula (4), Bob would have to pay

$$t_B^{\text{VCG}} = \Big[v_C(C) + v_D(C)\Big] - \Big[v_C(B) + v_D(B)\Big] = v_C + 5. \tag{9}$$

The case when $\tilde{v}_B > \tilde{v}_D > \tilde{v}_C$ is completely analogous. Hence, when Bob gets the object, he must pay the second highest value, plus a penalty equal to the size of the externality that his consumption will impose on Charlie.

Now, suppose that $\tilde{v}_D > \tilde{v}_B > \tilde{v}_C$. In that case, using formula (4), David would have to pay

$$t_D^{\text{VCG}} = \Big[v_B(B) + v_C(B)\Big] - \Big[v_B(D) + v_C(D)\Big] = v_B - 5. \tag{10}$$

When David gets the object, and Bob has the second highest adjusted value, then David pays Bob's value minus a discount equal to the size of the externality. In every other case, the VCG transfers coincide with the Vickrey auction transfers.

5. Second-Best Mechanisms

5.1. Impossibility of First Best

The VCG mechanism is efficient and incentive compatible in very general settings. However, it is not always perfect. Let us apply the VCG mechanism to the example from Section 1 about two roommates who must decide whether to *buy* or *not* an espresso machine. In order to do so, we must take into account Oscar. Oscar is the current owner of the espresso machine, which he values at $v_O = 1000$.

The VCG yields the efficient outcome, which is to buy the machine if and only if $v_F + v_G > v_O$. The transfers need to be determined on a case-by-case basis. Suppose that $v_F + v_G > v_O$. In that case, Frank should pay

$$t_F^{\text{VCG}} = \Big[v_G(\text{not buy}) + v_O(\text{not buy})\Big] - \Big[v_G(\text{buy}) + v_O(\text{buy})\Big] = 1000 - v_G. \quad (11)$$

Similarly, Gary should pay $t_G^{\text{VCG}} = 1000 - v_F$. The problem is that, in this case, the sum of the VCG ransfers is not enough to cover the price of the espresso machine:

$$t_F^{\text{VCG}} + t_G^{\text{VCG}} = 1000 + (1000 - v_F - v_G) < 1000. \quad (12)$$

The VCG mechanism applied to the roommate problem runs a deficit.

Can we find an efficient mechanism that does not have this problem? The answer is no. Let us see why. Thanks to the revelation principle (Claim 1), it suffices to pay attention to incentive-compatible direct mechanisms. Consider any direct mechanism which is both efficient and incentive compatible. Take an arbitrary number v_G^0, and let us analyze how the mechanism operates when Gary truthfully reports his value to be equal to v_G^0.

Let v_F^* be the number such that it is efficient to buy the espresso machine if an only if $v_F \geq v_F^*$. See Figure 4. Let p_F be the amount that Frank would have to pay if the machine is bought, still assuming that Gary reports $\hat{v}_G^0$. I claim that it must be the case that $p_F = v_F^*$.

The first thing to note is that p_F cannot depend on Frank's report. Suppose there were two different reports that Frank would make, such that both reports

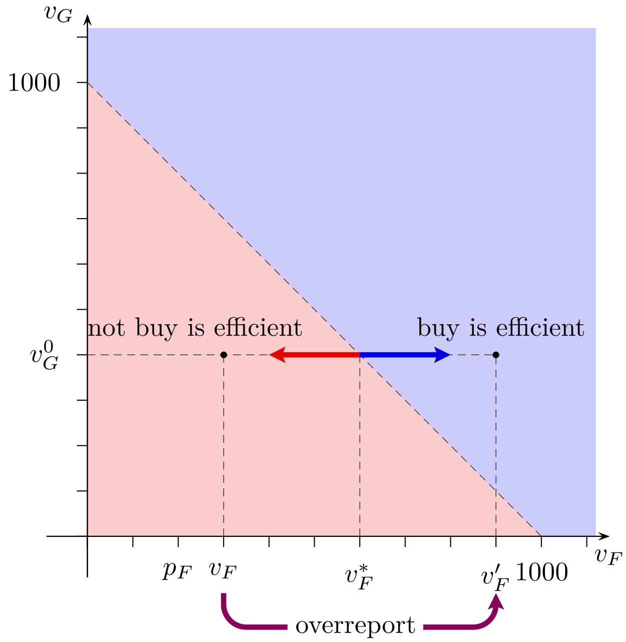

Figure 4 – Incentive compatibility for efficient mechanisms.

would lead to the machine being bough but would lead to different prices. Then, Frank would always prefer the report that leads to the lower price, and the mechanism would *not* be incentive-compatible.

Now, suppose that we had $p_F < v_F^*$, as in Figure 4. If Frank's true value was between p_F and v_F^*, we would have incentives to lie. In that case, if Frank reported truthfully, the machine would not be bought and Frank's utility would be zero. Instead, Frank could overreport and claim that his value equals $v_F' > v_F^*$, as in Figure 4. If he lied, the machine would be bought and he would only pay p_F. His utility from lying would thus equal $v_F - p_f > 0$. Hence, if $p_F < p^*$, then the mechanism would *not* be incentive compatible.

One can use a completely analogous argument to show that, if $p_F > v_F^*$, then the mechanism would also *not* be incentive compatible. Hence, for an efficient mechanism to be incentive compatible, it must be the case that $p_F = v_F^*$. Now, note that $v_F^* = 0$ if $v_G^0 \geq 1000$, and $v_F^* = 1000 - v_G^0$ if $v_G^0 < 1000$. This is precisely the same formula of the VCG mechanism transfer t_F^{VCG}. Hence, (essentially)[2]

[2] A careful reader might notice a small caveat. In our analysis, we assumed implicitly that the roommates do not pay anything when they do not buy the machine. In principle, we could also consider mechanisms in which this is not the case. Doing so would allow us to find additional incentive-compatible direct mechanisms, but any such mechanism runs a worse deficit than the

the *only* efficient and incentive-compatible direct mechanism for the provision of public goods is the VCG mechanism. Since the VCG mechanism runs a deficit, it is impossible to achieve efficiency while balancing the budget at the same time. When the value that each individual attaches to a public good is private information, there is *no* mechanism that guarantees efficient outcomes without an external source of funding.

Claim 4 *There is no efficient mechanism for the provision of public goods which never runs a deficit.*

A similar analysis can be carried out in more general situations. In general, among all efficient mechanisms, the VCG mechanism is the one that runs the smallest expected deficit. Hence, if the VCG always runs a deficit, then there is *no* mechanism that guarantees efficient outcomes without an external source of funding.

5.2. Second-Best Mechanisms

In situations for which there all efficient mechanisms require external funding, we can ask which is the *least* inefficient mechanism which never runs a deficit. Mechanisms which never run deficits are called *budget-balanced*. The least inefficient budget-balanced mechanisms are called *second-best* mechanisms. The first step is to precisely define what we mean when we say that 'a mechanism is more or less efficient than another'. We already know how to rank outcomes. Since we have assumed quasilinear utilities, it is reasonable to compare outcomes based on total social value.[3] We will rank mechanism based on the outcomes they generate, in accordance with the following definition.

Definition 3 Mechanism M_1 is *more efficient* than mechanism M_2, if M_1 always generates at least as much total social value as M_2, and sometimes it generates strictly more total social value.

The criterion from Definition 3 offers an incomplete ranking. There are many

VCG.

[3]Remember our discussion of the utilitatian criterion with quasilinear utility.

pairs of mechanisms such that neither one is more efficient than the other. Hence, this criterion will often not be able to pinpoint a single second-best mechanism. Instead, we will focus on the *set* of mechanism that cannot be improved upon.

Definition 4 A budget-balanced mechanism is a *second-best mechanism* if there is no other budget-balanced mechanism which is more efficient than it.

5.3. Public Goods

Let us proceed and find the set of second-best mechanisms for the roommate problem. Thanks to the revelation principle (Claim 1), it suffices to consider direct direct mechanisms which are both incentive-compatible and budget-balanced. There are only two alternatives for the roommate problem: buying the machine or not. Therefore, the choice rule of a direct mechanism can be characterize by specifying the set of possible profiles of reports that would lead to buying the machine. For example, for the 50-50 mechanism, this set corresponds to the blue region in Figure 2. For the VCG mechanism, this corresponds to the blue region in Figure 1. Incentive-compatibility imposes three restrictions on the shape that this region can have, as well as on the transfer rule.

The first restriction is that the region of reports where the machine is bought must extend to the north-east. See for instance the mechanism illustrated in Figure 5. That is, if the machine is bought for some profile of reports, then it must also be bought when both Frank and Gary report higher values. Mechanisms with this property are called *monotone*. To see why monotonicity is important, suppose that Frank is willing to pay a certain amount to buy the machine. If his value for the machine increases, he would still be happy to pay the same amount for the machine. If the mechanism would no longer allow him to buy the machine when his value is high, he would have incentives to lie and pretend that his value is low.

The second and third restrictions, are the same restrictions we discussed in the preceding section. The price p_j that roommate j pays cannot depend on his own report, conditional on the machine being bought. Also, p_j must be exactly equal to the minimum that j could make that would lead to buying the machine. This transfer rule is illustrated in Figure 5. It is also possible to verify that both the VCG and the 50-50 fixed-split mechanisms satisfy them. In fact, a

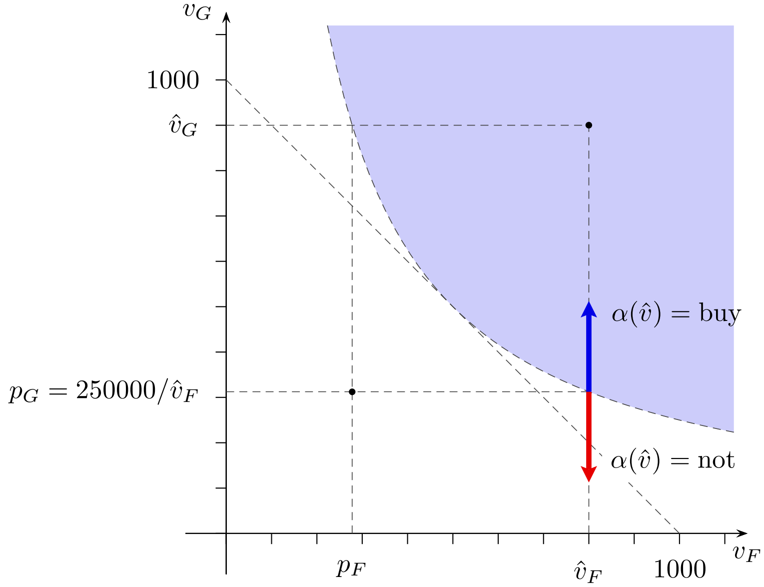

Figure 5 – A mechanism which is incentive compatible but not budget balanced. The machine is bought only if the product of the values is greater or equal that 250,000. If the machine is bought, each roommate pays $t_i = 250,000/\hat{v}_{-i}$.

direct mechanism is incentive compatible *if and only if* it satisfies these three restrictions.

Now let us turn our attention to balancing the budget. Note that the point (p_F, p_G) is below the budget line. This indicates that the mechanism is running a deficit. The only way to guarantee that this does not happen, is for the region where the machine is bought to be contained inside a rectangle northeast of the budget line.

Claim 5 *An incentive-compatible direct mechanism for the provision of public goods is a second-best mechanism if and only if it is a fixed-split mechanism.*

5.4. Bilateral Trade

Claim 6 *An incentive-compatible direct mechanism for bilateral trade is a second-best mechanism if and only if it is a fixed-price market mechanism.*

References

Clarke, E. H. (1971). Multipart pricing of public goods. *Public Choice*, *11*(Fall), 17–33.

Dasgupta, P., Hammond, P., & Maskin, E. (1979). The implementation of social choice rules: Some general results on incentive compatibility. *The Review of Economic Studies*, *46*(2), 185–216.

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica*, *41*(4), 587–601.

Groves, T. (1973). Incentives in teams. *Econometrica*, *41*(4), 617–631.

Holmström, B. R. (1977). *On incentives and control in organizations* (Unpublished doctoral dissertation). Stanford University.

Myerson, R. B. (1979). Incentive compatibility and the bargaining problem. *Econometrica*, *47*(1), 61–73.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, *16*(1), 8–37.